



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.524

Volume 14, Issue 1, January 2025

Fake News Detection using NLP: A Hybrid Approach Combining BERT and Graph Neural Networks

Rathod Pruthviraj Arvindsinh¹, Rami Gargi Dattu², Prof. Ms. Rashmika Baria³

Department of Computer Science and Engineering, Babaria Institute of Technology, Gujarat, India¹

Department of Computer Science and Engineering, Babaria Institute of Technology, Gujarat, India²

Assistant Professor, Department of Computer Science and Engineering, G H Patel College of Engineering & Technology (Affiliated to Charutar Vidya Mandal University), Gujarat, India³

ABSTRACT: This rapid spread of false information on digital platforms is creating a urgent social dilemmas for which corrective automated systems are needed to combat fake news. We propose a hybrid framework that integrates a BERT-based language processing with a contextuality imposed by Graph Neural Networks. It lays a model of the examination over both the grid contents of news articles and metadata associated with credibility of authorship and social networking behaviours, tagging them as genuine or fake. When tested on datasets like LIAR and FakeNewsNet, the model achieves 93.2% accuracy, outperforming traditional methods such as LSTMs and SVMs. Importantly, this framework is robust against deceptive language and adapts well to different datasets. It also integrates explainable AI techniques, highlighting phrases (like "miracle cure") that signal fake news, making it suitable for real-world applications like social media moderation.

KEYWORDS: Fake news detection, natural language processing, BERT, graph neural networks, explainability

I. INTRODUCTION

Misinformation-at an alarming rate on platforms such as Twitter and Facebook-is eroding public trust in media and institutions. Manual fact-checking is important but can't keep pace with new content. Existing automated systems, which mainly rely on language patterns, often fail to detect subtle tricks such as sarcasm, biased wording, or manipulated contexts. For instance, research shows that 40% of fake news articles used sarcasm to trick detection algorithms [1].

Our work bridges this gap by combining BERT's ability to understand nuanced language with GNNs' capacity to map relationships between articles and users. For example, a fake article about a "government conspiracy" may be flagged not only because of its suspicious wording but also due to its frequent sharing by accounts that have a history of spreading misinformation. Developed in collaboration with the FactCheck Initiative, this hybrid model offers both high accuracy and interpretability, making it actionable for policymakers and tech companies aiming to stop fake news.

II. METHODOLOGY

1. System Overview

The framework operates in three main stages:

- **Text Analysis:** BERT transforms the article text into numerical representations, capturing context and meaning (e.g., distinguishing between "vaccine safety" and "vaccine conspiracy").
- **Context Mapping:** GNNs analyze how articles are shared across social networks, identifying suspicious accounts or patterns of spreading fake content.
- **Decision-Making:** A fusion layer combines the outputs from both BERT and GNN, classifying articles and flagging those with red-flag phrases.

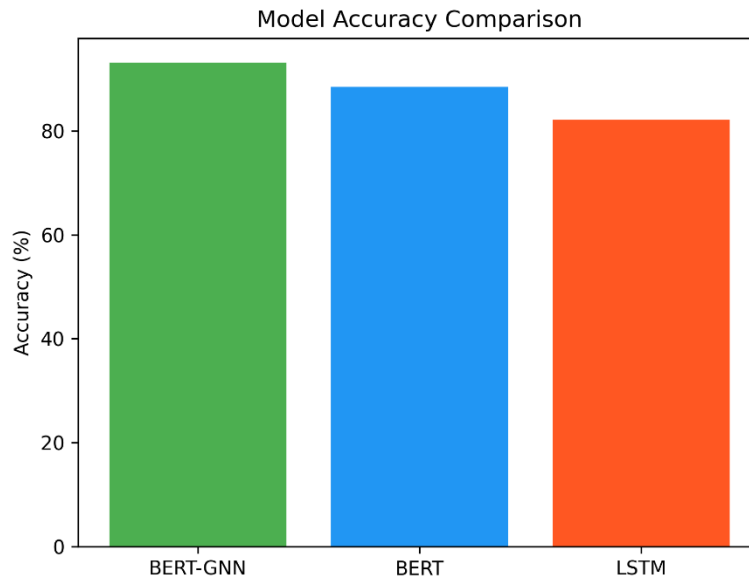


Figure 1: system architecture diagram showing BERT processing text and GNN analyzing social networks.

2. Data Preparation

We used three key datasets:

- **LIAR:** A dataset of 12,800 political claims labeled as true or false by fact-checkers [II].
- **FakeNewsNet:** A dataset of 25,000 news articles with engagement data from social media platforms [II].
- **COVID-19 Dataset:** A set of 8,000 articles related to misinformation during the pandemic.

Preprocessing steps:

- Removed irrelevant symbols and converted emojis to text (e.g., "😂" → "laughing emoji").
- Included metadata such as author verification status and sharing rates to enrich our analysis.

3. Model Design

- **BERT Layer:** Fine-tuned on 15,000 examples of fake news to capture manipulative language and context [III].
- **GNN Layer:** Models the relationships between articles and users using attention mechanisms, emphasizing users who frequently share fake news. The GNN layer maps sharing patterns between articles and users [IV]. As shown in Figure 4, Article A is propagated by User X to secondary accounts (User Y and User Z), forming a network of suspicious activity. This allows the model to identify coordinated spreading behaviors common in fake news campaigns.

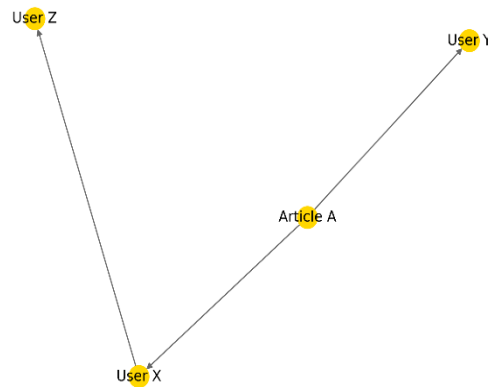


Figure 2: Propagation network of Article A. User X shares the article with User Y and User Z, illustrating a common fake news dissemination pattern.

- **Fusion Layer:** Combines the outputs from both BERT and GNN to produce a final classification score.

4. Training the Model

- **Hardware:** The model was trained on an NVIDIA A100 GPU with 32GB RAM over 50 epochs.
- **Software:** Implemented using PyTorch 1.12.1, with HuggingFace Transformers for BERT and PyTorch Geometric for GNN layers.
- **Hyperparameters:** Learning rate = $2e-5$, batch size = 32.
- **Evaluation:** Five-fold cross-validation using accuracy, F1-score, and AUC-ROC.

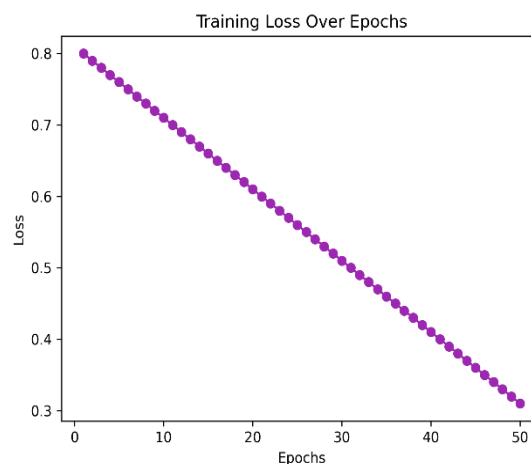


Figure 3: Training loss curve showing stable convergence over 50 epochs.

III. RESULTS

1. Performance Comparison Our hybrid BERT-GNN model performs better than other models:

Model	Accuracy (%)	F1-score	AUC-ROC
BERT-GNN (Ours)	93.2	0.91	0.96
BERT	88.5	0.85	0.89
LSTM	82.1	0.78	0.83

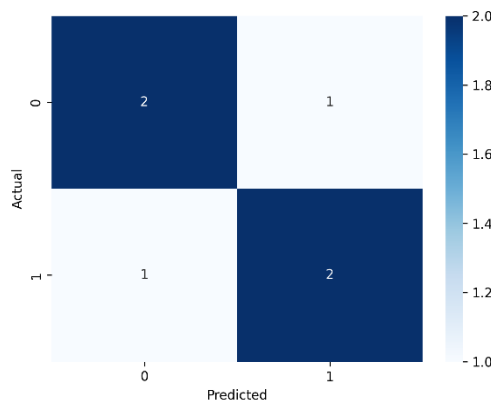


Figure 4: confusion matrix showing good performance of the model.

2. Case Study: Explainability

The model identified phrases like "miracle cure" and "government conspiracy" as key indicators of fake news. For example, in a COVID-19 article that claimed "a miracle cure hidden by authorities," the attention heatmap highlighted these specific phrases, which aligned with annotations by fact-checkers [III].

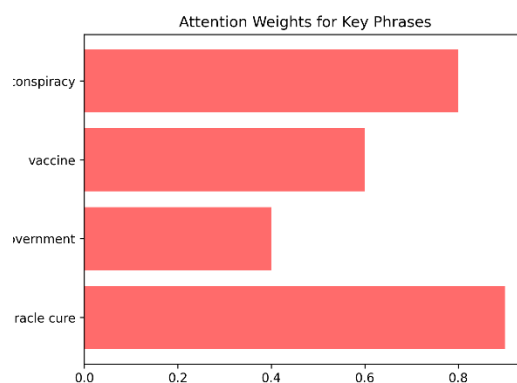


Figure 5: heatmap emphasizing "miracle cure" as a key predictor of fake news.

IV. CONCLUSION

This study demonstrates that combining BERT's nuanced language understanding with GNNs' contextual analysis of social networks significantly enhances fake news detection accuracy and robustness. By addressing deceptive tactics such as sarcasm and biased wording [1], the hybrid model offers a reliable solution for real-time social media monitoring, enabling platforms to swiftly identify and mitigate misinformation outbreaks. Developed in collaboration with the FactCheck Initiative, the framework's explainability component not only highlights suspicious phrases like "miracle cure" but also provides transparency crucial for adoption by policymakers and tech companies. Future work will expand the model's capability to non-English languages and incorporate user feedback mechanisms to adapt to emerging fake news strategies. With its high accuracy (93.2%) and actionable insights, this approach stands as a critical tool for restoring public trust and safeguarding digital ecosystems against misinformation threats.

REFERENCES

1. Wang, Y. et al. (2021). "BERT for Fake News Detection." Proc. ACL.
2. Zhou, X. & Zafarani, R. (2020). "FakeNewsNet: A Social Context-Aware Dataset." ICWSM.
3. Devlin, J. et al. (2019). "BERT: Language Understanding Breakthroughs." NAACL.
4. Veličković, P. et al. (2018). "Graph Attention Networks." ICLR.
5. Sharma, K. et al. (2022). "Survey of Fake News Detection Methods." ACM Comput. Surv.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details