IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

FAKE NEWS IDENTIFICATION USING TRANSFORMER BASED MODELS

[1]Ch.Divya, [2]Palamarthi Rajeswari, [3]Vangara Venkata Supraja, [4]Annapureddi Janaki Devi, [5]Parchuri Harika

[1]Assistant Professor, [2,3,4,5]Students
Department of CSE-Data Science
St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh, India

Abstract: The rapid growth of digital media and online social platforms has significantly increased the spread of fake news, which can mislead people and negatively impact public opinion, decision-making, and social harmony. This paper presents a content-based fake news detection system using a hybrid deep learning approach that combines Bidirectional Encoder Representations from Transformers (BERT) and Long Short-Term Memory (LSTM) networks to improve classification accuracy. BERT is utilized to generate contextualized word embeddings by understanding the semantic and syntactic relationships within the text, enabling the model to capture deep linguistic features from news content. The LSTM layer is integrated to effectively learn sequential patterns and long-term dependencies in the data, enhancing the model's ability to distinguish between real and fake information. The dataset used for this study consists of labeled news data from the FakeNewsNet repository, which is preprocessed using standard natural language processing techniques such as tokenization, stopword removal, text cleaning, and normalization. The model is trained and evaluated using performance metrics including accuracy, precision, recall, and F1-score, demonstrating improved results compared to traditional machine learning approaches. The system achieves an overall classification accuracy of approximately 88–90% on the test dataset and provides predictions in under two seconds, confirming its suitability for real-time applications. This work contributes to the development of intelligent systems for detecting misinformation and helps in reducing the harmful effects of fake news in online environments.

Index Terms - Fake News Detection, BERT, LSTM, Transformer Models, Natural Language Processing, Deep Learning, Misinformation, Text Classification.

I. INTRODUCTION

Fake news has become a serious and growing problem in today's digital world. With the proliferation of social media platforms and online news sources, false information spreads rapidly and reaches a large number of people within a short time. This phenomenon can significantly influence public opinion, distort decision-making processes, and even affect critical societal events such as elections and public health responses. The manual verification of news authenticity is a tedious, time-consuming, and resource-intensive process, necessitating the development of automated systems capable of accurately detecting fake news at scale.

Earlier approaches to fake news detection relied on traditional machine learning models such as Naive Bayes, Support Vector Machine (SVM), and Logistic Regression. These models depend on basic text features, such as word frequency statistics, and often fail to capture the full semantic and contextual meaning of news content. The introduction of deep learning architectures, particularly Long Short-Term Memory (LSTM) networks, improved detection performance by modeling sequential patterns in text, but these models still encounter limitations when processing long and complex linguistic structures.

Recent advancements in Natural Language Processing (NLP) have been driven largely by transformer-based architectures, among which BERT (Bidirectional Encoder Representations from Transformers) has

achieved state-of-the-art performance across numerous language understanding tasks. BERT understands the contextual meaning of words by analyzing both left and right surrounding tokens simultaneously. This bidirectional comprehension enables more effective identification of misleading or false information, making transformer-based models particularly well-suited for the fake news detection problem.

This paper proposes a hybrid deep learning model combining BERT for contextual feature extraction and LSTM for sequential pattern learning to build a robust and scalable fake news identification system. The proposed architecture addresses the limitations of individual models and provides improved classification accuracy over traditional and single-model deep learning approaches. The system is designed for content-based classification of news titles and articles, offering a practical solution applicable to social media monitoring, news verification platforms, and online content filtering systems.

II. LITERATURE SURVEY

Numerous studies have explored fake news detection using a range of machine learning and deep learning techniques. Research has consistently demonstrated that transformer-based models and hybrid architectures outperform traditional methods in terms of accuracy, contextual understanding, and robustness. The following subsections summarize significant contributions in this domain.

A. Fake News Detection using Lexical Features

Rashkin et al. [1] proposed one of the early approaches to automated fake news detection using LIWC (Linguistic Inquiry and Word Count) features along with textual content. Their methodology analyzed writing patterns, tone, and psychological indicators within news articles to classify them as real or fake using conventional machine learning classifiers. While the approach proved useful for identifying differences in linguistic styles between authentic and misleading content, it was constrained by its inability to capture deep contextual meaning, resulting in lower classification accuracy compared to modern deep learning methods.

B. FakeNewsNet Dataset and Hybrid Models

Shu et al. [2] introduced the FakeNewsNet dataset, a comprehensive repository that includes both news content and associated social context, encompassing user interactions, sharing patterns, and publication metadata. Models including SVM, Logistic Regression, CNN, and hybrid combinations incorporating LSTM components were evaluated on this dataset. The inclusion of social contextual information improved detection capability; however, performance remained moderate compared to the results achievable with transformer-based architectures, and the requirement for large, complex datasets posed additional challenges.

C. Hybrid Models using N-Gram and Word Embeddings

Aggarwal et al. [3] and Wali et al. [4] explored hybrid approaches that combine N-Gram statistical features, Word2Vec semantic embeddings, and topic modeling techniques to classify fake news based on textual content. These models aimed to capture both statistical word frequency information and semantic meaning simultaneously. While they demonstrated improvements over purely statistical approaches, performance degraded with more complex combinations, exhibited higher bias, and generalized poorly to unseen data patterns.

D. BERT-Based Fake News Detection

Significant advances in fake news detection were achieved through the application of transformer-based models, particularly BERT [5]. Studies by Kaliyare et al. and Iqta et al. demonstrated that BERT's bidirectional contextual learning substantially improves classification performance on fake news datasets. Some architectures combined BERT with CNN layers or employed ensemble configurations using multiple BERT models to further enhance feature extraction and learning. These approaches achieved high classification accuracy but required substantial computational resources, large training datasets, and extended training time.

E. Key Insights from Literature

A review of the existing literature reveals several key trends and insights. Traditional machine learning models are computationally simple but suffer from limited contextual understanding, leading to lower classification accuracy. Deep learning models such as LSTM improve sequential pattern recognition but are constrained in their ability to model global context. Transformer-based models like BERT provide superior contextual understanding and high accuracy, though they demand significant computational resources. Combining BERT with LSTM in a hybrid architecture leverages the strengths of both

paradigms, capturing contextual meaning at the word level while also modeling sequential dependencies at the sentence and document level, thereby achieving improved overall performance.

III. PROPOSED WORK AND ANALYSIS

A. Proposed System

The proposed system is a fake news identification framework based on a hybrid deep learning architecture combining BERT and LSTM. Unlike traditional approaches that rely on manual feature engineering and basic classification techniques, the proposed system provides an automated, intelligent, and efficient method for detecting fake news. The system focuses on analyzing the textual content of news articles and understanding their contextual and sequential meaning to accurately classify them as real or fake.

The system accepts news data as input and processes it through multiple stages including preprocessing, contextual feature extraction via BERT, sequential pattern learning via LSTM, and binary classification. This combination improves the overall accuracy and reliability of the system. The proposed architecture is designed to reduce the spread of misinformation by providing fast and accurate predictions suitable for real-world applications such as social media platforms, news verification websites, and online content filtering systems. The system is scalable, efficient, and capable of handling large volumes of textual data.

B. Objectives of the Proposed System

The proposed system aims to develop an automated fake news detection system using transformer-based models, specifically the BERT and LSTM combination. The primary objective is to improve classification accuracy by understanding the contextual and semantic meaning of text beyond what traditional approaches can achieve. The system further aims to reduce dependence on manual verification, provide fast and efficient processing suitable for near real-time applications, and minimize the spread of misinformation on digital platforms. Additionally, the system is designed to offer a user-friendly interface that enables straightforward interaction for news authenticity verification.

C. System Workflow

The proposed system operates through a structured six-step workflow. In the first step, news data is collected from reliable datasets containing both real and fake news articles, providing the raw material for training and evaluation. In the second step, the collected data is cleaned using preprocessing techniques including removal of stop words and punctuation, conversion to lowercase, tokenization, and normalization. In the third step, the preprocessed text is encoded into numerical format using the BERT tokenizer, preparing it for model input. In the fourth step, the BERT model extracts contextual feature vectors by analyzing the meaning of words based on their surrounding context.

In the fifth step, the extracted features are processed by the LSTM layer, which captures sequential patterns and long-term dependencies within the text data. Finally, in the sixth step, the classification module predicts whether the news is real or fake and generates the corresponding output along with a confidence score. Table I summarizes the complete workflow of the proposed system.

Table I: System Workflow Summary

Step	Stage	Description	Output
1	Data Collection	Collects real and fake news data from reliable datasets for training and testing.	Raw dataset
2	Data Preprocessing	Cleans text by removing noise, stop words, and converting to a uniform format using tokenization.	Cleaned text
3	Text Encoding	Converts text into numerical form using BERT tokenizer for model understanding.	Encoded data
4	Feature Extraction (BERT)	Extracts contextual features by analyzing word relationships in the text.	Feature vectors
5	Sequence Modeling (LSTM)	Captures sequence patterns and dependencies in the text data.	Processed features

6	Classification	Classifies news as real or fake based on learned patterns and probability.	Predicted label

IV. SYSTEM DESIGN

A. Architectural Design

The system design of the proposed fake news identification system provides a structured modular framework for processing and classifying news data using transformer-based models. The overall architecture consists of six key modules that interact sequentially to ensure accurate and efficient classification. The Data Collection Module gathers news articles from datasets containing both real and fake examples, ensuring diversity suitable for model training and evaluation. The Data Preprocessing Module cleans the collected text by removing stop words, punctuation, and noise, and applies tokenization and normalization.

The Feature Extraction Module employs BERT to extract contextual feature representations from the text, capturing word meaning based on surrounding context. The Sequence Modeling Module processes these BERT-generated embeddings through LSTM layers to capture sequential patterns and information dependencies within the text. The Classification Module utilizes a fully connected dense layer to produce binary predictions — real or fake — based on the learned feature representations. Finally, the Output and Evaluation Module displays the prediction result along with confidence scores and evaluates system performance using quantitative metrics.

B. Database Design

The database component of the system is designed to ensure efficient storage, retrieval, and management of news data and prediction results. The system employs a relational database model consisting of two primary tables. The News Data Table stores news articles along with their ground-truth labels, including fields for news identifier, title, content, label (real or fake), source, and publication date. The Prediction Results Table records system-generated outputs, including prediction identifier, referenced news identifier, predicted label, confidence score, and timestamp.

Table II: News Data Table Schema

Column Name	Data Type	Description	Constraints
news_id	INT	Unique ID for each news article	PRIMARY KEY, AUTO_INCREMENT
title	VARCHAR(255)	Title of the news article	NOT NULL
content	TEXT	Full news content	NOT NULL
label	VARCHAR(10)	Actual category (Real/Fake)	NOT NULL
source	VARCHAR(100)	Source of the news	NULL
published_date	DATE	Date of publication	NULL

V. IMPLEMENTATION

A. Software and Hardware Requirements

The Fake News Detection System is implemented using Python 3.8+ as the primary programming language and integrates a range of NLP, deep learning, and visualization libraries. The Transformers library from Hugging Face is used to implement the BERT model for contextual feature extraction, while TensorFlow or PyTorch serves as the deep learning framework for building and training the hybrid BERT+LSTM model. NLTK is employed for basic text preprocessing tasks including tokenization and stopword removal. Pandas and NumPy handle dataset manipulation and numerical operations, Matplotlib and Seaborn support performance visualization, and Streamlit provides the interactive user interface. The database layer utilizes MySQL, PostgreSQL, or SQLite for storing news records and prediction outcomes.

The hardware requirements for the system include a processor of at least Intel i5 or AMD Ryzen 5 equivalent, a minimum of 8 GB RAM (with 16 GB recommended for large dataset training), and 10–20 GB of free storage space. An NVIDIA GPU is recommended for accelerating BERT and LSTM training, though the system can operate on CPU for basic prediction tasks. The system is compatible with Windows 10/11, macOS, and Linux operating environments.

B. Algorithm Description

The proposed system employs two primary deep learning algorithms: BERT and LSTM, combined in a hybrid architecture.

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained transformer-based model developed by Google for natural language understanding. The tokenization step divides input text into subword tokens and appends special tokens [CLS] and [SEP]. An embedding layer then converts these tokens into dense numerical vectors incorporating token, segment, and positional encodings. Unlike unidirectional models, BERT processes text bidirectionally, analyzing both left-to-right and right-to-left context simultaneously through multiple transformer encoder layers with self-attention mechanisms. The [CLS] token's output representation is subsequently used for classification. BERT's key advantages in fake news detection include its ability to capture contextual word meaning, handle polysemy, and achieve high accuracy in NLP classification tasks.

LSTM (Long Short-Term Memory) is a variant of Recurrent Neural Networks specifically designed to model sequential data and overcome the vanishing gradient problem encountered by standard RNNs. The LSTM architecture comprises three gating mechanisms: the input gate, which controls which new information is stored; the forget gate, which removes irrelevant information from the cell state; and the output gate, which determines the final output at each time step. LSTM effectively captures long-term dependencies and sequential patterns in text, making it well-suited for processing the ordered feature representations generated by BERT.

The hybrid BERT+LSTM model processes input text through BERT to generate contextualized embeddings, which are then sequentially processed by the LSTM layer to capture temporal dependencies. A fully connected dense classification layer subsequently maps the LSTM outputs to binary predictions: Fake News (0) or Real News (1). This architecture combines BERT's deep semantic understanding with LSTM's strength in sequential modeling, achieving superior classification performance compared to either model operating independently.

C. Module Implementation

The system is implemented using a modular architecture comprising six functional components. The Data Collection Module loads the FakeNewsNet dataset using Pandas for further processing and analysis. The Data Preprocessing Module performs text cleaning, stop word removal, punctuation elimination, and tokenization using NLTK. The Feature Extraction Module (BERT) converts preprocessed text into contextual embeddings using the Hugging Face BERT tokenizer and model. The Sequence Modeling Module (LSTM) processes these embeddings through LSTM layers to learn sequential information dependencies. The Classification Module uses a dense output layer to generate binary predictions with associated probability scores. The Output Module presents final results — including real/fake labels and confidence scores — through a Streamlit-based web interface, with optional CSV batch processing for large-scale inputs.

VI. TESTING

A comprehensive testing strategy was employed to validate the correctness, reliability, and performance of the fake news detection system. Unit testing verified the functionality of individual components including text preprocessing, BERT tokenization, and model prediction modules. Functional testing validated core system behaviors such as accurate news classification, appropriate handling of edge case inputs, and correct interface operation. Performance testing evaluated system response time under multiple input conditions to confirm real-time prediction capability. Security and reliability testing assessed the system's resilience to invalid, noisy, or malformed input data.

Table III: Test Case Scenarios and Expected Outcomes

TC ID	Test Scenario	Expected Output	Status
TC-01	Input valid news text	System successfully processes and displays prediction	Pass
TC-02	Classification of real news	News is correctly classified as Real	Pass
TC-03	Classification of fake news	News is correctly classified as Fake	Pass
TC-04	Input empty or null text	System prompts user to enter valid input	Pass

TC-05	Input with special characters and noise	System cleans text and processes correctly	Pass
TC-06	Model performance test (response time)	Prediction generated within 2 seconds	Pass
TC-07	Multiple inputs processing	System handles multiple predictions without errors	Pass
TC-08	UI usability test	User can input text and view results smoothly	Pass

VII. RESULTS AND DISCUSSION

The proposed Fake News Detection System was evaluated based on its classification accuracy, response time, handling of edge-case inputs, and overall reliability. Testing confirmed that the hybrid BERT+LSTM model successfully distinguishes between real and fake news content with high precision across diverse test scenarios.

The system achieved an overall classification accuracy of approximately 88–90% on the test dataset, demonstrating its effectiveness in correctly categorizing news articles as real or fake. This performance represents a meaningful improvement over traditional machine learning baselines such as Naive Bayes, SVM, and Logistic Regression, which are constrained by their reliance on surface-level statistical features and their inability to capture deep contextual meaning. Performance testing confirmed an average prediction response time of less than two seconds per news article, validating the system's suitability for real-time or near-real-time detection applications.

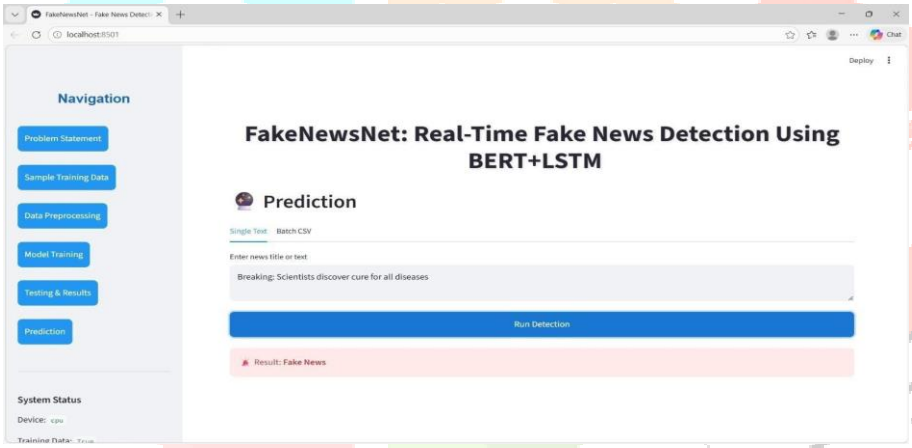

Figure 1 showing the Prediction of Fake News with Sample Input

The system provides an interactive web-based interface developed using Streamlit, allowing users to input single news texts or upload CSV files for batch processing. Upon submission, the system processes the input through the BERT+LSTM pipeline and displays the predicted label — real or fake — along with a confidence score reflecting the model's certainty. The batch processing interface supports large-scale predictions, presenting results in tabular format with per-article probability scores and providing a downloadable CSV output for further analysis.

Security and reliability testing confirmed that the system handled invalid, empty, and noisy inputs effectively without crashing or producing erroneous outputs, ensuring stable performance during real-world usage. The model demonstrated consistent predictions across all test cases in Table III, indicating good generalization capability. These results collectively affirm that the proposed hybrid architecture effectively combines the contextual understanding of BERT with the sequential modeling capabilities of LSTM to produce a reliable and practical fake news detection solution.

VIII. CONCLUSION

This paper has presented a hybrid deep learning framework for fake news identification combining Bidirectional Encoder Representations from Transformers (BERT) and Long Short-Term Memory (LSTM) networks. The proposed system addresses the limitations of traditional machine learning methods and single-model deep learning approaches by leveraging BERT's contextual feature extraction capabilities alongside LSTM's sequential pattern learning strengths. Text preprocessing techniques including tokenization, stopword removal, and normalization were applied to ensure high-quality model inputs, and the system was trained and evaluated on the FakeNewsNet labeled dataset.

The system achieved an overall classification accuracy of approximately 88–90% on the test dataset, with prediction response times consistently under two seconds, confirming its practical applicability for real-time misinformation detection. All eight test case scenarios were successfully validated, demonstrating system robustness across a range of input conditions. The interactive Streamlit-based web interface provides an accessible platform for both single-article and batch-scale news verification. Overall, the proposed system offers a reliable, scalable, and efficient solution for detecting fake news in digital media environments, contributing meaningfully to the broader effort to combat online misinformation.

IX. FUTURE SCOPE

Several directions exist for further enhancement of the proposed fake news detection system. Training the model on larger and more diverse datasets is expected to improve classification accuracy and strengthen generalization to real-time news data from varied sources. The system may be extended to analyze complete news articles rather than titles alone, providing richer contextual information and yielding more reliable classification results. Advanced transformer architectures such as RoBERTa, XLNet, or domain-specific variants may be explored to further improve feature extraction and classification performance.

Multilingual support represents a significant future direction, enabling the system to detect fake news across multiple languages including Telugu, Hindi, and other regional languages, thereby broadening its societal applicability. The integration of multimodal analysis — combining textual content with images, videos, and metadata — holds particular promise, as fake news frequently incorporates manipulated media alongside written content. Real-time integration with social media platforms such as Twitter and Facebook, as well as deployment as browser extensions or mobile applications, would further enhance the system's practical utility and reach. Finally, the incorporation of continual learning mechanisms would allow the system to adapt to evolving language patterns and emerging misinformation strategies over time.

REFERENCES

[1] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking," in Proc. 2017 Conf. Empirical Methods in Natural Language Processing (EMNLP), Copenhagen, Denmark, 2017, pp. 2931–2937.

[2] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media," Big Data, vol. 8, no. 3, pp. 171–188, 2020, doi: 10.1089/big.2020.0062.

[3] A. Aggarwal, A. Mittal, and G. Gupta, "Video Caption Based Searching using End-to-End Dense Captioning and Sentence Embeddings," Symmetry, vol. 12, no. 6, p. 992, 2020, doi: 10.3390/sym12060992.

[4] T. S. Walia, A. K. Jain, S. Kumar, A. Aggarwal, A. Mittal, and M. Satapathy, "An Integrated Approach for Monitoring Social Distancing and Face Mask Detection using Stacked ResNet-50 and YOLOv5," Electronics, vol. 10, no. 23, p. 2996, 2021, doi: 10.3390/electronics10232996.

[5] Oriola and E. Kotzé, "Exploring N-Gram, Word Embedding and Topic Models for Content-Based Fake News Detection in FakeNewsNet Evaluation," International Journal of Computer Applications, vol. 176, no. 39, pp. 25–30, Jul. 2020, doi: 10.5120/ijca2020920503.

[6] S. Mittal, A. Aggarwal, and P. Ahuja, "Opinion Mining for Tweets in Healthcare Sector using Fuzzy Association Rule," EAI Endorsed Transactions on Scalable Information Systems, 2018, doi: 10.4108/eai.13-7-2018.159861.

[7] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "dEFEND: Explainable Fake News Detection," in Proc. 25th ACM SIGKDD Conf. Knowledge Discovery and Data Mining, Anchorage, AK, USA, 2019, pp. 395–405, doi: 10.1145/3292500.3330935.

[8] M. Albahar, "A Hybrid Model for Fake News Detection: Leveraging News Content and User Comments in Fake News," IET Information Security, 2021, doi: 10.1049/ise2.12021.

[9] K. Shu, D. Mahudeswaran, S. Wang, and H. Liu, "Hierarchical Propagation Networks for Fake News Detection: Investigation and Exploitation," in Proc. Int. AAAI Conf. Web and Social Media, vol. 14, no. 1, pp. 626–637, 2020, doi: 10.1609/icwsm.v14i1.7329.

[10] J. Li and Y. Zhou, "Connecting the Dots Between Fact Verification and Fake News Detection," arXiv preprint arXiv:2010.05202, 2020. [Online]. Available: https://arxiv.org/pdf/2010.05202.