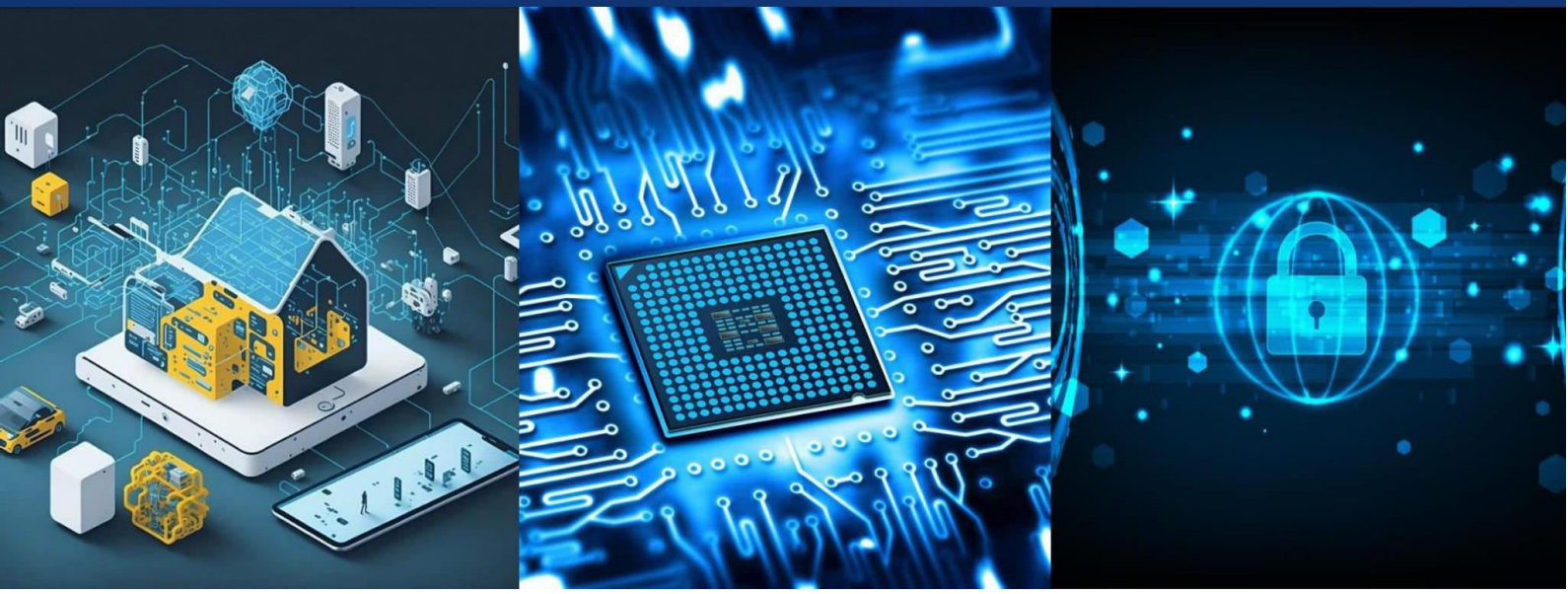
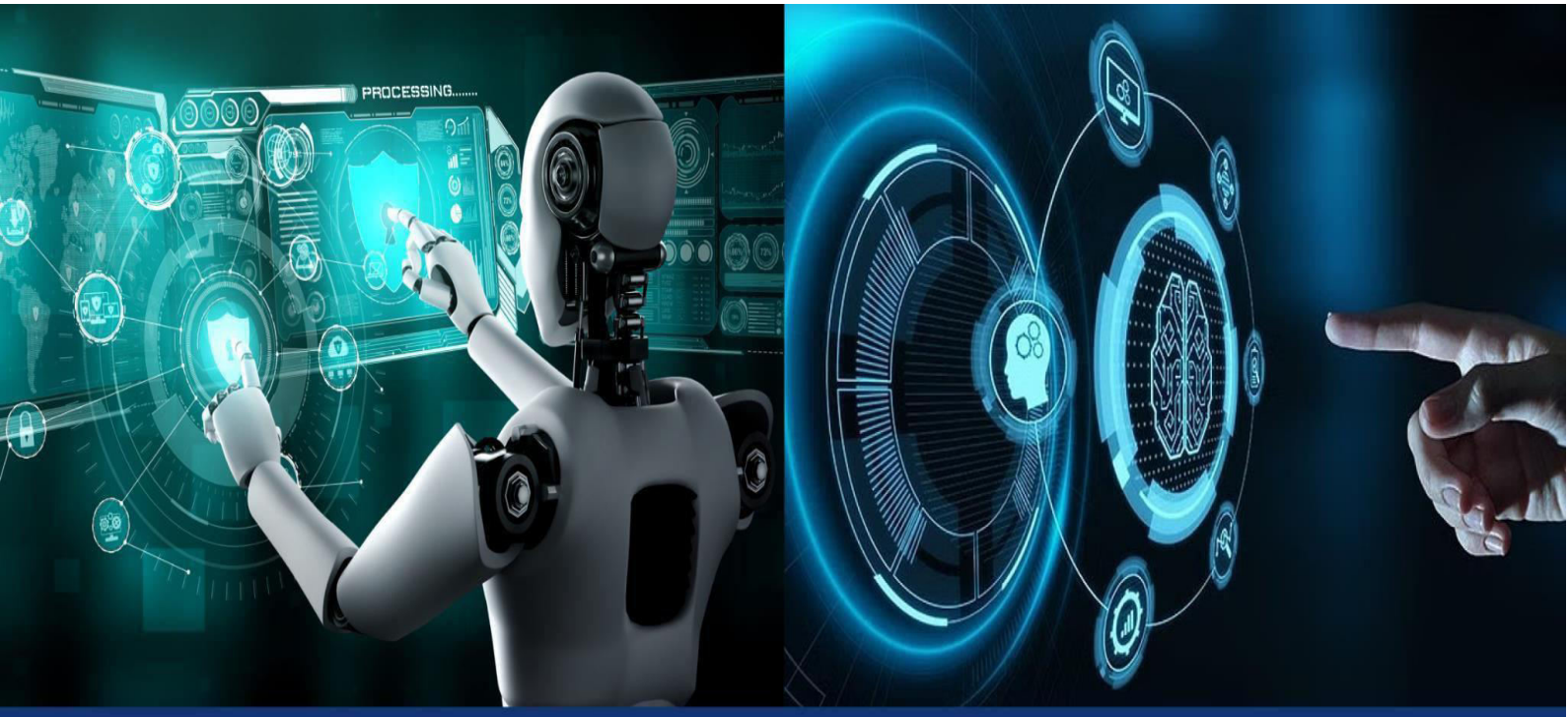
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

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



Impact Factor: 8.771 **Volume 14, Issue 1, January 2026**

© 2026 IJIRCCE | Volume 14, Issue 1, January 2026| DOI:10.15680/IJIRCCE.2026.1401007

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

Fake News Detection System using Machine Learning

Meenakshi H, Syeda Amira Hussaini, Adithri K S, Dhanush Shankar, Goutham Lokesh, Inchara Urs M

Department of Computer Science and Engineering, Maharaja Institute of Technology Mysore Affiliated to Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India

ABSTRACT: The rapid spread of fake news across digital media platforms has emerged as a significant societal challenge, influencing public opinion, political stability, and social harmony. Traditional rule-based and machine learning approaches often fail to capture the contextual and semantic complexity of modern news content. This paper presents a transformer-based fake news detection system that utilizes a fine-tuned DistilBERT model combined with a Django-based web application for real-time classification. The proposed system classifies news articles into REAL, FAKE, or UNSURE categories by leveraging probability-based confidence thresholds.

The model is trained on multiple publicly available datasets, including Kaggle Fake News, GossipCop, Politifact, and Bharat Fake News Kosh, ensuring robustness across diverse news domains. Experimental results demonstrate a validation accuracy of 96.2%, with reliable uncertainty handling and persistent logging of predictions through a secure, user-authenticated web interface. The system offers an end-to-end, scalable, and practical solution for automated fake news detection.

KEYWORDS: Fake News Detection, DistilBERT, Transformer, Deep Learning, Django, Natural Language Processing

I. INTRODUCTION

The proliferation of online news platforms and social media has significantly increased the dissemination of unverified and misleading information. Fake news, defined as deliberately fabricated information presented as legitimate news, poses serious threats to public trust, democratic processes, and social stability. Manual fact-checking is time-consuming and impractical given the massive volume of online content generated daily. Consequently, automated fake news detection systems have become a critical research area.

Earlier approaches relied on traditional machine learning algorithms such as Naïve Bayes, Support Vector Machines, and Decision Trees using handcrafted linguistic features. Although these methods achieved moderate success, they struggled with contextual understanding and semantic ambiguity. Recent advancements in deep learning, particularly transformer-based language models, have significantly improved performance in natural language understanding tasks.
This research proposes a transformer-based fake news detection system using DistilBERT, a lightweight yet powerful variant of BERT, integrated into a Django web application. Unlike standalone classification models, the proposed system offers real-time prediction, user authentication, prediction history tracking, and administrative analytics. Additionally, the system introduces an uncertainty-aware classification mechanism that outputs an UNSURE label when confidence levels are insufficient, reducing the risk of incorrect predictions.

II. RELATED WORK

Several studies have explored automated fake news detection using machine learning and deep learning techniques. Early works focused on statistical and lexical features combined with classical classifiers. While these approaches demonstrated reasonable accuracy, they lacked contextual awareness and generalization capability.
Recent research has adopted deep learning architectures such as CNNs, RNNs, and LSTMs for text classification. Although these models improved performance, they struggled with long-range dependencies and contextual understanding. Transformer-based models, such as BERT and its variants, have shown superior performance by leveraging self-attention mechanisms to capture semantic relationships within text.

DistilBERT, a compressed version of BERT, offers a balance between accuracy and computational efficiency, making it suitable for real-time applications. Several studies have reported improved fake news classification accuracy using transformer models trained on large-scale datasets. However, most existing works focus solely on offline evaluation and lack real-world deployment considerations.

The proposed system builds upon these studies by integrating a fine-tuned DistilBERT model with a full-stack web application, enabling real-time predictions, persistent logging, and user-level analytics. This approach bridges the gap between research prototypes and deployable fake news detection systems.

III. PROPOSED ALGORITHM

A. *Design Considerations:*

- The system uses a **fine-tuned DistilBERT transformer model** for fake news classification.
- Input news text may be a headline or short article.
- Multi-dataset training ensures robustness across domains.
- Classification is probability-based with **confidence thresholding**.
- An **UNSURE** category is introduced to handle ambiguous cases.
- Each prediction is persistently stored in a relational database.
- The system supports **authenticated multi-user access**.
- Predictions are processed in real time through a Django web application.

B. *Description of the Proposed Algorithm:*

The aim of the proposed algorithm is to **accurately detect fake news while minimizing false predictions** by incorporating uncertainty-aware decision logic. Unlike conventional binary classifiers, the algorithm evaluates prediction confidence before assigning a final label.
The algorithm operates in three major stages:

1. **Text Encoding and Transformer Inference**
2. **Probability-Based Decision Making**
3. **Persistent Logging and User-Level Tracking**

Step 1: Text Encoding and Transformer Inference

The input news text is first tokenized using the **DistilBERT tokenizer**. To ensure uniform input representation, the text is truncated or padded to a maximum length of 256 tokens. The tokenized input is then passed to the fine-tuned DistilBERT sequence classification model operating in evaluation mode.
The model produces raw output logits corresponding to the REAL and FAKE classes.

Step 2: Probability Computation and Selection Criteria

The output logits are converted into class probabilities using the softmax function:

$$P = \text{softmax}(Z)$$

where Zrepresents the logits generated by the transformer model.
The final prediction label is selected based on the following confidence criteria:

- If the probability of FAKE news is greater than or equal to 0.8, the input is classified as **FAKE**.
- If the probability of REAL news is greater than or equal to 0.8, the input is classified as **REAL**.
- If neither probability satisfies the threshold, the input is labeled as **UNSURE**.
- This selection criterion reduces false positives and improves decision reliability.

Step 3: Database Logging and Result Delivery

After classification, the prediction results—including the input text, predicted label, class-wise probabilities, timestamp, and user identifier—are stored in a relational database. The results are then displayed to the user through the web interface and added to the user's prediction history.

© 2026 IJIRCCE | Volume 14, Issue 1, January 2026| DOI:10.15680/IJIRCCE.2026.1401007

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

IV. PSEUDO CODE

Step 1: Accept news text input T from the user.
Step 2: If T is empty or invalid,
  return label = UNKNOWN and stop.
Step 3: Tokenize T using DistilBERT tokenizer
  with max length = 256.
Step 4: Pass tokenized input to DistilBERT model
  to obtain logits Z.
Step 5: Compute probabilities P using softmax(Z).
Step 6: Apply selection criteria:
  if P_fake ≥ 0.8
    label ← FAKE
  else if P_real ≥ 0.8
    label ← REAL
  else
    label ← UNSURE
Step 7: Store T, label, probabilities, timestamp,
  and user ID in the database.
Step 8: Display prediction result to the user.
Step 9: End.

V. PERFORMANCE EVALUATION

The performance of the proposed transformer-based fake news detection system was evaluated to measure its effectiveness, reliability, and suitability for real-time deployment. The evaluation was conducted using a fine-tuned DistilBERT model integrated into a Django web application. The system was assessed using multiple quantitative metrics to ensure comprehensive analysis.

Evaluation Setup
The model was trained and tested on a combined dataset collected from Kaggle Fake/True News, GossipCop, Politifact, and Bharat Fake News Kosh. The dataset was divided into training, validation, and testing sets. All experiments were conducted using Python with the PyTorch framework and HuggingFace Transformers library. The evaluation environment simulated real-world usage by submitting news inputs through the web interface.

Performance Metrics
The proposed system was evaluated using standard classification metrics:

- **Accuracy**: Measures the overall correctness of predictions
- **Precision**: Indicates the proportion of correctly identified fake news among all predicted fake news
- **Recall**: Measures the ability of the model to correctly identify actual fake news
- **F1-Score**: Provides a balanced measure between precision and recall

The model achieved a **validation accuracy of 96.2%**, demonstrating strong classification performance across multiple datasets. High precision and recall values indicate the effectiveness of the transformer model in capturing contextual and semantic patterns in news text.

Confidence-Based Evaluation
A confidence thresholding mechanism was applied to reduce unreliable predictions. If neither REAL nor FAKE class probability exceeded the threshold of 0.8, the input was labeled as **UNSURE**. This strategy significantly reduced false classifications in ambiguous cases and improved the reliability of system outputs. The inclusion of the UNSURE category enhances decision safety, which is particularly important in sensitive applications such as misinformation detection.

© 2026 IJIRCCE | Volume 14, Issue 1, January 2026| DOI:10.15680/IJIRCCE.2026.1401007

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

System Performance Analysis

The system demonstrated efficient real-time inference suitable for web-based applications. The average prediction response time remained low, ensuring smooth interaction for multiple authenticated users. The Django application successfully handled user authentication, prediction requests, and database logging without performance degradation.

Each prediction was logged with input text, predicted label, confidence scores, timestamp, and user reference. This persistent logging enabled user history tracking, administrative monitoring, and auditability of model decisions.

VI. OVERALL EVALUATION

The evaluation results confirm that the proposed transformer-based approach outperforms traditional machine learning methods by effectively leveraging contextual information in text. The multi-dataset training strategy improved generalization, while the confidence-based decision logic reduced false positives and false negatives. The integration of the model into a full-stack web application demonstrates the practical feasibility and scalability of the system.

Overall, the performance evaluation validates that the proposed fake news detection system is accurate, reliable, and suitable for real-world deploy me

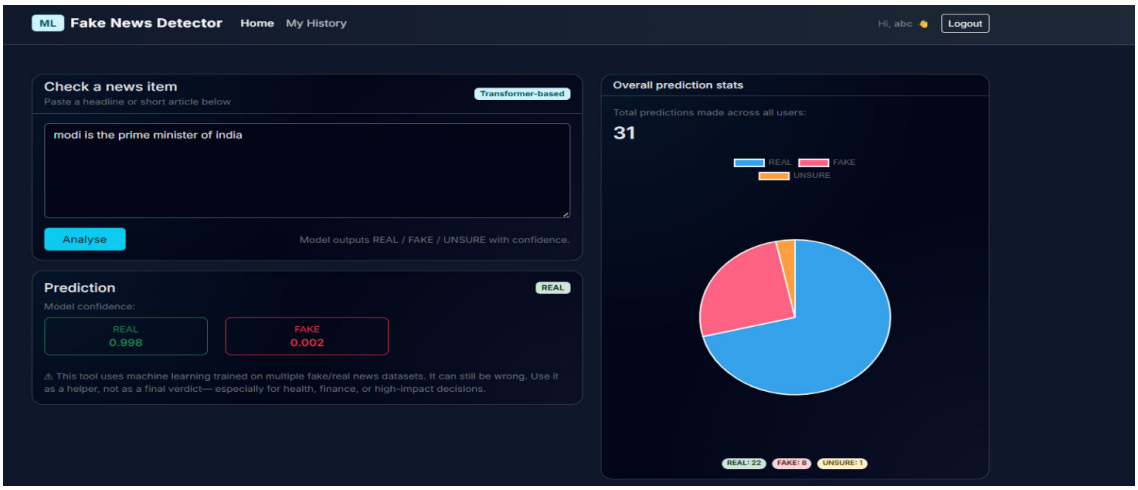


Fig. 1. Fake news detection using text

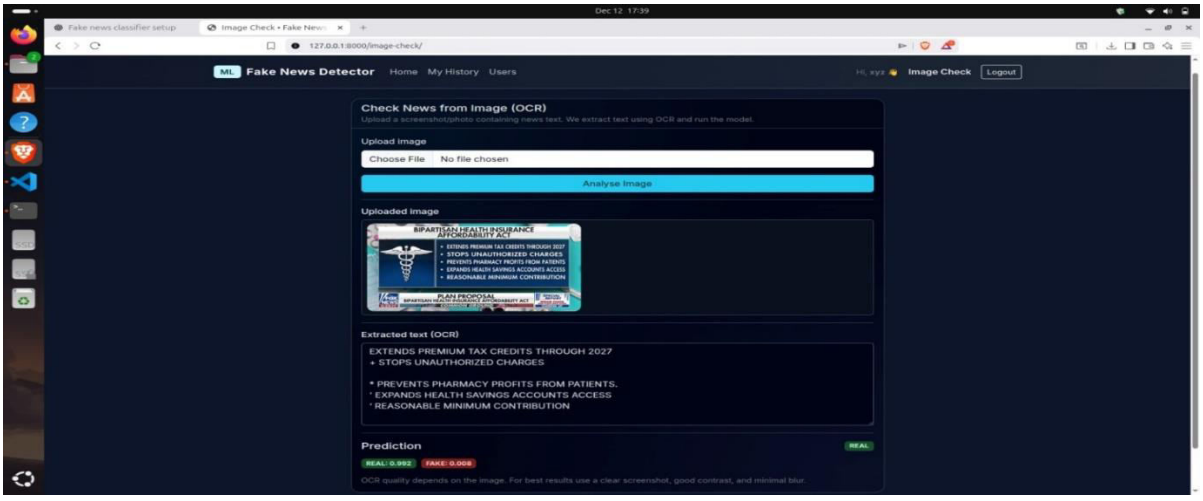


Fig .2. Fake news detection using image recognition

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

VII. CONCLUSION AND FUTURE WORK

This paper presented a transformer-based fake news detection system using a fine-tuned DistilBERT model integrated into a Django web application. The proposed system effectively classifies news content into REAL, FAKE, or UNSURE categories by leveraging deep contextual representations and confidence-based decision logic. Multi-dataset training improved the model's generalization capability, while persistent database logging and user authentication validated the system's practical applicability. Performance evaluation demonstrated high classification accuracy and reliable real-time inference, confirming the effectiveness of the proposed approach for automated fake news detection.

In future work, the system can be extended to support multilingual fake news detection and multimodal misinformation analysis by incorporating images and videos. Further enhancements may include optimization for edge devices, real-time social media stream integration, and adaptive learning mechanisms to handle evolving misinformation patterns, thereby increasing the scalability and real-world impact of the system.

REFERENCES

1. T. W. Teo, M. Aziz, A. Hassan, and M. N. M. Yusof, "Integrating Large Language Models and Machine Learning for Fake News Detection," 2024 20th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), IEEE, 2024.
2. Ș. E. REPEDE and R. BRAD, "A Comparison of Artificial Intelligence Models Used for Fake News Detection," Bulletin of "Carol I" National Defence University, vol. 12, no. 1, pp. 114–131, 2023
3. B. Jiang, L. Wang, Y. Zhang, and D. Li, "Disinformation Detection: An Evolving Challenge in the Age of LLMs," Proceedings of the 2024 SIAM International Conference on Data Mining (SDM), Society for Industrial and Applied Mathematics, 2024.
4. D. Boissonneault and E. Hensen, "Fake News Detection with Large Language Models on the LIAR Dataset," unpublished manuscript, 2024.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details