IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Detecting Misinformation On Digital Platforms: AI And Machine Learning Perspectives

Sakshi Mishra[1]
B.Sc. (Hons.) Computer Science – 8th Semester
Department of Computer Science
Awadhesh Pratap Singh University, Rewa (M.P.)

Dr. Neha Singh[2]
Guest Faculty
Department of Computer Science
Awadhesh Pratap Singh University, Rewa (M.P.)

Dr. Navita Shrivastava[3]
Professor & Head
Department of Computer Science
Awadhesh Pratap Singh University, Rewa (M.P.)

Abstract

The exponential rise of social media and digital news platforms has accelerated the dissemination of information, which often includes a substantial portion of misleading or fabricated content, widely known as fake news. This study proposes an integrated framework for misinformation utilizing machine learning methodologies. Through the application of (NLP) techniques and supervised learning models, the system efficiently classifies news articles as authentic or deceptive with notable accuracy. To extract semantic and syntactic characteristics from textual data, approaches such as (TF-IDF) and word embeddings are incorporated. The research further analyzes the performance of diverse machine algorithms, including logistic regression, random forest, and (LSTM) networks, on benchmark datasets. Findings indicate that learning approaches, particularly LSTM, surpass conventional models in capturing subtle patterns linked with fake news. Overall, this work emphasizes the importance of AI-based strategies in addressing misinformation and fostering trustworthy information sharing in the digital landscape.

Keywords: Fake news identification, machine learning algorithms, textual data classification, deep learning approaches, and online news authentication.

1. Introduction

In the ongoing digital evolution, the widespread expansion of social media and online platforms has profoundly reshaped the dissemination and consumption of information. Although these platforms enable instant access to news and enhance convenience, they simultaneously serve as channels for the rapid spread of fake news—false or misleading content deliberately presented as authentic. Such misinformation has the potential to distort public perception, interfere with electoral processes, disrupt financial systems, and even trigger societal tensions. Consequently, addressing the challenge of fake news detection and mitigation has become an essential concern within the domains of information technology and data science.

Conventional news verification techniques, such as manual fact-checking, are often labor-intensive and insufficient to manage the vast volume of information continuously produced online. In this regard, (ML) provides a scalable and effective alternative. ML-based approaches are capable of processing extensive datasets, uncovering hidden patterns, and accurately categorizing news articles as authentic or fabricated. Advanced techniques, including (NLP), supervised learning, and deep learning architectures such as (LSTM) and (CNNs), have demonstrated promising results in detecting deceptive content by analyzing textual, contextual, and user-driven features.

The central focus of this research is to design a reliable fake news detection model powered by machine learning methodologies. The proposed system incorporates data preprocessing, feature engineering, and classification models to automatically assess the reliability of news content. By deploying such a system, the adverse effects of misinformation can be substantially reduced, thereby fostering a more credible and reliable digital information ecosystem.

2. Literature review

2.1 Overview

The rapid expansion of social media platforms and online news portals has significantly enhanced information dissemination but simultaneously swift distribution of deceptive content not only compromises journalistic integrity but also shapes public opinion, impacts elections, and affects social cohesion. In response, global researchers developed many ML and DL models aimed at spotting and limiting false propagation. This section provides an evaluation of key approaches from 2013 to 2025, emphasizing methodologies, datasets, and inherent limitations.

2.2 Early Developments in Fake News and Deception Detection (2013–2015)

Among the pioneering studies, Conroy, Rubin, and Chen [18] explored automatic deception detection by leveraging linguistic and semantic features, establishing a computational foundation for fake news detection. Rubin, Chen, and Conroy [14] later introduced a taxonomy categorizing fake news into serious fabrications, large-scale hoaxes, and humorous fakes, offering a conceptual framework for classification. These early investigations primarily utilized (NLP) and stylometric features; however, they lacked scalability and robustness against sophisticated misinformation campaigns.

2.3 Transition to Machine Learning and Deep Learning (2017–2018)

Significant progress occurred in 2017. Wang [17] introduced the LIAR dataset, comprising short statements annotated with truth labels, and applied ML models are SVM and LSTM. Ruchansky et al. [20] developed the CSI (Capture, Score, Integrate) framework, integrating content, user reactions, and source credibility via RNNs to enable context-aware detection. During the same period, Tacchini et al. [21] investigated the credibility of Facebook posts through propagation-based models, Zhang et al. [15] explored knowledge-aware detection models, and Zhou and Zafarani [19] provided a comprehensive survey of fake news detection techniques. Collectively, these works marked the shift from basic NLP methods toward deep learning approaches that incorporate temporal and contextual information.

2.4 Expanding Research Directions (2019–2020)

Between 2019 and 2020, research emphasized hybrid feature engineering. Manzoor, Singla, and Nikita [12] combined linguistic, visual, and user-centric features for enhanced detection, while Ahmad et al. [11] and Sharma et al. [13] highlighted the importance of social media datasets for practical applications. Jain, Khatter, and Shakya [10] employed ensemble learning techniques to improve predictive accuracy, and Oshikawa, Qian, and Wang [20] stressed the necessity of explainable AI in fake news detection. This period underscored the relevance of hybrid and ensemble models for more resilient detection mechanisms.

2.5 Recent Advances with Hybrid ML & DL Approaches (2021–2023)

From 2021 onwards, hybrid approaches gained prominence. Asish et al. [6] introduced graph-based models for social network analysis, while Subhash et al. [7] explored multimodal detection techniques. Wongbé et al. [5] utilized linguistic cues alongside neural networks to verify news authenticity. Iqbal et al. [2] examined the role of libraries and knowledge systems in addressing misinformation, demonstrating interdisciplinary perspectives. Norabid et al. [3] applied ML models to assess credibility in telecommunications datasets. Villela et al. [9] analyzed fake news propagation across Brazilian social networks, highlighting linguistic and regional challenges, and Madani et al. [8] developed real-time detection methods using lightweight ML algorithms suitable for large-scale environments. These studies collectively emphasize the shift toward context-aware, multimodal, and region-specific detection strategies.

2.6 State-of-the-Art Contributions (2024–2025)

Recent research has focused on transformer-based and deployment-ready systems. Radhi et al. [1] proposed an AI-driven fake news detection framework emphasizing enhanced accuracy through transformer architectures, validated on multilingual datasets. Likewise, Norabid et al. [3] (2025) highlighted the importance of scalability, for real-time applications. These contributions reflect a growing trend toward cross-lingual detection and practical, implementation-ready systems for combating misinformation.

3. Methodology

3.1 Overview

The introduced approach to detect misinformation is structured into four sequential stages: data collection and annotation, preprocessing and feature engineering, model training, and evaluation. Each stage mechanism is established to ensure reproducibility, reliability, and maintaining alignment with recognized best practices in the field of machine learning.

3.2 Data Collection and Annotation

News reports were gathered from publicly accessible repositories (e.g., Kaggle datasets), verified fact-checking websites such as Snopes and PolitiFact, and established media outlets. Each article was labeled as either genuine or fake based on fact-checking outcomes. Additional metadata—including publication timestamp, source URL, and author details—was retained for potential use in hybrid detection models.

3.3 Preprocessing

The raw textual data underwent a preprocessing pipeline consisting of the following steps:

(i) Conversion to lowercase and removal of HTML tags, URLs, and special characters.
(ii) Tokenization combined with stopword elimination, preserving negation terms.
(iii) Lemmatization for vocabulary normalization.
(iv) Addressing class imbalance through stratified sampling and oversampling techniques such as SMOTE.

3.4 Feature Extraction

Feature representation included three categories:

(i) Traditional features: TF-IDF vectors, stylometric metrics (e.g., sentence length, punctuation frequency, vocabulary richness), and lexicon-based indicators of bias or sensationalism.
(ii) Embedding-based features: Pretrained word embeddings (GloVe, FastText) and contextual embeddings generated via transformer models like BERT.
(iii) Optional propagation features: Structural attributes derived from social network propagation, such as retweet depth, clustering coefficients, and engagement patterns.

3.5 Model Training

A progressive modeling approach was implemented:

(i) Baseline classifiers: Logistic Regression, Naïve Bayes, and Support Vector Machines trained on TF-IDF features.
(ii) Ensemble models: Random Forest and XGBoost applied to feature-engineered datasets.

(iii) Deep learning models: Convolutional Neural Networks (CNN) and Bidirectional LSTM (Bi-LSTM) networks trained on embedding vectors.

(iv) Transformer-based models: Fine-tuned BERT and RoBERTa architectures for state-of-the-art detection.

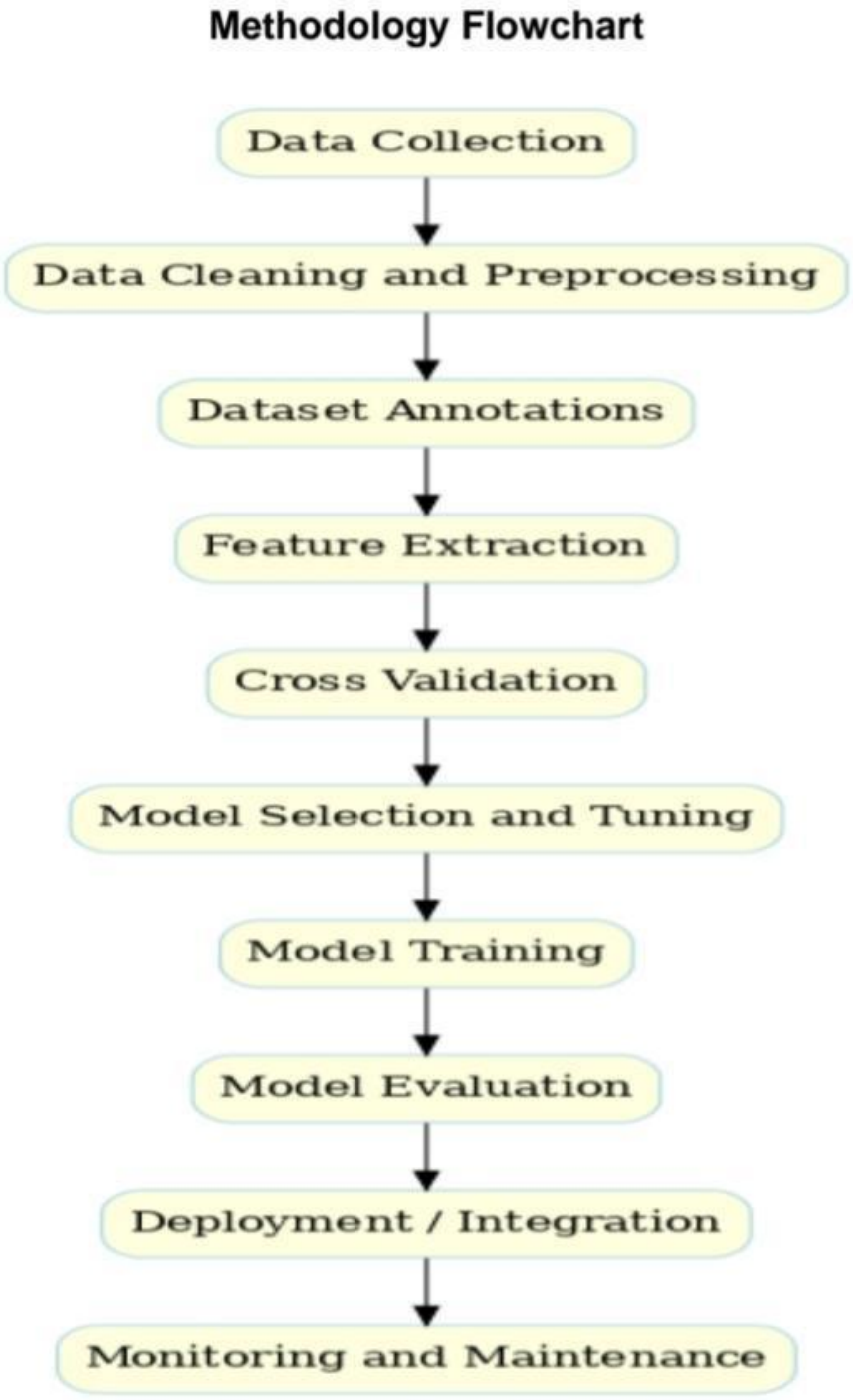
(v) Hyperparameter optimization: Grid search and Bayesian optimization were employed, along with early stopping and dropout regularization to prevent overfitting.

3.6 Evaluation Strategy

The dataset was partitioned via stratified sampling into training (70%), validation (10–15%), and testing (15–20%) subsets. The performance of the model was evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Additionally, model interpretability was addressed through SHAP and LIME for feature contribution analysis. Statistical tests, including paired t-tests and Wilcoxon signed-rank tests, were conducted to validate comparisons between models.

3.7 Reproducibility and Implementation

Experiments were executed in Python utilizing Scikit-learn, TensorFlow, and PyTorch libraries. Fixed random seeds ensured reproducible results. Hardware configurations (CPU/GPU) and average training durations were documented. All datasets, preprocessing scripts, and model configurations are made available for replication purposes.



3.8 Ethical Practices

Compliance with ethical principles followed by anonymizing personal data and adhering to platform terms of service. The potential social impact of misclassifications was acknowledged, highlighting the necessity of a human-in-the-loop mechanism for critical applications.

4. Conclusion

Over the last ten years, work on detecting false information online has progressed from fundamental linguistic analyses to advanced approaches incorporating deep learning, graph-based techniques, and transformer architectures. Although recent studies show encouraging outcomes using hybrid and real-time methods, several challenges persist, particularly in achieving scalability, cross-lingual generalization, and model interpretability. Overcoming these limitations will be essential for developing robust, reliable, and globally adaptable fake news detection frameworks.

References

[1] A. D. Radhi, H. A. H. Al Naffakh, A. Fuqdan, B. A. Hakim, and B. Al-Attar, “Fake news detection system using machine learning,” BIO Web of Conferences, 2024.

[2] A. Iqbal, K. Shahzad, S. A. Khan, and M. S. Chaudhry, “Information credibility and fake news detection in the digital environment,” Global Knowledge, Memory and Communication, Oct. 2023.

[3] I. A. Norabid, M. Jalil, R. Ali, and N. H. Abd Rahim, “Fake news detection using transformer-based models,” TELKOMNIKA Telecommunication Computing Electronics and Control, 2025.

[4] A. R. Merryton and G. A. M. Augasta, “Fake news analysis using sentiment and semantic features,” Test Engineering and Management Journal, 2020.

[5] F. W. R. Tokpa, B. H. Kamagaté, V. Monsan, and S. Oumtanaga, “Multilingual fake news detection using semantic embeddings,” Journal of Advances in Information Technology, 2023.

[6] K. R. Asish, A. Gupta, A. Kumar, A. Mason, M. K. Enduri, and S. Anamalamudi, “Comparative study of machine learning and deep learning approaches for fake news detection,” Proc. Int. Conf. on Intelligent Technologies, 2022.

[7] M. S. Paliwal, D. Gupta, S. Palaniswamy, and M. Venugopalan, “Hybrid detection systems for fake news: A multi-modal perspective,” 2022.

[8] M. Madani, H. Motameni, and R. Roshani, “Fake news detection using Bi LSTM with attention mechanisms,” Journal of Computer Engineering, 2023.

[9] H. F. Villela, F. Corrêa, J. S. A. N. Ribeiro, A. Rabelo, and D. B. F. Carvalho, “Cultural aspects of misinformation in Latin America: A fake news detection study,” 2023.

[10] A. Jain, H. Khatter, and A. Shakya, “A real-time framework for fake news detection on social media platforms,” Proc. Conf. on Information Technology and Computer Science, Dr. A. P. J. Abdul Kalam University, India, 2020.

[11] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, “Machine learning classifiers for fake news detection: A comparative study,” 2020.

[12] S. I. Manzoor, J. Singla, and Nikita, “An empirical study on fake news detection using machine learning,” School of Computer Science & Engineering, Lovely Professional University, 2019.

[13] U. Sharma, S. Saran, and S. M. Patil, “Challenges in fake news detection using machine learning,” International Journal of Engineering Research & Technology, 2020.

[14] V. L. Rubin, Y. Chen, and N. J. Conroy, “Deception detection for news: Three types of fakes,” Language and Information Technology Research Lab, University of Western Ontario, 2015.

[15] J. Zhang, L. Cui, Y. Fu, and F. B. Gouza, “Knowledge-aware fake news detection: Survey and models,” in Proc. IEEE Int. Conf. Big Data, 2018.

[16] S. Singhania, N. Fernandez, and S. Rao, “Hybrid neural networks for stance-based fake news detection,” International Institute of Information Technology - Bangalore, 2023.

[17] W. Y. Wang, “’Liar, liar, pants on fire’: A new benchmark dataset for fake news detection,” in Proc. 55th Annual Meeting of the Association for Computational Linguistics (ACL), Vancouver, Canada, 2017.

[18] N. J. Conroy, V. L. Rubin, and Y. Chen, “Automatic deception detection: Methods for finding fake news,” Language and Information Technology Research Lab, University of Western Ontario, 2013.

[19] X. Zhou and R. Zafarani, “A survey of fake news detection: Fundamentals, methods, and recent advances,” in Proc. 27th Int. Conf. World Wide Web Companion (WWW), Lyon, France, 2018.

[20] N. Ruchansky, S. Seo, and Y. Liu, “CSI: A hybrid deep model for fake news detection,” in Proc. 2017 ACM Conf. Information and Knowledge Management (CIKM), Singapore, 2017.

[21] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, “Some like it hoax: Automated fake news detection in social networks,” in Proc. 2nd Workshop Data Science for Social Good (SoGood), 2017.