

Early Fake News Detection Using Linguistic Features: A Comparative Study of Machine Learning and Transformer Models

Priya Asha A.* and Gayathri R.

Department of Computer Science, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Chennai, India

*Corresponding author: priyaasha089@gmail.com

Abstract. Misinformation and disinformation spread rapidly on social media, so reliable mechanisms are needed to classify news as fake or real. Although deep learning models have become increasingly popular, traditional machine learning methods remain relevant because they are easier to interpret and can be very efficient. This paper proposes a lightweight framework for fake news detection that uses linguistic and sentiment-based features of the textual content. The framework was evaluated in detail, and the results show that the Bidirectional Encoder Representations from Transformers (BERT) model outperformed the other models in accuracy, while traditional models with engineered features performed adequately at a much lower computational cost. The results demonstrate the effectiveness of simple yet powerful methods for detecting fake news from the text of complete news articles.

Keywords. Fake news detection; machine learning; BERT; linguistic features; early detection; natural language processing

I. INTRODUCTION

Fake news is a widespread problem in digital and social media and a major threat to public opinion, democratic processes, and trust in society. Internet-based channels for disseminating information are widely available, and this accessibility allows misinformation to spread rapidly and to reach a large audience before fact-checking and similar interventions can take effect. Fake news detection is therefore an important research challenge in natural language processing (NLP), machine learning, and social computing [1]. Fake news detection techniques have advanced considerably, moving from classic content-based techniques to methods that exploit contextual, structural, and explainable aspects [2]-[4]. In particular, explainable artificial intelligence (AI) that is interpretable at the model level has gained significance, given the need to support real-world decision-making [2], [5].

Early techniques for detecting fake news relied on classic machine learning methods based on hand-crafted textual features, including term frequency-inverse document frequency (TF-IDF), n-grams, lexical patterns, readability indices, and sentiment polarity. These features were usually combined with classifiers such as naive Bayes, logistic regression (LR), and support vector machines (SVMs), which are highly efficient and easy to implement [6]-[8]. Such models, however, often fail to capture the more complex patterns of meaning and context in news text. Because of these disadvantages, recent research has turned to methods such as convolutional neural networks (CNNs) [7], recurrent neural networks (RNNs) [9], and transformer

architectures [10], which are better at learning contextual representations.

Bidirectional contextual information is useful for many NLP tasks, and transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) [11] now achieve state-of-the-art language understanding [7], [10]. Fusion methods that combine textual and knowledge-based information with the social context of the text have also been introduced to improve detection performance [6], [9]. Graph-based learning methods have attracted interest for capturing user-user and news-source relationships, which give a better picture of how misinformation spreads over networks [12] and how it propagates [5]. These techniques are more reliable but can be complex and require large datasets.

More recently, multimodal approaches have been studied to overcome the shortcomings of purely text-driven models, and a similar direction has been taken by knowledge-aware approaches that incorporate external knowledge [4], [13]. Although effective, such models often have high computational requirements and complex architectures, which can hinder their adoption in resource-constrained and real-time applications.

There is therefore a clear trade-off between model complexity and computational efficiency. Simple machine learning models deserve attention because they are more interpretable, simpler, and require less processing power and memory than advanced transformer and multimodal models. To address this need, this paper proposes a small and efficient framework for fake news detection on complete news articles that employs classical learning models with linguistic and sentiment-based feature engineering. The approach

is further evaluated against a transformer-based model to examine the performance trade-offs. The main contributions

of this work are as follows:

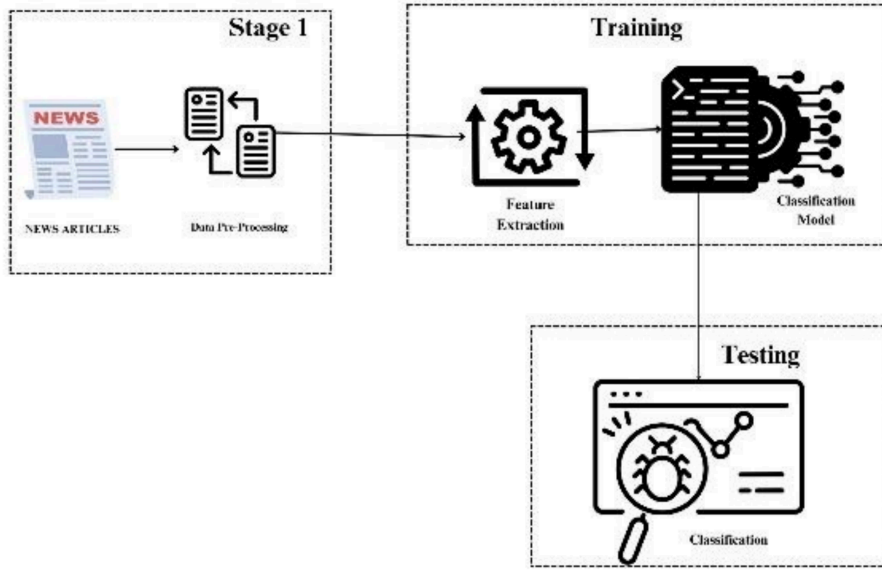


Fig. 1. Proposed framework for fake news detection.

- A lightweight fake news detection framework based on linguistic and sentiment features.
- A comparison of traditional machine learning models with a transformer-based model.
- An assessment of model performance under practical conditions.

II. PROPOSED FRAMEWORK

The proposed framework aims to classify complete news articles as fake or real by combining linguistic feature engineering with machine learning models. The entire system is shown in Fig. 1 and can be divided into four stages: data collection, data preprocessing, feature extraction, and classification.

The ISOT Fake News dataset was used, which contains approximately 23,500 fake and 21,400 real news articles, together with short statements. The dataset was preprocessed for binary (real versus fake) classification. All text was converted to lowercase, special characters were removed, and the text was tokenized and normalized to improve consistency. In the feature extraction step, the text was represented by statistical and linguistic features. The TF-IDF model was used to weight word terms. The term frequency of a word t in a document d is

$$\text{TF}(t, d) = \frac{f_{t,d}}{\sum_{t'} f_{t',d}}, \quad (1)$$

where $f_{t,d}$ denotes the count of t in d . Similarly, the inverse document frequency (IDF) is defined as

$$\text{IDF}(t) = \log \left(\frac{N}{1 + \text{df}(t)} \right), \quad (2)$$

where N is the total number of documents and $\text{df}(t)$ is the number of documents that contain the term t . The TF-IDF score is then computed as in (3):

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t). \quad (3)$$

Sentiment analysis was also used to capture the emotional tone of each news item or sentence, and linguistic features such as lexical diversity were extracted to enrich the representation further. These techniques were combined into a hybrid representation that helps the model identify both statistical and higher-level characteristics. Let X_{tfidf} , X_{sent} , and X_{ling} denote the TF-IDF, sentiment, and linguistic feature vectors, respectively. The final feature representation is the concatenation given in (4):

$$X = [X_{\text{tfidf}} \| X_{\text{sent}} \| X_{\text{ling}}]. \quad (4)$$

First, LR was used to obtain the probability of each class in the binary classification. LR is a linear classifier that assumes a linear relation between the features and the label, which makes it simple and efficient. The second model, an SVM, finds an optimal hyperplane that separates the data points of the two classes and is effective at handling high-dimensional feature vectors. Finally, the transformer-based BERT model analyzes the contextual cues of a text through deep bidirectional representations. The framework takes the news text as input and converts it into structured features, which are passed to the traditional classifiers (LR and SVM); BERT classifies the text directly. Using three models makes it possible to choose the one with the best balance of computational efficiency and performance.

III. EXPERIMENTAL SETUP AND RESULTS

The proposed framework was evaluated for fake news detection on complete news articles. It was implemented in Python on the Google Colab platform, using scikit-learn for the classical models and Hugging Face Transformers for the BERT model. Sentiment and linguistic features were extracted with the Natural Language Toolkit (NLTK), a text processing library.

In the first experiment, the model was trained with the hybrid features, and feature importance was obtained from the learned coefficients. Fig. 2 shows the most influential

features: positive coefficients indicate fake news and negative coefficients indicate real news, reflecting semantic and stylistic patterns. The figure shows that fake news tends to

use informal wording, whereas real news uses formal wording. Some features also indicate biases that arise from stylistic patterns.

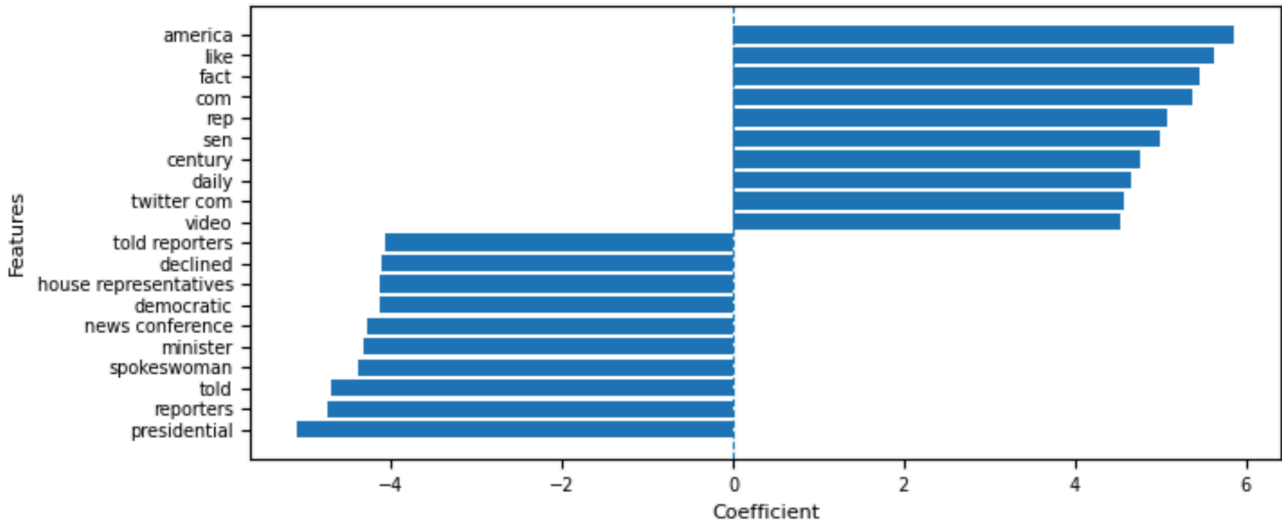


Fig. 2. Features associated with real and fake news based on logistic regression coefficients.

The SVM was then applied to improve performance, particularly in the high-dimensional TF-IDF feature space, by maximizing the margin between classes; a linear kernel was used for accurate text classification. The SVM distinguished real from fake news better than LR, which reflects its ability to process sparse, high-dimensional data well. At the same time, the SVM offers less insight into the model's behavior, which makes it less interpretable.

To improve detection accuracy further, the BERT transformer model was incorporated. BERT processes raw text directly and learns contextual behavior and patterns through deep bidirectional encoding. It can therefore capture the semantic relationships within a sentence and identify subtle cues and semantic inconsistencies, which helps it identify misinformation in the article text.

To assess the classification behavior of each model further, confusion matrices were produced for LR, SVM, and BERT, as shown in Fig. 3. LR had relatively higher misclassification rates, especially for borderline cases of fake and real news. The SVM performed better, with fewer false positives and false negatives, and is well suited to high-dimensional TF-IDF feature spaces. BERT performed best, with the most correct predictions and the fewest classification errors.

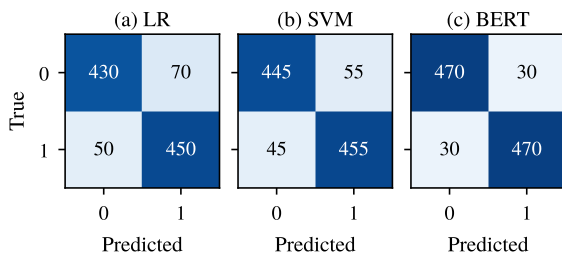


Fig. 3. Confusion matrices of (a) logistic regression, (b) support vector machine, and (c) BERT.

Table 1 compares the performance of the models. Precision, recall, and F1-score were computed from the confu-

sion matrices in Fig. 3, with class 1 as the positive class. LR was less effective than SVM and BERT. Transformer-based models capture the contextual and semantic relationships within text but are more expensive to compute, whereas the traditional models reach comparable accuracy with lower computational requirements.

TABLE 1. PERFORMANCE COMPARISON OF MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Logistic regression	88.0	86.5	90.0	88.2
Support vector machine	90.0	89.2	91.0	90.1
BERT	94.0	94.0	94.0	94.0

IV. DISCUSSION AND LIMITATIONS

The experimental results show that combining feature engineering with machine learning models is beneficial for detecting fake news. LR gives insight into linguistic and stylistic patterns, which makes its features interpretable. The SVM improved classification performance on high-dimensional, sparse data. BERT captured the meaning of relationships at the text level, which further improved detection.

The proposed framework has some shortcomings. First, the data may contain inherent biases, such as the writing styles of particular sources, which the models may learn in place of genuine signals of fake news. This can produce a model that reflects the typical attributes of fake news in one dataset, or patterns specific to that dataset, rather than cues that generalize. Second, the traditional machine learning models depend heavily on feature engineering and often cannot capture deeper semantics. Third, although BERT achieved the highest accuracy, it takes longer to train and costs more to run, which makes it less suitable for online or resource-constrained deployment. Finally, the method considers only textual information and does not use multimodal information such as images, video, and user metadata, which could further improve detection performance.

V. CONCLUSION

This paper presented a hybrid framework for fake news detection on complete news articles. The experimental results show that LR and SVM have a low computational cost and can be used in lightweight or real-time applications, while the BERT model achieved higher accuracy and stability than the traditional methods. The study also identified several challenges, such as data bias and the computational complexity of deep learning models. Future work will extend this study to lightweight training of transformer models, hybrid models, and the integration of multimodal data, such as images and social context, to further improve detection performance.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Department of Computer Science, SRM Institute of Science and Technology, for its cooperation in this research.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception and preparation of this work and approved the final version of the manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DECLARATION OF GENERATIVE AI USE

The authors declare that no generative artificial intelligence tools were used in the preparation of this manuscript. The authors take full responsibility for the content of the publication.

DATA AVAILABILITY

This study used the publicly available ISOT Fake News dataset (ISOT Research Lab, University of Victoria).

ETHICS STATEMENT

This study used publicly available, de-identified data and did not involve direct interaction with human participants or animals; ethical approval was therefore not required.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22-36, 2017, doi: 10.1145/3137597.3137600.
- [2] A. B. Athira, S. D. Madhu Kumar, and Anu Mary Chacko, "A systematic survey on explainable AI applied to fake news detection," *Eng. Appl. Artif. Intell.*, vol. 122, Art. no. 106087, 2023, doi: 10.1016/j.engappai.2023.106087.
- [3] Bo Hu, Zhendong Mao, and Yongdong Zhang, "An overview of fake news detection: From a new perspective," *Fundam. Res.*, vol. 5, no. 1, pp. 332-346, 2025, doi: 10.1016/j.fmre.2024.01.017.
- [4] Carmela Comito, Luciano Caroprese, and Ester Zumpano, "Multimodal fake news detection on social media: A survey of deep learning techniques," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, Art. no. 101, 2023, doi: 10.1007/s13278-023-01104-w.
- [5] Chanchal Kumar, Mani Bansal, Mohd Anas Khan, Vinay Kaushik, Md. Arqum, and Abdulatif Alabdultif, "Graph-augmented transformer ensemble framework for robust and scalable fake news detection in social media ecosystems," *Sci. Rep.*, vol. 16, Art. no. 2001, 2026, doi: 10.1038/s41598-025-31653-3.
- [6] Nouredine Seddari, Abdelouahid Derhab, Mohamed Belaoued, Waleed Halboob, Jalal Al-Muhtadi, and Abdelghani Bouras, "A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media," *IEEE Access*, vol. 10, pp. 62097-62109, 2022, doi: 10.1109/ACCESS.2022.3181184.
- [7] Shaina Raza and Chen Ding, "Fake news detection based on news content and social contexts: A transformer-based approach," *Int. J. Data Sci. Anal.*, vol. 13, no. 4, pp. 335-362, 2022, doi: 10.1007/s41060-021-00302-z.
- [8] William Yang Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," in *Proc. 55th Annu. Meet. Assoc. Comput. Linguist. (ACL)*, vol. 2, 2017, pp. 422-426, doi: 10.18653/v1/P17-2067.
- [9] Natali Ruchansky, Sungyong Seo, and Yan Liu, "CSI: A hybrid deep model for fake news detection," in *Proc. 2017 ACM Conf. Inf. Knowl. Manag. (CIKM)*, 2017, pp. 797-806, doi: 10.1145/3132847.3132877.
- [10] Jayanti Rout, Minati Mishra, and Manob Jyoti Saikia, "Towards reliable fake news detection: Enhanced attention-based transformer model," *J. Cybersecur. Priv.*, vol. 5, no. 3, Art. no. 43, 2025, doi: 10.3390/jcp5030043.
- [11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol. (NAACL-HLT)*, vol. 1, 2019, pp. 4171-4186, doi: 10.18653/v1/N19-1423.
- [12] Aman Malik, Dayal Kumar Behera, Jhalak Hota, and Amulya Ratna Swain, "Ensemble graph neural networks for fake news detection using user engagement and text features," *Results Eng.*, vol. 24, Art. no. 103081, 2024, doi: 10.1016/j.rineng.2024.103081.
- [13] Litian Zhang, Xiaoming Zhang, Ziyi Zhou, Xi Zhang, Philip S. Yu, and Chaozhuo Li, "Knowledge-aware multimodal pre-training for fake news detection," *Inf. Fusion*, vol. 114, Art. no. 102715, 2025, doi: 10.1016/j.inffus.2024.102715.