Fake News Detection: A Review of Machine Learning and Deep Learning Methods

Wuyue Li

Faculty of Science and Technology, University of Macau, Macau, China

dc22911@um.edu.mo

Abstract. Fake news is spreading quickly in the digital era, endangering media trust and social cognition. The Internet and social media platforms have made information spread rapidly, but they have also led to a large number of misleading contents, some plantforms keeping publish news with wrong information to the public. This widespread false information distorts the public's understanding of the event, and it threatens social stability. The manual identification of true and false news faces enormous challenges, mainly due to the large amount and diverse forms of online information. Therefore, developing an automated fake news detection system has become necessary. With an emphasis on machine learning and deep learning techniques, this article reviews the state of the art in false news detection research. Several approaches about fake news detection are included in this article. The performance of various models in experiments is covered, along with important elements including feature extraction, classification techniques, and performance evaluation.

Keywords: Fake news detection, machine learning, deep learning, information authenticity, natural language processing.

1. Introduction

The escalating problem of fake news significantly impacts social cognition and media credibility in the digital age. The ease of information dissemination through the internet and social media platforms has led to an unprecedented spread of information, much of it misleading [1,2]. This widespread misinformation contributes to misinterpretation of events, threatens social integrity, and fosters chaos [3]. Manually discerning between authentic and fabricated content is challenging due to the sheer volume and diverse nature of online information [4]. This highlights the critical need for automated fake news detection systems [3,4].

The identification of fake news is a new field of study that is hindered by a lack of resources, such as processing methods and datasets. Manually identifying fake news is difficult and requires extensive understanding of news [3,4]. Therefore, machine learning classifiers are essential for automatic identification of false news [4]. This necessity has driven significant research into leveraging machine learning and deep learning for fake news detection [1,2].

Researchers have explored various machine learning approaches in recent years. Baarir and Djeffal proposed a system using machine learning [3], employing Term Frequency-Inverse Document Frequency (TF-IDF) of bag-of-words and n-grams for feature extraction, and Support Vector Machine (SVM) as a classifier. They also developed a dataset for training, demonstrating the system's efficiency [4]. Chitti et al. focused on identifying hoaxes using an XGBoost machine learning-based classification method. Their research aimed to measure the accuracy of deceptive detection using algorithms like Random Forest (RF), Decision Trees (DT), Logistic Regression (LR), Passive Aggressive Classifier (PAC), Extra Trees Classifier (ETC), and boosting algorithms [4]. This work underscores the utility of ensemble and boosting techniques in improving detection accuracy. Similarly, Ahmad et al. proposed a machine learning ensemble approach [1], investigating textual properties to distinguish fake content and training combinations of algorithms with various ensemble methods, showing superior performance on real-world datasets compared to individual learners.

Bailke et al. introduced an advanced multi-functional fake news detection system [5]. Their system uses Gradient Boosting Classifier and Decision Tree Classifier for textual news articles and integrates

Error Level Analysis (ELA) with Convolutional Neural Networks (CNN) for image forgery detection. Notably, it can analyze news articles in both English and Hindi, addressing linguistic diversity.

More recent advancements involve deep learning. Xiong et al. proposed a fake news detection method based on a comment tree structure [6], optimizing comment selection using cosine similarity filtering and a cross-attention mechanism. This method fuses news sentences and comment information features, improving detection accuracy on public datasets [6]. Ayyalasomayajula et al. explored leveraging feature extraction and fine-tuning with BERT for enhanced fake news detection [7]. Their fine-tuned BERT model outperformed other classifiers, achieving 98.6% accuracy on the test dataset.

Using federated learning and trusted authority techniques, Djenouri et al. introduced a unique method for detecting false news in Internet of Things (IoT) applications, emphasizing data security [8]. They used user clustering and convolution transformers to handle multi-modality difficulties. With an average accuracy of 0.85, their approach performed better when tested using data from Twitter.CNN and LSTM in false news identification were compared by Kumar et al. [9], who trained models using both deep learning methods (LSTM, CNN) and conventional machine learning algorithms (Logistic Regression, Decision Tree Classifier, Naïve Bayes Classifier, SVM). At 95%, CNN and logistic regression demonstrated the best accuracy.

Srivastav et al. provided an in-depth review [10], noting that deep learning models like LSTM and BERT generally outperform traditional machine learning techniques, achieving higher average accuracies. They emphasized the reliance of these methods on textual properties and NLP techniques like TF-IDF and word embeddings. While deep learning is superior, challenges persist, particularly the lack of labeled data. Manzoor et al. also reviewed various machine learning approaches for fake news detection, highlighting their limitations and improvements through deep learning [2].

The purpose of this study is to summarize the state of fake news detecting techniques, focusing on machine learning and deep learning approaches. The paper will analyze strategies for feature extraction, classification, and performance evaluation, providing an overview of advancements, challenges, and future directions.

2. Overview of mainstream methods

2.1. Traditional Machine Learning Approaches

One of a foundational methodology for fake news detection involves the application of some traditional machine learning algorithms, which in conjunction with traditional text feature extraction techniques. To be more specific, researchers have used classifiers like Support Vector Machines (SVM), Logistic Regression (LR), Decision Trees (DT), and Naive Bayes in conjunction with conventional feature extraction methods like Bag-of-Words (BoW) or n-gram models weighted by Term Frequency-Inverse Document Frequency (TF-IDF). Although this approach is conceptually and logic straightforward and could provide a relatively simple framework in this task, it is not as efficacy as contemporary deep learning techniques.

2.2. Ensemble Methods

To improve the accuracy of a single classifier, a wide range of researchers adopt ensemble methods. The basic idea of it is to combine the prediction results of multiple isolated machine learning methods to create a more powerful and robust classification model. The studies mentioned in the research indicate that this methodology could perform better than any individual learner by using algorithms such as random forest (RF), XGBoost, gradient boosting classifiers, etc. This approach reduces error rates by made best use of the advantages of different methods and bypass the disadvantages of them.

2.3. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) as a deep learning model that automatically extract hierarchical features, are applied in fake news detection to process both textual and multimodal information. For instance, Bailke et al. in 2025 integrated a CNN with Error Level Analysis (ELA) to validate news content by identifying manipulated images [8]. Furthermore, CNNs have demonstrated text classification accuracies as high as 95% in comparative studies. These applications underscore the efficacy and versatility of CNNs as a tool for automated disinformation detection.

2.4. Transformer Model

The Transformer approach, especially Bidirectional Encoder Representation from Transformers (BERT), represents the most cutting-edge technology in the field of natural language processing (NLP). Unlike traditional models, BERT can deeply understand more complex relationship of the context through its self-attention mechanism. Moreover, in false news detection, research achieved an amazing accuracy of 98.6% on the test dataset by fine-tuning pre-trained BERT models. This aligns with broader findings that deep learning architectures like BERT and LSTM consistently outperform classical machine learning approaches in this task. As such, they are currently regarded as one of the most effective methodologies for this purpose.

3. Analysis of research results

3.1. Dataset

These studies each have their own focus on datasets used, reflecting different source features. The dataset of traditional machine learning methods is constructed by merging two specified Kaggle datasets, one fake news and one real news, and its core relies on highly feature engineering. This process involves converting textual content into representations such as Bag-of-Words or n-grams and analyzing structured metadata like authors and publication dates. Conversely, deep learning models like CNNs and Transformers largely bypass this requirement, as they are designed to learn salient features directly from raw materials with relatively fewer preprocessing.

3.2. Evaluation Index

Based on a review of the provided articles, the evaluation metrics employed are standard for classification tasks. The paper on geometric deep learning and the article detailing a multi-functional detection system [5] both utilize a wide range of measurements, such as F1-score, recall, accuracy, and precision, to evaluate their models. The comparative analysis of CNN and LSTM [8] primarily utilizes accuracy as its main performance indicator and employs a confusion matrix to provide a more detailed view of classification results.

3.3. Analysis of Experimental Results

As shown in table 1, the comparative analysis by Kumar et al. establishes a baseline, finding that a CNN and a traditional Logistic Regression model performed best among their tested methods, both achieving a high accuracy of 95%. This highlights the effectiveness of both deep learning and well-tuned classical algorithms. Pushing performance further, the work by Ch et al. demonstrates the power of model combination. Their proposed ensemble, which integrates CNN, LSTM, and GRU models optimized with a Whale Algorithm, achieved an exceptionally high accuracy of 99.50%, significantly outperforming the individual deep learning models.

Other studies show the value of incorporating non-textual context. Bailke et al. successfully developed a multi-functional system that analyzes both text by using Gradient Boosting and Decision Trees and images by using a CNN with Error Level Analysis, proving the efficacy of a multi-modal approach. Similarly, Kam et al. used a geometric deep learning model to analyze the social propagation structure of news articles. Their model achieved strong results across all metrics (accuracy, precision, recall, and F1-score), confirming that analyzing how news spreads is a powerful and distinct signal for identifying misinformation.

Table 1. Experimental results data of different model methods

Method name	Dataset	Accuracy (%)
Traditional ML (Logistic Regression)	Merged datasets (Fake + Real News)	27.7-88.46
Ensemble Methods	Combined six separate datasets	85-96.17
CNN (Convolutional Neural Network)	Raw materials	95-99.5
Transformer Model	Raw materials(training (80%)+ validation(10%)+testing (10%) subsets	83.4-98.6

4. Discussion

The main challenges in current research on fake news detection include but are not limited to: limitations in dataset size and quality, insufficient model generalization ability, and difficulty in multimodal information fusion. To address these issues, on the one hand, the training effectiveness of the model can be improved by constructing larger and higher quality datasets; On the other hand, it is particularly important to explore how to effectively combine information beyond text, such as images and videos, to improve detection accuracy. Through the data we mentioned before, we can see the amazing accuracy on CNN method and the stability on Transformer method. In addition, further research is needed to develop a universal model architecture that can adapt to the characteristics of different types of news, in order to enhance the robustness and practicality of the system.

5. Conclusion

From fundamental, conventional machine learning techniques to the newest Transformer model-based technologies, this study offers a thorough analysis of the uses of machine learning and deep learning in the field of false news identification. Through analysis of existing research, we found that although deep learning models have shown significant advantages in accuracy, they still need to overcome issues such as high data annotation costs and poor model interpretability in practical deployment. Future work should focus on addressing the aforementioned challenges while exploring more efficient and transparent algorithm designs, in order to successfully reduce the harm caused by fake news worldwide and maintain a healthy online environment and social order.

References

[1] I. Ahmad, M. Yousaf, S. Yousaf, & M. O. Ahmad. Fake News Detection Using Machine Learning Ensemble Methods. Computational Intelligence and Neuroscience, 8885861. (2020).

[2] S. I. Manzoor, J. Singla, et al. Fake News Detection Using Machine Learning Approaches: A Systematic Review. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 230–234). IEEE. (2019).

[3] N. F. Baarir, & A. Djeffal. Fake news detection using machine learning. In 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)* (pp. 125–130). IEEE. (2021). https://doi.org/10.1109/IHSH51661.2021.9378748

[4] S. Chitti, D. R. Rinku, T. K. Juluru, P. R. Rao, & S. Gouthami. Identifying hoaxes in fake spotter using XG boost machine learning based classification method. In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 990-996). IEEE. (2024). https://doi.org/10.1109/ICUIS64676.2024.10866064

[5] P. Bailke, T. Kulkarni, Y. Kulkarni, A. Kulkarni, A. Jadhav, & D. Jadhav. Advanced fake news detection: Multi-lingual analysis and image forgery detection integration. In *2025 IEEE 14th International

Conference on Communication Systems and Network Technologies (CSNT)* (pp. 321-326). IEEE. (2025). https://doi.org/10.1109/CSNT64827.2025.10969051

[6] S. Xiong, T. Huang, H. Xie, & J. Shen. Sentence-guided comment tree fusion for fake news detection. In *2024 4th International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC)* (pp. 32-35). IEEE. (2024). https://doi.org/10.1109/ICAIRC64177.2024.10900273

[7] M. M. T. Ayyalasomayajula, S. Bussa, S. S. N. Kowsalya, N. Mishra, S. Prasad, & A. Mehra. Leveraging feature extraction and fine-tuning techniques for enhanced detection of fake news using BERT. In *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N) * (pp. 1630–1635). IEEE. (2024). https://doi.org/10.1109/ICAC2N63387.2024.10895840

[8] K. J. Kumar, K. Shreya, L. P. Divakarla, P. C. Nair, & N. Sampath. Interpretable AI insights in fake news detection: A comparative analysis of CNN and LSTM. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1–6). IEEE. (2024). https://doi.org/10.1109/ICCCNT61001.2024.10724705

[9] Y. Djenouri, A. N. Belbachir, T. Michalak, & G. Srivastava. A federated convolution transformer for fake news detection. IEEE Transactions on Big Data, 10(3), 214–225. (2024). https://doi.org/10.1109/TBDATA.2023.3325746

[10] A. K. Srivastav, V. Ranga, & D. K. Vishwakarma. Machine learning and deep learning in fake news detection: An in-depth review. In *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N) * (pp. 898–903). IEEE. (2024). https://doi.org/10.1109/ICAC2N63387.2024.10894957