Grand Asian journal of Computing and Emerging Technologies (GJCET)

Fake News Detection via Textual BERT Embeddings and Knowledge-Aware Graph Neural Networks

Issue (1) & volume (1)

Issue Date August 31, 2025

Wajahat Arshad[1], Ayesha Manzoor[1], Muhammad Hassan[1]*

[1] *Department of Computer Science, University of Management and Technology*

**Corresponding Author:*

wajahatali4564@gmail.com

Received data: 10 July 2025 Revised date: 30 July 2025 Accepted date:14 August 2025 Published date: 31 August 2025

Abstract: Fake news generation and propagation is a huge challenge of the digital era, resulting in different social impacts, namely bandwagon, validity content, and deceiving the public with spam and much more. The rapid spread of fake news not only fosters misinformation but also degrades the credibility of news sources. To comprehend the critical need for addressing this persuasive issue, this research presents a framework for detecting fake news using a knowledge-based approach by combining GCN and GNN only for text data. An automatic fact-checking process is applied using concepts like information retrieval, NLP, and Graph theory. The knowledge base is generated using a Twitter dataset, which contains. These attributes serve as pivotal indicators for the development of a knowledge base, subsequently employed to detect prevalent patterns and traits linked to deceptive or false information. We have employed Named Entity Recognition (NER) model to extract SPO triples and Latent Dirichlet Allocation (LDA) for topic modeling, thereby contributing to knowledge base generation. To evaluate the efficacy and efficiency of our proposed model, we utilize deep learning algorithms like GPT-3 and BERT Transformer, providing an acceptable level of accuracy. This research paper delivers valuable insights into addressing the proliferation of fake news on Twitter.

Keywords: Fake news classification, Graph Neural Network, LLM, BERT, social context, GCN.

1. Introduction

The rapid proliferation of misinformation across digital platforms poses a growing threat to societal trust and the reliability of information. The use of social networks brings many benefits, but it also presents a significant issue: the widespread dissemination of fake news. It refers to intentionally false or misleading content related to people, topics, and events, whether emotionally or politically charged. Most of the time, such untrusted content goes viral without proper fact-checking. This increasing trend highlights the urgent need for effective and efficient solutions to find and reduce the impact of fake news on social media.

The term “Fake news” refers to fabricated or intentionally misleading information presented as factual news. It is a phenomenon that poses significant challenges to society and public discourse. Due majority of social media platforms and the ease of sharing information over online media have contributed to the rapid spread of false content. This introduction sets the stage to delve deeper into the origins, mechanisms, consequences, and potential solutions in the fight against fake news.

Fake news dissemination enforces to misinformation and undermines the credibility of news sources. To understand this increasingly broad issue, this research presents a framework for identifying fake news on Twitter using a knowledge-based approach. I have collected a dataset for evaluating our state-of-the-art model on the Twitter platform for both text and images, and combined the GCN and GNN for testing our model. Our proposed framework has formulated four different attributes, namely: SPO triple, SPO sentiment polarity, topic modeling, and combining both GCN and GNN for that purpose. These attributes are the key indicators contributing to knowledge base creation and used for fetching patterns and properties, enabling a deeper and more nuanced analysis of Twitter news content in multi-modality. The framework uses deep learning algorithms, including BERT transformer for text and ResNet50 for image analysis, to improve the accuracy and effectiveness of fake news identification by combining GCN and GNN using a knowledge base graph. These algorithms play a key role in the classification and fine-tuning processes, allowing the framework to separate fake and real news with greater accuracy. Extensive testing has been done to check the effectiveness in identifying fake news on Twitter, where accuracy is employed as a key measure to gauge its performance and reliability. Overall, the results of this study have important implications for combating the spread of fake news on Twitter, advancing the understanding of a knowledge-based approach by combining GCN and GNN, and the use of advanced deep learning algorithms to combat fake news.

Our main Contribution of this study are summarized as follows:

- We present a deep learning-based approach that combines analyzing both text and visual features for identifying fake articles using BERT for text and Res-Net for images.
- In addition to multimodal features, we specifically incorporate knowledge-based properties such as SPO triple and Sentiment polarity to enhance the input data for identification.
- For model testing, we evaluate the model's performance using different classification metrics such as accuracy, F1-score, and confusion matrices.

2. Related Work

The work done in the past on fake news detection includes different methods that help to detect and mitigate misinformation. Early research in this area has focused mainly on extracting fake news from a single modality, only in text, images, or videos. This literature provides an overview of recent research and approaches on the detection of single- and multi-model fake news.

2.1 Machine Learning Based Approaches

Several studies have explored the use of traditional machine learning (ML) techniques for fake news detection, focusing mainly on misinformation spread through different social media platforms. This study employed classical ML models such as SVM, Decision Trees, Logistic regression, neural networks, and ensemble methods to identify pandemic-related misinformation on Facebook, Twitter, Instagram, and YouTube (Naeem et al., 2024). The model has achieved the accuracies between 83% and 89%, showing their effectiveness in identifying fake content. However, their dependency on shallow features limited their ability to interpret linguistic nuances, and the use of an imbalanced or estimated dataset size embeds challenges for real-world implementation.

Another study (D'Ulizia et al., 2024) focused on multimodal deception detection using supervised ML algorithms such as Neural Networks, Random Forests, SVMs, and K-Nearest Neighbors. The authors evaluated these algorithms across diverse datasets such as TRuLie, Bag-of-Lies, and real-time courtroom videos. Performance of these models ranges from 75% to 85%. Despite offering comparative insights, the study suffered from limitations such as a small dataset size and inconsistencies in input modalities, which reduced the generalizability of the model.

A novel ML-based model known as Multi-model Aggregation Portrait Model (MAPM) was introduced for detecting fake user profiles on the Weibo platform(Li et al., 2024). This approach used multi-dimensional behavioral features and showed the accuracy of approximately 90.2% in differentiating user categories such as normal, reproduce, and lottery users. Nonetheless, its performance was evaluated using Weibo data, limiting the transferability of the model to other platforms with different user behavior patterns.

(Ellam et al., 2025) focused on using NLP with traditional ML models using the ISOT fake News Dataset. The dataset has 25,000+ articles collected from two sources, such as Reuters and Politi-fact. The models, including SVMs and Logistic Regression, achieved the prominent accuracy of 88% to 92%. Although the results were good but it focused only on English; it didn't work well in other languages or cultural contexts.

2.2 Deep Learning Based Approaches

(Vineela et al., 2025) proposed hybrid deep learning model that used TF-IDF features for textual data with visual features extracted using MobileNetV2 and VGG-19. The model evaluated on 20,015 news articles and achieved the accuracy of 89%, showing the benefit of combining multimodal data. However, this study lacked on generalizing cross-domain dataset.

(Yadav & Gupta, 2024) introduced a model using vision Transformer (ViT) to embed emotional properties into multimodal classification and achieved the remarkable accuracy ranging from 94% to 98% across five datasets. The emotional sentiment fusion offered a unique direction or approach, yet its dependency on emotion limited its application to more neutral content. Similarly, another study developed a framework which combined LSTM for text and CLIP for image classification, achieved the accuracy of 99% on text data and 93.12% for joined input (Kumari & Singh, 2024). It supported multiple languages and showed the benefits of multimodal fusion. But the study lacked testing under real-time adversarial environments.

Another study tested deep learning models, including RNN, LSTM, GRU, BERT, and GPT-3 (Nair et al., 2024). While GPT-3 reached 81% accuracy, the performance of BERT remained lower at 61%. This variability indicated model sensitivity to data type, and the study did not address scalability or real-time processing.

(Su et al., 2025) utilized both semantic information and user credibility, leveraging hypergraph structures, and achieved an accuracy of up to 90% across various datasets. The dual-channel structure architecture improved relational reasoning, though reliance on user metadata raised concerns regarding privacy and robustness. Another study used an RNN-LSTM framework on the LIAR dataset; this model achieved 99.1% accuracy. Its strong text classification capability was evident; however, the approach lacked validation in streaming data environments, were fake news spreads rapidly.

2.3 Large Language Models (LLMs) and Hybrid Approaches

(Wang et al., 2024) introduced an LLM-based system (FND-LLM) evaluated on Weibo, Gossip-cop, and Politi-fact and achieved the accuracies of 91.2%, 90.5%, and 92.6%, respectively. The model showed the power of LLM for multilingual and cross-domain fake news classification. But it lacked dynamic changes in rapidly evolving fake news streams. Another study used a proposed hold for LLMs and Vision-Language Models in Italian language misinformation detection (Bondielli et al., 2024). Although resources offered rich annotation for both detection and relation categorization. But the lack of experimental results evaluation limited its utility for benchmarking.

Similarly, another study presented a modular framework that used different modalities (Liu et al., 2025). The model achieved the accuracy of 92.8% on Weibo and 95% on Weibo-21, highlighting effectiveness. But this model needed more testing under inconsistent or noisy data conditions for adversarial robustness.

3. Materials and Methods

3.1 Overview

Our proposed model integrates BERT for semantic embeddings with GCN and GNN using a knowledge graph constructed by Subject-Predicate-Object (SPO) triple fetched from tweets. These combined features used to classify tweets as fake or real.

3.2 Text Embedding with BERT

Given a tweet t, we use BERT to calculate a contextual embedding:

$$h_t = \mathrm{BERT}(t) \in \mathrm{R}^{d\text{bert}}$$

where $d_{\text{bert}} = 768$.

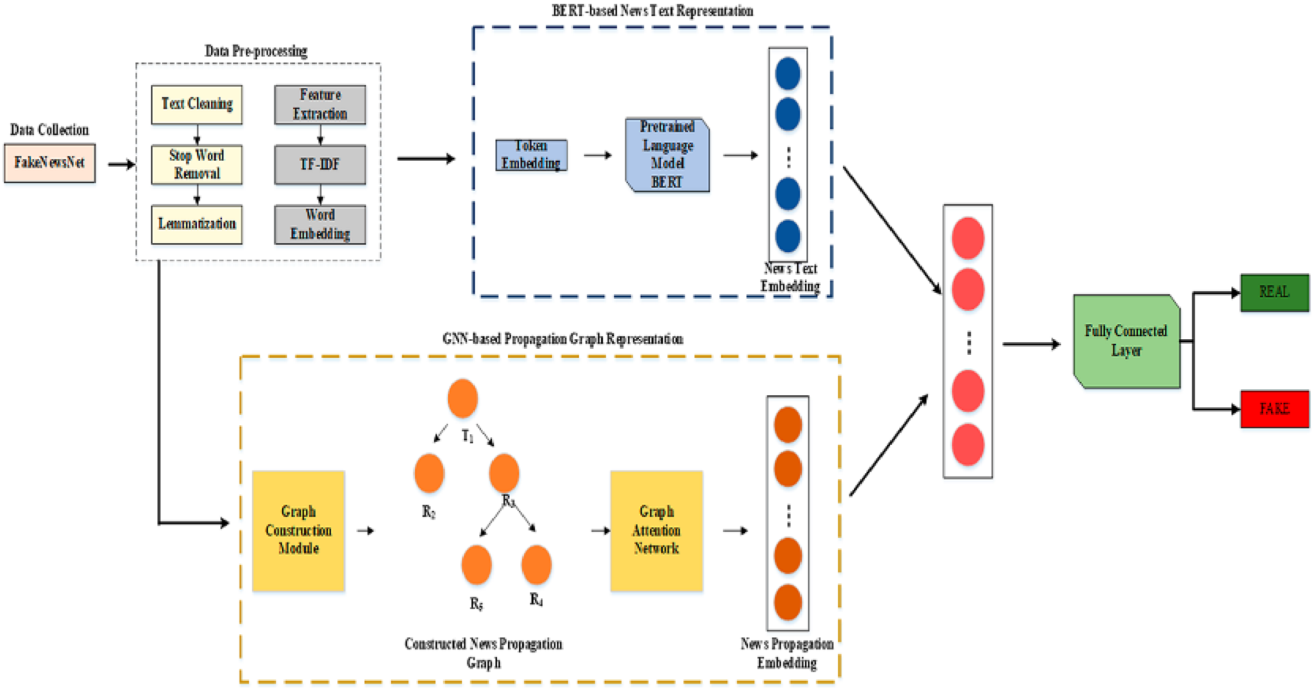

Fig. 1: Proposed Methodology of multi model **(GCN and GNN)**

3.3 Constructing Knowledge Graph

Extract SPO triples $\{(s_i, p_i, o_i)\}$ from tweet text using dependency parsing. Construct a directed knowledge graph:

$$G = (V, E)$$

- **Nodes** V: entities $s_i \cup o_i$
- **Edges** E: from s_i to o_i, labeled by p_i

X ∈ R N×dnode

A^ is the normalized adjacency matrix, σ = ReLU & Output graph embedding (via mean).

3.5 Fusion & Classification

Concatenate BERT and GCN embeddings:

$$z = [h_t \| g] \in \mathrm{R}^{d\text{bert}+d\text{gcn}}$$

Initial node features $x_v \in \mathrm{R}^{d\text{node}}$ are randomly initialized.

3.4 GCN Encoding

Using a two-layer GCN:

$$H^{(1)} = \sigma \;\; \hat{A} X W^{(0)}, H^{(2)} = \mathrm{GCN} \;\; H^{(1)}, \hat{A}$$

Where:

Then pass through an MLP:

$$y\hat{} = \mathrm{Softmax}\,(W_2 \cdot \mathrm{ReLU}(W_1 z + b_1) + b_2)$$

3.6 Evaluation

We are testing our proposed model based on the following metrics:

- Accuracy
- F1 Score

- Confusion Matrix
- Precision
- Recall

4. Datasets

To test the correctness of our proposed model, we used three commonly used datasets in the area of fake News Detection: Weibo, PolitiFact and Gossip Cop. Each dataset gives novel features, languages and content types that enhance the generalizability of our proposed model across different domains.

Table 1: Dataset statistics of Weibo, Gossip Cop, PolitiFact

Dataset	Tab	Quantities	Aggregate
Weibo	Fake news	4779	9528
	Real news	4749	
Gossip Cop	Fake news	16817	22140
	Real news	5323	
PolitiFact	Fake news	624	1056
	Real news	432	

5. Conclusion

To sum up, our experiment showcased encouraging performance, with BERT achieving 61% accuracy and GPT-3 surpassing all with an accuracy rate of 81% in classifying fake news articles. These deep learning methods were applied on our formulated knowledge base which includes attributes, namely Subject-Predicate-Object (SPO) triplet, SPO sentiment polarity, SPO occurrence, and topic modelling, generated as key indicators of knowledge base to detect patterns and traits related to misinformation or fake news. The deep learning models effectively capture both textual and numeric features, providing a comprehensive understanding of the veracity of information and their patterns. Additionally, we discussed the potential future scope of integrating Graph Convolutional Networks (GCN) and Graph Neural Networks (GNN) using knowledge graphs for images dataset, which can further enhance the performance of fake news detection systems.

6. Future recommendation

In this research, we proposed and implemented a knowledge-based fake news detection model that uses BERT for textual feature extraction and combines it with Graph Convolutional Networks (GCN) and Graph Neural Networks (GNN) to leverage the structural relationships among entities in a knowledge graph. Our state-of-the-art approach mainly focused on textual data from Twitter, improved by SPO triples to model semantic and contextual association between claims. As part of our future work, we plan to enhance our fake news detection system by incorporating Graph Convolutional Networks (GCN) and Graph Neural Networks (GNN) for visual dataset. This will involve the creation of knowledge graphs using the SPO triples extracted from our data. By leveraging the power of graph-based algorithms, we aim to capture and analyze the intricate relationships and dependencies among entities in the knowledge base. By employing GCN and GNN, we anticipate gaining deeper insights into the complex network of fake news propagation and detection. These models have the potential to uncover hidden patterns, identify influential entities, and enhance the overall understanding of information flow within the knowledge base. Our future work will focus on integrating these same advanced graph-based techniques and evaluating their effectiveness in enhancing the accuracy on images dataset and robustness of our fake news detection system.

Acknowledgement: I am grateful to my supervisor for their regular and continued support and insightful guidance throughout this research. Their expertise in AI helps in depth and quality of our investigation. I am also grateful to Colleagues for their collaborative efforts and insightful discussions that greatly contributed to the success of this research.

Funding: There is no funding for this project

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: Data will be available on request

Conflicts of Interest: The authors declare no conflicts of interest.

References

Bondielli, A., Dell'Oglio, P., Lenci, A., Marcelloni, F., & Passaro, L. (2024). Dataset for multimodal fake news detection and verification tasks. *Data in Brief*, *54*, 110440. https://doi.org/10.1016/j.dib.2024.110440

D'Ulizia, A., D'Andrea, A., Grifoni, P., & Ferri, F. (2024). Analysis, Evaluation, and Future Directions on Multimodal Deception Detection. *Technologies*, *12*(5). https://doi.org/10.3390/technologies12050071

Ellam, I. V., Okorie, K. M., & Okebanama, U. F. (2025). Fake News Detection System Using Natural Language Processing: An Optimized Approach. *European Journal of Applied Science, Engineering and Technology*, *3*(2), 162–184. https://doi.org/10.59324/ejaset.2025.3(2).15

Kumari, S., & Singh, M. P. (2024). A Deep Learning Multimodal Framework for Fake News Detection. *Engineering, Technology and Applied Science Research*, *14*(5), 16527–16533. https://doi.org/10.48084/etasr.8170

Li, J., Jiang, W., Zhang, J., Shao, Y., & Zhu, W. (2024). Fake User Detection Based on Multi-Model Joint Representation. *Information (Switzerland)*, *15*(5). https://doi.org/10.3390/info15050266

Liu, Y., Liu, Y., Li, Z., Yao, R., Zhang, Y., & Wang, D. (2025). Modality Interactive Mixture-of-Experts for Fake News Detection. *WWW 2025 - Proceedings of the ACM Web Conference*, 5139–5150. https://doi.org/10.1145/3696410.3714522

Naeem, J., Gul, O. M., Parlak, I. B., Karpouzis, K., Salman, Y. B., & Kadry, S. N. (2024). Detection of Misinformation Related to Pandemic Diseases using Machine Learning Techniques in Social Media Platforms. *EAI Endorsed Transactions on Pervasive Health and Technology*, *10*, 1–20. https://doi.org/10.4108/eetpht.10.6459

Nair, V., Pareek, D. J., & Bhatt, S. (2024). A Knowledge-Based Deep Learning Approach for Automatic Fake News Detection using BERT on Twitter. *Procedia Computer Science*, *235*(2023), 1870–1882.

https://doi.org/10.1016/j.procs.2024.04.178

Su, X., Yang, J., Wu, J., & Qiu, Z. (2025). Hy-DeFake: Hypergraph neural networks for detecting fake news in online social networks. *Neural Networks*, *187*(March), 107302. https://doi.org/10.1016/j.neunet.2025.107302

Vineela, A., Bhavani, A., Krishna, B. V., & Sankar, A. B. (2025). An artful multimodal exploration in discerning fake news through text and image harmony. *Multimedia Tools and Applications*, *0123456789*. https://doi.org/10.1007/s11042-025-20695-4

Wang, J., Zhu, Z., Liu, C., Li, R., & Wu, X. (2024). LLM-Enhanced multimodal detection of fake news. *PLoS ONE*, *19*(10 October), 1–21. https://doi.org/10.1371/journal.pone.0312240

Yadav, A., & Gupta, A. (2024). An emotion-driven, transformer-based network for multimodal fake news detection. *International Journal of Multimedia Information Retrieval*, *13*(1), 1–16. https://doi.org/10.1007/s13735-023-00315-3