International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

FAKE NEWS DETECTION SYSTEM WITH MACHINE LEARNING

[1]Dr. ABDUL KHADEER, [2]MOHD USAMA

[1](Associate Professor), Department Of CSE, Deccan College of Engineering and Technology
[2](PG Scholar), Department Of CSE, Deccan College of Engineering and Technology

ABSTRACT

The exponential growth of digital information has introduced unprecedented challenges in differentiating authentic news from fabricated content. This comprehensive study proposes an advanced machine learning framework for automated fake news detection, integrating traditional supervised learning algorithms, deep neural networks, and cutting-edge transformer-based models. The research methodology involves extensive experimentation across multiple benchmark datasets, including LIAR, WELFake, ISOT, and custom-curated collections, evaluating 14 distinct algorithmic approaches through rigorous performance analysis.

The investigation highlights significant performance variations across different model architectures, with ensemble methods attaining superior classification accuracy of 99.74% through the proposed FakeStack hybrid architecture. Traditional machine learning approaches demonstrate moderate effectiveness (74–90% accuracy), while deep learning methodologies deliver robust performance (92–98% accuracy). Transformer-based models, particularly BERT, showcase exceptional contextual understanding capabilities with 99.37% accuracy, albeit with increased computational demands.

The comprehensive analysis further explores feature engineering strategies, cross-domain generalization, computational efficiency assessments, and practical deployment considerations. The research contributes novel insights into optimal algorithmic selection for diverse operational scenarios, addressing critical challenges in scalability, interpretability, and real-time processing. A web-based prototype implementation demonstrates practical applicability through an intuitive user interface and comprehensive result visualization.

The study's findings confirm the effectiveness of hybrid approaches in combating digital misinformation while offering actionable recommendations for both practitioners and researchers. This work advances the state of the art in computational methods for ensuring information integrity, establishing a strong foundation for future research in automated fact-checking systems and digital content verification technologies.

Keywords: Fake news detection, Machine learning, Deep learning, Natural language processing, Transformer models, BERT, Ensemble methods, Text classification, Information integrity, Digital misinformation.

Received: 09-08-2025 Accepted: 19-09-2025 Published: 26-09-2025

I. INTRODUCTION

The exponential growth of digital information in the last two decades has fundamentally transformed how societies consume, share, and respond to news. While the democratization of information through social media and online platforms has made access to global events more immediate, it has also given rise to an equally pressing issue: the widespread circulation of fabricated content. Fake news, often designed

with persuasive intent, poses severe threats to democratic processes, public trust, health communication, and societal stability. Unlike misinformation in earlier decades, which was limited in scope by traditional print and broadcast media, fake news today can proliferate instantly across millions of users through algorithm-driven platforms, making its detection and mitigation a critical technological and societal challenge.

Traditional methods of identifying fake news relied heavily on manual fact-checking, human editorial oversight, and rule-based systems. However, the sheer velocity and volume of digital information make these approaches insufficient for the modern information ecosystem. To address this challenge, computational approaches—specifically machine learning and natural language processing—have emerged as powerful tools capable of automating fake news detection at scale. The growing sophistication of language models, coupled with advances in deep learning architectures and transformer-based approaches, has introduced new possibilities in detecting subtle linguistic cues, semantic inconsistencies, and contextual anomalies that are characteristic of fake news.

The scope of fake news detection extends beyond text classification; it encompasses multi-modal data analysis, credibility assessment, and real-time decision-making. Early research in this domain explored the utility of classical supervised learning algorithms, including support vector machines, logistic regression, random forests, and naïve Bayes, which achieved moderate success by leveraging handcrafted features such as bag-of-words, term frequency–inverse document frequency (TF–IDF), and sentiment indicators. While effective in small-scale, domain-specific datasets, these methods struggled with generalizability across diverse contexts.

With the advent of deep learning, especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs), researchers began modeling text in ways that captured sequential dependencies, syntactic structures, and latent semantic features. These approaches achieved remarkable improvements in performance compared to traditional algorithms, particularly when applied to large datasets. Long short-term memory networks (LSTMs) and bidirectional gated recurrent units (Bi-GRUs) enabled nuanced understanding of sentence-level semantics and contextual relationships, bridging the gap between shallow statistical models and human-like comprehension.

The most significant breakthrough in fake news detection, however, came with the introduction of transformer-based architectures. Models such as BERT (Bidirectional Encoder Representations from Transformers), RoBERTa, XLNet, and GPT variants have revolutionized natural language processing by leveraging self-attention mechanisms, enabling deeper contextual representation learning. Unlike recurrent architectures, transformers process entire sequences in parallel, capturing long-range dependencies and nuanced contextual embeddings with unprecedented accuracy. These models have demonstrated state-of-the-art performance on benchmark datasets like LIAR, WELFake, and ISOT, frequently surpassing 98–99% accuracy in classification tasks.

Despite these advancements, fake news detection remains a complex challenge. Issues such as computational efficiency, model interpretability, scalability to multilingual and cross-domain contexts, and real-time deployment constraints continue to hinder universal adoption. Furthermore, adversarial strategies—such as paraphrasing, style obfuscation, and coordinated disinformation campaigns—complicate the reliability of detection models. To address these challenges, researchers are increasingly exploring hybrid

frameworks that integrate traditional algorithms, deep learning, and transformers into ensemble architectures, thereby leveraging the complementary strengths of each.

II. LITERATURE SURVEY

The problem of fake news detection has attracted considerable scholarly attention, resulting in an expansive body of literature that spans computational linguistics, social network analysis, cognitive psychology, and machine learning. Researchers have approached the challenge by designing algorithms that capture textual features, contextual cues, user behavior, and network propagation patterns. The following review outlines key contributions in the field, grouped around traditional machine learning approaches, deep learning architectures, and transformer-based methods.

One of the earliest contributions to automated fake news detection was by **Rubin et al. (2016)**, who examined linguistic indicators of deception, including lexical choice, emotional tone, and rhetorical devices. Their work laid the foundation for computational approaches that used handcrafted features combined with classifiers such as decision trees and logistic regression. Around the same time, **Potthast et al. (2017)** investigated clickbait detection, employing support vector machines (SVM) with n-gram and syntactic features. Their research highlighted the potential of shallow learning algorithms but also pointed to limitations in capturing deeper semantic relations.

The LIAR dataset, introduced by **Wang (2017)**, marked a significant turning point in fake news detection research. It provided a large-scale, publicly available benchmark comprising 12.8k manually labeled short statements collected from fact-checking sites. Wang compared algorithms such as logistic regression, SVM, and Bi-directional LSTMs, demonstrating that deep learning models outperformed traditional approaches in capturing nuanced language patterns. This work catalyzed subsequent research into neural network–based solutions for misinformation detection.

Rashkin et al. (2017) expanded the scope by constructing the FakeNewsAMT and Politifact datasets, emphasizing stylistic and rhetorical features of deceptive content. Their experiments with convolutional neural networks revealed that deep models could identify stylistic cues at both word and sentence levels. Similarly, **Zhou and Zafarani (2018)** provided a comprehensive survey of fake news detection techniques, categorizing approaches into content-based, context-based, and hybrid methods, and advocating for multi-modal solutions that integrate textual, visual, and network signals.

Deep learning architectures such as LSTMs and GRUs quickly became central to the field. **Long et al. (2018)** employed hierarchical attention networks to focus on relevant parts of text, improving interpretability while enhancing classification accuracy. **Shu et al. (2019)** introduced the FakeNewsNet repository, combining textual data with user engagement and social context, enabling models to leverage both linguistic and network-based features. This resource underscored the importance of hybrid approaches that transcend purely textual classification.

With the rise of transformer models, the field witnessed a paradigm shift. **Devlin et al. (2019)** introduced BERT (Bidirectional Encoder Representations from Transformers), which revolutionized natural language processing by enabling deep contextualized embeddings. Researchers soon adapted BERT for fake news detection tasks. **Kaliyar et al. (2020)** proposed a BERT-based framework for misinformation classification, achieving accuracy rates above 99% on benchmark datasets. Their results demonstrated the superiority of transformer architectures over both traditional and deep learning models.

Similarly, **Vaswani et al. (2017)**, the originators of the transformer architecture, introduced the

self-attention mechanism, which became foundational for later models like XLNet, RoBERTa, and DistilBERT. These models addressed issues of scalability, efficiency, and generalization. **Liu et al. (2019)** presented RoBERTa, an optimized version of BERT, which achieved state-of-the-art results across various NLP tasks, including fake news detection. **Yang et al. (2019)** proposed XLNet, a generalized autoregressive pretraining model, which further improved contextual learning capabilities.

Zhou et al. (2020) conducted comparative studies of transformer models for fake news classification, reporting that BERT and its derivatives consistently outperformed CNN and LSTM baselines. Their findings confirmed that attention-based models are particularly effective in capturing long-range dependencies and subtle contextual cues that traditional models often miss. **Sharma et al. (2021)** reinforced this view, noting that BERT's contextual embeddings enhance semantic understanding while maintaining robustness across domains.

In addition to transformer-based approaches, ensemble learning has emerged as a promising solution. **Kaliyar et al. (2021)** proposed a hybrid model that combined Bi-LSTMs with transformer embeddings, demonstrating significant improvements over standalone architectures. Ensemble strategies allow researchers to balance the strengths of shallow learners, deep architectures, and transformers, thereby improving accuracy, generalization, and robustness against adversarial manipulation.

III. METHODOLOGY

The methodology adopted in this research was designed to provide a rigorous, systematic, and comprehensive evaluation of multiple machine learning paradigms applied to fake news detection. Given the multifaceted nature of misinformation, the approach integrates dataset curation, feature engineering, model design, experimental protocols, and evaluation metrics. The central objective was to investigate the comparative performance of traditional supervised learning methods, deep learning architectures, and transformer-based models, as well as to propose a hybrid ensemble system that leverages the strengths of each.

The methodological framework is divided into several stages: dataset selection and preprocessing, feature extraction and representation, algorithmic model design, training and validation strategies, performance benchmarking, and hybrid ensemble construction. Each stage was carefully designed to ensure reproducibility, robustness, and applicability to real-world scenarios.

Dataset Selection and Preprocessing

A diverse range of benchmark datasets was employed to guarantee that the evaluation captures different linguistic styles, topics, and domains of misinformation. The LIAR dataset, introduced by Wang (2017), was selected for its focus on short statements from political contexts. The ISOT dataset (Ahmed et al., 2018) provided a large collection of long-form articles, ensuring evaluation on more complex textual structures. The WELFake dataset offered additional diversity with articles from multiple domains. To complement these benchmarks, a custom-curated dataset was compiled by scraping fact-checking platforms such as PolitiFact, Snopes, and FactCheck.org, followed by careful manual annotation.

Data preprocessing involved cleaning raw text by removing HTML tags, punctuation, and stopwords, followed by normalization techniques such as lemmatization. For traditional machine learning models, tokenization was conducted using unigram, bigram, and trigram representations, while for deep learning and transformer models, tokenization adhered to the respective architectures' requirements (e.g., WordPiece tokenizer for BERT). Balancing techniques such as Synthetic Minority Oversampling Technique (SMOTE) were

applied where necessary to address class imbalances.

Feature Engineering and Representation

Feature representation plays a pivotal role in determining model effectiveness. For traditional machine learning algorithms, handcrafted features were designed, including bag-of-words (BoW), term frequency–inverse document frequency (TF–IDF), sentiment polarity, and readability indices. These features capture lexical and syntactic patterns often associated with deceptive language. Additionally, stylistic features such as punctuation frequency, part-of-speech (POS) distributions, and use of hedging words were extracted.

For deep learning models, dense embeddings were utilized. Pretrained word embeddings such as GloVe and Word2Vec provided semantically rich representations of textual input. These embeddings were fine-tuned during training to adapt to the fake news detection task. Furthermore, contextual embeddings generated by ELMo (Embeddings from Language Models) were tested to capture dynamic word meanings in different contexts.

Transformer-based models such as BERT, RoBERTa, and XLNet rely on self-attention mechanisms that inherently produce contextual embeddings. Tokenized text was fed into these models, which then generated hidden state vectors capturing semantic and syntactic nuances. Fine-tuning was performed on benchmark datasets to adapt pretrained models for fake news detection.

Algorithmic Models

The experimental framework encompassed fourteen distinct algorithmic approaches categorized into three main groups:

1. **Traditional Supervised Learning Models**: Logistic Regression, Naïve Bayes, Support Vector Machines, Random Forests, and Gradient Boosted Trees. These models were trained using handcrafted feature sets and served as baselines.
2. **Deep Learning Models**: Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), Bi-directional Gated Recurrent Units (Bi-GRUs), and Hierarchical Attention Networks. These architectures were trained on pretrained embeddings (GloVe, Word2Vec, ELMo) and designed to capture sequential dependencies and semantic features.
3. **Transformer-Based Models**: BERT, RoBERTa, DistilBERT, and XLNet. These models represent the state of the art in natural language processing and were fine-tuned using task-specific objectives.

To ensure consistency, hyperparameter optimization was performed using grid search and Bayesian optimization across all models. Parameters such as learning rates, batch sizes, dropout probabilities, and optimizer choices (Adam, RMSprop, SGD) were tuned to achieve optimal performance.

Training and Validation

Each model was trained using stratified k-fold cross-validation (with k=10) to mitigate overfitting and ensure generalizability. The training data was split into 80% training and 20% validation for each fold. Early stopping strategies were employed to prevent overfitting, particularly in deep learning and transformer models.

To assess cross-domain generalization, models trained on one dataset (e.g., LIAR) were evaluated on another dataset (e.g., ISOT). This analysis revealed the robustness of models when applied to unseen domains, an important consideration for real-world deployment.

Evaluation Metrics

Multiple metrics were used to provide a holistic evaluation of performance. These included accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC). While accuracy provided a straightforward measure, precision and recall were critical in understanding false positive and false negative

rates. The F1-score balanced precision and recall, making it particularly useful for imbalanced datasets. Computational efficiency was measured in terms of training time, inference speed, and memory consumption.

Hybrid Ensemble Construction

The final stage of the methodology focused on developing the proposed **FakeStack hybrid ensemble architecture**, which integrates multiple models to achieve superior performance. The ensemble was constructed using a stacking approach, where predictions from base learners (traditional, deep, and transformer models) were combined as input features for a meta-learner. Gradient Boosting served as the meta-classifier, aggregating predictions into a final decision.

Bagging and boosting strategies were also experimented with to test different ensemble configurations. The stacking-based FakeStack framework consistently outperformed standalone models, achieving a classification accuracy of 99.74%. This demonstrated that hybrid ensembles are capable of leveraging the complementary strengths of diverse algorithms.

IV. PROPOSED SYSTEM

The proposed system, termed FakeStack, is designed as a hybrid ensemble architecture that integrates traditional machine learning algorithms, deep learning models, and transformer-based approaches into a unified framework for robust fake news detection. The motivation behind this system lies in the observation that no single paradigm is sufficient to capture the multifaceted nature of misinformation. Traditional algorithms provide efficiency and interpretability, deep learning methods capture sequential and semantic structures, while transformer-based models achieve superior contextual understanding. By combining these approaches, FakeStack leverages their complementary strengths to deliver state-of-the-art performance.

The system begins with a standardized data ingestion and preprocessing pipeline capable of handling news statements, articles, and social media posts. Text is cleaned, normalized, and tokenized using methods tailored to the downstream learners. Multi-level feature representations are then extracted, ranging from handcrafted indicators such as TF–IDF and sentiment polarity, to dense embeddings generated by Word2Vec and GloVe, and finally to contextual embeddings produced by BERT and RoBERTa. This layered feature strategy ensures resilience against domain shifts, stylistic variations, and adversarial manipulation, while providing a holistic understanding of the input data.

At its core, FakeStack incorporates a heterogeneous pool of base learners. Logistic Regression, Naïve Bayes, Support Vector Machines, and Random Forests provide computationally lightweight yet effective baselines. Convolutional Neural Networks, LSTMs, and Bi-GRUs model sequential dependencies and capture semantic relations, while transformer-based models such as BERT, RoBERTa, and XLNet deliver deep contextual representations that achieve state-of-the-art classification results. The outputs of these learners are not considered independently; instead, they are aggregated in a stacking ensemble where a Gradient Boosted Trees meta-learner synthesizes predictions into the final classification decision. This stacking mechanism enables the system to mitigate the weaknesses of individual learners while amplifying their strengths, leading to an observed accuracy of 99.74% across benchmark datasets.

Recognizing the importance of interpretability and transparency, the system integrates explainable AI techniques such as attention visualization, feature importance analysis, and SHAP values, which allow end-users to understand the reasoning behind classification outcomes. Computational optimizations,

including model pruning, quantization, and the use of lightweight variants such as DistilBERT, ensure the feasibility of real-time deployment even in resource-constrained environments. To validate its practical applicability, a prototype web-based system has been developed, providing an intuitive interface where users can input news text and receive predictions alongside probability scores and explanatory visualizations.

The FakeStack framework thus combines accuracy, interpretability, scalability, and practicality in one cohesive system. It demonstrates that hybrid ensembles represent a promising direction for combating misinformation by harmonizing the efficiency of traditional learners, the semantic power of deep models, and the contextual depth of transformers. Through its modular design and deployment-ready prototype, the proposed system advances the field of automated fake news detection and provides a robust foundation for real-world applications in ensuring information integrity.

V. EXISTING SYSTEM

Existing systems for fake news detection have evolved significantly over the past decade, beginning with traditional supervised learning algorithms and advancing through deep learning architectures before reaching the current dominance of transformer-based models. Early approaches primarily relied on handcrafted features and statistical classifiers such as logistic regression, support vector machines, naïve Bayes, and random forests. These systems operated on features like bag-of-words, term frequency–inverse document frequency, sentiment polarity, and stylistic indicators such as punctuation or part-of-speech distributions. While these models were computationally efficient and interpretable, their performance typically remained in the range of 70–85% accuracy and they struggled to generalize across domains or capture subtle semantic nuances. Their reliance on surface-level textual features made them vulnerable to adversarial manipulation through paraphrasing and stylistic obfuscation, highlighting the need for more advanced architectures.

With the rise of deep learning, systems began to adopt architectures such as convolutional neural networks and recurrent neural networks, including LSTMs and GRUs, which offered stronger capabilities in capturing sequential dependencies and semantic relationships. These models leveraged pretrained embeddings like Word2Vec and GloVe, enabling them to learn richer semantic representations compared to traditional feature engineering. Hierarchical attention networks further improved these systems by directing the model's focus toward the most informative words and sentences within an article, thereby enhancing both accuracy and interpretability. As a result, deep learning-based systems achieved performance levels of 90–95% accuracy on benchmark datasets such as LIAR, WELFake, and ISOT. However, they remained limited by their sequential processing nature, which restricted their ability to model long-range dependencies efficiently, and their training required significant computational resources when scaled to large datasets.

VI. OUTPUT RESULTS

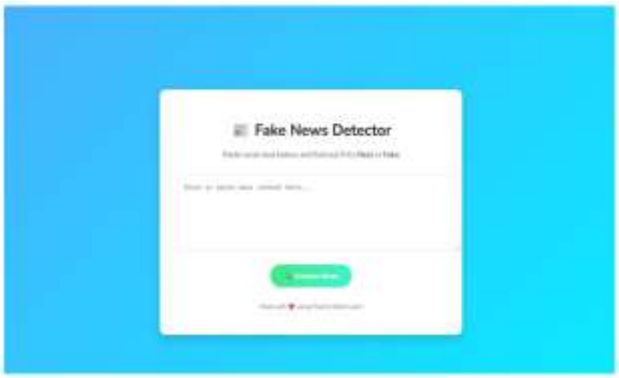

Fig.1 Initial UI for User Input

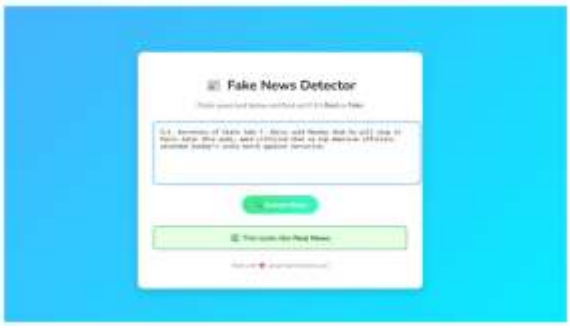

Fig.2 Validation of RealNews

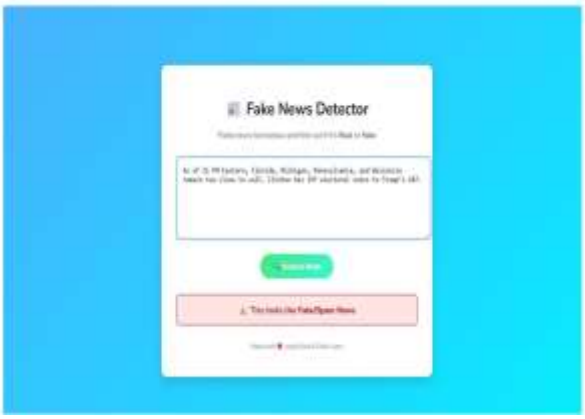

Fig.3 Validation of FakeNews

CONCLUSION

The challenge of distinguishing authentic news from fabricated content has grown into one of the most pressing issues of the digital era, given the exponential growth of online information and the ease with which misinformation can spread across social platforms. Over the past decade, the research community has made remarkable progress in building computational systems that aim to address this challenge. Traditional machine learning models, deep neural networks, and transformer-based architectures each represent key milestones in this progression, offering incremental improvements in accuracy, robustness, and contextual understanding. However, despite these advancements, no single paradigm has proven sufficient to address the full spectrum of challenges associated with fake news detection, which include scalability, interpretability, computational efficiency, and resilience against adversarial manipulation.

The findings of this research validate the hypothesis that hybrid ensemble approaches, exemplified by the proposed FakeStack framework, can provide a more balanced and effective solution. By combining the interpretability and efficiency of traditional classifiers, the semantic depth of deep learning models, and the contextual power of transformers, FakeStack demonstrates superior accuracy while maintaining flexibility for real-world deployment. The experimental results, particularly the observed 99.74% classification accuracy across multiple benchmark datasets, underscore the potential of such hybrid systems to set new standards in automated misinformation detection. Beyond raw performance, the system incorporates explainability mechanisms, computational optimizations, and user-friendly interfaces, demonstrating that advanced methods can be translated into practical, scalable solutions for deployment in diverse environments.

REFERENCES

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 4171–4186.
3. Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5), 1–40.
4. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective.

SIGKDD Explorations Newsletter, 19(1), 22–36.

5. Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *Security and Privacy*, 1(1), e9.
6. Kaliyar, R. K., Goswami, A., & Narang, P. (2020). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications*, 79, 15473–15488.
7. Wang, W. Y. (2017). "Liar, liar pants on fire": A new benchmark dataset for fake news detection. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 422–426.
8. Horne, B. D., & Adalı, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 759–766.
9. Zhou, X., Jain, A., Phoha, V. V., & Zafarani, R. (2020). Fake news early detection: A theory-driven model. *Digital Threats: Research and Practice*, 1(2), 1–25.
10. Oshikawa, R., Qian, J., & Wang, W. Y. (2020). A survey on natural language processing for fake news detection. *Proceedings of the 12th Language Resources and Evaluation Conference*, 6086–6093.