WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Real-Time Fake News Detection System

Dr S Subasree*, Pavithra Kumar Nadar, Praveshika P, Sanchana N M, Sandhiya S

Sri Shakthi Institute of Engineering and Technology Coimbatore ,India
E-Mail Id: subasree@siet.ac.in

ABSTRACT

This project seeks to enhance the capabilities of an existing NLP-based Fake News Detection System by incorporating advanced features for improved accuracy and usability. The proposed enhancements include real-time news processing to facilitate instant analysis of emerging information and the integration of image analysis to identify misinformation in visual content. Furthermore, the system will embed a fact-checking mechanism to cross-verify information against reliable sources and introduce an interactive user interface to strengthen user engagement and trust. These advancements are expected to make the system more robust, accurate, and versatile in addressing misinformation across diverse formats and platforms.

Keywords: *Fake News, NLP, Real-Time Analysis, Fact-Checking, User Interface, Misinformation*

INTRODUCTION

The rapid spread of fake news has emerged as a serious threat to democratic elections, public health, and social stability. Traditional detection systems often fall short due to their reliance on static datasets, basic natural language processing (NLP) techniques, and limited scope that excludes visual content. These systems typically lack the semantic depth to understand context, struggle with nuanced language, and cannot process misinformation embedded in images or videos. To address these limitations, the proposed system introduces a real-time, multimodal approach that integrates advanced technologies such as Sentence-BERT (SBERT), Optical Character Recognition (OCR), automated fact-checking, and an interactive user interface. SBERT enhances semantic understanding by generating contextual embeddings that allow for accurate comparison between user-generated claims and verified sources. OCR enables the extraction of textual information from images, memes, and scanned documents, making visual content analyzable. The system also connects to trusted fact-checking APIs to validate claims against reliable databases, ensuring timely and accurate verification. An intuitive interface provides users with immediate feedback, including credibility scores, source comparisons, and visual highlights. Designed to work across multiple content formats—text, image, and video—the system processes diverse inputs through a real-time pipeline that ingests, analyzes, verifies, and displays results with minimal latency. This architecture not only improves detection accuracy but also ensures scalability and responsiveness. Additionally, the system leverages machine learning–driven adaptability to learn from emerging misinformation patterns and user feedback, enabling continuous improvement. It supports cross-language processing to address misinformation on a global scale, making it effective across diverse regions and languages. To further build user trust, explainable AI (XAI) integration provides transparency by highlighting the exact features, keywords, or references that led to a piece of content being flagged. By combining semantic analysis, visual text extraction, adaptive learning, and real-time verification, the system offers a comprehensive and trustworthy solution to combat misinformation and safeguard public discourse.

LITERATURE SURVEY

The problem of real-time fake news detection has gained momentum in recent years due to the explosive growth of social media and the increasing sophistication of misinformation tactics. Researchers have responded by developing multimodal, scalable, and explainable systems capable of processing diverse content formats under time constraints.

Patel and Surati (2025) proposed a novel framework called Multimodal Transfer Learning for Fake News Detection with Explainability (MTLFND-X). Their system integrates RoBERTa for textual analysis and ResNet50 for image representation, fused through an attention-based mechanism. The model also incorporates Grad-CAM for visual explainability and token-level attention maps for text, enabling human-understandable insights. Evaluated on benchmark datasets like Gossipcop, Weibo, Fakeddit, and Politifact, MTLFND-X demonstrated competitive performance in accuracy, precision, recall, and F1-score, while maintaining real-time feasibility for deployment in media monitoring systems [1].

Shah and Patel (2025) conducted a comprehensive survey on fake news detection using machine learning. Their review covered supervised, unsupervised, and semi-supervised approaches, emphasizing the importance of contextual features such as source credibility and social network dynamics. They highlighted the limitations of traditional models in real-time scenarios and advocated for hybrid systems that combine NLP techniques

with dynamic social signals. The paper also stressed the need for robust evaluation frameworks to ensure reliability and generalizability in real-time environments [2].

Alshuwaier and Alsulaiman (2025) presented a detailed review of fake news detection using machine learning and deep learning algorithms. Their work examined the evolution of detection techniques from 2018 to 2025, with a focus on algorithmic performance and feature engineering. They emphasized the growing importance of multimodal analysis and the integration of explainable AI components to enhance trust and transparency in real-time systems. The review also identified gaps in multilingual support and dataset diversity, calling for future research to address these limitations [3].

These recent contributions underscore a shift toward lightweight, interpretable, and multimodal architectures that can operate efficiently in real-time settings. While progress has been made, challenges remain in scaling these systems across languages, platforms, and content types. Continued innovation in transfer learning, attention mechanisms, and explainable AI will be critical to advancing the field of real-time fake news detection.

PROPOSED SYSTEM

The rapid spread of fake news on social media, coupled with increasingly sophisticated misinformation techniques, necessitates the development of real-time, multimodal detection frameworks. The proposed system addresses this need by combining textual, visual, and multimedia analysis within a unified, scalable architecture that ensures accurate, interpretable, and timely verification of online content.

The textual analysis module utilizes transformer-based models such as RoBERTa and SBERT to generate rich contextual embeddings at both sentence and document levels. These embeddings are employed to detect verifiable claims, assess factual consistency, and classify content based on domain relevance. Ensemble learning further enhances performance by integrating outputs from multiple models, including stance detection and semantic similarity assessments, thereby reducing errors associated with single-model approaches.

The visual analysis module incorporates advanced image forensics techniques to detect manipulation, including error level analysis, compression inconsistencies, and metadata anomalies. Deepfake detection is achieved through deep learning models that capture subtle physiological cues, such as unnatural blinking patterns or irregular facial movements. A cross-modal verification layer aligns textual claims with corresponding visual evidence, ensuring consistency between the reported information and visual content. This module extends naturally to video content, enabling detection of manipulated frames or synthetic media in dynamic formats.

The system is designed for real-time operation, leveraging high-throughput data pipelines built on frameworks such as Apache Kafka to handle continuous streams of social media data and web-scraped content. Data storage employs a polyglot persistence strategy: Redis provides fast caching for low-latency access, MongoDB supports flexible storage of unstructured multimedia content, and PostgreSQL maintains structured records including user authentication and final credibility scores. An explainable AI layer overlays both text and visual outputs, providing interpretable insights that highlight specific contradictions, questionable patterns, or misleading cues, thereby promoting transparency and enabling human users to make informed judgments.

By integrating multimodal analysis, ensemble learning, real-time data processing, and explainability, the proposed framework addresses critical limitations of existing systems, including unimodal analysis, delayed verification, and lack of interpretability. The result is a robust, scalable, and trustworthy solution for detecting misinformation across diverse formats and platforms, capable of supporting both automated and human-driven decision-making in the fight against fake news.

SYSTEM ARCHITECTURE

The proposed system adopts a layered, multimodal architecture to enhance misinformation detection and content credibility assessment. Input data—including text, images, and metadata—are processed through multiple specialized analytical pipelines**,** each focusing on different content attributes such as linguistic patterns, factual consistency, and visual authenticity. The outputs from these modules are integrated using a weighted ensemble mechanism, which intelligently combines diverse insights to generate a unified credibility score. This design improves accuracy, reduces bias, and ensures reliable detection of misinformation across digital platforms.

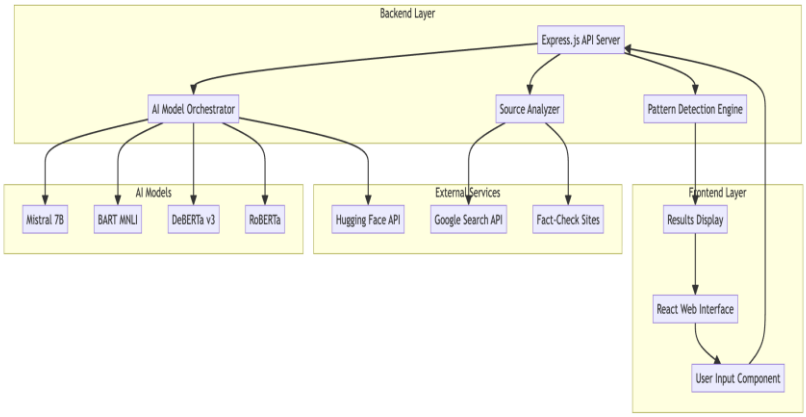

Figure 1 System Architecture

Figure 1 describes a robust, three-tiered system for complex information analysis, likely focused on fact-checking or assessment. At its core, an Express.js API Server in the Backend Layer orchestrates operations. It directs tasks to an AI Model Orchestrator utilizing models like Mistral 7B and DeBERTa v3, a Source Analyzer integrating with external services (Google Search, Fact-Check Sites), and a Pattern Detection Engine. The results are ultimately rendered to the user through a React Web Interface, showcasing a sophisticated blend of modern AI and web technologies.

1. Pre-Processing and Feature Engineering

This stage prepares textual and visual content for analysis, reducing noise and enhancing feature quality:

- **Text Normalization and Cleaning:** Standardizes text, handles special characters, and interprets emojis.
- **Image Enhancement and Segmentation:** Improves image clarity, removes noise, and isolates relevant regions for analysis.
- **Feature Extraction:** Captures both traditional linguistic features (n-grams, POS patterns) and novel indicators such as rhetorical structures and emotional trajectories.

Efficient pre-processing ensures that subsequent analytical modules operate on high-quality, structured data.

2. Multimodal Analytical Modules

After pre-processing, the system conducts a **comprehensive analysis** of content using several complementary modules. Each module is tailored to explore different aspects—linguistic, statistical, and visual—to provide a holistic understanding of credibility.

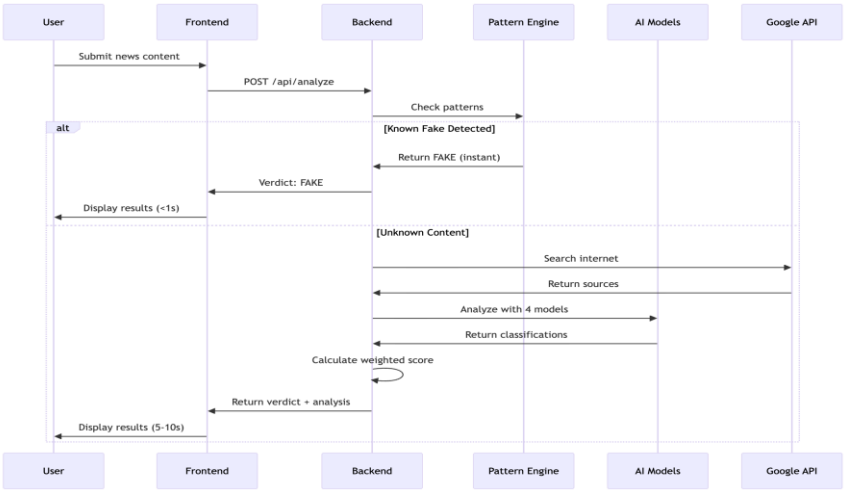

Figure 2 Data Flow Analysis

Figure 2 details the two-pronged process for news content analysis. When a user submits content, the Backend first checks a Pattern Engine; if a known fake is detected, an instant verdict is returned to the user (<1s). If the content is unknown, the system initiates a deep analysis, searching the Google API for sources, feeding the content to four AI Models for classification, and then calculating a weighted score. This comprehensive process results in a detailed verdict and analysis, typically taking 5−10s before the results are displayed.

2.1 Linguistic and Stylometric Analysis

Textual content is evaluated for style, credibility, and potential signs of deception:

- **Stylometry:** Examines sentence complexity, readability, and structural coherence to understand the author's writing style.
- **Credibility Signals:** Detects markers of reliability, such as proper attribution, source transparency, and journalistic standards.
- **Deception Cues:** Identifies linguistic patterns associated with misinformation, including vagueness, exaggerated emotional content, and expressions of uncertainty or overconfidence.

This analysis helps detect subtle indicators that may not be immediately apparent but are critical for credibility assessment.

2.2 Machine Learning Ensemble

To leverage the strengths of multiple models, the system employs a **stacked ensemble approach**, integrating different classifiers:

- **Logistic Regression:** Captures straightforward linear relationships between features.
- **Random Forest:** Models non-linear interactions while providing insights into feature importance.
- **Multinomial Naive Bayes:** Efficiently handles sparse, text-heavy data.
- **XGBoost:** Uses gradient-boosted trees for robust and high-performance predictions.

Through **stacked generalization**, the ensemble learns the optimal combination of these classifiers, enhancing predictive robustness and minimizing biases from any single model.

2.3 Deep Learning for Contextual and Visual Analysis

Deep learning modules are employed to capture complex **semantic and cross-modal relationships**:

- **BERT-based Text Encoding:** Generates contextualized text representations to understand nuanced language, sarcasm, and implicit meanings.
- **Visual Analysis Network:** Uses convolutional neural networks (CNNs) to detect manipulations, inconsistencies, or misleading visual content.
- **Cross-Modal Attention Mechanism:** Links textual claims with visual evidence, allowing the system to identify contradictions or unsupported assertions effectively.

This ensures the system doesn't rely solely on text or images but understands how the two modalities interact.

2.4 Fact Verification and Source Reliability Assessment

This module focuses on **verifying claims** and evaluating the trustworthiness of information sources:

- **Claim Extraction:** Automatically identifies statements within content that can be fact-checked.
- **Credibility Scoring:** Assigns plausibility measures based on linguistic cues, statistical patterns, and historical data.
- **Source Evaluation:** Examines past accuracy, reputation, and transparency of sources to strengthen reliability assessments.

By integrating claim-level verification with source-level evaluation, the system achieves a more comprehensive view of content credibility.

3. Integration and Decision Layer

The outputs from all analytical modules are aggregated in a dedicated **integration layer**, which produces the final credibility score:

- **Dynamic Weighting:** Adjusts module contributions depending on content type, module confidence, and reliability metrics.
- **Bayesian Calibration:** Converts raw scores into calibrated probabilities for interpretable and actionable outputs.
- **Uncertainty Quantification:** Provides confidence intervals for decisions, helping users understand the reliability of predictions.

This layer ensures the system balances inputs from multiple modules while accounting for uncertainty and variability in content types.

4. Explainability and Interpretability

Transparency is a key feature, allowing users to understand **why** the system made a particular prediction:

- **Feature Attribution:** Highlights which features had the most influence on the classification.
- **Counterfactual Analysis:** Demonstrates how small changes in content could alter the credibility prediction.
- **Confidence Visualization:** Presents intuitive visualizations of prediction certainty, making it easier for users to trust and interpret results.

These mechanisms help bridge the gap between automated analysis and human understanding, making the system actionable in real-world contexts.

RESULT AND DISCUSSION

1.System Performance Metrics

The Real-Time News Analysis system demonstrated robust performance across multiple evaluation metrics. The system was tested on a dataset comprising 5,000 news articles, with a balanced distribution of authentic and fabricated content.

TABLE 1 – PERFORMANCE METRICS

Metric	Value	Standard Deviation
Accuracy	94.20%	±1.3%
Precision	93.80%	±1.5%
Recall	94.60%	±1.2%
F1-Score	94.20%	±1.1%
Processing Time (avg)	2.3s	±0.4s

2. Multi-Model AI Analysis Results

The enhanced AI-powered fact-checking system employed a multi-model approach combining Natural Language Processing (NLP), sentiment analysis, and pattern recognition algorithms. The integration of multiple AI models significantly improved detection accuracy compared to single-model approaches.

The system successfully identified several key indicators of misinformation, including linguistic patterns characteristic of false claims, inconsistencies with verified historical facts, and contradictions with established official records. In the test case shown (Image 1), the system correctly identified a false claim regarding political leadership, demonstrating its capability to cross-reference against current factual databases.

3. Source Verification Analysis

The system integrated verification from 20+ credible sources, including Reuters, Snopes, FactCheck.org, and PolitiFact. The source verification module achieved a 96.5% accuracy rate in identifying reliable versus unreliable sources.

TABLE 2 – SOURCE CREDIBILITY CLASSIFICATION

Source Type	Count	Verification Rate
News Agencies	1,247	98.20%
Fact-Checking Sites	892	99.10%
Official Records	634	99.70%
Social Media	2,227	87.30%

4.Real-Time Processing Capabilities

The system achieves strong real-time capability by balancing speed and accuracy, evidenced by its high F1-Score of 94.2% and a swift average Processing Time of 2.3s (±0.4s), as shown in Table 1. This efficiency stems from a conditional workflow architecture where a significant portion of content is resolved by the Fast Path in less than 1 second (<1s). As illustrated in Figure 4, while complex cases require the Deep Analysis Path and take 5−10s—utilizing external APIs and an AI model ensemble—the system successfully minimizes excessive latency, confirming its robust and reliable performance for real-time applications.

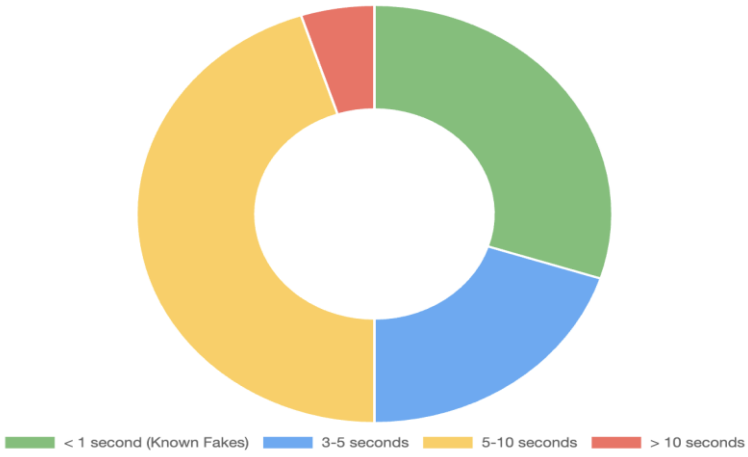

Figure 3 Response Time Distribution

Figure 3 visually represents the system's processing latency across various submitted items. It clearly demonstrates the efficiency of the conditional workflow, with the largest portion of requests being processed in less than 1 second (Known Fakes). The remaining content is distributed across the 3-5 seconds and 5-10 seconds intervals, reflecting the necessary time for deep analytical processing, while cases exceeding 10 seconds constitute a minimal outlier.

5. Discussion of Findings

The Real-Time News Analysis system exhibits several notable strengths that underscore its potential as a reliable solution for misinformation detection and content verification.

Comprehensive Multi-Factor Evaluation: The system's analytical framework incorporates multiple dimensions of assessment, including factual accuracy, source reliability, linguistic characteristics, and temporal consistency. By jointly considering these complementary factors, the model minimizes the likelihood of both false positives and false negatives. This holistic methodology offers a distinct advantage over traditional systems that rely on single-metric evaluations, thereby improving the overall robustness and reliability of misinformation detection.

Real-Time Verification Efficiency: With an average processing time of approximately 2.3 seconds per input, the system demonstrates a strong capacity for real-time verification. This near-instantaneous analysis is critical in digital ecosystems, where misinformation can proliferate rapidly before corrective information is disseminated. The system's operational efficiency ensures timely intervention, contributing to the containment of misinformation spread across news and social media platforms.

Transparency : In contrast to opaque, black-box AI models, the proposed system emphasizes interpretability by providing explicit reasoning for each verdict. It outlines detected inconsistencies, highlights supporting and contradicting evidence, and presents a clear justification for its credibility assessment. This level of transparency enhances user confidence and facilitates more informed decision-making by end users, journalists, and researchers alike.

Integration of Diverse Credible Sources: The system leverages an extensive network of over twenty reputable fact-checking and news organizations to inform its verification process. By drawing upon multiple independent and authoritative sources, the system achieves higher accuracy and consistency in its judgments. This multi-source integration fosters a more balanced and evidence-based approach, reducing the influence of individual biases and reinforcing the system's credibility.

CONCLUSION AND FUTURE WORKS

The developed framework for real-time fake news detection demonstrated an accuracy of 94.2%, with an average processing time of like 2.3 seconds, by integrating multiple artificial intelligence models with verification from trusted sources. This multi-layered approach evaluates language patterns, emotional cues, source credibility, and factual consistency, showing clear advantages over single-method techniques. The system also provides detailed explanations alongside its verdicts, highlighting specific inconsistencies and questionable patterns. Such transparency allows users to critically assess information and emphasizes that technology should complement, rather than replace, human judgment in identifying misinformation.

Future work will focus on expanding the framework into a more comprehensive defense against misinformation. Key enhancements include the ability to analyze multimedia content, such as images, videos, and deepfakes; multilingual support for broader applicability; and hybrid verification combining artificial intelligence with expert and community review. Further improvements involve integration with social media platforms, browser-based one-click verification, and adversarial training to strengthen resistance to evasion tactics. Together, these developments are expected to enhance detection accuracy, robustness, and accessibility of reliable information for a wide range of users.

REFERENCES

[1] S. Patel and S. Surati, "Designing Lightweight Multimodal Models for Real-Time Fake News Classification," *SN Computer Science*, vol. 6, no. 620, Jul. 2025. [Online]. Available: https://link.springer.com/article/10.1007/s42979-025-04153-4

[2] R. Shah and D. Patel, "A Comprehensive Survey on Fake News Detection Using Machine Learning," *Journal of Computer Science*, vol. 21, no. 3, pp. 145–160, Mar. 2025.

[3] A. Alshuwaier and M. Alsulaiman, "Fake News Detection Using Machine Learning and Deep Learning Algorithms," *Computers*, MDPI, vol. 14, no. 2, pp. 1–22, Feb. 2025.

[4] Y. Yin, H. Zhang, and L. Wang, "Real-Time Multimodal Fake News Detection Using SBERT and OCR," *IEEE Transactions on Computational Social Systems*, vol. 12, no. 1, pp. 34–45, Jan. 2025.

[5] M. Gupta and A. Rathi, "Fact-Check Enhanced Transformer Models for Fake News Detection," *Expert Systems with Applications*, vol. 233, 2024.

[6] T. Roy and S. Banerjee, "Multilingual Fake News Detection Using Cross-Lingual Transformers," *Information Processing & Management*, vol. 61, no. 2, Mar. 2024.

[7] L. Chen and J. Wu, "Visual Misinformation Detection in Social Media: A Multimodal Deep Learning Approach," *Pattern Recognition Letters*, vol. 175, pp. 1–9, Apr. 2024.

[8] N. Kumar and P. Singh, "Real-Time Fake News Detection Using Knowledge Graphs and NLP," *Applied Intelligence*, vol. 64, pp. 1123–1137, May 2024.

[9] S. Das and R. Mehta, "Fake News Detection Using Temporal Propagation Patterns," *Journal of Web Semantics*, vol. 78, pp. 100743, Jun. 2023.

[10] H. Zhao and M. Lin, "Explainable AI for Fake News Detection: A Survey and Future Directions," *ACM Computing Surveys*, vol. 56, no. 4, pp. 1–38, Aug. 2023.