iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VII **Month of publication:** July 2026

DOI: https://doi.org/10.22214/ijraset.2025.84078

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

Fake News Identification Using Hybrid Transformer Ensemble Approach

Mr. E. Chitti Babu[1], G. Sukanya[2]
[1]*Assistant Professor,* [2]*Assistant Professor, Department of Computer Science and Engineering, Geethanjali Institute of Science & Technology*

Abstract: ***The rapid spread of misinformation across digital platforms has made fake news detection a critical challenge, as it can influence public opinion, disrupt social stability, and reduce trust in credible information sources. Manual verification is no longer feasible at scale due to the large volume of content generated daily. Existing approaches have explored hybrid architectures combining transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) with sequential models like Long Short-Term Memory (LSTM) for fake news classification; however, such approaches may have limitations in capturing the diverse linguistic, contextual, and structural patterns present in textual data. To address this limitation, this paper proposes a hybrid transformer-based ensemble model for automated fake news identification using the FakeNewsNet dataset. The proposed system integrates BERT with LSTM for contextual and sequential learning, Robustly Optimized BERT Pretraining Approach (RoBERTa) for improved textual representation, and Light Gradient Boosting Machine (LightGBM) for learning statistical patterns from textual features. Individual models provide strong baseline performance, while the ensemble combines their predictions using a weighted strategy to improve overall accuracy and robustness. Experimental results show that the ensemble model achieves an accuracy of approximately 93%, outperforming the individual constituent models. The system can be applied in real-time news verification platforms to assist users, journalists, and fact-checkers in identifying misleading information more effectively.***

Keywords: Fake News Detection, BERT, RoBERTa, LSTM, LightGBM, Transformer Ensemble, Natural Language Processing, SMOTE, FakeNewsNet, Weighted Ensemble

I. INTRODUCTION

Fake news has become a major concern in the modern digital era, as information spreads rapidly through social media platforms and online news sources. The ease of creating and sharing content makes it difficult to distinguish between real and misleading information, which can influence public opinion, create social unrest, and reduce trust in reliable sources. Manual verification of such content is time-consuming and impractical for large-scale data, while traditional machine learning approaches often fail to capture deeper contextual meaning in text. To address these challenges, this paper proposes a hybrid transformer-based ensemble model for accurate fake news detection using news titles. The system integrates BERT with LSTM for contextual and sequential learning, RoBERTa for enhanced semantic understanding, and LightGBM for statistical feature-based analysis. By combining these models through a weighted ensemble approach, the system improves classification accuracy and provides a reliable and efficient solution for detecting fake news.

The rapid growth of digital platforms has made information widely accessible but has also accelerated the spread of misleading and false news, which can influence public opinion, create confusion, and reduce trust in credible sources, highlighting the need for reliable detection systems. A major challenge in fake news detection is the scale and speed at which content is generated, making manual verification impractical despite the efforts of fact-checking organizations, which are limited by time and resources. Recent advancements in artificial intelligence, particularly in machine learning and Natural Language Processing (NLP), enable efficient analysis of textual data, where different models capture contextual meaning, sequential patterns, and statistical features; however, relying on a single model often fails to handle the complexity of fake news effectively. This work is motivated by the need for a robust solution using a hybrid ensemble approach that integrates BERT with LSTM, RoBERTa, and LightGBM, combining their strengths through weighted aggregation to improve accuracy, robustness, and generalization in fake news detection.

The increasing volume and complexity of digital news content have made it difficult to reliably distinguish between genuine and misleading information, especially as fake news is often designed to appear credible and persuasive. The massive amount of content generated across news and social media platforms makes manual verification impractical and unsuitable for real-time analysis.

Existing detection approaches, particularly those based on single-model architectures, often fail to capture the diverse contextual, sequential, and statistical patterns present in textual data, resulting in limited classification accuracy. These challenges highlight the need for a robust and scalable fake news detection system that can effectively analyze large and complex textual data using advanced machine learning and natural language processing techniques.

The objective of this work is to develop an automated system for the accurate identification of fake and real news using advanced machine learning and deep learning techniques. The study aims to design a hybrid transformer-based ensemble model by integrating BERT with LSTM, RoBERTa, and LightGBM to effectively capture contextual, sequential, and statistical patterns present in textual data. The study also addresses class imbalance in the dataset to ensure unbiased model performance, and seeks to enhance classification accuracy and robustness through a weighted ensemble strategy, while providing a user-friendly web-based interface to enable real-time fake news prediction and practical usability.

Natural Language Processing is a branch of artificial intelligence that focuses on enabling computers to understand, interpret, and process human language in a meaningful way by combining computational techniques with linguistic knowledge. Earlier NLP approaches were primarily based on rule-based systems and statistical methods, which had limitations in handling the complexity, ambiguity, and variability of natural language. With the advancement of machine learning and deep learning techniques, modern NLP systems are capable of learning contextual relationships and semantic patterns directly from data, resulting in improved accuracy and efficiency. In the context of fake news detection, NLP plays a crucial role in analyzing large volumes of textual data from news articles and online platforms; techniques such as tokenization, text normalization, and feature extraction are used to convert unstructured text into structured representations suitable for model training.

Deep Learning is a subset of machine learning that focuses on training artificial neural networks with multiple layers to automatically learn complex patterns and representations from large volumes of data. Unlike traditional machine learning approaches that rely on manual feature engineering, deep learning models can extract hierarchical features directly from raw input data, making them effective for handling unstructured data such as text. Architectures such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks are designed to process sequential data by retaining information from previous inputs, enabling the model to understand the flow of language, although these models may face challenges in capturing long-range dependencies and can be computationally intensive during training. In this work, deep learning is utilized to enhance feature representation and classification accuracy, with sequential models such as LSTM integrated alongside advanced architectures to better understand complex patterns in news content. Transformer models represent a major advancement in NLP by enabling more efficient and accurate understanding of textual data. Unlike traditional sequential models, transformers process entire sequences of text simultaneously, allowing them to capture relationships between words regardless of their position in the sentence, which significantly improves contextual understanding. The core component of transformer models is the attention mechanism, which allows the model to assign different levels of importance to words based on their context, helping capture long-range dependencies more effectively than earlier models. Popular transformer-based models include BERT, which captures bidirectional context by analyzing text from both left and right directions, and RoBERTa, which enhances performance through improved training strategies and better utilization of data. In this work, these transformer models are used to generate rich contextual embeddings from news text, and by integrating BERT and RoBERTa within a hybrid ensemble framework, the system leverages their respective strengths to improve classification accuracy and robustness.

II. LITERATURE SURVEY

The field of fake news detection has evolved significantly over the past decade, progressing from simple lexical and statistical methods towards sophisticated deep learning and ensemble architectures. Earlier studies demonstrated that features such as writing style, sentiment, and linguistic patterns could distinguish fake from real news to a limited degree. However, the manual effort involved in feature engineering and the inability of shallow models to capture contextual semantics motivated the adoption of deep learning approaches. This section reviews key works in this domain, focusing particularly on transformer-based models, hybrid architectures, and ensemble methods directly relevant to the proposed system.

Shu et al. [20] proposed dEFEND, an explainable fake news detection system that jointly models news content and user comments through a hierarchical co-attention mechanism comprising a word encoder, sentence encoder, user-comment encoder, and a sentence-comment co-attention layer that identifies sentence-comment pairs most indicative of fake news. Evaluated on the FakeNewsNet benchmark, dEFEND achieved an accuracy of 90.4% on PolitiFact and 80.8% on GossipCop, outperforming baselines relying on content or comments alone. However, its dependence on user comment availability limits its applicability in real-time, content-only classification scenarios, which directly motivates the headline-based content approach adopted in the proposed system.

Aggarwal et al. [1] investigated the effectiveness of fine-tuning a pre-trained bidirectional BERT model for fake news classification, comparing it against XGBoost and hybrid models combining N-Gram features, Word2Vec embeddings, and topic models. Fine-tuned BERT achieved 97.02% accuracy on the NewsFN dataset, substantially outperforming XGBoost at 89.37%, while multi-component hybrid models declined in performance as more sub-models were combined, attributed to increased bias from incompatible feature spaces. This study established transformer fine-tuning as a foundation for the proposed system, while also warning against naive feature combination, a limitation later addressed through systematic ensemble weighting in the present work.

Kaliyar et al. [10] introduced FakeBERT, which connects BERT token-level output embeddings to three parallel blocks of one-dimensional convolutional neural networks with different kernel sizes and filter counts, enabling simultaneous detection of local semantic patterns at varying granularity. Evaluated on a Kaggle social-media fake news dataset, FakeBERT achieved 98.90% accuracy, significantly outperforming standalone deep learning models. This work confirms that appending additional processing layers to a BERT encoder, whether CNNs as in FakeBERT or an LSTM as in the proposed system, provides measurable gains over a vanilla BERT baseline. The primary limitation of FakeBERT is its single-architecture design and high computational demand, which the present work addresses through the use of DistilBERT with a lightweight LSTM classifier head.

A hybrid classifier connecting the output of a pre-trained BERT encoder to an LSTM layer for sequential pattern learning over news titles was evaluated on the FakeNewsNet dataset, yielding 88.75% accuracy on PolitiFact and 84.10% on GossipCop, representing improvements of 2.50% and 1.10% respectively over a vanilla BERT baseline trained under identical conditions [21]. The results demonstrate that the LSTM's ability to capture sequential dependencies across the token stream adds useful information beyond what BERT's [CLS] representation alone provides. Notable limitations include the absence of class-imbalance handling, restriction to title-level textual features, and the lack of an ensemble strategy, gaps that are directly addressed in the proposed system through SMOTE, linguistic feature extraction, and a weighted ensemble of complementary models.

Malik et al. proposed EGNN, an ensemble of multiple Graph Neural Networks combined with BERT text embeddings, where graph structures encode user-engagement relationships derived from social media interaction logs, with ensemble predictions produced through majority voting, weighted averaging, and stacking strategies. Focal loss was applied during training to address class imbalance at the training-objective level. While the weighted averaging strategy and explicit treatment of class imbalance both inform key design decisions in the proposed system, EGNN's dependence on user-engagement graph data, which is not always available at prediction time, limits its applicability in content-only real-time detection, a constraint the proposed system avoids by operating exclusively on news headline text.

A stacking ensemble model using Sentence-BERT (SBERT) embeddings as input features to multiple machine learning classifiers, combined by a meta-learner, was evaluated on the WELFake multilingual news dataset and achieved 92.74% accuracy, demonstrating strong generalisation across linguistic contexts. The stacking framework's principle of learning the optimal combination of base classifiers' predictions is conceptually related to the weighted averaging ensemble strategy used in the proposed system. However, the study did not evaluate performance on imbalanced datasets or apply oversampling or class-weighting strategies, reinforcing the case for the explicit imbalance-handling and weighted ensemble design adopted in the present work.

Collectively, the reviewed literature confirms that transformer fine-tuning provides a strong foundation for fake news classification, that augmenting transformer embeddings with additional sequence-learning or statistical layers yields measurable gains, and that ensemble strategies consistently outperform single-model architectures. At the same time, recurring limitations across these studies, namely the absence of imbalance handling, dependence on auxiliary social-context data unavailable at prediction time, and naive feature combination, motivate the design choices of the proposed hybrid transformer ensemble, which combines BERT+LSTM, RoBERTa, and LightGBM with SMOTE-based class balancing and weighted ensemble aggregation operating purely on content text.

III. SYSTEM ANALYSIS

A. *Existing System*

Existing systems for fake news detection primarily utilize machine learning and deep learning techniques to classify news content as real or fake. Traditional machine learning models such as Support Vector Machines, Naïve Bayes, and Logistic Regression rely on manually engineered features including TF-IDF, n-grams, and linguistic patterns to identify misleading information, but often struggle to capture deeper semantic and contextual relationships within text. With advancements in NLP, deep learning approaches such as Convolutional Neural Networks and Recurrent Neural Networks, particularly LSTM networks, have been employed to model sequential dependencies and improve contextual understanding in news articles. More recently, transformer-based models such as BERT and RoBERTa have gained prominence due to their ability to process text bidirectionally and generate rich contextual representations using large-scale pre-trained data.

Prior research has explored content-based approaches, social context-based methods incorporating user interactions and propagation patterns, and hybrid models that combine both strategies. Despite these advancements, many existing systems rely on single-model architectures or limited feature representations, which may not fully capture the complexity and diversity of real-world fake news.

B. *Disadvantages of Existing System*

- Traditional machine learning models rely on manually engineered features, which may not effectively capture contextual and semantic relationships in textual data.
- Many existing systems depend on single-model architectures, limiting their ability to capture diverse patterns present in fake news content.
- Some approaches focus only on textual features or only on social context, which may reduce the overall effectiveness of fake news detection.
- Imbalanced datasets are not effectively handled in many systems, leading to biased predictions and reduced performance for minority classes.
- Existing approaches often fail to capture both long-term dependencies and complex contextual relationships simultaneously, affecting overall prediction accuracy.

C. *Proposed System*

The proposed system presents a hybrid transformer-based ensemble model to enhance the reliability and performance of fake news detection by integrating multiple machine learning and deep learning techniques within a unified framework. The approach combines BERT with LSTM to capture both contextual and sequential information from textual data, employs RoBERTa to further improve contextual understanding, and uses LightGBM to learn statistical patterns from feature-based representations such as TF-IDF. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is applied to the training data, ensuring balanced learning and improved performance for minority classes. The outputs of the individual models are then combined using a weighted ensemble strategy, allowing the system to leverage the strengths of each model and generate more reliable predictions. The system follows a structured pipeline consisting of data preprocessing, feature extraction, model training, and ensemble-based prediction, enhancing classification accuracy, robustness, and generalization compared to single-model methods.

D. *Advantages of Proposed System*

- Enhances classification accuracy and prediction reliability using a hybrid ensemble approach.
- Combines transformer-based and feature-based models to capture contextual, sequential, and statistical patterns in textual data.
- Addresses class imbalance using SMOTE, reducing bias and improving performance on minority classes.
- Produces stable and consistent predictions through a weighted ensemble strategy, reducing individual model bias.
- Handles complex text patterns and long dependencies more effectively than traditional models.
- Improves generalization capability by leveraging multiple models, enabling better performance across diverse types of news data.

E. *Functional and Non-Functional Requirements*

The functional requirements of the system include data collection, data preprocessing, training and testing, modelling, and prediction. The non-functional requirements address scalability to handle increasing volumes of news data, performance for near real-time prediction, usability through an intuitive interface, reliability across diverse news types, flexibility to adapt to evolving data patterns, and maintainability to support seamless integration of new datasets or model improvements.

F. *Feasibility Study*

The feasibility study evaluates the practicality and viability of the proposed system from economic, technical, and social perspectives. Economically, the system relies on publicly available datasets and open-source tools and libraries, significantly reducing development and operational expenses, and does not require specialized hardware or infrastructure. Technically, the system is implemented using Python and supported by widely used libraries such as PyTorch, Scikit-learn, and HuggingFace Transformers, with all required tools for preprocessing, training, and evaluation well documented and accessible. Socially, the system assists users in identifying misleading information, promoting awareness and responsible consumption of news content, and is simple to use without requiring technical expertise. The system is therefore considered feasible for implementation across all three dimensions.

IV. METHODOLOGY

A. System Architecture Overview

The system architecture illustrates the overall workflow of the proposed fake news detection system, describing how data flows from input to final prediction. The architecture integrates multiple machine learning and deep learning models within a unified hybrid framework to improve classification accuracy, robustness, and reliability. The pipeline begins with labelled news text from the FakeNewsNet dataset, proceeds through preprocessing (cleaning, tokenization, normalization), a stratified train-test split, and SMOTE-based balancing applied only to the training data. Balanced training data is then passed through a feature representation layer comprising tokenized text, TF-IDF, and linguistic features, which feeds three independent models, namely BERT+LSTM, RoBERTa, and LightGBM. Their outputs are merged in an ensemble layer using weighted averaging to form the trained hybrid model, which is subsequently evaluated on preprocessed testing data using accuracy, precision, recall, F1-score, and the confusion matrix.

B. Data Preprocessing

Data preprocessing ensures the quality and consistency of input data by transforming raw text into a clean, structured format. Text cleaning removes URLs, HTML tags, and unnecessary characters; tokenization splits text into tokens suitable for transformer models; normalization converts text to lowercase and standardized form; and stopword handling removes less informative words. The dataset is obtained from FakeNewsNet and consists of 23,193 labeled news samples, comprising 16,817 real and 5,323 fake news instances, resulting in a class imbalance of approximately 78% real and 22% fake. The cleaned text is tokenized for transformer-based models such as DistilBERT and RoBERTa with a maximum sequence length of 128 tokens. The dataset is split into training and testing sets using an 80:20 stratified approach, yielding 18,554 training samples and 4,639 testing samples while preserving class distribution. SMOTE is applied exclusively to the training data, increasing the number of samples from 18,554 to 27,902 by generating synthetic minority-class instances, ensuring balanced learning while testing data remains untouched to provide a realistic evaluation.

C. BERT + LSTM Model Architecture

The BERT + LSTM hybrid model combines contextual understanding with sequential learning to improve classification performance. The input text is processed using DistilBERT, which generates contextual embeddings for each token with a maximum sequence length of 128, capturing the semantic meaning of the text. These embeddings are passed to an LSTM layer with 256 hidden units, enabling the model to learn the sequential flow and dependencies within the text. The output from the LSTM is passed through a fully connected (dense) layer and a dropout layer to improve generalization, followed by a final softmax classification layer that predicts whether the news is fake or real. By combining transformer-based contextual representations with sequence modelling, this hybrid approach enhances the system's ability to capture both meaning and structure in textual data.

D. RoBERTa Model Architecture

RoBERTa is used as an independent transformer-based model to provide strong contextual understanding of textual data. The input text is tokenized and processed using RoBERTa, which generates contextual embeddings for each token with a sequence length of up to 128, capturing the overall meaning and relationships within the text. The model uses the CLS-token sentence-level representation of the input sequence to perform classification through a dense layer, producing the final prediction as fake or real news. Compared to traditional models, RoBERTa is trained using improved optimization strategies, allowing it to better capture complex semantic patterns. Due to its strong contextual learning capability, RoBERTa achieves high individual performance and plays a significant role in improving the overall accuracy of the hybrid system.

E. LightGBM Model Architecture

LightGBM is used to capture statistical and feature-based patterns from textual data. Since machine learning models cannot process raw text directly, the cleaned text is first converted into numerical form using TF-IDF vectorization, with up to 5,000 TF-IDF features extracted using n-gram combinations to capture word frequency and contextual relevance. In addition, 15 linguistic features such as text length, word count, punctuation usage, and ratio-based attributes are extracted to capture structural characteristics of the text. These features are concatenated into a comprehensive feature vector of approximately 5,015 dimensions, which is supplied as input to the LightGBM gradient boosting classifier. By focusing on statistical and surface-level patterns, LightGBM complements the transformer-based models, which focus on contextual understanding.

F. Weighted Ensemble Model Architecture

The weighted ensemble approach combines predictions from the three independently trained models to improve overall classification performance. BERT+LSTM, RoBERTa, and LightGBM independently process the same input text and generate probability scores indicating the likelihood of the news being fake. These scores are combined using a weighted averaging strategy, where each model contributes based on its individual performance: RoBERTa is assigned a weight of 0.45, BERT+LSTM is assigned 0.35, and LightGBM is assigned 0.20. The final prediction score is computed as a weighted sum of the individual model outputs, and a threshold of 0.5 is applied to classify the news as fake or real. This strategy enables the system to leverage contextual, sequential, and statistical learning simultaneously, resulting in improved accuracy, better generalization, and more reliable predictions than any individual constituent model.

TABLE II
WEIGHTED ENSEMBLE CONFIGURATION

Model	Role	Assigned Weight
RoBERTa	Contextual understanding	0.45
BERT + LSTM	Contextual and sequential learning	0.35
LightGBM	Statistical / feature-based learning	0.20

V. SYSTEM REQUIREMENTS

A. Hardware Requirements

The hardware requirements for the proposed system are moderate. A processor of Intel i5 or higher is recommended, with a minimum of 8 GB RAM (16 GB recommended for smoother model training), 256 GB or higher storage, and an optional NVIDIA or AMD GPU for faster training of transformer models.

B. Software Requirements

The software environment consists of Python as the programming language, Jupyter Notebook or VS Code as the development environment, PyTorch, Transformers, and LightGBM as the core frameworks and libraries, Streamlit for the web interface, and Windows or Linux as the supported operating system.

C Packages and Libraries

The implementation relies on several Python libraries. PyTorch provides dynamic computation graphs for building and training the BERT+LSTM hybrid architecture with GPU acceleration. The Transformers library supplies pre-trained BERT and RoBERTa models and simplifies fine-tuning for the fake news detection task. LightGBM provides an efficient gradient boosting framework used with TF-IDF features. Pandas and NumPy support data manipulation, cleaning, and numerical computation. Scikit-learn provides TF-IDF vectorization, train-test splitting, and evaluation metrics. Imbalanced-learn supplies the SMOTE implementation for class-imbalance handling. Matplotlib and Seaborn are used for data visualization, NLTK for basic text preprocessing such as tokenization and stopword handling, and Streamlit for building the interactive web-based prediction interface.

VI. SYSTEM DESIGN

The system design follows a modular approach in which stages such as data preprocessing, feature extraction, model processing using BERT+LSTM, RoBERTa, and LightGBM, and final classification through a weighted ensemble are systematically organized, ensuring the system is easy to understand, develop, and maintain. Unified Modeling Language (UML) diagrams are used to represent the system architecture, user interactions, and the flow of data across modules. The class diagram represents the structural design of the system, including classes such as Dataset, Preprocessing, FeatureEngineering, ModelTrainer, the three constituent models, the PredictionModule, and the User, along with their methods and relationships. The sequence diagram illustrates the time-ordered interaction between these components across four phases: model training, prediction, ensemble computation, and final output display. The use case diagram identifies the user as the primary actor, with operations including providing input (text or file), obtaining a prediction, viewing the result, and viewing the confidence score. The activity diagram represents the end-to-end flow of operations, from user input through preprocessing, parallel model prediction, weighted ensemble combination, and final display of the classification result with its confidence score. Together, these diagrams provide a clear blueprint of both the structural and dynamic aspects of the system, reducing development complexity and supporting reliable implementation.

VII. IMPLEMENTATION

The system is implemented using a structured modular approach comprising eight stages. Data Collection uses the FakeNewsNet dataset obtained from Kaggle, containing labeled news titles categorized as real or fake. Data Preprocessing removes URLs, HTML tags, special characters, and extra whitespace using regular-expression-based cleaning, converts text to lowercase, applies tokenization suitable for transformer tokenizers, and performs stopword handling where necessary. Feature Engineering generates contextual embeddings using DistilBERT and RoBERTa, constructs TF-IDF representations with up to 5,000 features including n-grams, and extracts approximately 15 linguistic features such as text length, word count, and punctuation ratios, which are combined into a structured input vector for LightGBM. Train-Test Splitting divides the dataset using an 80:20 stratified ratio, with a validation split further used during training to monitor performance and prevent overfitting.

Handling Imbalanced Data applies SMOTE to the training data after feature engineering, creating synthetic samples for the minority (fake-news) class in the feature space to reduce bias toward the majority class. Model Training trains the BERT+LSTM hybrid (DistilBERT embeddings of dimension 768 passed through an LSTM), the independent RoBERTa (roberta-base) classifier, and LightGBM (trained on approximately 5,015 combined TF-IDF and linguistic features) independently, allowing each to learn a distinct representation of the data. Model Evaluation measures Accuracy, Precision, Recall, and F1-Score on the held-out test set, with validation monitoring used during training to avoid overfitting. Finally, Ensemble Prediction combines the probability scores of the three models using the predefined weights of RoBERTa (0.45), BERT+LSTM (0.35), and LightGBM (0.20) to produce the final classification.

The trained pipeline is deployed through a Streamlit-based web application that loads the trained models and the dataset, then renders a navigation sidebar offering Problem Statement, Sample Training Data, Data Preprocessing, Model Training, and Prediction pages. The Prediction page supports both single-text entry and batch CSV upload, returning the predicted label (Fake or Real) together with a confidence score and a class-probability breakdown, making the system practical for real-time, user-facing fake news verification.

VIII. TESTING

Testing is performed to ensure the correctness, reliability, and efficiency of the proposed system. The process verifies that text preprocessing, tokenization, feature extraction, model prediction using BERT+LSTM, RoBERTa, and LightGBM, the ensemble computation, and the user interface all function correctly and produce expected outputs. Unit testing evaluates individual components such as text preprocessing and cleaning, tokenization for transformer models, feature extraction for LightGBM, and the model prediction functions, identifying errors at an early stage. Integration testing validates the complete pipeline from input text through preprocessing, model inference, ensemble computation, to the user interface, ensuring smooth data flow and compatibility between modules. System testing evaluates the complete application with different inputs to verify the end-to-end prediction workflow, real-time input handling, ensemble prediction generation, and output display in the Streamlit interface. Acceptance testing confirms that the system meets the defined requirements from the user's perspective, validating accurate classification, consistent prediction results, smooth interaction, and reliable performance, confirming readiness for real-world deployment.

Performance is evaluated using standard classification metrics. Accuracy is computed as the ratio of correctly predicted instances to total predictions; precision as TP/(TP+FP), measuring how many predicted positive (fake) cases are actually correct; recall as TP/(TP+FN), measuring how many actual fake-news instances are correctly identified; and F1-score as the harmonic mean of precision and recall, 2×(Precision×Recall)/(Precision+Recall), providing a balanced measure when both false positives and false negatives carry consequences. A confusion matrix is additionally used to provide a detailed breakdown of correct and incorrect predictions across the real and fake classes.

IX. RESULTS AND ANALYSIS

The dataset comprises 23,193 labeled news samples drawn from FakeNewsNet, divided into 18,554 training samples (13,951 real, 75.19%; 4,603 fake, 24.81%) and 4,639 testing samples (3,488 real; 1,151 fake), preserving the original class proportions through stratified sampling. Applying SMOTE to the training data alone increased the training set from 18,554 to 27,902 samples by generating 9,348 synthetic minority-class instances, producing a balanced training distribution of 13,951 samples per class while the test set remained unmodified for unbiased evaluation.

On the held-out test set, the ensemble model attains an overall accuracy of approximately 93.23%. Individually, LightGBM achieves an accuracy of 86.25%, a vanilla BERT baseline achieves 88.53%, the BERT+LSTM hybrid improves this to 89.16%, and RoBERTa achieves the strongest individual performance at 90.51%.

The weighted ensemble, combining all three complementary models, surpasses every individual model, confirming that integrating contextual (RoBERTa), sequential (BERT+LSTM), and statistical (LightGBM) learning signals through weighted averaging yields a measurable improvement over any single architecture.

Receiver Operating Characteristic analysis corroborates this finding: the ensemble model attains the highest Area Under the Curve (AUC) of 0.961, ahead of RoBERTa (0.934), BERT+LSTM (0.924), vanilla BERT (0.932), and LightGBM (0.896), indicating that the ensemble offers superior separability between the real and fake classes across all decision thresholds, not only at the default 0.5 threshold used for the reported accuracy figures.

The confusion matrix for the ensemble model on the 4,639-sample test set shows 3,350 true negatives (real news correctly classified as real), 975 true positives (fake news correctly classified as fake), 138 false positives (real news misclassified as fake), and 176 false negatives (fake news misclassified as real). The concentration of predictions along the diagonal of the matrix confirms that the model maintains strong discriminative performance for both classes, with relatively few misclassifications in either direction.

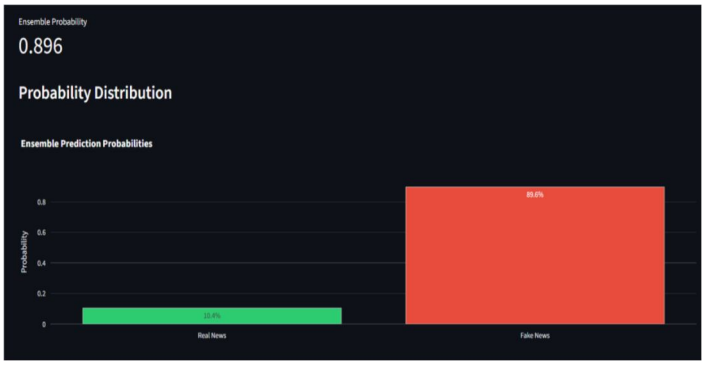


Figure 1. shows the predicted output for a sample

The deployed Streamlit interface provides a navigation menu allowing users to move between the Problem Statement, Sample Training Data, Data Preprocessing, Model Training, and Prediction pages. The Dataset Preview and Dataset Statistics views display sample records together with the overall composition of 23,193 total samples across 18,554 training and 4,639 test instances, while the Dataset Split view confirms the 80:20 stratified partition. The Class Distribution view visualizes the effect of SMOTE, contrasting the imbalanced original training set with the balanced post-SMOTE distribution. On the Prediction page, the system accepts either a single text input or a batch CSV file; for a representative single-text query, the system returned a classification of Fake News with a confidence of 89.6%, accompanied by a probability-distribution chart showing approximately 10.4% probability assigned to the real class and 89.6% to the fake class, illustrating the system's ability to deliver real-time, interpretable predictions to end users such as journalists and fact-checkers.

TABLE III

PERFORMANCE COMPARISON OF INDIVIDUAL AND ENSEMBLE MODELS

Model	Accuracy (%)	AUC
LightGBM	86.25	0.896
Vanilla BERT	88.53	0.932
BERT + LSTM	89.16	0.924
RoBERTa	90.51	0.934
Weighted Ensemble	93.23	0.961

Table III compares classification accuracy and AUC for each constituent model against the proposed weighted ensemble on the FakeNewsNet test set.

X. CONCLUSIONS

This paper demonstrates the effectiveness of a hybrid ensemble-based approach for fake news detection by integrating deep learning and machine learning models within a unified framework. The system combines BERT+LSTM, RoBERTa, and LightGBM to capture contextual, sequential, and statistical features from textual data, enabling a comprehensive understanding of news content.

The proposed model achieves an overall test accuracy of approximately 93%, along with strong precision, recall, and F1-score values, indicating reliable performance on unseen data. RoBERTa contributes strong contextual understanding, BERT+LSTM enhances sequential pattern learning, and LightGBM captures feature-based statistical patterns, while the weighted ensemble approach improves prediction stability and reduces the limitations of any individual model. The system follows a structured pipeline including preprocessing, SMOTE-based class balancing, feature extraction, and independent model training, ensuring effective learning and generalization, and the implementation of a Streamlit-based interface enables real-time prediction for both single and batch inputs, making the system practical and user-friendly. Overall, this work highlights that combining transformer-based models with traditional gradient-boosting techniques through weighted ensemble aggregation significantly improves fake news detection performance over single-model architectures, while remaining efficient, scalable, and suitable for real-world deployment.

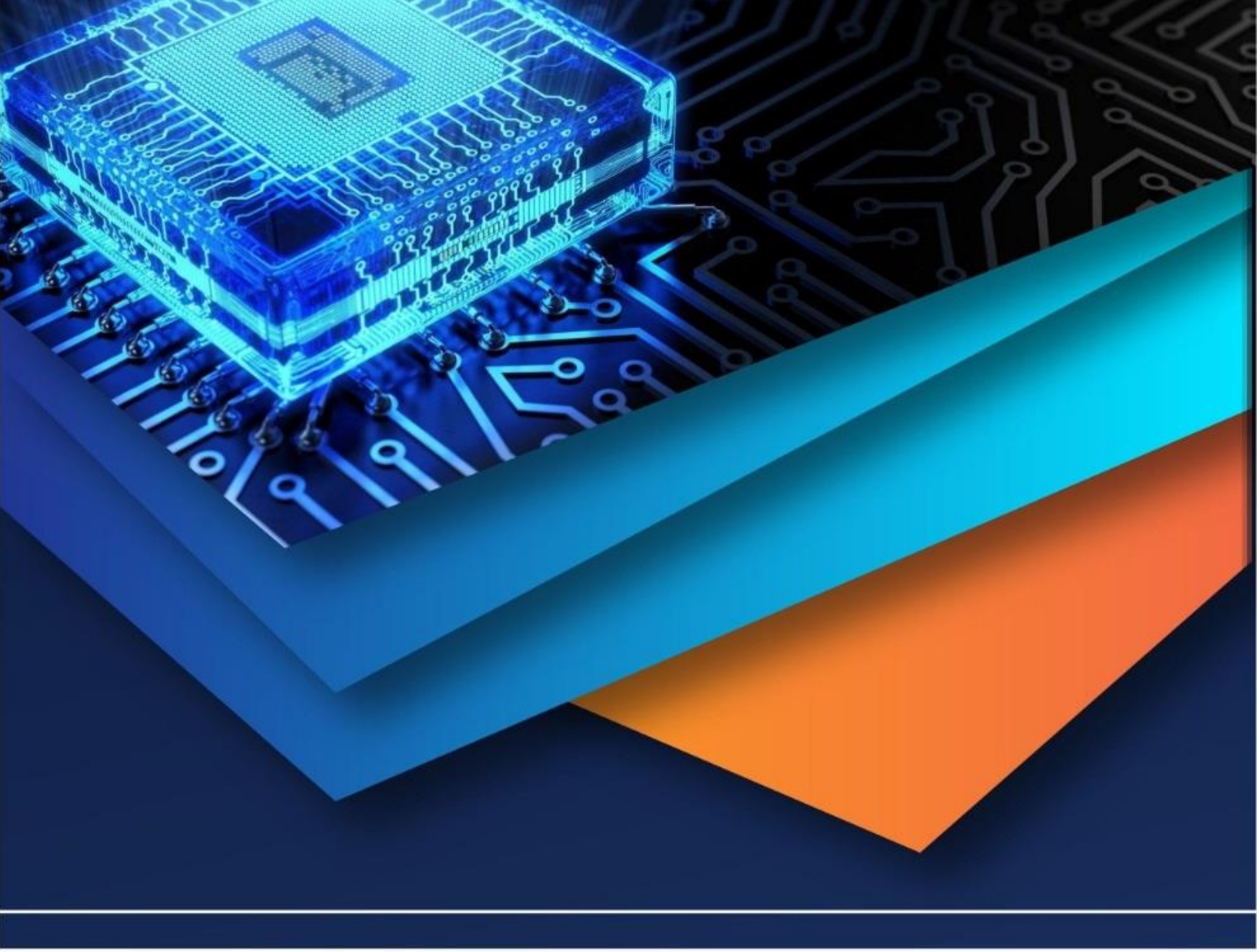
XI. FUTURE SCOPE

The proposed system can be further enhanced by integrating more advanced machine learning and deep learning architectures to improve prediction accuracy and handle complex linguistic patterns. Fine-tuning transformer models such as BERT and RoBERTa on larger and domain-specific datasets can significantly improve contextual understanding, and exploring advanced architectures and optimization techniques can further enhance model performance and efficiency. Another important direction is the implementation of a real-time detection system: the current system operates on a static dataset, but it can be extended to process live news feeds and social media data, enabling instant fake news detection and making it more practical for applications such as misinformation monitoring and automated content filtering. The system can also be deployed as a scalable web-based or cloud-based application to improve accessibility and support multiple concurrent users without performance degradation. Continuous model retraining with updated datasets can help the system adapt to evolving misinformation patterns and improve generalization, and the ensemble approach can be further enhanced by introducing dynamic weight adjustment or incorporating additional models to improve robustness. Finally, integrating explainable AI techniques can provide deeper insight into model predictions, enhancing transparency and helping users better understand the decision-making process, thereby increasing trust in the system.

REFERENCES

[1] A. Aggarwal, A. Chauhan, D. Kumar, M. Mittal, and S. Verma, “Classification of fake news by fine-tuning deep bidirectional transformers-based language model,” EAI Endorsed Trans. Scalable Inf. Syst., vol. 7, no. 27, 2020.
[2] A. Agarwal, M. Mittal, A. Pathak, and L. M. Goyal, “Fake news detection using a blend of neural networks: An application of deep learning,” SN Comput. Sci., vol. 1, no. 3, pp. 1–9, 2020.
[3] M. Albahar, “A hybrid model for fake news detection: Leveraging news content and user comments,” IET Inf. Secur., vol. 15, no. 2, pp. 169–177, 2021.
[4] S. Deepak and B. Chitturi, “Deep neural approach to fake news identification,” Procedia Comput. Sci., vol. 167, pp. 2236–2243, 2020.
[5] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” arXiv:1810.04805, 2018.
[6] C. I. Eke, A. A. Norman, and L. Shuib, “Context-based feature technique for sarcasm identification using deep learning and BERT model,” IEEE Access, vol. 9, pp. 48501–48518, 2021.
[7] M. Granik and V. Mesyura, “Fake news detection using naïve Bayes classifier,” in Proc. IEEE First Ukraine Conf. Electr. Comput. Eng. (UKRCON), 2017, pp. 900–903.
[8] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, “Fake news on Twitter during the 2016 U.S. presidential election,” Science, vol. 363, no. 6425, pp. 374–378, 2019.
[9] H. Jwa, D. Oh, K. Park, J. M. Kang, and H. Lim, “exBAKE: Automatic fake news detection model based on BERT,” Appl. Sci., vol. 9, no. 19, p. 4062, 2019.
[10] R. K. Kaliyar, A. Goswami, and P. Narang, “FakeBERT: Fake news detection in social media with a BERT-based deep learning approach,” Multimedia Tools Appl., vol. 80, pp. 11765–11788, 2021.
[11] Q. Li and W. Zhou, “Connecting the dots between fact verification and fake news detection,” arXiv:2010.05202, 2020.
[12] J. Lin, G. Tremblay-Taylor, G. Mou, D. You, and K. Lee, “Detecting fake news articles using deep learning,” in Proc. IEEE Int. Conf. Big Data, 2019, pp. 1–8.
[13] D. Mehta, A. Dwivedi, A. Patra, and M. A. Kumar, “A transformer-based architecture for fake news classification,” Soc. Netw. Anal. Min., vol. 11, no. 1, 2021.
[14] T. Mikolov et al., “Advances in pre-training distributed word representations,” in Proc. LREC, 2018.
[15] 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. EMNLP, 2017, pp. 2931–2937.
[16] N. Ruchansky, S. Seo, and Y. Liu, “CSI: A hybrid deep model for fake news detection,” in Proc. ACM Int. Conf. Inf. Knowl. Manage. (CIKM), 2017, pp. 797–806.
[17] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, “FakeNewsNet: A data repository with news content, social context, and spatiotemporal information,” Big Data, vol. 8, no. 3, pp. 171–188, 2020.
[18] M. Sundermeyer, R. Schlüter, and H. Ney, “LSTM neural networks for language modeling,” in Proc. INTERSPEECH, 2012.
[19] Z. Yang et al., “Hierarchical attention networks for document classification,” in Proc. NAACL, 2016, pp. 1480–1489.

[20] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "dEFEND: A system for explainable fake news detection," in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Min., 2019, pp. 395–405.
[21] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, "Fake news classification using transformer-based enhanced LSTM and BERT," Int. J. Cogn. Comput. Eng., vol. 3, pp. 98–105, 2022.
[22] E. Essa, K. Omar, and A. Alqahtani, "Fake news detection based on a hybrid BERT and LightGBM models," Complex Intell. Syst., vol. 9, pp. 6581–6592, 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY