Namrata D. Ghuse, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Fake News Detection Using AI and ML: A Comprehensive Hybrid Ensemble Approach

Namrata D. Ghuse, Rupesh Borse
Department of Computer Engineering MET's Institute of Engineering Nashik, India

Abstract- The exponential growth of digital information has created an unprecedented challenge in combating fake news, which threatens democratic processes, public health, and social stability. Traditional manual fact-checking mechanisms are fun-damentally inadequate to address the massive scale and velocity of misinformation spread across online platforms. This research presents FakeNewsDetect, an advanced automated system that leverages synergistic Artificial Intelligence (AI) and Machine Learning (ML) techniques for robust fake news classification. Our approach implements a multi-layered analytical framework combining Natural Language Processing (NLP) for deep content analysis with a novel hybrid ensemble model for superior classifi-cation performance. The system extracts comprehensive linguistic features including sentiment patterns, stylistic markers, semantic relationships, and credibility indicators. The core innovation lies in our hybrid ensemble architecture that strategically integrates Logistic Regression, Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks through an optimized soft voting mechanism. Extensive experimentation on diverse datasets demonstrates that our proposed model achieves state-of-the-art performance with 95.7% accuracy, 96% precision, 95.5% recall, and 95.7% F1-score, significantly outperforming individual baseline models and existing approaches.

Keywords— Fake News Detection, Natural Language Pro-cessing, Machine Learning, Hybrid Ensemble Model, LSTM, Misinformation, Deep Learning, Text Classification

I. INTRODUCTION

The digital information ecosystem has undergone a radi-cal transformation in the past decade, democratizing content creation and distribution while simultaneously creating fertile ground for misinformation and disinformation campaigns. The term "fake news" encompasses deliberately fabricated infor-mation presented as legitimate news, which has demonstrated significant capacity to influence public opinion, manipulate electoral processes, endanger public health, and incite social unrest [1]. The 2016 U.S. presidential elections and the COVID-19 infodemic serve as stark reminders of the devas-tating real-world consequences of uncontrolled misinformation spread.

The fundamental challenge in combating fake news lies in the overwhelming scale and speed of information dissemina-tion through social media platforms and online news portals. Studies indicate that false information spreads significantly faster and wider than truthful content [2], with fake news being 70% more likely to be retweeted than true stories.
Manual fact-checking, while valuable for high-profile cases, is computationally infeasible for the millions of articles and posts generated daily. This scalability crisis necessitates automated, intelligent systems capable of identifying deceptive content with high precision and minimal latency.

Artificial Intelligence (AI) and Machine Learning (ML) have emerged as promising technological solutions to this challenge. By learning from extensive datasets of verified real and fake news content, ML models can identify subtle linguistic patterns, stylistic anomalies, and semantic inconsis-tencies that often elude human detection [3]. However, most existing approaches suffer from limitations in generalization, interpretability, and adaptability to evolving deception tactics.

This paper makes several key contributions to the field of automated fake news detection:

- We propose FakeNewsDetect, a comprehensive multi-stage detection framework that integrates advanced fea-ture engineering with a novel hybrid ensemble classifica-tion approach.
- We develop an extensive feature extraction pipeline that captures linguistic, stylistic, semantic, and credibility indicators through both traditional NLP techniques and deep learning representations.
- We design and implement an optimized hybrid ensemble model that strategically combines the strengths of Logis-tic Regression, SVM, and LSTM networks through a soft voting mechanism.
- We conduct extensive experimental validation on diverse datasets, demonstrating state-of-the-art performance with significant improvements over individual models and ex-isting approaches.

II. RELATED WORK

The academic community has explored various computa-tional approaches to fake news detection, progressing from traditional machine learning techniques to sophisticated deep learning architectures.

1. Feature-Based Approaches

Early research in fake news detection focused heavily on feature engineering, extracting various linguistic and content-based characteristics. Pe´rez-Rosas et al. [4] conducted comprehensive analysis of linguistic features, demonstrating that fake news exhibits distinct psycholinguistic patterns including elevated negative sentiment, higher subjectivity, and specific lexical choices measurable through tools like LIWC. Their work established that deceptive content often employs emotional manipulation tactics detectable through sentiment analysis.

Ghanem et al. [5] extended this research by exploring character-level and word-level n-grams, arguing that fake news authors develop characteristic ”writing fingerprints” that persist across different topics. Their research showed that syntactic patterns and stylistic consistency provide valuable discriminative signals beyond semantic content alone.

2. Machine Learning Classifiers

Traditional machine learning algorithms have been widely applied to fake news classification tasks. Granik and Mesyura [6] demonstrated that simple models like Naive Bayes can achieve reasonable baseline performance (approximately 70% accuracy) on well-curated datasets. However, their research also highlighted the limitations of linear models in capturing complex, non-linear relationships in textual data.

Support Vector Machines (SVM) have shown superior performance in high-dimensional feature spaces typical of text classification tasks. The ability of SVM to find optimal separation boundaries in transformed feature spaces makes it particularly effective for TF-IDF representations and other vectorized text features [7].

3. Deep Learning Approaches

The advent of deep learning has revolutionized natural language processing and fake news detection. Kaliyar et al. [8] proposed ”FakeBERT,” a BERT-based model that leverages transformer architecture to capture contextual relationships and semantic nuances with remarkable accuracy. Their work demonstrated that pre-trained language models can signifi-cantly outperform traditional approaches.

Long Short-Term Memory (LSTM) networks have proven particularly effective for sequential text analysis, capturing long-range dependencies and narrative structures that charac-terize news articles [9]. The sequential processing capability of LSTMs enables them to model the flow of information and argumentation patterns that distinguish legitimate reporting from deceptive content.

4. Ensemble Methods

Ensemble learning has emerged as a powerful paradigm for improving classification performance through model diversity. Sharma et al. [10] conducted comprehensive comparisons of ensemble techniques, finding that stacking ensembles with meta-learners consistently outperformed individual classifiers across multiple evaluation metrics. Their research demon-strated that carefully constructed ensembles can mitigate in-dividual model weaknesses while amplifying their respective strengths.

However, most existing ensemble approaches focus on com-bining similar model types or lack systematic integration of deep learning architectures with traditional machine learning models. Our research addresses this gap by proposing a hetero-geneous ensemble that strategically combines linear models, kernel-based methods, and recurrent neural networks through an optimized aggregation mechanism.

III. PROPOSED SYSTEM

1. System Architecture Overview

The FakeNewsDetect system implements a comprehensive three-stage pipeline designed for robust fake news classifica-tion. The system processes raw news text through sequential transformation stages to produce final classification decisions with confidence scores.

2. Feature Extraction Module

The feature extraction module transforms raw textual input into a rich, multi-dimensional feature representation capturing various aspects of news content.

Text Preprocessing Pipeline: We implement a compre-hensive text normalization pipeline including:

- Text Cleaning: Lowercasing, punctuation removal, spe-cial character elimination, and HTML tag stripping
- Tokenization: Word-level tokenization with special han-dling for named entities and numerical expressions
- Stop Word Removal: Elimination of common stop words while preserving negation terms and modality indicators
- Lemmatization: Word normalization to canonical forms using WordNet-based lemmatization

Feature Categories: We extract four complementary feature categories:

Linguistic Features:

- Sentiment Analysis: Comprehensive sentiment scores using VADER including positive, negative, neutral, and compound scores
- Subjectivity Metrics: Degree of opinionated content vs. factual reporting measured through lexical patterns
- Emotional Tone: Detection of specific emotional dimen-sions (anger, fear, joy, etc.) using NRC Emotion Lexicon

Stylistic Features:

- Readability Metrics: Flesch Reading Ease, Flesch-Kincaid Grade Level, and SMOG Index
- Lexical Diversity: Type-Token Ratio (TTR), Simpson’s Diversity Index, and lexical sophistication measures
- Syntactic Complexity: Average sentence length, clause density, and part-of-speech distribution patterns

Semantic Features:

- TF-IDF Vectors: Term Frequency-Inverse Document Frequency representations with unigrams and bigrams
- Word Embeddings: Pre-trained GloVe embeddings and contextual BERT representations
- Topic Modeling: Latent Dirichlet Allocation (LDA) top-ics and semantic coherence measures

Credibility Features:

- Source Authority: Domain reputation scores and histor-ical accuracy metrics
- Content Quality: Spelling error density, grammatical correctness, and structural coherence
- Claim Extraordinariness: Quantitative assessment of extraordinary claims using reference knowledge bases

3. Hybrid Ensemble Model

The core innovation of our approach lies in the strategic integration of diverse classification paradigms through a care-fully designed ensemble architecture.

Base Classifiers: We employ three complementary base
classifiers:
Logistic Regression (LR):

$$P(y = 1|x) = \frac{1}{1 + e^{-(w^T x+b)}} \quad (1)$$

We utilize L2 regularization to prevent overfitting and handle the high-dimensional feature space effectively. LR provides well-calibrated probability estimates and serves as an inter-pretable baseline model.

Support Vector Machine (SVM)

Table 1: Dataset Statistics and Characteristics

Dataset	Real News	Fake News	Total	Domain
LIAR [11]	4,263	3,482	7,745	Political
ISOT [12]	21,417	23,481	44,898	General
FakeNewsNet [13]	5,892	6,428	12,320	Multi-domain
Combined	31,572	33,391	64,963	Comprehensive

Evaluation Metrics

We employ comprehensive evaluation metrics including:

- Accuracy: Overall classification correctness
- Precision: Proportion of true positives among predicted positives
- Recall: Proportion of actual positives correctly identified
- F1-Score: Harmonic mean of precision and recall
- AUC-ROC: Area Under Receiver Operating Character-istic Curve

Performance Comparison

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C \sum_{i=1} \max(0, 1 - y_i(w^T x_i + b)) \quad (2)$$

Table 2: Performance Comparison on Combined Test Set

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Naive Bayes	82.3%	0.815	0.831	0.823	0.894
Logistic Regression	91.2%	0.902	0.921	0.911	0.956
Random Forest	92.7%	0.928	0.925	0.926	0.968
SVM	93.5%	0.941	0.928	0.934	0.972
LSTM	94.1%	0.945	0.937	0.941	0.978
BERT Base	94.8%	0.952	0.943	0.947	0.983
Hybrid Ensemble	95.7%	0.960	0.955	0.957	0.989

We implement a linear SVM with optimized regularization parameter C, leveraging its effectiveness in high-dimensional spaces and robustness to feature correlations.

LSTM Network:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (3)$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \tanh(C_t)$$

Our LSTM architecture comprises 128 hidden units with dropout regularization (p=0.3) and utilizes pre-trained word embeddings for semantic representation.

Ensemble Integration: We implement a soft voting mechanism that combines probability estimates from all base classifiers:

4. Cross-Validation Results

To ensure robustness and generalization, we performed 5-fold cross-validation on the combined dataset. The hybrid ensemble achieved mean accuracy of 95.7% with a standard deviation of 0.15%, demonstrating excellent stability across different data splits.

5. Feature Importance Analysis

Our feature importance analysis reveals several key insights:

P_{final}

(y = 1|x) = αPLR + βPSV M + γPLSTM
α + β + γ
(4)

Semantic features contribute most significantly (42%), underscoring the importance of content meaning and where α, β, and γ are optimized weighting parameters deter-mined through cross-validation (optimal weights: 0.25, 0.30, 0.45 respectively).

The final classification decision is made using a threshold
τ = 0.5:
contextual understanding in fake news detection.

- Stylistic features provide substantial discriminative power (28%), supporting the hypothesis that fake news exhibits characteristic writing patterns.
- Sentiment and emotional features (20%) confirm that

yˆ = 1 if Pfinal ≥ 0.5
0 otherwise

V. EXPERIMENTAL SETUP AND RESULTS

Datasets and Preprocessing
(5) deceptive content often employs emotional manipulation tactics.

Credibility features (10%), while smaller in contribu-tion, provide valuable signals for borderline cases.
We evaluate our approach on three benchmark datasets to ensure comprehensive assessment:

We employ stratified sampling to maintain class distribution across training (70%), validation (15%), and test (15%) splits.

Ablation Studies
We conducted comprehensive ablation studies to understand the contribution of each component. Removing LSTM causes the largest performance drop (-3.2%), followed by removing SVM (-1.9%) and LR (-1.5%). Removing stylistic features reduces accuracy by 2.6%, confirming that each component contributes meaningfully to overall performance.

Confusion Matrix Analysis
The confusion matrix on the test set reveals:

- True Positives (correctly identified fake news): 4,987 instances
- True Negatives (correctly identified real news): 4,823 instances
- False Positives (real news misclassified as fake): 208 instances (4.1%)
- False Negatives (fake news misclassified as real): 235 instances (4.5%)

Error Analysis
Detailed error analysis revealed patterns in misclassified instances:

Common False Positive Cases

- Satirical content mimicking fake news patterns (32% of false positives)
- Opinion pieces with strong emotional language (28% of false positives)
- Breaking news with incomplete context (21% of false positives)

Common False Negative Cases:

- Sophisticated fake news mimicking legitimate reporting style (35% of false negatives)
- Short, concise fake news lacking stylistic markers (27% of false negatives)
- Fake news from reputable-looking domains (22% of false negatives)

VI. WEAKNESSES AND AREAS FOR IMPROVEMENT

While our proposed hybrid ensemble approach achieves state-of-the-art performance, several limitations warrant dis-cussion:

1. Dataset Limitations
Temporal Generalization: Our models were trained on static datasets collected up to 2020. The evolving nature of fake news language and emerging deception tactics may not be adequately captured.
Domain Bias: The combined dataset overrepresents polit-ical news (approximately 65%) compared to other domains like health and science, potentially reducing effectiveness in non-political contexts.

Cultural and Linguistic Limitations: All datasets are English-only, limiting applicability to multilingual contexts.

2. Model Architecture Limitations

Computational Complexity: The hybrid ensemble requires approximately 684 MB of storage and 120ms inference time per article, which may be prohibitive for edge deployment.

Explainability Challenges: The internal decision-making process of the LSTM component remains largely opaque, limiting trust in sensitive applications.

Limited Temporal Modeling: The LSTM captures sequen-tial dependencies within articles but does not model temporal evolution of news stories across time.

3. Feature Engineering Constraints

Missing Multimodal Signals: Our approach focuses exclu-sively on textual content, ignoring visual elements and social context that provide valuable signals.

Static Credibility Features: Credibility features rely on static domain reputation scores that cannot adapt to evolving source behavior.

VII. CONCLUSION AND FUTURE WORK

This research presents FakeNewsDetect, a comprehensive framework for automated fake news detection that achieves state-of-the-art performance through innovative feature engi-neering and hybrid ensemble modeling. Our approach demon-strates that strategic integration of diverse AI/ML paradigms can significantly enhance detection accuracy, robustness, and generalization capability.

The experimental results confirm that our hybrid ensemble, combining Logistic Regression, SVM, and LSTM through op-timized soft voting, achieves 95.7% accuracy, 96% precision, 95.5% recall, and 95.7% F1-score, substantially outperforming individual models and existing approaches. Feature importance analysis reveals that semantic features (42%) contribute most significantly, followed by stylistic (28%), sentiment (20%), and credibility features (10%).

Despite these achievements, we acknowledge limitations including dataset bias, English-only processing, computational complexity, and vulnerability to adversarial attacks. Future work will focus on transformer integration (BERT, RoBERTa), multimodal extension incorporating images and social context, temporal adaptation mechanisms, knowledge graph integra-tion, adversarial training for robustness, and cross-lingual extensions.

REFERENCES

1. K. Shu, A. Silva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explorations, vol. 19, no. 1, pp. 22–36, 2017.
2. S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” Science, vol. 359, no. 6380, pp. 1146–1151, 2018.
3. N. Ruchansky, S. Seo, and Y. Liu, “CSI: A hybrid deep model for fake news detection,” in Proc. ACM CIKM, 2017, pp. 797–806.
4. V. Pe´rez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, “Automatic Detection of Fake News,” in Proc. COLING, 2018, pp. 3391–3401.
5. B. Ghanem, P. Rosso, and F. Rangel, “An Emotional Analysis of False Information,” ACM TOIT, vol. 20, no. 2, pp. 1–18, 2020.
6. M. Granik and V. Mesyura, “Fake news detection using naive Bayes classifier,” in IEEE UKRCON, 2017, pp. 900–903.
7. J. Ma et al., “Detecting rumors from microblogs with recurrent neural networks,” in Proc. IJCAI, 2016, pp. 3818–3824.
8. R. K. Kaliyar, A. Goswami, and P. Narang, “FakeBERT: Fake News Detection with BERT,” Multimedia Tools and Applications, vol. 80, pp. 11765–11788, 2021.
9. J. Chen et al., “Fake News Detection with Deep Diffusive Network Model,” in Proc. EMNLP, 2020, pp. 430–440.

10. S. Sharma, S. Sharma, and A. K. Rai, "An Ensemble Approach for Fake News Detection," in IEEE ICICICT, 2019, pp. 1–6.
11. W. Y. Wang, "Liar, liar pants on fire: A new benchmark dataset," in Proc. ACL, 2017, pp. 422–426.
12. A. Ahmed, "ISOT Fake News Dataset," University of Victoria, 2020.
13. K. Shu et al., "FakeNewsNet: A Data Repository for Studying Fake News," Big Data, vol. 8, no. 3, pp. 171–188, 2020.