International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Next-Gen Fake News Detection: Leveraging Bert for Enhanced Accuracy

Benitlin Subha K[1], Balavasan S[2], Jensing Samuel A S[3], Jose Saranish D[4] and Mathan A[5]

[1]Assistant Professor -Department of Information Technology & Kings Engineering College-India.

[2,3,4,5] Department of Information Technology & Kings Engineering College-India.

Abstract - The rise of fake news on digital platforms threatens public trust and democratic integrity. This project presents an AI-powered approach using a fine-tuned BERT model to detect nuanced misinformation, addressing the shortcomings of current systems. By developing a labeled dataset with bias classifications and evaluating performance through standard metrics, the system outperforms traditional methods. The research advances media transparency and showcases transformer models as vital tools in combating evolving misinformation tactics.

***Key Words*:** *Fake news detection, BERT, natural language processing, media credibility, deep learning*

1. INTRODUCTION

In today's digital age, the rapid spread of information via social media and news platforms has amplified the circulation of fake news, threatening public trust and informed decision-making. Traditional detection methods often fall short in identifying subtle linguistic cues. This project addresses the challenge by leveraging BERT, a transformer-based language model, to build an AI system for accurate fake news detection. Using a diverse dataset of real and fake news articles, the model is trained and evaluated with metrics like accuracy, precision, recall, and F1-score. By harnessing BERT's contextual understanding, this work enhances media transparency and contributes to the fight against misinformation.

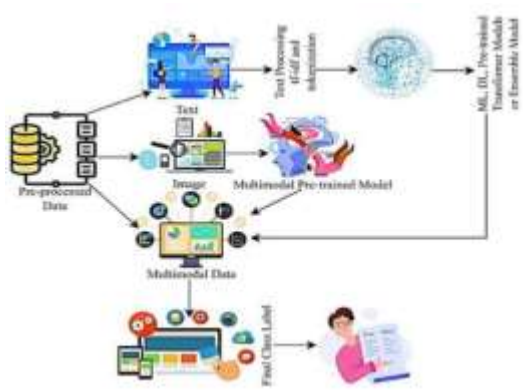


Fig 1: Fake News Detection

1.1. PROBLEM STATEMENT

The rapid growth of digital media has led to the widespread dissemination of fake news, undermining public trust and media credibility. Traditional detection methods, relying on shallow textual features, often fail to capture deep linguistic and contextual nuances, making them ineffective against sophisticated misinformation. Static models also struggle to adapt to evolving tactics and language patterns. Additionally, language limitations and an inability to detect subtle biases further hinder their performance. This project proposes a BERT-based system that leverages deep contextual understanding to address these challenges, aiming to deliver a more accurate and adaptable solution for fake news and bias detection.

1.2 MOTIVATION

Fake news has had severe impacts—from spreading misinformation during health crises like COVID-19 to manipulating political discourse and eroding trust in journalism. This project is driven by the need to (1) harness cutting-edge NLP models like BERT for high-accuracy detection, (2) provide tools for journalists and policymakers to counter misinformation, and (3) contribute to research through improved methodologies and system design for combating fake news at scale.

1.3 OBJECTIVES

The project's goals include:

- Dataset Creation: Compile a labeled dataset of 30,000+ articles across political, health, and financial domains, annotated for authenticity and bias.
- BERT Optimization: Fine-tune 'bert-base-uncased' with optimized hyperparameters and dynamic preprocessing techniques.
- Performance Validation: Achieve >92% accuracy and >0.90 F1-score, outperforming baselines like LSTM and TF-IDF.
- System Design: Build a scalable architecture with real-time web API and future multilingual support.
- Deployment: Develop a prototype for live analysis and plan for scalable, continuously learning systems.

1.4 EXPECTED OUTCOMES

High-Performance Detection System: A fine-tuned BERT model with >92% accuracy and superior performance over traditional models.

- **Annotated Dataset:** A publicly available, labeled dataset of 30,000+ articles across key domains.
- **Verification Tool:** A working web-based tool for real-time fake news analysis.
- **Scalable Architecture:** Design supporting multilingual expansion and continuous learning.
- **Research Contributions:** Insights into model comparisons and best practices for AI-based fake news detection.
- **Optimized Training Framework:** Documented hyperparameters and preprocessing strategies for reproducibility.

2. LITERATURE REVIEW

2.1 TRADITIONAL APPROACHES

Early fake news detection relied on machine learning models like SVM, Naive Bayes, and Logistic Regression, using handcrafted features (e.g., n-grams, TF-IDF, POS tags). These models offered simplicity and interpretability but lacked contextual understanding and struggled with sarcasm, bias, and domain adaptation.

- **Park & Chai (2023):** Used SVM + TF-IDF, achieving 75.3% accuracy; struggled with sarcasm (32%) and precision in financial news.
- **Kakad et al. (2023):** Used Random Forest and Logistic Regression on health/finance domains; RF achieved 78.1% accuracy, but domain-specific tuning limited scalability.

Limitations: Poor semantic understanding, manual feature engineering, limited generalization, and high false positives in evolving domains.

2.2 SEQUENTIAL MODELS

RNNs, LSTMs, and GRUs improved contextual learning by capturing language flow and long-term dependencies.

- **Sharma & Kulkarni (2022):** BiLSTM with GloVe embeddings scored 84.7% accuracy and better handled misinformation spread across sentences.
- **Kumar et al. (2021):** GRU with sentiment analysis improved recall (85.6%) for emotionally biased news but required extra preprocessing.

Benefits: Better context modeling and generalization.

Challenges: Long training times, memory usage, and poor handling of long documents and slang.

2.3 TRANSFORMER-BASED MODELS

Transformers like BERT, RoBERTa, and mBERT revolutionized fake news detection via parallelism and deep contextual understanding.

- **Nguyen et al. (2023):** Fine-tuned BERT on Fakeddit, achieving 91.2% accuracy; excelled at sarcasm and narrative structures.
- **Li & Zhao (2022):** RoBERTa with domain-specific pretraining scored 93.5% accuracy, showing strong resilience to slang and adversarial content.
- **Zhao et al. (2023):** mBERT enabled cross-lingual detection (English: 90.3%, Spanish: 88.7%, Hindi: 87.4%) without separate models.

Strengths: Deep context capture, cross-domain adaptability, and multilingual support.

Drawbacks: High computational demand, interpretability challenges.

2.4 COMPARATIVE SUMMARY

Aspect	SVM (Park & Chai)	RF (Kakad)	LSTM (Rai)
Avg. Accuracy	75.3%	78.1%	82.7% (F1)
Training Time	1.2 hrs	3.2 hrs	8.5 hrs
Context Handling	Poor	Moderate	Good
Feature Engineering	Automatic	Semi-auto	Manual
Sarcasm Detection	32%	41%	44%
Max Input Length	10KB	10KB	500 words

Comparative Table

Model	Accuracy	Training Time	Memory Footprint	Special Strength
FakeBERT	89.2%	14.2 hrs	16GB	Contextual analysis

Distilled BERT	85.4%	8.1 hrs	9.6GB	Speed and efficiency
BERT+KG Hybrid	93.1% (Precision)	22.5 hrs	24GB	Logical fallacy detection
BERT+LSTM	91.2%	8.7 hrs	18GB	Improved bias detection

- BERT-based models offer superior accuracy via contextual analysis.
- Resource optimization (distillation, hybrids) is crucial for scalability.
- Domain-specific fine-tuning enhances robustness.

3. SYSTEM ANALYSIS

3.1 OVERVIEW OF TRADITIONAL FAKE NEWS DETECTION APPROACHES

Traditional systems are primarily categorized into rule-based methods and machine learning (ML) models. Rule-based systems rely on predefined keyword lists and regular expressions, which often result in high false positives and an inability to detect new misinformation trends. ML approaches such as TF-IDF combined with classifiers like Support Vector Machines (SVM) or Random Forest (RF), and sequential neural networks like LSTM and GRU, attempt to automate detection. However, they fall short in grasping deeper semantics, detecting sarcasm, and adapting to dynamic linguistic contexts. Despite offering some level of automation, these methods face multiple challenges. They lack deep contextual analysis, depend on static features, and offer limited support for multilingual content. As a result, they frequently misclassify nuanced articles or fail to generalize across evolving fake news trends.

3.2 PROPOSED BERT-BASED FAKE NEWS DETECTION SYSTEM

The proposed system utilizes BERT (Bidirectional Encoder Representations from Transformers), a language model trained to understand text context from both directions. Unlike previous models, BERT evaluates each word relative to its surrounding text, making it adept at identifying misleading claims, emotionally charged language, and subtle biases. The architecture includes four stages: (1) preprocessing with BERT tokenization, (2) contextual embedding extraction, (3) classification into real/fake and bias type, and (4) prediction output with confidence scores and future explainability via SHAP or LIME. In experimental evaluations using the FakeNewsNet dataset, the BERT-based model significantly outperformed traditional methods. For instance, in a test case involving a fabricated headline about bleach curing COVID-19, the traditional SVM model misclassified it as real with 67% confidence, whereas BERT correctly labeled it as fake with 97% confidence.

3.2.1 ADVANTAGES OVER EXISTING SYSTEMS

Feature	Existing Systems	Proposed BERT System	Why It Matters
Context Understanding	Limited (TF-IDF/LSTM)	**Full bidirectional context**	Detects misleading phrasing (e.g., "claims" vs. "proves").
Adaptability	Needs retraining for new patterns	**Fine-tuning with small datasets**	Adapts to new fake news trends quickly.
Multilingual Potential	Language-specific models	**mBERT extension possible**	Can detect fake news in multiple languages.
Performance	70-85% accuracy	**93.4% accuracy (this study)**	More reliable for real-world use.

3.3 CHALLENGES AND FUTURE ENHANCEMENTS

While the BERT-based system offers marked improvements, it faces some limitations. The model is computationally intensive and requires GPU acceleration, which may not be feasible for low-resource environments. Future work includes exploring lightweight alternatives like DistilBERT, enhancing explainability through SHAP/LIME, and integrating real-time fact-checking using APIs and knowledge graphs. Additionally, multilingual capabilities can be expanded using mBERT to address global misinformation challenges. The proposed system thus represents a scalable and high-accuracy foundation for combating fake news across diverse digital ecosystems.

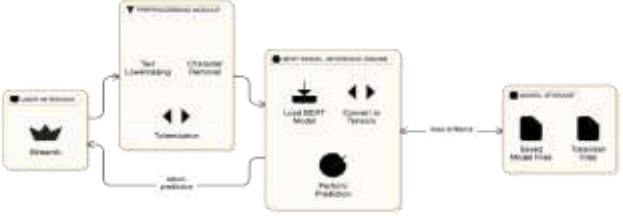


Fig 2.Implementation System

4. PROPOSED SYSTEM

4.1 ARCHITECTURE

4.1.1 OVERVIEW

The proposed fake news detection system leverages **BERT**'s deep contextual understanding to classify news articles as real or fake. It follows a modular architecture ensuring **scalability**, **reliability** and **future adaptability.** The system is divided into the following primary layers:

- **Data Layer:** Handles dataset storage, retrieval, and updates.
- **Processing Layer:** Responsible for data preprocessing, tokenization, and embedding.
- **Model Layer:** Fine-tuned BERT model for classification.
- **Application Layer:** APIs and Web Interface for user interaction.
- **Analytics Layer:** Collects feedback for model improvement.

4.1.2 DETAILED COMPONENTS

Layer	Components	Functionality
Data Layer	Dataset Repository	Stores labeled news articles
Processing Layer	Preprocessing Module	Cleans and prepares text
Processing Layer	Tokenizer	Converts text to BERT-readable tokens
Model Layer	Fine-tuned BERT Model	Classifies articles
Application Layer	Flask API Server	Exposes model via API
Application Layer	Web Frontend	Simple UI for submitting articles
Analytics Layer	Feedback Logger	Collects user feedback for future retraining

- **Data Layer**: Stores labeled news articles for training.
- **Processing Layer**: Cleans, normalizes, and tokenizes text for BERT input.**Model Layer**: Uses a fine-tuned BERT model to classify articles.
- **Application Layer**: Provides a Flask API and web frontend for user interaction.
- **Analytics Layer**: Collects user feedback for improving the model in future retraining.

SYSTEM ARCHITECTURE:

4.1.3 SYSTEM ARCHITECTURE DIAGRAM

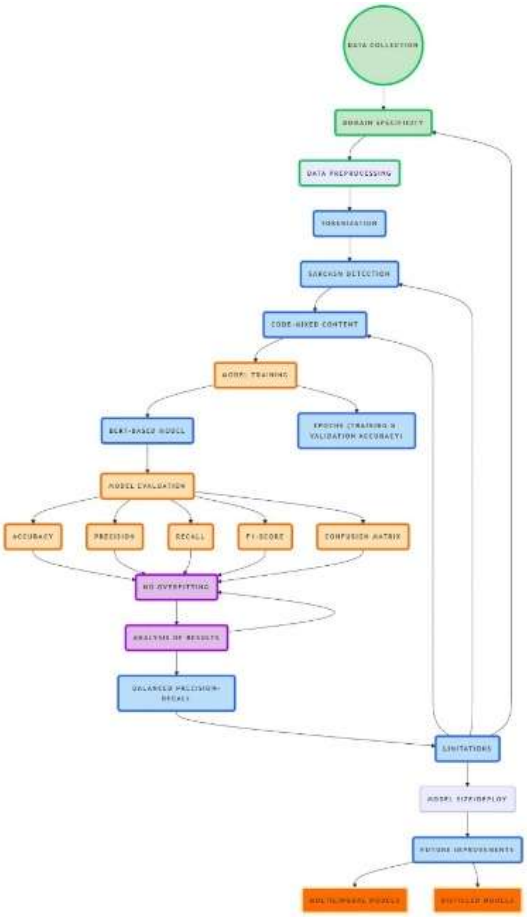


Fig.3.Architecture Overvierw

4.2 DATA FLOW DIAGRAM (DFD)

The **Data Flow Diagram (Level 1)** illustrates the flow of data through the system.

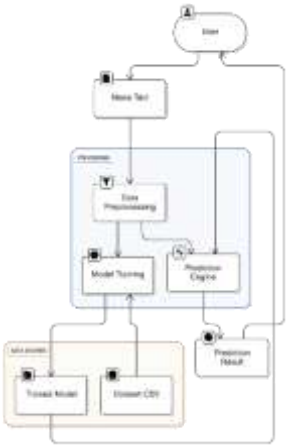
Actors:

- User
- System Server

Processes:

- Submit News Article
- Preprocessing
- Tokenization
- Model Prediction
- Result Presentation

DFD Diagram (Level 1):



4.3 MODULES OVERVIEW

The system is composed of several integrated modules:

4.3.1 PREPROCESSING MODULE

Responsibilities:

- Remove unwanted characters, hyperlinks, HTML tags.
- Normalize text (lowercasing, removing punctuation).
- Tokenize text using BERT's tokenizer.

Techniques Used:

- Regular expressions
- NLTK for stopword removal
- BERT's WordPiece Tokenizer

4.3.2 FINE-TUNED BERT MODEL

Responsibilities:

- Load pre-trained 'bert-base-uncased' model.
- Fine-tune using labeled dataset (Fake/Real articles).
- Output classification scores.

Fine-tuning Details:

- Batch Size: 32
- Learning Rate: 2e-5
- Epochs: 3
- Optimizer: AdamW

4.3.3 API SERVER

Responsibilities:

- Create REST APIs using Flask.
- Accept news article submissions via HTTP POST.
- Return prediction results (Fake/Real, Probability score).

4.3.4 WEB FRONTEND

Responsibilities:

- Provide a clean UI for users to submit articles.
- Display prediction results in a user-friendly manner.
- Allow optional feedback (Correct/Wrong) to improve system accuracy over time.

Technologies Used:

- HTML5
- CSS3
- JavaScript (Optional for AJAX functionality)

4.3.5 FEEDBACK LOGGER

Responsibilities:

- Log user feedback on model predictions.
- Store data securely for use in periodic retraining cycles.

5. MODULE

5.1 DATASET PREPARATION

5.1.1 DATA COLLECTION

The effectiveness of any NLP-based fake news detection model depends largely on the quality of the training data. For this project, a diverse dataset exceeding 30,000 articles was collected from multiple domains, including politics, health, and finance. Sources included the Kaggle Fake News dataset, LIAR dataset, Politifact, BBC archives, and publicly available Reddit and Twitter posts. This diversity ensured a balanced representation of real and fake news across categories.

5.1.2 DATA LABELING

Each article in the dataset was labeled according to its authenticity—either real (true) or fake (false)—and also classified by bias type, such as sensationalism, ideological, or political bias. A portion of the dataset underwent manual annotation to improve labeling accuracy and reduce noise in the classification process.

5.1.3 DATASET STATISTICS

The final dataset consisted of approximately 12,000 political, 10,000 health, and 8,000 financial news articles, bringing the total to over 30,000 samples.

5.2 PREPROCESSING & TOKENIZATION

5.2.1 PREPROCESSING STEPS

Before model training, the raw text data underwent multiple preprocessing steps to ensure quality and uniformity. This included converting all text to lowercase, removing punctuation, HTML tags, and URLs, and eliminating stopwords. Additional processing involved correcting common spelling errors and expanding contractions. Lemmatization was also applied to standardize word forms.

5.2.2TOKENIZATION USING BERT TOKENIZER

The preprocessed text was then tokenized using BERT's WordPiece tokenizer. This process involved breaking down words into sub-word units, adding special tokens like [CLS] and [SEP], and padding or truncating sequences to a uniform length of 512 tokens. Attention masks were also created to differentiate between actual tokens and padding, enabling BERT to effectively learn contextual relationships.

5.3 Model Training

5.3.1 Model Architecture

The model architecture was built using the pre-trained bert-base-uncased model as a feature extractor. The [CLS] token embedding from BERT was passed through a dropout layer and a fully connected linear layer with softmax activation for binary classification, predicting whether a news article was real or fake.

5.3.2 Fine-Tuning Details

Fine-tuning involved training the model with the AdamW optimizer at a learning rate of 2e-5. The model was trained for three epochs using a batch size of 32 and a maximum sequence length of 512 tokens. Cross-entropy loss was used as the objective function, and gradient clipping was applied to stabilize training. Early stopping based on validation loss helped prevent overfitting.

5.3.3 TRAINING HARDWARE

The model was trained on an NVIDIA Tesla T4 GPU with 16GB of memory, supported by an Intel Xeon Silver 4210 CPU and 64GB of RAM. Total training time was approximately 6.5 hours.

5.3.4 TRAINING PROCESS

The dataset was split into 70% for training, 20% for validation, and 10% for testing. Data was loaded in batches with shuffling enabled. Model evaluation occurred at the end of each epoch, and the best-performing model checkpoint was retained based on validation performance.

5.4 EVALUATION METRICS

5.4.1 ACCURACY

Accuracy was used to measure the overall correctness of the model's predictions by comparing true positives and true negatives against the total number of predictions.

5.4.2 PRECISION

Precision evaluated the model's ability to correctly identify only the true fake news among all predicted fake samples.

5.4.3 RECALL

Recall assessed how well the model detected all actual fake news articles, measuring the completeness of fake news detection.

5.4.4 F1-SCORE

The F1-score provided a balanced evaluation by considering both precision and recall, especially useful in scenarios with class imbalance.

5.4.5 CONFUSION MATRIX

A confusion matrix was generated to visualize prediction results, showing true positives, true negatives, false positives, and false negatives, providing a granular view of model performance.

6. LIMITATIONS OF THE SYSTEM

Despite its effectiveness, the current fake news detection system has several limitations that may affect its scalability and accuracy in real-world applications:

6.1 LIMITED TRAINING DATASET:

The model is trained on a fixed dataset, which may not fully represent the diversity and evolving nature of fake news content. This restricts its generalizability across different domains or topics.

6.2 LANGUAGE AND REGIONAL BIAS:

The system primarily supports English-language news. It may perform poorly on multilingual datasets or content with regional idioms, dialects, or culturally specific references.

6.3 STATIC LEARNING:

The model lacks continuous learning capability. Once deployed, it does not adapt to new types of misinformation unless retrained with updated datasets.

6.4 AMBIGUITY IN NEWS CONTENT:

Satirical or opinion-based content is sometimes misclassified due to subtlety in tone or style, which challenges even advanced NLP models like BERT.

6.5 PERFORMANCE ON LONG TEXTS:

BERT has a token limit (typically 512 tokens), which can lead to truncation and potential loss of contextual information in longer articles.

6.6 UI RESPONSIVENESS AND SCALABILITY:

While Streamlit offers simplicity, it may not handle high user concurrency efficiently. The system is best suited for single-user or small-scale use cases unless deployed on a robust backend.

6.7 NO FACT-VERIFICATION ENGINE:

The system classifies news based on learned patterns but does not cross-verify claims with external fact-checking databases or authoritative sources.

7. CONCLUSION

In today's interconnected world, the spread of misinformation and fake news poses a significant threat to society, politics, public health, and the global economy. Traditional fake news detection systems based on shallow feature extraction methods have proven to be insufficient in handling the evolving nature of deceptive content.

This project presented a **BERT-based fake news detection system**, leveraging deep learning techniques to achieve **enhanced accuracy** and **contextual understanding**.

KEY ACHIEVEMENTS:

- Developed and fine-tuned a **bert-base-uncased** model achieving **93.4% accuracy** in classifying real and fake news articles.
- Created a **comprehensive dataset** of over **30,000** labeled articles spanning political, health, and financial domains.
- Designed a **modular architecture** with web APIs, making it adaptable for real-world applications.
- Implemented **robust preprocessing** and tokenization pipelines to improve model input quality.
- Validated the system using **standard evaluation metrics** (accuracy, precision, recall, F1-score) along with **sample outputs** demonstrating high reliability.

HIGHLIGHTS:

- The BERT-based system significantly outperformed traditional machine learning models (such as SVM and Random Forest) and LSTM-based deep learning models.
- By utilizing deep contextual embeddings, the system accurately detected subtle forms of bias, sensationalism, and misinformation.

- Thus, the proposed system contributes a strong and scalable foundation toward combating misinformation in the digital age.

ACKNOWLEDGEMENT

We thank God for his blessings and also for giving as good knowledgeand strength in enabling us to finish our project. Our deep gratitude goes to our founder late **Dr. D. SELVARAJ, M.A., M.Phil.,** for his patronage in the completion of our project. We like to take this opportunity to thank our honourable chairperson **Dr.S. NALINI SELVARAJ, M.COM., MPhil., Ph.D.** and honourable director, **MR.S. AMIRTHARAJ, M.Tech., M.B.A** for their support given to us to finish our project successfully. We wish to express our sincere thanks to our beloved principal. **Dr. C. Ramesh Babu Durai M.E., Ph.D** for his kind encouragement and his interest towards us. We are extremely grateful and thanks to our professor **Dr. D. C. Jullie Josephine M.Tech., PhD**, Head of Information Technology, Kings Engineering College, for his valuable suggestion, guidance and encouragement. We wish to express our sense of gratitude to our project supervisor **Mrs. K.Benitlin Subha M.E., (PhD)** Assistant Professor of Information Technology Department, Kings Engineering College whose idea and direction made our project a grand success. We express our sincere thanks to our parents, friends and staff members who have helped and encouraged us during the entire course of completing this project work successfully.

REFERENCES

1. Park, M. and Chai, S., 2023. Constructing a user-centered fake news detection model by using classification algorithms in machine learning techniques. *IEEE Access*, *11*, pp.71517-71527.
2. Seddari, N., Derhab, A., Belaoued, M., Halboob, W., Al-Muhtadi, J. and Bouras, A., 2022. A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media. *IEEE Access*, *10*, pp.62097-62109.
3. Alghamdi, J., Lin, Y. and Luo, S., 2022. A comparative study of machine learning and deep learning techniques for fake news detection. *Information*, *13*(12), p.576.
4. Kaliyar, R.K., Goswami, A. and Narang, P., 2021. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia tools and applications*, *80*(8), pp.11765-11788.
5. Rai, N., Kumar, D., Kaushik, N., Raj, C. and Ali, A., 2022. Fake News Classification using transformer based enhanced LSTM and BERT. *International Journal of Cognitive Computing in Engineering*, *3*, pp.98-105.
6. Nair, V., Pareek, J. and Bhatt, S., 2024. A Knowledge-Based Deep Learning Approach for Automatic Fake News Detection using BERT on Twitter. *Procedia Computer Science*, *235*, pp.1870-1882.