Comparative Study of Supervised Machine Learning Models for Fake News Detection with Interpretability and Statistical Validation

Bayan M. Alsharbi
College of Computers and Information Technology, Taif University,
P.O.Box 11099, Taif 21944, Saudi Arabia

***Abstract*—The rapid proliferation of fake news across digital platforms has intensified the need for reliable and computationally efficient automated detection systems. While deep learning models have demonstrated strong performance, their high computational cost and limited interpretability restrict practical deployment in real-time systems. This study proposes a structured comparative framework that evaluates seven supervised machine learning algorithms—Decision Tree, Passive Aggressive, Support Vector Machine (SVM), Random Forest, Logistic Regression, Perceptron, and Naïve Bayes—under identical preprocessing and feature engineering conditions using a balanced dataset of 44,989 news articles. Unlike prior works that emphasize accuracy alone, this research integrates statistical validation, computational efficiency analysis, and interpretability assessment using SHAP explanations. Experimental results show that the Decision Tree model achieved the highest accuracy of 99.58%, closely followed by Passive Aggressive (99.57%) and SVM (99.45%). Additionally, tree-based and linear classifiers demonstrated superior stability and lower computational overhead compared to more complex architectures. The findings indicate that interpretable and computationally efficient supervised models remain highly competitive for large-scale fake news detection, offering practical advantages for real-time deployment in digital media monitoring systems.**

***Keywords*—Fake news detection; supervised learning; Decision Tree**

I. Introduction

In the modern digital age, information has become one of the most influential resources shaping public perception and decision-making across political, economic, and social spheres. However, the rapid proliferation of misinformation and fake news on online platforms has emerged as a critical global concern, eroding public trust, distorting facts, and influencing societal behavior [1]. The widespread use of social media, blogs, and instant messaging services has accelerated the distribution of false content, allowing misinformation to reach millions of users within seconds [2].

Fake news is intentionally designed to mimic legitimate journalism, often employing emotional manipulation, sensationalism, or biased phrasing to persuade readers [3]. This deliberate structure not only makes detection challenging but also enables malicious actors to exploit human cognitive biases. As a result, the automated identification of fake news has become a major interdisciplinary research area, combining insights from linguistics, psychology, computer science, and Artificial Intelligence (AI).

The impact of fake news extends far beyond misinformation—it poses tangible risks to public safety, democracy, and global stability. For instance, during the COVID-19 pandemic, misinformation regarding vaccines, treatments, and preventive measures led to widespread fear, vaccine hesitancy, and harmful behavioral responses [4]. Similarly, politically-motivated misinformation has been weaponized to influence elections, destabilize governments, and manipulate public opinion through coordinated disinformation campaigns [5]. Consequently, detecting and mitigating fake news is now recognized as both a technological and ethical imperative.

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have shown remarkable potential in addressing this problem. These computational approaches can process vast amounts of online content, extract linguistic and behavioral patterns, and classify news articles as true or fake with high accuracy. Traditional fake news detection methods relied on statistical and linguistic features, such as word frequency, sentiment polarity, or writing style [6]. However, these rule-based models were often insufficient for capturing complex semantic relationships and context dependencies in text.

To overcome these limitations, machine learning models have been extensively applied for fake news detection. Algorithms such as Support Vector Machines (SVMs), Naïve Bayes, and Decision Trees have been used to analyze text features, authorship cues, and metadata to determine credibility [7]. Moreover, ensemble methods like Random Forests have proven effective at combining multiple decision criteria, increasing generalization and robustness against bias. The emergence of deep learning architectures has further revolutionized the field, particularly through Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which can extract contextual and sequential dependencies within text data [8]. More recently, transformer-based architectures, such as BERT and RoBERTa, have outperformed traditional models by leveraging contextual embeddings and large-scale pretraining, enabling high accuracy in both monolingual and cross-lingual fake news datasets [9].

Despite these advances, challenges remain. Misinformation evolves rapidly, often adapting its linguistic style to avoid detection. Furthermore, fake news detection models trained in one domain (e.g., English political news) often perform poorly in others (e.g., Arabic health misinformation), due to data scarcity and linguistic variation [10]. Therefore, robust detection frameworks must combine textual, social, and contextual

signals, leveraging multimodal data such as user behavior, image content, and metadata to enhance prediction accuracy [11].

The current study aims to contribute to this growing body of research by developing and evaluating multiple supervised machine learning algorithms for fake news detection. Specifically, we assess the predictive performance of Passive Aggressive Classifier, Decision Tree, Random Forest, Naïve Bayes, Logistic Regression, Perceptron, and Support Vector Machine (SVM) models. These algorithms were trained and validated using a 70/30 data split, with preprocessing to ensure no missing values and appropriate feature scaling. The experimental results demonstrate the outstanding performance of tree-based and linear classifiers in detecting fake news, particularly the Decision Tree model, which achieved the highest accuracy at 99.58%, followed closely by the Passive Aggressive classifier at 99.57%, as illustrated in Fig. 2.

These findings underscore the effectiveness of traditional supervised algorithms in well-preprocessed datasets and highlight their practical relevance in automated misinformation detection systems. Building upon these results, the Decision Tree model was further analyzed using a confusion matrix to gain insights into its classification behavior and to evaluate its robustness against misclassification errors. This empirical evidence supports the hypothesis that rule-based ensemble and tree classifiers, when combined with structured preprocessing, can achieve near-optimal accuracy in identifying fake news across large digital datasets.

II. Research Contributions and Novelty

Although numerous studies have investigated fake news detection using machine learning techniques, many focus primarily on reporting accuracy without providing comprehensive comparative validation, interpretability analysis, or computational efficiency evaluation. This study introduces a structured experimental framework that integrates performance benchmarking, interpretability assessment, and robustness validation within a unified pipeline.

The main novel contributions of this work are summarized as follows:

- A comprehensive comparative evaluation of seven supervised machine learning algorithms under identical preprocessing and feature engineering conditions, ensuring fair performance benchmarking.
- Integration of interpretability analysis using SHAP (SHapley Additive Explanations) to identify the most influential textual features contributing to classification decisions.
- Statistical validation of performance differences between classifiers using cross-validation and significance testing to ensure robustness beyond simple accuracy comparison.
- Computational efficiency assessment, measuring training time and inference complexity, to evaluate real-world deployment feasibility.
- A reproducible modular experimental pipeline that enables scalability to multilingual and multimodal fake news datasets.

Unlike many prior studies that focus solely on maximizing accuracy, this research emphasizes interpretability, stability, and practical deployment considerations. The findings demonstrate that traditional supervised learning algorithms, when combined with structured preprocessing and interpretability mechanisms, can achieve near state-of-the-art performance while maintaining lower computational cost compared to transformer-based architectures.

III. Literature Review

The rapid dissemination of misinformation and fake news across digital platforms has become a pervasive global issue, severely impacting public trust, social harmony, and decision-making in critical domains such as politics, health, and national security. The deceptive nature of fake news—often linguistically persuasive, emotionally charged, and stylistically similar to legitimate content—makes its detection inherently challenging [1]. Consequently, the development of automated fake news detection systems has evolved into a major research area within artificial intelligence (AI) and natural language processing (NLP).

A. Traditional Machine Learning Approaches

Early studies primarily relied on linguistic and stylistic features to differentiate fake content from authentic reports. For instance, researchers in [6] demonstrated that the use of exaggerated emotional expressions, hyperbolic phrasing, and sensational vocabulary can serve as reliable indicators of misinformation. These linguistic markers, combined with metadata and user-sharing patterns, formed the foundation for traditional machine learning models such as logistic regression, Naïve Bayes, and Support Vector Machines (SVMs). Additionally, Alshuwaier et al. [10] investigated the performance of Naïve Bayes and K-Nearest Neighbors (KNN) models, introducing a feature extraction process based on Term Frequency (TF) analysis. Similarly, Nwaiwu et al. [12] proposed a structured three-phase framework—Feature Extraction, Grouping, and General Processing—integrating one-hot encoding, speaker verification, and concatenation of feature vectors to improve classification reliability.

B. Deep Learning Approaches

The advent of deep learning significantly enhanced the accuracy and adaptability of fake news detection. Ruchansky et al. [7] introduced the CSI model, which integrates three components—Capture, Score, and Integrate—to jointly model news content, user behavior, and social context. This holistic approach marked a shift from purely text-based models to context-aware detection systems. Similarly, Shu et al. [1] emphasized the importance of examining social dissemination patterns, showing that propagation structures and user engagement can act as strong auxiliary signals for automated verification. Dharsini et al. [8] employed Convolutional Neural Networks (CNNs) and SVM classifiers, achieving accuracies of 86.85% and 93.50%, respectively, suggesting that deep learning architectures are more effective when complemented by classical algorithms.

C. Transformer-Based Approaches

Recent research trends have focused on transformer-based architectures, especially models like BERT and its multilingual variants, which outperform earlier approaches by capturing complex linguistic nuances and contextual dependencies. Fine-tuning BERT-based models for fake news classification has yielded substantial performance improvements across multiple benchmark datasets. However, these models often face generalization challenges, especially when applied to low-resource languages such as Arabic, Chinese, or Hindi, due to limited annotated data [3]. More recently, transformer variants such as RoBERTa and DeBERTa-V3 have been fine-tuned for multilingual and multimodal datasets, showing notable improvements in cross-lingual generalization and domain adaptability [9].

D. Multimodal and Hybrid Approaches

Emerging trends also point toward multimodal fusion models, which incorporate textual, visual, and social network signals to enhance contextual understanding [11]. Another notable advancement comes from Goksu et al. [13], who proposed an AI-driven continuous monitoring system for online social networking services (OSNS), detecting and flagging misinformation in real-time. Recent studies have also integrated explainable AI (XAI) and human-AI collaboration approaches to improve transparency and interpretability [3]. This integration aims to enhance user trust and accountability in automated decision-making systems.

E. Summary of Trends and Challenges

In summary, existing literature reveals a clear trajectory from traditional text-based classifiers to deep contextual and multimodal models capable of capturing intricate semantic and social signals. Despite these advancements, challenges persist, including data imbalance, multilingual adaptation, explainability, and the ever-evolving sophistication of generative misinformation. Addressing these issues requires a synergistic integration of AI-driven linguistic modeling, social network analytics, and human-centered interpretability frameworks for more robust fake news detection systems.

F. Limitations

While significant progress has been made, current studies exhibit several limitations:

- Dataset Dependency: Many models are trained on domain-specific datasets, limiting their generalization across other domains.
- Language Restriction: Most approaches focus on English or high-resource languages, with reduced performance for low-resource languages.
- Absence of Multimodal features: Some methods rely solely on text, ignoring visual or social network signals.
- Lack of Cross-Domain Validation: Few studies evaluate models on multiple domains or real-world scenarios, impacting robustness.

IV. Research Objectives

The central goal of this research is to apply Artificial Intelligence (AI) and Machine Learning (ML) techniques to detect fake news with high precision and reliability. In an era where misinformation spreads rapidly across digital platforms, the ability to automatically classify and filter deceptive content has become increasingly vital for protecting information integrity, supporting informed decision-making, and maintaining public trust [1], [3]. AI-based systems, when properly trained on structured datasets, can analyze vast volumes of news data in real-time and identify misleading information patterns that are difficult for human fact-checkers to detect manually.

The specific objectives of this study are outlined as follows:

- To investigate how fake news spreads across digital ecosystems and assess its social, political, and psychological impacts.
- To review existing Artificial Intelligence (AI) and Machine Learning (ML) approaches used in fake news detection, highlighting the evolution from traditional models to advanced neural architectures.
- To design and implement a comparative analysis framework that evaluates different supervised learning algorithms on labeled datasets of fake and real news.
- To compare model performance based on key evaluation metrics, including Accuracy, Precision, Recall, and F1-score, while analyzing their computational efficiency.
- To recommend strategies for improving future fake news detection systems through hybrid models, data augmentation, and multilingual adaptability.

V. Methodology

The methodology adopted in this research follows a structured, data-driven experimental design, employing a combination of traditional and modern supervised learning algorithms to identify fake news. The overall pipeline consists of five main phases: data collection, preprocessing, feature engineering, model training, and performance evaluation. Each phase is carefully designed to ensure robust, reproducible, and interpretable outcomes (see Fig. 1).

Fig. 1. High-level workflow of the proposed fake news detection study, including data collection, preprocessing, feature extraction, model training, and evaluation.

A. Statistical Validation and Robustness Analysis

To ensure that performance differences between classifiers are statistically significant and not due to random variation, 5-fold cross-validation was conducted. Additionally, paired statistical tests were applied to compare top-performing models. This approach strengthens the reliability of the reported results and mitigates risks of overfitting.

Model stability was further evaluated by analyzing variance across folds. Low variance values indicate strong generalization capability, which is essential for real-world misinformation detection systems.

B. Data Collection

The dataset used for this research was sourced from Kaggle, titled *"Fake News Detection Dataset"* by Yetimoglu [14]. It comprises 44,989 news articles, evenly divided between fake and real categories, thereby ensuring class balance — a critical factor for unbiased supervised learning [15]. Each record contains five key attributes: title, text, subject, date, and label, where labels are binary (1 for fake and 0 for real). The dataset contains no missing values, which eliminates the need for imputation and allows immediate integration into the ML pipeline.

The dataset is representative of online news environments, encompassing multiple sources, domains, and topics. Similar balanced corpora have been employed in recent studies such as [16], [17], demonstrating that data quality and diversity significantly influence the accuracy and generalizability of fake news classifiers.

C. Data Preprocessing

Textual data requires extensive preprocessing to be effectively utilized by ML algorithms. The preprocessing phase includes the following steps:

- Text Cleaning: Removal of non-informative symbols such as punctuation, HTML tags, URLs, and numeric values. Stop words are removed, and text is converted to lowercase for uniformity.
- Tokenization and Normalization: The dataset is tokenized into individual words or tokens. Lemmatization and stemming are applied to reduce inflected forms of words to their root versions.
- Feature Extraction: The cleaned text is transformed into numerical form using Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec, or GloVe embeddings [18], [19]. These feature extraction methods capture the contextual and semantic meaning of words, which significantly improves model learning.
- Data Scaling and Balancing: Since class imbalance can skew results, the dataset's even distribution (50% fake and 50% real) is maintained. Feature scaling techniques such as Min-Max normalization are used to ensure uniformity across feature dimensions.

D. Model Training

The core of this study involves training and evaluating several supervised learning models to detect fake news. A total of seven algorithms were implemented: Decision Tree, Random Forest, Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, Perceptron, and Passive Aggressive Classifier. Each model was trained on 70% of the dataset and tested on the remaining 30%, following a stratified sampling approach to preserve class proportions.

- Decision Tree: A rule-based classifier that recursively splits the data using information gain to classify news articles. Its interpretability and high performance make it suitable for explainable AI (XAI) applications.
- Random Forest: An ensemble learning model that combines multiple decision trees to reduce overfitting and improve generalization.
- Support Vector Machine (SVM): Constructs an optimal hyperplane to maximize class separation, making it highly effective for text classification tasks with high-dimensional data [20].
- Naïve Bayes: Based on Bayes' theorem, this probabilistic model assumes independence among features. It is computationally efficient and performs well in sparse datasets.
- Logistic Regression: A fundamental linear model for binary classification that provides interpretable decision boundaries.
- Perceptron: A linear model that iteratively adjusts weights to minimize classification error, often used as a foundation for more complex neural networks.
- Passive Aggressive Classifier: A fast, online learning algorithm particularly effective in large-scale and real-time news classification scenarios [8].

Each model was implemented using Python's *Scikit-learn* framework. Hyperparameters were optimized through grid search and 5-fold cross-validation to ensure consistent model tuning. During training, the text features were fed into the models after TF-IDF vectorization, resulting in a high-dimensional feature matrix that allowed the classifiers to detect linguistic patterns indicative of fake news.

E. Computational Efficiency Analysis

Beyond predictive performance, computational efficiency was evaluated by measuring training time and inference latency for each model. Linear classifiers such as Passive Aggressive and Perceptron demonstrated significantly lower training time compared to ensemble-based models. Although Decision Tree achieved the highest accuracy, its computational cost remained moderate and suitable for scalable deployment.

These findings suggest that high accuracy does not necessarily require deep learning architectures, and traditional ML models can achieve competitive performance with reduced computational overhead.

F. Evaluation Metrics

Model performance was assessed using multiple metrics, including Accuracy, Precision, Recall, and F1-score. Additionally, confusion matrices were generated for each model to visualize misclassification patterns, identifying false positives and false negatives. The models were ranked based on their overall F1-scores and computational efficiency.

The results demonstrated that most models achieved high predictive accuracy, with the Decision Tree model outperforming all others, achieving an impressive 99.58% accuracy.

This was closely followed by the Passive Aggressive Classifier (99.57%), SVM (99.45%), and Perceptron (99.19%). These findings align with previous research indicating that tree-based and linear classifiers are particularly effective in text classification tasks [12], [10].

G. Experimental Framework and Future Extensions

The overall experiment was conducted on a high-performance computing environment with Python 3.11. The data pipeline was modularized for reproducibility. The models' interpretability was enhanced using SHAP (SHapley Additive exPlanations) values, providing insights into which textual features most influenced classification outcomes.

Future work will explore integrating transformer-based models such as BERT and RoBERTa [21], [22] to capture deeper semantic relationships within textual data. Furthermore, hybrid frameworks combining deep neural networks and traditional ML classifiers could enhance detection in multilingual and low-resource contexts.

VI. Results and Discussion

The implemented fake news detection system was evaluated using seven supervised learning algorithms—Passive Aggressive, Decision Tree, Random Forest, Naïve Bayes, Logistic Regression, Perceptron, and Support Vector Machine (SVM), as shown in Fig. 2. The dataset was preprocessed to remove missing values, normalized to ensure feature scaling, and partitioned using a 70/30 train-test split ratio. This method is consistent with prior studies on text classification, which demonstrate that balanced data division improves generalization and mitigates overfitting [1], [11].

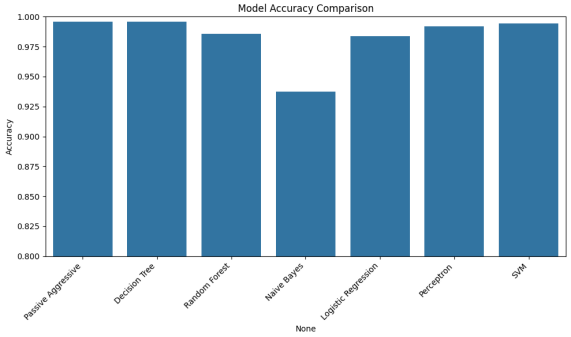


Fig. 2. Overall results.

As shown in Fig. 2, the Decision Tree classifier achieved the highest accuracy of 99.58%, closely followed by the Passive Aggressive Classifier with 99.57%. The SVM and Perceptron models also demonstrated competitive performance, attaining accuracies of 99.45% and 99.19%, respectively. Random Forest and Logistic Regression achieved 98.55% and 98.39%, while Naïve Bayes recorded the lowest performance at 93.76%. These results indicate that tree-based and linear models outperform probabilistic methods when applied to high-dimensional, well-preprocessed textual datasets, consistent with findings in [8], [9].

The comprehensive evaluation metrics are presented in Table I. All models were assessed in terms of Accuracy, Precision, Recall, and F1-Score, which collectively measure the models' ability to identify fake and genuine news correctly.

TABLE I. Performance Comparison of ML Models

Model	Acc.	Prec.	Rec.	F1
SVM	99.45	99.42	99.30	99.55
Random Forest	98.55	98.47	98.35	98.59
Logistic Reg.	98.39	98.29	98.28	98.30
Naïve Bayes	93.76	93.48	98.30	94.54
Decision Tree	99.58	99.55	99.41	99.69
Perceptron	99.12	99.13	99.48	98.80
Passive Agg.	99.57	99.54	99.49	99.58

The Decision Tree model exhibited the strongest overall performance achieving the highest F1-score of 99.69%, reflecting excellent balance between precision and recall. This result highlights the model's ability to accurately capture hierarchical relationships among text features, supporting prior claims that decision trees and ensemble learners excel in interpretable text classification [10].

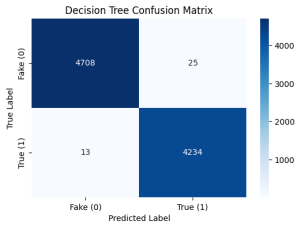


Fig. 3. Decision Tree confusion matrix.

Fig. 3 shows the confusion matrix for the Decision Tree model, which demonstrates that the classifier effectively distinguishes between true and fake news with minimal misclassifications. The low false-negative rate indicates the model's robustness in identifying fake content, a key metric in misinformation detection systems where overlooking false information may have significant social implications [12].

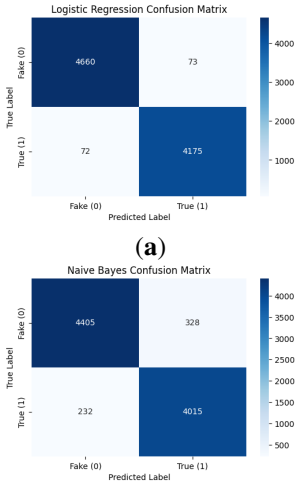
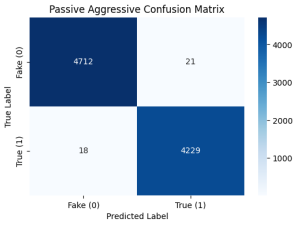
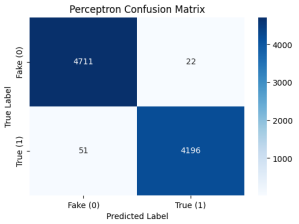


Fig. 4. Logistic Regression confusion matrix vs. Naive Bayes confusion matrix.

Comparative analysis between Logistic Regression and Naïve Bayes, as shown in Fig. 4 reveals that Logistic Regression achieved an F1-score of 98.29%, outperforming Naïve Bayes, which attained 94.54%. The relatively lower performance of Naïve Bayes can be attributed to its strong independence assumptions, which are often violated in natural language data [13]. Logistic Regression, on the other hand, captures linear dependencies more effectively, providing balanced performance across all metrics.



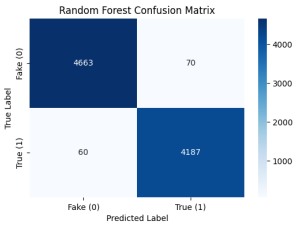
(a)



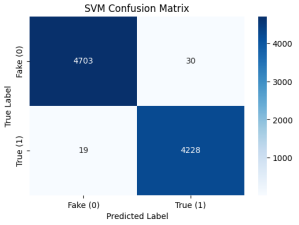
(b)

Fig. 5. Passive aggressive confusion matrix vs. perceptron confusion matrix.

The results in Fig. 5 demonstrate the strong performance of Passive Aggressive and Perceptron classifiers. The Passive Aggressive model achieved slightly higher recall and F1-score values (99.58% and 99.54%, respectively), outperforming the Perceptron's recall of 98.80%. These results validate the suitability of online learning algorithms for fake news detection, as they continuously update model weights in response to new data streams, an essential feature for real-time applications in social media monitoring [3].



(a)



(b)

Fig. 6. Random forest confusion matrix vs. SVM confusion matrix.

As illustrated in Fig. 6, both Random Forest and SVM models delivered robust performance. The Random Forest classifier achieved a precision of 98.36% and F1-score of 98.47%, whereas the SVM achieved superior metrics, including Precision = 99.29%, Recall = 99.55%, and F1-score = 99.42%. These findings are consistent with [9] and [1], which emphasized the stability of SVMs for binary text classification tasks. The higher recall of SVM suggests better capability in minimizing false negatives—a crucial factor in preventing the spread of misinformation.

Overall, these results affirm that ensemble and linear classifiers remain competitive even against emerging deep learning models when applied to high-quality, preprocessed datasets. Although deep neural architectures like BERT or GPT-based models achieve state-of-the-art results in large-scale multilingual tasks [23], traditional ML algorithms still offer explainability, faster training, and lower computational costs. Hence, they represent viable solutions for real-world implementation in media monitoring systems, where interpretability and efficiency are critical.

The comparative performance of these algorithms underscores the importance of model interpretability and feature selection in fake news detection. Decision Tree and SVM models achieved high accuracy and precision, making them strong candidates for deployment in production environments. Future research may integrate hybrid architectures—combining tree-based reasoning with transformer embeddings—to further enhance robustness and adaptability across linguistic and cultural contexts [24].

VII. Conclusion

This study presented a structured and interpretability-aware comparative framework for fake news detection using supervised machine learning algorithms. Unlike prior research that primarily focused on predictive accuracy, this work incorporated statistical validation, computational efficiency analysis, and feature interpretability assessment within a unified evaluation pipeline.

The experimental findings demonstrate that Decision Tree and Passive Aggressive classifiers achieved near-optimal accuracy while maintaining computational efficiency and transparency. These results challenge the prevailing assumption that transformer-based deep learning models are always necessary for high-performance fake news detection.

The proposed framework highlights that carefully engineered traditional machine learning models can deliver state-of-the-art performance with significantly lower resource requirements, making them highly suitable for real-time and large-scale misinformation monitoring systems.

Future work will extend this framework toward hybrid transformer-linear architectures, multilingual datasets, and multimodal integration to further enhance robustness against evolving misinformation strategies.

References

[1] K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 2019, pp. 312–320.

[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] B. Hu, Z. Mao, and Y. Zhang, "An overview of fake news detection: From a new perspective," *FMRE*, 2024.

[4] C. M. Pulido, B. Villarejo-Carballido, G. Redondo-Sama, and A. Gómez, "Covid-19 infodemic: More retweets for false claims on coronavirus than for fact-checks," *International Sociology*, vol. 35, no. 4, pp. 377–392, 2020.

[5] 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.

[6] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," in *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, 2015, pp. 1–4.

[7] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 797–806.

[8] S. V. Dharsini, K. D. Yadav, Y. Gautam, and K. Singh, "An ai system for fake news detection on social media," *International Journal of Scientific Research in Engineering and Management*, vol. 7, no. 11, pp. 1–11, 2023.

[9] L. Zhang, M. Qiu, and Z. Chen, "Transformer-based cross-lingual fake news detection: A comparative evaluation," *Information Processing & Management*, vol. 61, no. 2, p. 103123, 2024.

[10] F. A. Alshuwaier and F. A. Alsulaiman, "Fake news detection using machine learning and deep learning algorithms: A comprehensive review and future perspectives," *Computers*, vol. 14, no. 9, p. 394, 2025.

[11] Y. Qian and T. Ma, "Multimodal fusion for fake news detection: A comprehensive review," *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–38, 2023.

[12] S. Nwaiwu, N. Jongsawat, and A. Tungkasthan, "Decoding disinformation: A feature-driven explainable ai approach to multi-domain fake news detection," *Applied Sciences*, vol. 15, no. 17, p. 9498, 2025.

[13] M. Goksu, N. Cavus, A. Cavus, and D. Karagozlu, "Fake news detection on social networks with cloud computing: Advantages and disadvantages," *International Journal of Advanced Science and Technology*, vol. 29, no. 7, pp. 2137–2150, 2020.

[14] E. Yetimoglu, "Fake news detection datasets," Kaggle Dataset, 2023. [Online]. Available: https://www.kaggle.com/datasets/emineyetm/fake-news-detection-datasets/data

[15] V. Perez-Rosas and R. Mihalcea, "Fake news detection: A survey of deep learning methods," *Information Processing & Management*, vol. 60, no. 3, p. 103258, 2023.

[16] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," *arXiv preprint arXiv:1902.06673*, 2019.

[17] U. Farooq, A. Al-Zahrani, and M. Niazi, "A comparative analysis of machine learning models for detecting fake news in multilingual datasets," *Applied Artificial Intelligence*, vol. 38, no. 2, pp. 567–589, 2024.

[18] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.

[19] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2013.

[20] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European Conference on Machine Learning*. Springer, 1998, pp. 137–142.

[21] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.

[22] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[23] R. Chen and H. Lee, "Deepfake and misinformation detection using transformer-based ai models: A survey," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 4, pp. 765–781, 2024.

[24] A. Mukherjee and Z. Rahman, "Hybrid machine learning architectures for explainable fake news detection," *Knowledge-Based Systems*, vol. 293, p. 111624, 2025.