

A Comparative Study of Machine Learning, Deep Learning, and Transformer-Based Models for Fake News Detection

Harish Chauhan

School Of Computing Science
And Engineering
Galgotias University
Greater Noida, India

harish.23SCSE1010992@galgotiasuniversity.ac.in

Ankit Raj

School Of Computing Science
And Engineering
Galgotias University
Greater Noida, India

ankit.22SCSE1011957@galgotiasuniversity.ac.in

Rishabh Prasad

School Of Computing Science
And Engineering
Galgotias University
Greater Noida, India

Rishabh.prasad@galgotiasuniversity.edu.in

Abstract—The widespread use of digital news platforms and social media has significantly changed how people consume information. While online access to news has improved information availability, it has also increased the circulation of misleading and fabricated content. Fake news spreads rapidly due to sensational headlines and unverified sharing, which can influence public opinion and decision-making. Detecting such misinformation has therefore become a critical challenge. This paper presents a comprehensive text-based fake news detection system using machine learning and deep learning techniques. Multiple classical machine learning models are evaluated using TF-IDF feature representation, including Naive Bayes, Logistic Regression, Support Vector Machines, Random Forest, and SGD Classifier. In addition, deep learning architectures such as Convolutional Neural Networks (CNN), Bidirectional LSTM (BiLSTM), and a fine-tuned BERT transformer model are implemented to capture semantic and contextual information from text. The performance of all models is compared using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Experimental results demonstrate that transformer-based models achieve superior performance due to their contextual understanding, while traditional models provide competitive results with lower computational cost. This study highlights the effectiveness of combining classical and modern NLP approaches for robust misinformation detection.

I. INTRODUCTION

In the modern digital era, news dissemination has shifted largely from traditional media outlets to online platforms and social networks. Information can be published and shared instantly, reaching millions of users within minutes. Although this rapid distribution offers convenience, it also enables the uncontrolled spread of fake or misleading news. Many users consume and forward information without verifying its credibility, which amplifies misinformation across digital ecosystems. Fake news has serious implications, including misinformation during public health crises, political manipulation, social unrest, and damage to public trust. False narratives often rely on emotionally appealing language and misleading headlines, making them difficult to distinguish from authentic news. As a result, automated systems capable of identifying fake news have become increasingly important. Manual verification of online content is neither scalable nor efficient due to the massive volume of textual data generated daily. From a computer science perspective, machine learning and natural language processing (NLP) techniques provide effective tools to analyze and classify large-scale text data. By learning linguistic patterns, word distributions, and contextual relationships, these models can distinguish between real and fake news articles. This research focuses on developing a text-based fake news detection framework that leverages both traditional

machine learning models and advanced deep learning architectures. The objective is to analyze the strengths and limitations of different approaches and identify models that offer a balance between accuracy, efficiency, and scalability. We delineate the main scientific contributions of our study focused on intelligent system engineering as follows:

- We propose
- The proposed model
- The advanced
- it attained an accuracy of 96% and 99% accuracy.

The rest of the paper is organized as follows: [Section 2](#) summarizes related work for Fake News Detection techniques. [Section 3](#) followed by datasets and media processing. [Section 4](#) section delves into the proposed model. The results and analysis are discussed in [Section 5](#). Finally, the conclusion and future direction are in [Section 6](#).

II. LITERATURE REVIEW

[1] previous studies about manipulated content on social media highlight multimodal learning shows that accuracy for manipulated content detection improves by combining text, image instead of unimodal approaches. But these also suffer from poor generalization because these models are trained on small or domain-specific datasets.

[2] On datasets like ISOT, Bidirectional Long Short term memory (BiLSTM) outperforms as traditional machine learning models. BiLSTM is used very frequently in fake news classification because it has ability to learn long-range of dependencies in text. ISOT refers Information Security and Object technology, a research lab of University of Victoria.

In the paper [3], the author says that to detect fake news from social media content, a good way is to use attention mechanisms with deep learning models because this approach focuses only on words that have more semantic importance. Here, the author has used approaches like multi-head attention, co-attention, and graph-aware attention networks. This approach is beneficial both for unimodal and multimodal tasks and helps improve accuracy by capturing contextual knowledge, long-distance dependencies, and relationships between multiple modalities, compared to traditional CNN/LSTM models.

In the paper [4], the author says that there is too much misinformation spread in the Arabic language because there are limited tools that are specific to Arabic. In the early days, the author explored deep learning models like CNN, LSTM, and BiLSTM on Arabic text datasets, and then he found that deep neural networks perform better because they capture semantic

and sequential patterns from text. However, recent studies say that hybrid methods combined with CNN/LSTM show more precise feature extraction and also very good classification results.

The paper [5] states that the combination of attention mechanisms with large language models (LLMs) increases the accuracy significantly because they have a rich contextual understanding of text for misinformation detection.

The paper [6] states that modern deep learning models and transformer-based approaches are highly precise because these approaches capture complex semantic and contextual patterns from the given dataset. Modern studies are also using multimodal approaches that integrate visual and social features. By using context-based embeddings, some other pretrained models like BERT are also adding more accuracy.

In the paper [7], the author says that during a multimodal approach, combining different types of data like text, image, and social context using deep learning models significantly improves accuracy as compared to a unimodal approach.

[8] Recent studies on fake news detection with LLMs show that large language models markedly improve the identification of misinformation by modeling rich contextual and semantic relationships, surpassing traditional machine learning and earlier deep learning techniques. To tackle issues such as nuanced linguistic signals and constantly changing deceptive strategies, researchers have integrated LLMs with attention mechanisms, knowledge graphs, and multimodal information sources. Although these approaches achieve strong performance, persistent challenges—including model bias, scalability limits, lack of transparency, and susceptibility to adversarial attacks—continue to motivate advancements and guide future work toward more resilient and interpretable detection systems.

[9] Existing studies indicate that artificial intelligence (AI) is fundamental to fake news detection, as it enables automated examination of linguistic, semantic, and behavioral patterns in online content and surpasses traditional rule-based approaches. Research employing machine learning, deep learning, and hybrid AI methods reports higher detection accuracy, with recent work focusing on transformer-based models and multimodal systems that integrate text, images, and social context. Nonetheless, persistent challenges—including data quality, model interpretability, and the ability to adapt to continuously changing misinformation strategies—remain central topics for future research on AI-based fake news detection.

[10] Previous studies on fake news detection with machine learning indicate that traditional models such as SVM, Random Forest, Naive Bayes, and Logistic Regression are effective at spotting misleading content by learning patterns from linguistic and metadata features, and they often outperform simple keyword-based approaches. This line of work emphasizes feature engineering methods like TF-IDF, n-grams, and sentiment analysis scores as key to strong model performance, even though these techniques can have difficulty capturing more complex semantic subtleties. Although machine learning methods offer strong baselines and efficient training, they generally fall short of deep learning models in modeling longrange dependencies and contextual meaning, which has led recent research to explore hybrid and neural-based approaches.

III. DATASET AND PREPROCESSING

The dataset used in this study was obtained from a publicly available Kaggle repository containing labeled fake and real news articles. Two separate files representing fake and authentic

news were merged into a single dataset. Each news instance was assigned a binary label, where fake news was labeled as 0 and real news as 1.

A. Preprocessing and Feature Engineering

Text preprocessing was applied to clean and standardize the input data. This included converting text to lowercase, removing unnecessary symbols, and handling tokenization. Stop words were removed to reduce noise, and the processed text was prepared for feature extraction and model input. These steps ensured consistency and improved model learning efficiency.

For classical machine learning models, TF-IDF vectorization was used to transform textual data into numerical feature vectors. Unigrams and bigrams were considered to capture both individual words and short phrase patterns. This representation highlights important terms while reducing the influence of commonly occurring words. For deep learning models, text was tokenized and converted into padded sequences to ensure uniform input length. Word embeddings were learned during training to capture semantic relationships.

IV. PROPOSED METHODOLOGY

This research proposes a multi-model text-based fake news detection framework using machine learning and deep learning techniques. The methodology is structured into sequential stages, including data collection, preprocessing, feature extraction, model training, and evaluation. Text preprocessing was applied to clean and standardize the input data. This included converting text to lowercase, removing unnecessary symbols, and handling tokenization. Stop words were removed to reduce noise, and the processed text was prepared for feature extraction and model input. These steps ensured consistency and improved model learning efficiency.

A. Model Training

Multiple models were trained independently to compare their effectiveness like:

Machine Learning Models:

- Multinomial Naive Bayes
- Logistic Regression
- Linear Support Vector Classifier
- Random Forest Classifier
- SGD Classifier

Deep Learning Models:

- TextCNN
- Bidirectional LSTM

Additionally, a BERT-based transformer model was fine-tuned on the dataset to leverage contextual embeddings from pretrained language models.

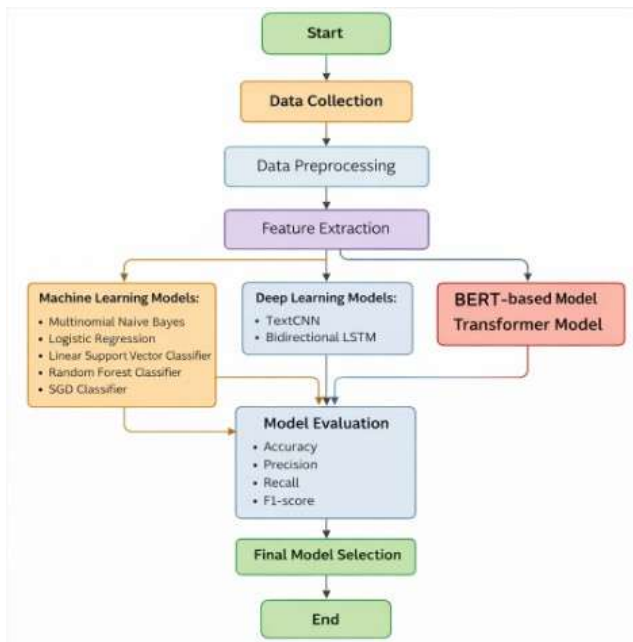


Fig. 1

Multinomial Naive Bayes is widely used for text classification problems as it considers word frequency information. It performs efficiently on TF-IDF-based features and handles high-dimensional sparse data effectively. Due to its strong performance, simplicity, and low computational cost, this model was selected as the final classifier for misinformation detection.

Logistic Regression is a linear classification model that estimates the probability of a given input belonging to a particular class. It performs well on large-scale textual datasets and provides good interpretability for understanding feature importance.

Linear SVC is a margin-based classifier that separates classes by maximizing the distance between decision boundaries. It is highly effective for high-dimensional text data and performs well with TF-IDF representations.

Random Forest is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and reduce overfitting. It captures non-linear relationships in the data and provides robust performance across different datasets.

SGD Classifier is an efficient optimization-based model suitable for large datasets. It is particularly effective for linear classifiers such as SVM and Logistic Regression when trained using stochastic gradient descent. A deep learning model that applies convolutional filters to extract local n-gram features from text. It effectively captures key phrases useful for text classification.

A recurrent neural network that processes text in both forward and backward directions. It captures long-term dependencies and contextual relationships within sequences.

A transformer-based language model that learns deep contextual embeddings using self-attention. It provides superior performance by understanding semantic and contextual meaning in text.

All models were evaluated on the test dataset using accuracy, precision, recall, and F1-score. These metrics provide a comprehensive understanding of model performance, particularly in distinguishing fake news from real news.

Based on experimental evaluation, the BERT-based transformer model was selected as the final classification model. It achieved the highest accuracy, precision, recall, and F1 score

among all evaluated models. BERT effectively captures contextual and semantic information in news text, enabling more accurate discrimination between real and fake news compared to traditional and deep learning models. The finetuned BERT model and its corresponding tokenizer were saved after training. During deployment, incoming news text is tokenized using the same pretrained tokenizer and passed to the trained model for classification as real or fake. This approach ensures consistency between training and inference and allows practical deployment in real-world news verification systems.

V. RESULTS AND ANALYSIS

The experimental results demonstrate notable performance differences among the evaluated models. Classical machine learning models using TF-IDF features achieved strong baseline results, with Logistic Regression and Linear SVC performing consistently well. These models benefited from high-dimensional sparse representations and efficient optimization. Deep learning models showed improved performance by learning semantic representations directly from data. The BiLSTM model outperformed CNN-based approaches by effectively capturing sequential context within news articles.

The BERT model achieved the highest overall performance across all evaluation metrics. Its ability to understand contextual meaning and long-range dependencies enabled more accurate classification of complex news content. However, this improvement comes at the cost of higher computational complexity.

The comparative analysis confirms that while transformer-based models provide superior accuracy, traditional and deep learning models remain valuable due to their efficiency and lower resource requirements.

The comparative evaluation of all implemented models highlights clear differences in their ability to detect fake news from textual data. Classical machine learning models trained using TF-IDF features, such as Logistic Regression, Linear SVC, and SGD Classifier, demonstrated strong and consistent performance, confirming their effectiveness for high-dimensional text classification tasks. Probabilistic models like Multinomial Naive Bayes achieved reliable results with low computational cost, making them suitable for large-scale applications.

Deep learning models, including TextCNN and BiLSTM, showed improved performance by capturing semantic and sequential patterns in text, with BiLSTM outperforming convolution-based approaches. The BERT-based model achieved the best overall performance across all evaluation metrics. Its contextual representation of text enabled more accurate classification compared to both traditional and deep learning models.

Based on the experimental results, the BERT-based transformer model was selected as the final model for fake news classification. It consistently achieved the highest accuracy and F1-score compared to traditional machine learning and deep learning models. BERT's ability to understand contextual and semantic relationships within news text allows it to identify subtle patterns commonly present in misleading content. Although it requires higher computational resources, its reliable and robust performance makes it the most suitable choice for accurate fake news detection in this project.

VI. CONCLUSION

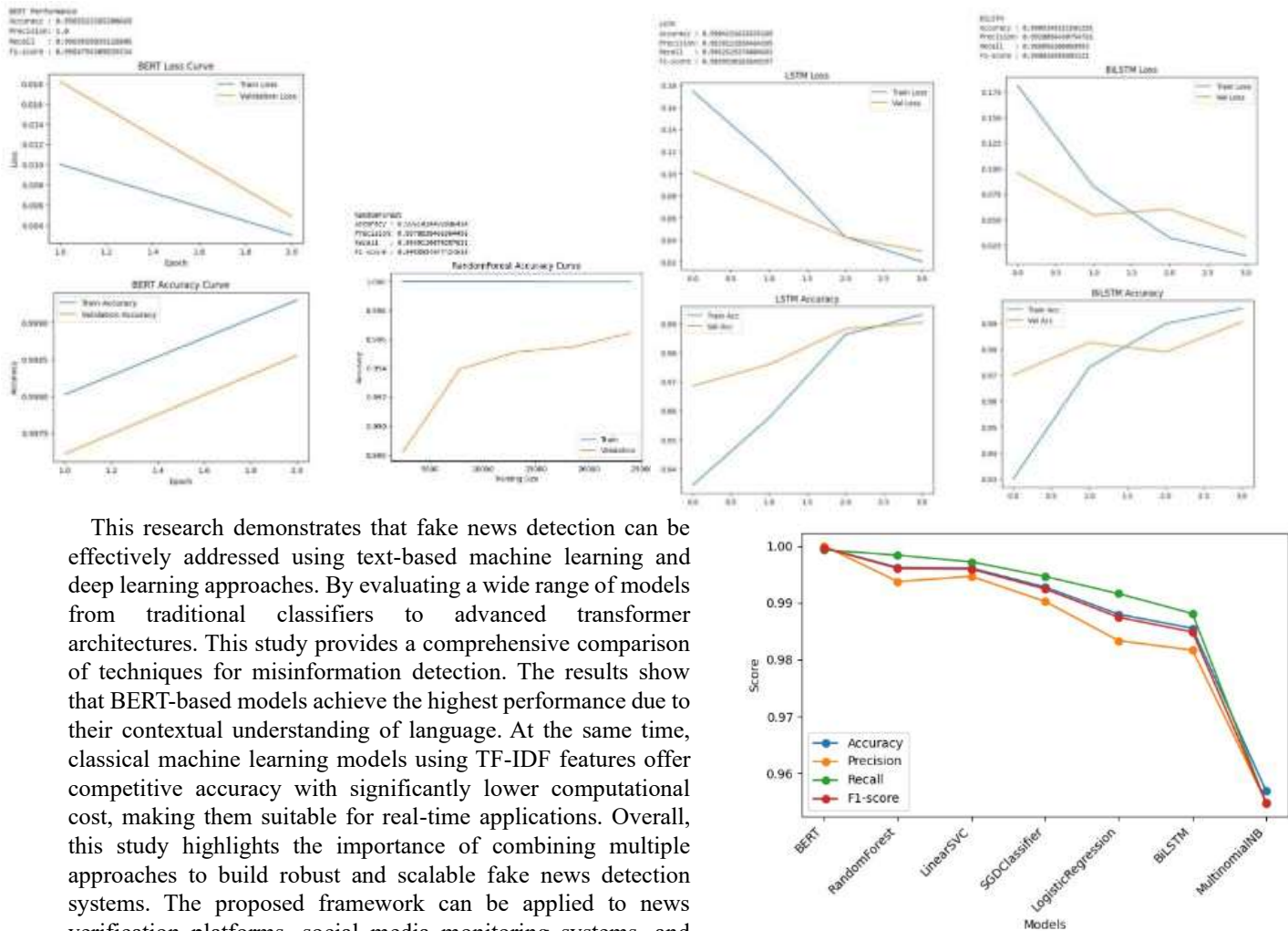


Fig. 3: Comparative Performance of ML/DL models

Fig. 4

time deployment strategies to further enhance misinformation detection.

REFERENCES

- [1] O. Prakash and R. Kumar, "A unified generalization enabled ml architecture for manipulated multi-modal social media," *Multimedia Tools and Applications*, vol. 83, no. 8, pp. 22749–22771, 2024.
- [2] M. M. Lawal and A. Abdulrauf, "Fake news detection using bi-lstm architecture: A deep learning approach on the isot dataset," *Journal of Computing Theories and Applications*, vol. 3, no. 2, pp. 104–114, 2025.
- [3] O. Prakash and R. Kumar, "Fake news detection in social networks using attention mechanism," in *Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 2*. Springer, 2023, pp. 453–462.
- [4] K. M. Fouad, S. F. Sabbeh, and W. Medhat, "Arabic fake news detection using deep learning," *Computers, Materials & Continua*, vol. 71, no. 2, 2022.
- [5] X. Xu, P. Yu, Z. Xu, and J. Wang, "A hybrid attention framework for fake news detection with large language models," in *2025 5th International Conference on Neural Networks, Information and Communication Engineering (NNICE)*, 2025, pp. 587–590.
- [6] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar, and M. S. Rahman, "A comprehensive review on fake news detection with deep learning," *IEEE Access*, vol. 9, pp. 156151–156170, 2021.
- [7] M. Nasser, N. I. Arshad, A. Ali, H. Alhussian, F. Saeed, A. Da'u, and I. Nafea, "A systematic review of multimodal fake news detection on social media using deep learning models," *Results in Engineering*, vol. 26, p. 104752, 2025.
- [8] J. Yi, Z. Xu, T. Huang, and P. Yu, *Challenges and Innovations in LLM-*

2025, p. 87–93. [Online]. Available: <https://doi.org/10.1145/3728725.3728739>

- [9] A. Iqbal, K. Shahzad, S. A. Khan, and M. S. Chaudhry, “The relationship of artificial intelligence (ai) with fake news detection (fnd): a systematic literature review,” *Global Knowledge, Memory and Communication*, vol. 74, no. 5-6, pp. 1617–1637, 10 2023. [Online]. Available: <https://doi.org/10.1108/GKMC-07-2023-0264>
- [10] Z. Khanam, B. N. Alwasel, H. Sirafi, and M. Rashid, “Fake news detection using machine learning approaches,” *IOP Conference Series: Materials Science and Engineering*, vol. 1099, no. 1, p. 012040, mar 2021. [Online]. Available: <https://doi.org/10.1088/1757-899X/1099/1/012040>