Intelligent Multimodal Deep Learning Framework for Fake News and Rumor Detection in Online Social Networks

Dr. Saravanan K [1], Dr. Arvind Kumar Tiwari [2]

[1] Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia

[2] Lincoln University College, Malaysia /
Professor, Computer Science & Engineering Department, Kamla Nehru Institute of Technology (KNIT), Sultanpur, U.P, India.

Email ID: [1] pdf.saravanan@lincoln.edu.my / saravanankv1706@gmail.com

[2] pdfsv.arvind@lincoln.edu.my / arvind@knit.ac.in

Abstract

The rapid growth of online social networks (OSNs) has greatly improved global communication and information sharing, but it has also accelerated the spread of fake news and rumors. This misinformation can threaten political stability, public health, and social harmony. Conventional fake news detection methods that rely on rule-based techniques or basic machine learning models are often unable to address the complex and evolving nature of online misinformation. To overcome these challenges, this study proposes an **Intelligent Multimodal Fake News Detection System (IMFNDS)** that integrates different types of information for more reliable detection. The proposed framework analyzes textual, visual, propagation, and metadata features to enhance accuracy. Transformer-based models such as BERT are used for text understanding, while Convolutional Neural Networks (CNNs) extract visual patterns from images and videos. In addition, Graph Neural Networks (GNNs) model how information spreads across social network communities. These diverse features are combined using an attention-based multimodal fusion mechanism, allowing the system to focus on the most relevant signals. An ensemble classification approach is further applied to increase robustness and reduce bias. Experimental evaluation shows that the proposed model performs better than unimodal methods, achieving higher accuracy, improved F1-score, and effective early detection of misinformation. Overall, IMFNDS offers a scalable and interpretable solution for identifying fake news and supporting trustworthy online information ecosystems.

Keywords- Fake News Detection, Deep Learning, Multimodal Framework, Online Social Networks, Graph Neural Networks, Rumor Detection

1. Introduction

Online social networks (OSNs) such as Twitter, Facebook, Instagram, YouTube, and Reddit have become dominant platforms for communication and information exchange in the modern digital era. These platforms allow billions of users worldwide to generate, share, and consume content instantly, transforming the way people access news, form opinions, participate in social and political discourse [1]. At the same time, however, the rapid growth of these networks has introduced serious challenges, particularly the widespread dissemination of misinformation, including fake news and rumors [2] [3]. Deep neural networks reduce reliance on manual feature engineering and offer improved generalization capabilities [4]. Transformer-based language models such as BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa have achieved remarkable success in understanding contextual semantics, sarcasm and linguistic subtleties commonly present in deceptive narratives. These models capture long-range dependencies in text, making them particularly effective for analyzing fake news articles, social media posts, and rumor discussions [5].

2. Methodology

The IMFNDS framework uses a multimodal architecture to detect misinformation in online social networks. It combines text, visual content, propagation patterns, and user metadata to improve detection reliability. This approach overcomes the limitations of single-modality models. The system consists of data collection, preprocessing, feature extraction, multimodal fusion, and classification modules.

2.1 Data Collection and Preprocessing Layer

The **Data Collection Layer** forms the foundation of IMFNDS by gathering raw multimodal data from diverse OSN platforms such as **Twitter, Facebook, Reddit, and Instagram**. Fake news and rumors manifest differently across platforms depending on user behavior, media format, and engagement mechanisms. The collected raw data is inherently noisy, incomplete, and heterogeneous, making preprocessing a crucial step before feature extraction. The **Preprocessing Layer** standardizes and cleans each modality to ensure consistent input representations for deep learning models.

2.2 Multimodal Fusion and Classification Layer

IMFNDS integrates heterogeneous features through a **Multimodal Fusion Layer** using an attention-based mechanism to capture inter-modal dependencies. This method assigns adaptive weights to each modality based on contextual importance. The fused representation is then classified using an **ensemble-based classifier** to detect fake or real content.

F=Attention(T,V,G,M) F = Attention (T, V, G, M) F=Attention (T,V,G,M)

where **T** represents textual features, **V** visual features, **G** graph diffusion features, and **M** metadata features.

3. Results and Analysis

The IMFNDS model was evaluated using the FakeNewsNet and Twitter15 datasets containing labeled fake and real news. Its performance was compared with unimodal and baseline models. Metrics such as Accuracy, Precision, Recall, and F1-score were used to measure detection performance. Results demonstrate the effectiveness of multimodal feature integration for misinformation detection.

3.1 Quantitative Results

The first set of experiments compared IMFNDS against individual unimodal models as well as a basic multimodal fusion approach. The **BERT text model** achieved **89.2% accuracy**, while the **CNN image model** reached **81.4%**, indicating limits of **unimodal methods**. The **GNN propagation model** improved performance to **85.6%** by capturing **diffusion patterns**. **Multimodal fusion** increased accuracy to **93.8%**, and the **IMFNDS ensemble** achieved the highest performance with **95.4% accuracy and 0.94 F1-score**. The results are summarized in Table 1.

Table 1: Performance Comparison

Model Approach	Accuracy (%)	Precision	Recall	F1-score
Text-only (BERT)	89.2	0.88	0.87	0.87
Image-only (CNN)	81.4	0.79	0.80	0.79
Graph-only (GNN)	85.6	0.84	0.83	0.83
Multimodal Fusion	93.8	0.93	0.92	0.92
IMFNDS (Ensemble)	**95.4**	**0.95**	**0.94**	**0.94**

3.2 Early Detection Performance

Early detection is a critical requirement in misinformation mitigation, as fake news spreads rapidly before fact-checking mechanisms can respond. To test IMFNDS under early-stage conditions, model performance was evaluated within the first two hours of news diffusion. The results are presented in Table 2.

Table 2: Early Detection Results (First 2 Hours)

Model	Early Accuracy (%)
Text-only	82.5
Graph-only	80.2
Multimodal Fusion	88.7
IMFNDS	**91.9**

The findings show that IMFNDS maintains strong predictive capability even in early detection scenarios, achieving **91.9% early accuracy**. This demonstrates that combining content-based and propagation-based signals enables the model to detect misinformation before it reaches large-scale virality.

3.3 Discussion

The results show that detecting fake news using only one type of data is unreliable. Text, image, or graph-based models alone cannot fully capture misinformation patterns. By combining text, visuals, metadata, and propagation features, IMFNDS provides more accurate detection. The ensemble and attention-based fusion methods also improve robustness and transparency.

4. Conclusion

The proposed **IMFNDS** applies a deep learning–based multimodal framework for fake news detection. It integrates textual, visual, propagation, and metadata features using **Transformer, CNN, and GNN models**. An attention-based fusion mechanism combines these heterogeneous features. Experimental results on **FakeNewsNet** and **Twitter15** show improved accuracy over unimodal models.

References

1. 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.
2. H. Zhang, Z. Ma, and X. Jin, “BERT-based fake news detection on social media,” *IEEE Access*, vol. 9, pp. 35678–35691, 2021.
3. 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.
4. D. Roy and M. Chakraborty, “Hybrid graph and transformer model for rumor detection,” *IEEE Trans. Comput. Social Syst.*, early access, 2023.
5. P. Zhou, X. Zhou, and J. Yang, “Fake news detection with multi-model ensemble learning,” *IEEE Access*, vol. 9, pp. 109345–109356, 2021.