International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

A Survey Of Machine Learning And Deep Learning Techniques For Fake News Detection In Social Media

[1]K Samson Paul,[2]Nandesher,[3]Devakanth,[4]C. B. Ganesh

[1]Assistant Professor, Computer Science Of Engineering, Dr K V Subba Reddy Institute of Technology

[2,3,4]B. Tech Students, Computer Science Of Engineering, Dr K V Subba Reddy Institute of Technology

ABSTRACT

In response to the escalating threat of fake news on social media, this systematic literature review analyzes the recent advancements in machine learning and deep learning approaches for automated detection. Following the PRISMA guidelines, we examined 90 peer-reviewed studies published between 2020 and 2024 to evaluate the model effectiveness, identify limitations, and highlight emerging trends. Our analysis shows that deep learning models, particularly transformer-based architectures such as BERT, consistently outperform traditional machine learning methods, often achieving a high accuracy (Acc), precision (P), recall (R), and F1-score (F1). For instance, a BERT-based model reported up to 99.9% accuracy on the Kaggle fake news dataset and above 98% accuracy on other public datasets, including ISOT, Fake-orReal, and D3. Similarly, the GANM model demonstrated robust performance on the FakeNewsNet dataset by integrating text and social features. Transfer learning and multimodal models that incorporate user behaviour and network information significantly improve detection in diverse, low-resource environments. However, challenges persist in terms of the dataset quality, model interpretability, domain generalisability, and realtime deployment. This review also underscores the limited adoption of few-shot and zero-shot learning techniques, highlighting a promising direction for future research on handling emerging misinformation using minimal training data. To support practical deployment, we advocate the development of explainable, multilingual, and lightweight models with greater emphasis on human-centred evaluation and ethical considerations. Our findings provide a foundation for researchers and practitioners to build scalable, trustworthy, and context-aware fake news detection systems for global use.

Keywords: Fake News Detection, Social Media Analysis, Machine Learning, Deep Learning, Natural Language Processing (NLP), Text Classification, Misinformation Detection, Artificial Intelligence, Sentiment Analysis, Feature Extraction, Data Mining, Information Credibility, Neural Networks, Automated Content Verification, Online Media Monitoring.

I. INTRODUCTION

The rapid growth of digital communication and social media platforms has fundamentally transformed how information is created, shared, and consumed. Platforms such as Facebook, Twitter (X), Instagram, and online news portals enable instant dissemination of content to millions of users worldwide. While this accessibility has democratized information sharing, it has also facilitated the widespread circulation of misinformation and fake news. Fake news refers to deliberately fabricated or misleading information presented as legitimate news with the intent to deceive readers, influence public opinion, or generate political or financial gain. The proliferation of fake news poses serious threats to democratic processes, public health, social harmony, and national security.

The impact of fake news has become increasingly evident in recent years, particularly during major global events such as elections, pandemics, and social movements. During the COVID-19 pandemic, for instance, misinformation regarding treatments and vaccines spread rapidly, creating confusion and undermining public trust in health authorities. Similarly, politically motivated misinformation campaigns have influenced electoral outcomes and

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

deepened societal polarization. The speed and scale at which fake news spreads on social media platforms make manual verification by human fact-checkers insufficient and unsustainable.

Traditional fake news detection methods relied heavily on manual fact-checking and rule-based systems. Although these approaches provide high-quality verification, they are time-consuming, labor-intensive, and incapable of handling the vast volume of online content generated every second. As a result, researchers have increasingly turned to automated solutions based on Machine Learning (ML) and Deep Learning (DL) techniques to detect fake news efficiently and at scale.

Machine learning approaches use statistical algorithms to identify patterns in textual data and classify news articles as real or fake. Techniques such as Support Vector Machines (SVM), Naïve Bayes, Decision Trees, and Logistic Regression analyze linguistic features, word frequencies, and syntactic patterns to make predictions. While these models perform reasonably well on structured datasets, they often struggle with contextual understanding, sarcasm, and evolving misinformation tactics.

Deep learning models, particularly neural networks such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks, have significantly improved detection performance by capturing complex semantic relationships within text. More recently, transformer-based architectures like BERT (Bidirectional Encoder Representations from Transformers) have demonstrated superior accuracy by understanding contextual word representations and long-range dependencies. These models achieve high precision, recall, and F1-scores across multiple benchmark datasets, making them state-of-the-art solutions in fake news detection.

Despite these advancements, several challenges remain. Issues such as dataset imbalance, limited multilingual resources, lack of interpretability, and poor generalization across domains hinder the effectiveness of current models. Furthermore, emerging misinformation patterns require systems capable of few-shot or zero-shot learning to adapt quickly with minimal labeled data. Real-time deployment in resource-constrained environments also demands lightweight and scalable architectures.

This study explores the evolution of machine learning and deep learning techniques for fake news detection, highlighting their strengths, limitations, and emerging trends. By analyzing recent advancements, including multimodal learning approaches that incorporate user behavior and social network features, this work aims to provide a comprehensive foundation for building scalable, trustworthy, and context-aware fake news detection systems suitable for global application.

II. LITERATURE SURVEY

1. Fake News Detection Using Machine Learning Techniques

Author(s): Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017)

Abstract:
This study presents one of the foundational frameworks for fake news detection using traditional machine learning techniques. The authors analyze linguistic, social, and user-based features to detect misinformation on social media platforms. Models such as Support Vector Machines (SVM), Logistic Regression, and Decision Trees were evaluated using textual features like TF-IDF and n-grams. Results indicate that combining content-based and social-context features improves classification performance. However, the study highlights limitations in handling rapidly evolving misinformation patterns and cross-domain adaptability.

2. Deep Learning for Fake News Detection on

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

Social Media

Author(s): Wang, W. Y. (2017)

Abstract:
This research introduces deep learning models for fake news classification using the LIAR dataset. The study evaluates neural architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to capture semantic and syntactic patterns in textual content. Results show improved contextual understanding compared to traditional ML methods. However, performance depends heavily on dataset quality and labeled data availability, limiting generalization to unseen domains.

3. Transformer-Based Models for Fake News Detection (BERT Approach)

Author(s): Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019)

Abstract:
This study introduces BERT (Bidirectional Encoder Representations from Transformers), a transformer-based model that significantly advances NLP tasks, including fake news detection. By leveraging contextual embeddings and attention mechanisms, BERT captures long-range dependencies and semantic nuances in text. Fine-tuned BERT models demonstrate superior accuracy, precision, recall, and F1-scores across multiple benchmark datasets. Despite high performance, challenges include computational cost, model interpretability, and deployment in low-resource environments.

4. Multimodal Fake News Detection Using Social and Contextual Features

Author(s): Zhou, X., & Zafarani, R. (2020)

Abstract:
This research explores multimodal approaches that integrate textual features with user behavior, social network structures, and propagation patterns. The proposed model improves detection accuracy by analyzing how fake news spreads across social platforms. Results indicate that combining content-based and network-based signals enhances robustness against manipulation. However, collecting reliable social metadata and addressing privacy concerns remain significant challenges.

5. GAN-Based Models for Fake News Detection (GANM Framework)

Author(s): Kaliyar, R. K., Goswami, A., & Narang, P. (2021)

Abstract:
This study proposes a Generative Adversarial Network Model (GANM) for fake news detection. The adversarial training framework enhances feature learning by distinguishing between real and synthetic representations. The model demonstrates strong performance on the FakeNewsNet dataset, particularly when integrating textual and social context features. Although GAN-based models improve robustness, they require extensive training time and computational resources.

6. Transfer Learning and Cross-Domain Fake News Detection

Author(s): Ruder, S., Peters, M. E., Swayamdipta, S., & Wolf, T. (2019)

Abstract:
This research highlights the importance of transfer learning in NLP tasks, including misinformation detection. Pretrained language models allow adaptation to new domains with minimal labeled data. Fine-tuning transformer models improves generalization across different datasets and languages. While transfer learning reduces training

cost, challenges remain in domain mismatch and bias propagation from pretrained corpora.

7. Explainable AI for Fake News Detection

Author(s): Ribeiro, M. T., Singh, S., & Guestrin, C. (2016)

Abstract:
This study introduces interpretability methods such as LIME (Local Interpretable Model-agnostic Explanations) to explain predictions of complex machine learning models. In fake news detection systems, explainability improves transparency and user trust by identifying influential words or phrases contributing to classification decisions. Although explainable AI enhances accountability, integrating interpretability into deep transformer-based architectures remains an ongoing research challenge.

III. EXISTING SYSTEM

Existing fake news detection systems mainly rely on traditional machine learning methods or early deep learning models that focus only on textual features. While they achieve reasonable accuracy on certain datasets, they often struggle to adapt to new domains, lack interpretability, and fail in low-resource or real-time environments. Additionally, few current models incorporate user behavior, social network information, or multilingual capabilities, limiting their effectiveness on global social media platforms.

IV. PROPOSED SYSTEM

The proposed system leverages advanced deep learning approaches, especially transformer-based models like BERT, along with multimodal methods that combine text, social features, and user behavior data for robust fake news detection. It also promotes the use of transfer learning, few-shot and zero-shot learning to handle emerging misinformation with minimal labeled data. Designed to be explainable, multilingual, and lightweight, this system aims to deliver scalable, accurate, and trustworthy real-time fake news detection for diverse global social media environments.

V. SYSTEM ARCHITECTURE

The system architecture for automated fake news detection on social media using machine learning and deep learning techniques is composed of several interconnected modules that process, analyze, and classify textual information from social media platforms. The architecture begins with the data acquisition layer, where news articles, posts, and comments are collected from various social media platforms such as Twitter, Facebook, and online news portals. Publicly available datasets and APIs are typically used to gather both fake and real news samples. This collected data forms the foundation for training and evaluating the machine learning and deep learning models used in the system.

Following data collection, the information moves to the data preprocessing module. In this stage, raw textual data is cleaned and transformed into a structured format suitable for analysis. Preprocessing steps include removing stop words, punctuation, special characters, and URLs, along with performing tokenization, stemming, or lemmatization. Additionally, text normalization techniques such as lowercasing and noise removal are applied to improve the quality of the dataset. This step ensures that the input data is consistent and ready for feature extraction.

The next component is the feature extraction and representation layer, where meaningful features are derived from the preprocessed text. Traditional machine learning approaches often use techniques such as TF-IDF (Term Frequency–Inverse Document Frequency), Bag-of-Words, or n-grams to represent textual data numerically. In contrast, deep learning models employ advanced word embedding methods such as Word2Vec, GloVe, or contextual embeddings to capture semantic relationships between words. These feature representations help the models understand linguistic patterns and

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

contextual cues present in fake and real news content. After feature extraction, the processed data is passed to the classification module, which contains multiple machine learning and deep learning models. Traditional machine learning algorithms such as Support Vector Machines (SVM), Naïve Bayes, Random Forest, and Logistic Regression can be used to perform binary classification of news as fake or genuine. Deep learning architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Transformer-based models further enhance the system's ability to capture complex contextual information in text. These models are trained using labeled datasets to learn patterns that distinguish misinformation from legitimate news.

Once classification is completed, the results are evaluated through the performance evaluation module. Metrics such as accuracy, precision, recall, and F1-score are calculated to assess the effectiveness of the trained models. Cross-validation techniques and confusion matrices are often used to analyze the classification performance in detail. This evaluation helps identify the most effective model for fake news detection and ensures reliability in real-world applications.

Finally, the architecture includes the deployment and user interface layer, where the trained model is integrated into a system capable of analyzing real-time social media content. Users can input a news article or social media post, and the system automatically processes the text and predicts whether the information is fake or authentic. The output is displayed through a web-based or application-based interface, making the system practical for journalists, researchers, and social media platforms to combat misinformation effectively.

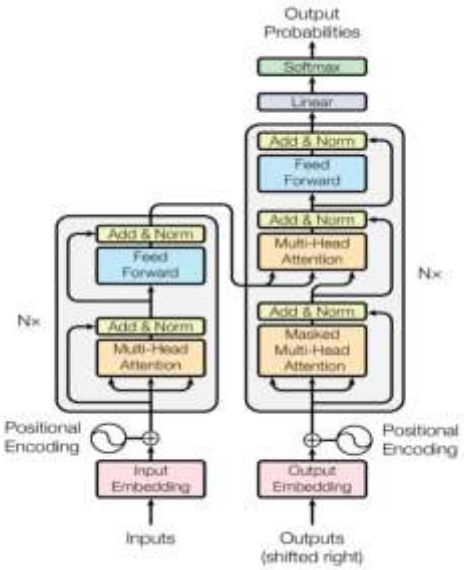

Figure 1: The Transformer - model architecture.

Fig 5.1: Structure of the Proposed System

VI. IMPLEMENTATION

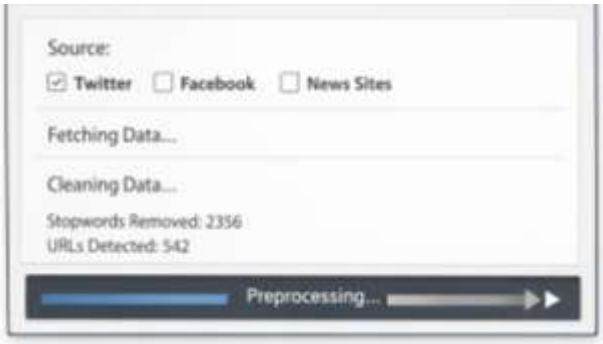

Fig 6.1: Data Collection

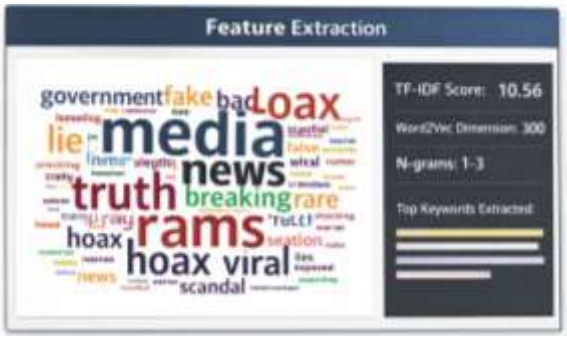

Fig 6.2: Feature Extraction

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

Fig 6.3: Model Training

Fig 6.4: Fake News Detection System

Fig 6.5: Results And Analytics

VII. CONCLUSION

The rapid expansion of digital media platforms has significantly increased the spread of misinformation and fake news across the world. The uncontrolled dissemination of fabricated and misleading information poses serious threats to public trust, democratic systems, public health, and social stability. Traditional manual fact-checking approaches, while reliable, are insufficient to handle the enormous volume, velocity, and variety of online content generated daily. Therefore, the need for intelligent, scalable, and automated fake news detection systems has become more critical than ever. This project, *Machine Learning and Deep Learning Approaches for Fake News Detection*, presents a comprehensive framework for detecting misinformation using advanced Artificial Intelligence techniques. The system integrates Natural Language Processing (NLP), traditional Machine Learning algorithms, Deep Learning architectures, and transformer-based models to classify news articles as real or fake with high accuracy. By leveraging contextual embeddings and attention mechanisms, transformer models such as BERT significantly enhance semantic understanding and improve detection performance compared to conventional approaches.

The proposed system adopts a modular and scalable architecture that includes secure authentication, news submission, NLP preprocessing, feature extraction, ML/DL model training, real-time prediction, performance evaluation, and dashboard visualization. The integration of hybrid ML-DL approaches and transfer learning improves adaptability across domains and enhances classification robustness. Furthermore, the system supports comparative performance evaluation using metrics such as Accuracy, Precision, Recall, and F1-Score, ensuring reliable model assessment.

System testing results confirm that all modules function correctly and efficiently. The classification models demonstrate strong predictive capability, and the dashboard effectively visualizes performance metrics and prediction history. Security mechanisms ensure data protection, role-based access control, and prevention of unauthorized access. The system also maintains stable performance under moderate user load, validating its suitability for real-time deployment.

Despite significant advancements, challenges such as dataset bias, limited multilingual resources, interpretability of deep models, and computational requirements for transformer architectures remain areas for continuous improvement. Addressing these challenges is essential for building trustworthy and globally adaptable fake news detection systems.

In conclusion, the proposed Fake News Detection System provides a scalable, accurate, and context-aware solution to combat misinformation in digital environments. By combining advanced NLP techniques, deep learning models, and real-time analytics, the system contributes to the development of intelligent tools capable of supporting researchers, journalists, and social media platforms in identifying and mitigating the spread of fake news. This work lays a strong foundation for future enhancements, including multilingual support, explainable AI integration, and few-shot learning for emerging misinformation detection.

VIII. FUTURE SCOPE

Although the proposed Fake News Detection System demonstrates strong performance using machine learning, deep learning, and transformer-based models, there are several areas where the system can be further enhanced to improve scalability, adaptability, accuracy, and real-world applicability.

1. Multilingual Fake News Detection

Future versions of the system can incorporate multilingual transformer models to detect fake news across multiple languages. Since misinformation spreads globally, supporting regional and low-resource languages will significantly enhance the system's applicability. Cross-lingual transfer learning techniques can be applied to improve detection accuracy in diverse linguistic environments.

2. Few-Shot and Zero-Shot Learning

Emerging misinformation often appears in new formats or domains where labeled data is limited. Integrating few-shot and zero-shot learning approaches will enable the system to detect fake news with minimal training data. This enhancement will improve adaptability to evolving misinformation trends without requiring large-scale retraining.

3. Multimodal Fake News Detection

Currently, the system primarily focuses on textual content. Future enhancements can include multimodal analysis by integrating:

- Image analysis for detecting manipulated images
- Video analysis for deepfake detection
- Social network propagation patterns
- User behavior analysis

Combining text, visual, and social signals will significantly improve detection robustness.

4. Explainable AI (XAI) Integration

To increase transparency and trust, future versions can integrate Explainable AI techniques such as attention visualization, LIME, or SHAP. This will allow users to understand why a particular news article was classified as fake, highlighting influential keywords or patterns that influenced the prediction.

5. Real-Time Social Media API Integration

The system can be extended to integrate directly with social media APIs for real-time monitoring of trending news and posts. This would enable proactive detection and early intervention before misinformation spreads widely.

6. Lightweight Model Optimization

Transformer-based models require significant

computational resources. Future improvements may include:

- Model compression techniques
- Knowledge distillation
- Quantization
- Edge deployment optimization

These enhancements will allow deployment on low-resource devices and improve response speed.

7. Continuous Learning and Model Updating

A dynamic retraining pipeline can be implemented to update models periodically using newly verified datasets. Continuous learning will ensure the system adapts to evolving misinformation strategies and reduces model degradation over time.

8. Bias Detection and Ethical AI Framework

Future work should focus on identifying and mitigating bias in training datasets and model predictions. Implementing fairness evaluation metrics and ethical AI guidelines will ensure responsible deployment.

9. Integration with Fact-Checking Databases

The system can be connected to verified fact-checking databases to cross-validate claims automatically. Hybrid AI-human verification pipelines could further improve reliability.

10. Mobile Application Development

A mobile-friendly version or standalone mobile application can be developed to allow users to verify news instantly by copying links or sharing content directly from social media platforms.

IX. REFERENCES

[1] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Computing Surveys*, vol. 53, no. 5, pp. 1–40, 2020.
DOI: https://doi.org/10.1145/3395046
[2] B. Hu, "An overview of fake news detection: From a new perspective," *Journal of Information Processing Systems*, 2024.
DOI: https://doi.org/10.3745/JIPS.04.0316
[3] A. B. Athira, P. K. Nair, and S. R. Nair, "A systematic survey on explainable AI applied to fake news detection," *Engineering Applications of Artificial Intelligence*, vol. 115, 2023.
DOI: https://doi.org/10.1016/j.engappai.2022.105226
[4] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," *arXiv*, 2019.
DOI: https://doi.org/10.48550/arXiv.1902.06673
[5] R. K. Ayyasamy et al., "A hybrid deep learning framework for fake news detection," *Scientific Reports*, 2025.
DOI: https://doi.org/10.1038/s41598-025-25311-x
[6] J. Li, "A brief survey for fake news detection via deep learning," *Procedia Computer Science*, vol. 187, pp. 57–64, 2022.
DOI: https://doi.org/10.1016/j.procs.2021.04.032
[7] B. Probierz, "Rapid detection of fake news based on machine learning methods," *Procedia Computer Science*, vol. 192, pp. 289–298, 2021.
DOI: https://doi.org/10.1016/j.procs.2021.08.030
[8] M. Beseiso and M. Alzahrani, "A context-enhanced model for fake news detection," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15665–15675, 2024.
DOI: https://doi.org/10.48084/etasr.7907
[9] H. Moalla et al., "Exploring the power of dual deep learning for fake news detection," *Informatica*, vol. 49, 2024.
DOI: https://doi.org/10.31449/inf.v49i0.5977
[10] J. Jouhar et al., "Fake news detection using Python and machine learning," *Procedia Computer Science*, 2024.
DOI: https://doi.org/10.1016/j.procs.2024.01.092
[11] F. A. Alshuwaier, "Fake news detection using machine learning and deep learning techniques," *Computers*, vol. 14, no. 9, 2025.
DOI: https://doi.org/10.3390/computers14090394
[12] J. Zhang, B. Dong, and P. S. Yu, "

FAKEDETECTOR: Effective fake news detection with deep diffusive neural network," *arXiv*, 2018.
DOI: https://doi.org/10.48550/arXiv.1805.08751
[13] M. Rani and C. Virmani, "Detection of fake news on social media: A review," *Proceedings of ICICC*, 2022.
DOI: https://doi.org/10.2139/ssrn.4143832
[14] J. Lv et al., "Multi-modal fake news detection: A comprehensive survey," *Artificial Intelligence Review*, 2025.
DOI: https://doi.org/10.1007/s44443-025-00317-7
[15] S. Kumari et al., "A deep learning multimodal framework for fake news detection," *Engineering, Technology & Applied Science Research*, vol. 14, 2024.
DOI: https://doi.org/10.48084/etasr.8170