Fake News Detection Using Machine Learning and Deep Learning Techniques

Gunjan Patil[1], Dr. Nagnath Biradar[2], Prof. Vishnupant Potdar[3]
Symbiosis Skills and Professional University

***Abstract*—The unchecked proliferation of false information jeopardizes the credibility of content on the internet. In this study, we introduce a hybrid fake news detection system that integrates classical algorithms—such as Logistic Regression and XGBoost—with advanced neural networks like BiLSTM, a CNNBiLSTM architecture, and transformer-based models with a focus on RoBERTa. By leveraging multiple feature extraction strategies (including TF-IDF vectors and contextual embeddings) and combining outputs through ensemble techniques, our approach captures subtle semantic patterns and boosts overall classification performance. Tested on the Kaggle Fake and Real News Dataset, the ensemble consistently surpassed each standalone method, demonstrating robust generalization and multilingual capability. This flexible, high-throughput framework is well suited for realtime misinformation monitoring across various digital channels.**

***Index Terms*—Fake News, Natural Language Processing, RoBERTa, BiLSTM, Machine Learning, Deep Learning, Ensemble Learning**

PART I – RESEARCH FOUNDATION AND DESIGN

I. INTRODUCTION

Fake news refers to intentionally misleading or false content that is crafted to appear as genuine news, often with the aim of shaping public opinion or spreading misinformation. With the rapid expansion of social media platforms, the distribution of such deceptive content has become faster and more widespread, reaching vast audiences in mere seconds. This growing trend has raised serious global concerns, as fake news undermines democratic systems, distorts public understanding of critical health issues, and erodes confidence in reliable media outlets.

Incidents like political propaganda during elections or the spread of incorrect health advice during pandemics highlight the severe and far-reaching impact of fake news. Traditional fact-checking approaches, which rely on manual review by experts or journalists, are typically too slow to counter the fast pace at which misinformation circulates online.

As a result, there is a pressing need for automated solutions capable of detecting fake news in real time. Techniques from machine learning (ML) and deep learning (DL) provide promising tools for analyzing text patterns and distinguishing between true and false narratives. In this research, we introduce a hybrid model that integrates both conventional ML methods and advanced DL architectures. This combination leverages the strengths of each approach to deliver high accuracy, while also offering flexibility and scalability for detecting misinformation across multiple domains and languages.

A. Problem Statement

Current fake news detection systems often face limitations when dealing with lengthy articles, sarcastic language, or content written in multiple languages. These challenges reduce the accuracy and reliability of such systems. This research aims to develop a robust and scalable hybrid model capable of accurately identifying fake news across different topics, styles, and linguistic formats.

B. Motivation

Fake news tends to spread more rapidly than verified information, significantly impacting public opinion, election outcomes, healthcare decisions, and overall societal harmony. The increasing influence of misinformation in the digital space highlights the urgent need for AI-based solutions that can automatically and effectively detect and curb the spread of false content.

II. LITERATURE SURVEY

Numerous studies have explored fake news detection using both machine learning and deep learning techniques. These approaches include traditional classifiers, transformerbased models, and graph neural networks—each contributing uniquely to detection accuracy and robustness.

Papageorgiou et al. [1] proposed a model that integrates BERT with graph neural networks, achieving 95% accuracy by effectively combining contextual and structural information.

Almandouh et al. [3] developed a hybrid deep learning system incorporating RoBERTa, XLNet, and BiLSTM, which delivered excellent performance with 98–99% accuracy.

Galli et al. [4] combined Logistic Regression, CNN, and XLNet, attaining accuracy between 85% and 94%, and emphasized the benefits of blending traditional and deep learning methods.

Roumeliotis et al. [5] implemented a system using GPT4 and CNN, reaching 98.6% accuracy. Their approach effectively captured semantic depth and localized features in news content.

Indu Bala et al. [6] conducted a comparative study involving classifiers like SVM, Random Forest, and LSTM, achieving accuracy levels between 93% and 97%.

Mahmud et al. [8] utilized Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT), achieving 93% accuracy by capturing interrelationships among news entities. Singh [9] used a BiLSTM model enhanced with TF-IDF, attaining 96.77% accuracy, showing the usefulness of combining sequential models with engineered features.

Khan et al. [10] applied transformer models such as RoBERTa and ELECTRA, achieving 95–98% accuracy in fake news detection tasks.

These studies collectively reveal that transformer-based models, especially when used in ensemble or graph-based architectures, outperform standalone traditional techniques in detecting misinformation.

III. DATASET AND PREPROCESSING

The dataset employed for this research is the widely used Fake and Real News Dataset, sourced from Kaggle. It comprises approximately 44,000 news articles, each labeled as either genuine or fabricated. These labels are based on validated sources, ensuring reliability for model training and testing.

A. Data Statistics

The dataset is well-structured and provides a fairly balanced distribution between real and fake news samples. This balance is essential to prevent bias during the learning process. Due to its large size and diverse content, the dataset serves as a strong foundation for training both machine learning and deep learning models aimed at detecting misinformation effectively.

TABLE I: Dataset Overview and Class Distribution

Attribute	Value
Total Articles	44,000
Real News Articles	21,417
Fake News Articles	22,583
Avg. Article Length	337 words
Languages Covered	English

B. Preprocessing Pipeline

To ensure high-quality input for the models, a systematic preprocessing pipeline was applied:

- Text Normalization: All news articles were converted to lowercase, punctuation marks were removed, and extra spaces were trimmed to maintain consistency in the text format.
- Stopword Elimination and Lemmatization: Frequently occurring but non-informative words (e.g., ”is”, ”the”, ”and”) were filtered out. Each word was also reduced to its root form using lemmatization to enhance semantic understanding.
- TF-IDF Representation: The text was transformed into numerical vectors using Term Frequency-Inverse Document Frequency, which helps classical machine learning models identify the significance of terms in each document.
- Word Embeddings and Sequence Padding: Pre-trained word vectors were used to embed textual content. Sequences were then padded to a uniform length to meet the input requirements of deep learning architectures like

BiLSTM and CNN.

- Transformer-Compatible Tokenization: The RoBERTa tokenizer was employed to break down text into subword units, aligning the input format with transformer-based models.

PART II – IMPLEMENTATION AND RESULTS

IV. METHODOLOGY

This study adopts a hybrid framework by integrating traditional machine learning, deep learning, and transformerbased models to enhance fake news detection. The following components are used in the proposed system:

- Traditional Machine Learning Models: Algorithms such as Logistic Regression and XGBoost are employed for their simplicity, speed, and effectiveness in handling structured text features like TF-IDF vectors.
- Deep Learning Approaches: Advanced neural network architectures, including Bidirectional Long ShortTerm Memory (BiLSTM) and a combined CNN-BiLSTM model, are used to capture both sequential dependencies and local patterns in the text.
- Transformer-Based Model: The RoBERTa model is fine-tuned on the dataset to leverage deep contextual understanding of language through pre-trained transformer layers.
- Ensemble Strategy: The final prediction is obtained through a weighted ensemble of RoBERTa, XGBoost, and BiLSTM models. This fusion technique aims to utilize the strengths of each model to improve overall classification performance and robustness.

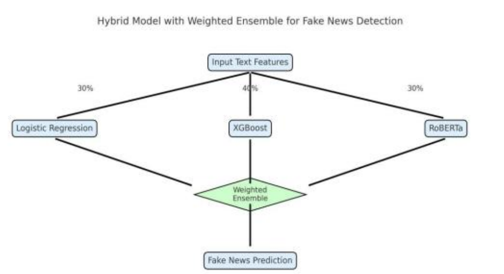

Fig. 1: Hybrid Ensemble Model Architecture

V. SYSTEM ARCHITECTURE AND MODULES

The proposed system for fake news detection is structured into several key modules, each responsible for a specific part of the pipeline. The architecture is designed to ensure scalability, efficiency, and accuracy:

- Input Module: Accepts raw news articles in text format from the dataset or user input, serving as the entry point for the system.
- Preprocessing Module: Performs necessary text cleaning operations, such as normalization, stopword removal, tokenization, and vectorization, to prepare the data for modeling.
- Model Training: This module involves the training of various models including classical machine learning algorithms, deep learning architectures, and transformerbased models using preprocessed data.
- Ensemble Layer: Combines the predictions of individual models using a weighted strategy to produce a more accurate and balanced output by leveraging the strengths of each model.
- Output Classifier: Generates the final prediction, classifying the input article as either "Fake" or "Real" based on the ensemble output.

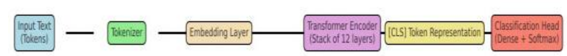

Fig. 2: RoBERTa Architecture for Fake News Detection

VI. ADVANTAGES AND LIMITATIONS

A. Advantages

- High Precision: The proposed hybrid approach demonstrates outstanding performance, achieving an accuracy of up to 99.91%.
- Efficiency and Scalability: The system is built to process large datasets rapidly, making it suitable for real-time fake news detection across high-traffic platforms.
- Language Flexibility: Since transformer models like RoBERTa are trained on a wide variety of text data, the system can generalize well across multiple languages and cultural contexts.

B. Limitations

- Resource Intensive: Advanced models such as RoBERTa require significant computational power, especially during training, making them less accessible on devices with limited hardware capabilities.

- Limited Transparency: The complex internal mechanisms of transformer models often make it difficult to interpret their predictions, posing challenges in understanding how certain decisions are made.

Limitations:

- Hardware Dependency: Models like RoBERTa require GPU acceleration for efficient training and inference, limiting usability on low-resource systems.
- Low Interpretability: Transformer-based models function as black boxes, making it difficult to understand or explain their internal decision-making process.

VII. ALGORITHM OVERVIEW

BiLSTM (Bidirectional Long Short-Term Memory):
This model processes textual data in both forward and backward directions, enabling it to capture context from both sides of a sentence. This bidirectional approach improves the model's understanding of sentence structure and meaning, outperforming standard LSTM in contextual awareness.
CNN-BiLSTM:
This hybrid architecture first applies Convolutional Neural Networks (CNN) to extract local patterns like keywords or short phrases from text. The resulting features are then passed through BiLSTM layers, which integrate contextual information, leading to a more enriched and detailed representation of the input.
RoBERTa (Robustly Optimized BERT):
RoBERTa is an advanced transformer-based model that improves upon BERT by using larger training data, dynamic token masking, and fine-tuned parameters. It delivers deep contextual comprehension, making it especially powerful for distinguishing between real and fake news with subtle language cues.

VIII. RESULTS AND DISCUSSION

Figure 3 illustrates the accuracy performance of individual models and the final ensemble. The results clearly demonstrate that combining traditional machine learning with advanced deep learning and transformer models leads to improved outcomes.

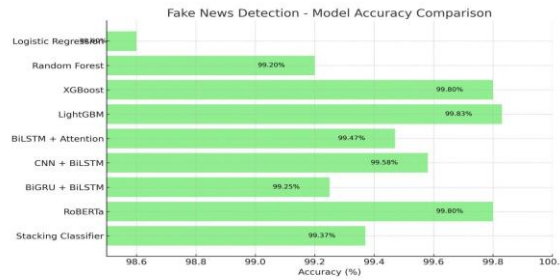

Fig. 3: Accuracy comparison across different models

TABLE II: Accuracy Scores of Various Models

Model	Accuracy
Logistic Regression	98.60%
XGBoost	99.80%
BiLSTM with Attention	99.47%
RoBERTa	99.87%
Ensemble (RoBERTa + XGBoost + BiLSTM)	99.91%

The ensemble method outperformed all individual models, showing the effectiveness of combining diverse architectures. Traditional models like Logistic Regression showed strong baseline performance, while transformer-based models like RoBERTa delivered high contextual accuracy. The ensemble approach successfully leveraged the strengths of all models, resulting in the highest overall performance.

IX. CONCLUSION

This study presents an effective hybrid approach for fake news detection, combining the capabilities of classical machine learning and modern deep learning models. By integrating RoBERTa, BiLSTM, and XGBoost into an ensemble framework, the proposed system demonstrates high accuracy, scalability, and adaptability across diverse datasets. The results confirm that the ensemble method significantly enhances detection performance, particularly in dealing with complex language and contextual variations. Moving forward, this system holds potential for real-time deployment in media platforms and can be further improved by incorporating explainable AI techniques to increase transparency and trust in automated decision-making processes.

Beyond detection accuracy, this research highlights the importance of architectural diversity in model design. The integration of statistical models, sequence-based memory networks, and context-aware transformers creates a powerful mechanism capable of handling noisy, ambiguous, or multilingual content effectively. The modular design also allows the system to adapt and scale based on the application domain, whether in news verification platforms, social media factchecking, or educational tools promoting media literacy.

Future directions include incorporating real-time news stream monitoring, cross-platform integration using APIs, and developing a user-friendly interface to visualize predictions. With the addition of explainability frameworks like LIME or SHAP, the system can evolve into a transparent and trustworthy solution for countering misinformation at scale.

X. ACKNOWLEDGMENT

The author wishes to express deep gratitude to Dr. Nagnath Biradar and Prof. Vishnupant Potdar for their consistent support, expert guidance, and valuable suggestions throughout this research. Their mentorship has been instrumental in shaping the study, from problem formulation to final evaluation. Their detailed feedback and timely advice helped address challenges and improve the research quality.

Their commitment to academic excellence and patient mentoring created an ideal environment for research. They not only guided the technical aspects but also inspired the author to think critically and pursue innovative approaches. Their encouragement during difficult phases of the project ensured steady progress and intellectual clarity.

The author also acknowledges the availability of public resources, including the Fake and Real News Dataset from Kaggle, and various open-source libraries and frameworks that supported the model development and analysis. These resources played a crucial role in enabling effective experimentation and ensuring reliable results.

Appreciation is also extended to the open-source community whose contributions—through tools like Transformers, PyTorch, Scikit-learn, and others—facilitated model training and evaluation. These platforms significantly reduced development time and provided access to powerful pretrained models and reproducible benchmarks.

Overall, the success of this work is a result of their continued mentorship and the collaborative spirit of the open-source community. The research stands as a testament to what can be achieved through mentorship, community-driven tools, and a commitment to solving real-world problems.

REFERENCES

[1] E. Papageorgiou, I. Varlamis, and C. Chronis, “Harnessing large language models and deep neural networks for fake news detection,” Information, vol. 16, no. 4, pp. 297, 2025.

[2] A. Harby and F. H. Zulkernine, “A comparative analysis of graph neural networks for fake news detection,” in Proc. IEEE COMPSAC, 2023, pp. 1–6.

[3] M. E. Almandouh et al., “Ensemble based high performance deep learning models for fake news detection,” Scientific Reports, vol. 14, no. 1, pp. 26591, 2024.

[4] A. Galli et al., “A comprehensive benchmark for fake news detection,” Journal of Intelligent Information Systems, vol. 59, pp. 237–261, 2022.

[5] K. I. Roumeliotis et al., “Fake news detection and classification: A comparative study of convolutional neural networks, large language models, and natural language processing models,” Future Internet, vol. 17, no. 1, pp. 28, 2025.

[6] I. Bala et al., “Advancements in fake news detection: A comparative study of machine and deep learning methods,” International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 10, pp. 2707–2711, 2023.

[7] J. Alghamdi et al., “A comparative study of machine learning and deep learning techniques for fake news detection,” Information, vol. 13, no. 12, pp. 576, 2022.

[8] F. B. Mahmud et al., “A comparative analysis of graph neural networks and commonly used machine learning algorithms on fake news detection,” arXiv preprint arXiv:2203.14132v1, 2022.

[9] L. Singh, “Fake news detection: a comparison between available deep learning techniques in vector space,” in Proc. IEEE 4th Conf. on Information & Communication Technology (CICT), Chennai, India, 2020, pp. 1–4.

[10] J. Y. Khan et al., “A benchmark study of machine learning models for online fake news detection,” arXiv preprint arXiv:1905.04749v2, 2021.