Fake News Detection Using Natural Language Processing (NLP)

Mahesh Maharjan[a] and Suman Mahato[b]

[a] Department of Computer Science and Information Technology, Amrit Campus, Tribhuvan University
Kathmandu, Nepal
E-mail: maheshmanmaharjan@gmail.com

[b] International School of Management and Technology, University of Sunderland
Kathmandu, Nepal
E-mail: kit23g.sum@ismt.edu.np

DOI: https://doi.org/10.3126/mvicjmit.v1i2.85885

Abstract

Fake news has become one of the most critical challenges in today’s information-driven world. Social media, online news platforms, and instant messaging apps make it easy for misinformation to spread rapidly, often with serious consequences for politics, public health, and society. This report examines how Natural Language Processing (NLP) techniques, supported by Machine Learning (ML) and Deep Learning (DL), can be used to automatically detect fake news. A literature-based review highlights the effectiveness of models such as SVM, Random Forest, LSTM, Bi-LSTM, GRU, and CNN. The study also explores feature extraction techniques like TF-IDF, Word2Vec, and GloVe, alongside the role of dataset diversity and multilingual contexts. A real case study of fake news during the COVID-19 pandemic is discussed to show the real-world impact of misinformation. The findings suggest that while ML models provide efficient solutions, DL approaches offer superior accuracy and contextual understanding, and challenges remain in reproducibility, bias, and multilingual detection.

Keywords: Fake News, Natural Language Processing, Machine Learning, Deep Learning, Misinformation

1. Introduction

In the digital age, information is more accessible than ever before. Social media platforms, online news outlets, and instant messaging applications have become the primary sources of news for millions of people around the world. However, alongside credible reporting, the internet has also become a breeding ground for misinformation and disinformation, more commonly known as fake news**.**

Fake news refers to deliberately false or misleading information presented as legitimate news. Its impact is far-reaching: it can shape political outcomes, influence public opinion, damage reputations, and even create health crises. During the COVID-19 pandemic, for instance, the rapid spread of false medical claims contributed to confusion and panic among the public (Vadakkethil et al., 2024).

Detecting fake news manually is impractical given the vast amount of content generated daily. As a result, researchers have turned toward automation using Natural Language Processing (NLP)**.** NLP provides computational methods to process and analyze human language, allowing systems to identify linguistic patterns, semantic cues, and textual features that may indicate whether a piece of news is fake or real.

Traditional machine learning (ML) techniques such as Support Vector Machines (SVM) and Random Forests have shown promise in classifying news articles. However, with the rise of deep learning (DL) models such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), systems have gained the ability to capture deeper contextual and semantic relationships. Despite these advances, challenges remain, including dataset limitations, multilingual applicability, reproducibility, and ethical concerns like algorithmic bias.

2. Aim and Objectives

The aim of this research is to critically investigate NLP techniques for automated fake news detection, analyzing the strengths, limitations, and best practices of ML and DL approaches, with the goal of improving the accuracy, reliability, and robustness of misinformation classification. The research objectives are as follows:

i. To systematically review recent research on fake news detection using NLP, ML, and DL.
ii. To compare the strengths and weaknesses of machine learning and deep learning models.
iii. To analyze different feature extraction and representation methods such as TF-IDF, Word2Vec, and GloVe.
iv. To evaluate the impact of dataset size, diversity, and multilingual contexts on model performance.
v. To explore one real-world case where fake news had a major social impact.
vi. To identify current challenges such as reproducibility, dataset bias, and adversarial manipulation.

3. Literature Review

Fake news detection has been studied widely in recent years, and researchers from computer science, linguistics, and social sciences have proposed different approaches to tackle the problem. The literature shows a clear evolution from early machine learning approaches toward more sophisticated deep learning models. Alongside this, much attention has been paid to feature extraction techniques, dataset challenges, and ethical issues.

3.1 Machine Learning Approaches

Early studies on fake news detection relied heavily on traditional machine learning (ML) models. These methods typically involved manually extracting features from text, such as word frequency or sentiment, and then training classifiers to distinguish fake from real news.

For example, Alomari et al. (2022) tested several ML classifiers, including Support Vector Machines (SVM) and Random Forests (RF), achieving accuracies of up to 98% on benchmark datasets. SVM, in particular, performed strongly when combined with TF-IDF (Term Frequency–Inverse Document Frequency), which converts text into weighted numerical values. Similarly, Rathore et al. (2024) conducted a comparative analysis of SVM, RF, and Logistic Regression (LR). Their results showed that SVM with TF-IDF achieved the highest accuracy of 99.13%, outperforming other ML models.

These findings suggest that ML methods are effective for structured, English-only datasets and require less computational power compared to deep learning. However, they often fail to capture contextual nuances in language, such as sarcasm or subtle wordplay, and tend to struggle with domain adaptation when applied to different types of data.

3.2 Deep Learning Approaches

As artificial intelligence advanced, researchers began exploring deep learning (DL) models, which can automatically learn features from raw text without heavy manual preprocessing. DL techniques are now widely used because they can model complex linguistic patterns and contextual relationships.

One popular approach is the use of Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks. Ushashree et al. (2023) applied LSTM to hoax news detection and reported that sequential models outperform ML approaches by capturing temporal and contextual dependencies. Bidirectional LSTM (Bi-LSTM) models go a step further, analyzing text from both directions, and have been shown to achieve even higher accuracy. Kadhar and Rajakumari (2025) found that Bi-LSTM outperformed SVM and Random Forest in terms of precision and recall, reaching close to 99% accuracy.

Deep learning technique applied to text is Convolutional Neural Networks (CNNs). Although originally designed for image recognition, CNNs are effective in capturing local patterns in text, such as recurring phrases or word combinations that are common in fake news. Kong et al. (2020) showed that CNNs combined with preprocessing techniques like tokenization and lemmatization achieved competitive performance.

Hybrid models that combine CNN, GRU (Gated Recurrent Units), and embeddings have been tested. Keya et al. (2021) demonstrated that using pre-trained embeddings such as Word2Vec and GloVe with hybrid deep learning architectures significantly boosts accuracy, particularly in low-resource and multilingual settings.

3.3 Feature Extraction and Representation

A key factor in the success of fake news detection systems is how textual data is represented for the models. The simplest and still widely used approach is TF-IDF, which measures the importance of words in a document relative to a collection. While TF-IDF works well with ML models, it does not capture the deeper meaning of words.

Researchers have turned to word embeddings, which map words into dense vector spaces where semantic relationships are preserved. Word2Vec and GloVe are popular embedding methods that capture similarities between words (e.g., “king” and “queen”). FastText goes further by considering subword information, which improves performance in morphologically rich languages.Keya et al. (2021) further argued that embeddings help deep learning models capture semantic and syntactic relationships that TF-IDF misses, making them more effective for detecting subtle patterns in fake news.

3.4 Dataset Challenges and Multilingual Contexts

Datasets play a crucial role in evaluating models. Most studies rely on benchmark datasets such as ISOT Fake News, WELFake, and LIAR. While these datasets are useful, they are mostly limited to English, raising concerns about the global applicability of models.

Çoban and Bakal (2025) stressed that multilingual fake news detection remains a major challenge, as most models fail when applied to other languages or cultural contexts. Vadakkethil and Kumar (2024) also highlighted that many models are overfitted to English datasets, and there is limited research into low-resource languages where fake news is equally problematic.

Dataset imbalance is another recurring issue. Fake news datasets often contain fewer fake samples than real ones, leading to bias in model performance. Varma et al. (2025) noted that reproducibility is a concern because many studies use different preprocessing pipelines, metrics, and datasets, making it difficult to compare results fairly.

3.5 Ethical and Social Considerations

Beyond technical performance, several studies have highlighted the ethical implications of automated fake news detection. Bhogi et al. (2023) argued that dataset bias is a major risk, as models trained on biased data may unfairly misclassify certain types of content. Privacy concerns also arise when using user-generated content from social media platforms.

Furthermore, there is a risk that detection systems could be misused for censorship or political purposes. For this reason, researchers emphasize the need for transparency, fairness, and accountability in designing such systems.

4. Real Case Study: COVID-19 Fake News

During the COVID-19 pandemic**,** fake news spread rapidly across social media platforms, often faster than official health information. For instance, false claims such as “drinking hot water can kill the virus” or “5G towers spread COVID-19" went viral worldwide. According to Vadakkethil et al. (2024), such misinformation created confusion, encouraged harmful behaviors, and undermined trust in public health authorities. In some regions, this even led to violent acts, such as the destruction of 5G towers.

Researchers quickly applied NLP-based detection models to address this. For example, Bhogi et al. (2023) applied ML methods to social media text during COVID-19 and found that SVM combined with TF-IDF achieved strong results in detecting pandemic-related fake news. However, they also highlighted the limitation of detecting misinformation in real-time, as fake stories evolved rapidly.

This case shows the real-world importance of automated fake news detection systems and the risks when misinformation spreads unchecked.

6. Results

- ML models such as SVM and Random Forest achieve 95–98% accuracy but lack contextual depth.
- DL models such as Bi-LSTM, GRU, and CNN outperform ML, often achieving 99%+ accuracy.
- Embeddings like Word2Vec and GloVe outperform TF-IDF by capturing semantic meaning.
- Dataset diversity improves robustness, while reliance on a single dataset limits generalization.
- Multilingual fake news detection remains an underexplored research area.

Table 1: Model Performance Comparison

Model	**Accuracy (%)**
SVM (ML)	~97.5
Random Forest (ML)	~96.8
Logistic Regression	~95.0
LSTM (DL)	~98.5
Bi-LSTM (DL)	~99.2
GRU (DL)	~99.0
CNN (DL)	~98.8

The bar graph highlights that deep learning models (Bi-LSTM, GRU, CNN, LSTM) generally outperform traditional machine learning models (SVM, Random Forest, Logistic Regression)**,** with Bi-LSTM achieving the highest accuracy (~99.2%).

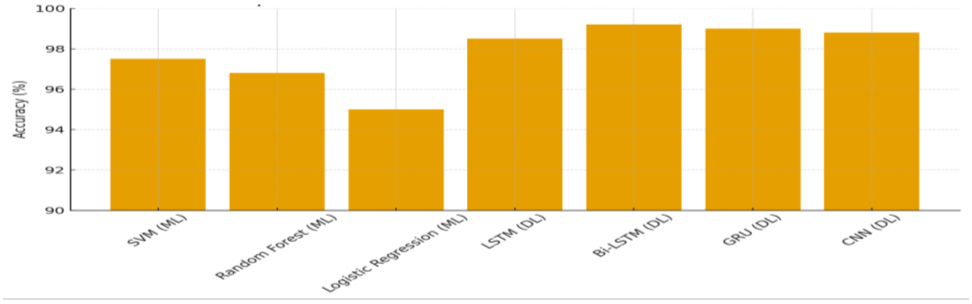

Figure 1:Accuracy comparison of ML vs DL models for fake news detection

7. Conclusion

This study confirms that NLP techniques are highly effective in detecting fake news. While machine learning methods remain strong baselines, deep learning models deliver higher accuracy and contextual understanding. The real case of COVID-19 misinformation shows the urgency of implementing robust and scalable detection systems.

However, challenges remain in terms of reproducibility, dataset bias, adversarial robustness, and multilingual applicability. Future research should focus on building diverse datasets, improving real-time detection, and ensuring ethical design to prevent misuse. Developing transparent, fair, and reliable systems is essential for restoring trust in digital information ecosystems.

References

Alomari, D. M., Aboulnour, M., & Aljabri, M. (2022). Fake news detection using machine learning models. *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE. https://doi.org/10.1109/CICN56167.2022.10008340

Bhogi, A., Dasari, A., Garg, N., & Sharma, N. (2023). Machine learning for fake news detection on social media text. *2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*. IEEE. https://doi.org/10.1109/ICAICCIT60255.2023.10465880

Çoban, M. K., & Bakal, G. (2025). NLP-driven fake news detection: A machine learning perspective. *2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA)*. IEEE. https://doi.org/10.1109/ICHORA65333.2025.11017210

Kadhar, S. A. A., & Rajakumari, S. B. (2025). Fake news detection in social media using bidirectional recurrent neural network method. *2025 Global Conference in Emerging Technology (GINOTECH)*, Pune, India. IEEE. https://doi.org/10.1109/GINOTECH63460.2025.11076804

Keya, A. J., Afridi, S., Maria, A. S., Pinki, S. S., Ghosh, J., & Mridha, M. F. (2021). Fake news detection based on deep learning. *2021 International Conference on Science & Contemporary Technologies (ICSCT)*. IEEE. https://doi.org/10.1109/ICSCT53883.2021.9642565

Kong, S. H., Tan, L. M., Gan, K. H., & Samsudin, N. H. (2020). Fake news detection using deep learning. *2020 IEEE Conference Paper*. IEEE. Available at: https://ieeexplore.ieee.org/document/9108841

Rathore, S. P. S., Juneja, A. K., Shaktawat, N., Choudhary, H. D., Anitha, K., & Yadav, S. (2024). Comparative analysis of machine learning models for fake news detection using natural language processing. *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*. IEEE. https://doi.org/10.1109/ICAC2N63387.2024.10895423

Ushashree, P., Naik, A., Gurav, S., Kumar, A., Chethan, S. R., & Madhumala, B. S. (2023). Fake news detection using neural network. *2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS)*. IEEE. https://doi.org/10.1109/ICICACS57338.2023.10100208

Vadakkethil, S. E., & Kumar, V. M. (2024). Leveraging natural language processing for detecting fake news: A comparative analysis. *2024 2nd International Conference on Disruptive Technologies (ICDT)*. IEEE. https://doi.org/10.1109/ICDT61202.2024.10489386

Varma, M. J., Rohit, M. S., & Selvi, G. S. G. (2025). Fake news detection using natural language processing. *2025 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*. IEEE. https://doi.org/10.1109/ICCECE61355.2025.10940549