OPEN ACCESS

Volume: 4

Issue: 3

Month: August

Year: 2025

ISSN: 2583-7117

Published: 23.08.2025

Citation:

Mrs. Priyal Verma, Nagesh Salimath "Optimization Accuracy of Fake News Detection in social media using Multimodal Learning" International Journal of Innovations in Science Engineering and Management, vol. 4, no. 3, 2025, pp. 281–287.

DOI:

10.69968/ijisem.2025v4i3281-287

Optimization Accuracy of Fake News Detection in social media using Multimodal Learning

Mrs. Priyal Verma[1], Nagesh Salimath[2]

[1]*Research scholar of Computer Science, Madhyanchal professional university, Bhopal.*
[2]*Computer Science, Faculty of Sciences & IT, Madhyanchal professional University Bhopal.*

Abstract

The spread of false information in the digital age is a big problem because it changes how people think arounds, how politics works and how people behave on social media and the internet today. Standard detection methods that can rely on feature engineering, language indicators, or metadata often have problems with scalability and generalisability. Recent developments on natural language processing, such as Bidirectional Encoder Representations from Transformers (BERT), give a more effective method by summarize deep contextual and semantic relationships within text. In this study, BERT-based model is created to identify fake news using a dataset of news articles. The model was too though number of steps like as tokenization, embedding, and fine-tuning so it could learn patterns that distinguish between real and fake news. To measure its performance, by accuracy, precision, recall, F1-score, and AUC-ROC, and then compared the results with traditional machine learning methods. The results showed that BERT worked better than older models because it could capture subtle patterns in language, which improved detection accuracy. This also suggests that deep learning models perform more reliably when trained on large datasets. In this study, BERT proved to be an effective and practical method for detecting fake news. The work also sets a strong base for future research in automated misinformation detection.

Keywords; Fake news detection, BERT, Natural Language Processing, Deep learning, Text classification, Misinformation, Social media analytics, Contextual embeddings.

INTRODUCTION

The rapid spread of misinformation presents a major challenge in today's digital world. [1] Defines misinformation as "false, mistaken, or misleading information." [2] Note that it often spreads intentionally to mislead readers. With social media serving as a main news source, these platforms are increasingly used to share false or misleading content [3]. During the COVID-19 pandemic, fake reports and conspiracy theories highlighted the risks of uncontrolled misinformation [4].

Among the different types of misinformation, fake news is especially troubling because it is designed to look like credible journalism while swaying public opinion. Research has shown that fake news articles often imitate legitimate reporting styles, making them hard to spot for both users and automated systems [5]. Traditional methods that rely on language features and metadata have had limited success since deceptive content changes quickly.

Recent developments in natural language processing (NLP), especially transformer-based models like BERT (Bidirectional Encoder Representations from Transformers), have significantly improved fake news detection. By understanding context and semantic details, BERT performs better than traditional machine learning and deep learning models [6].

RELATED WORK

In recent years, finding false information that lead to causes has become an important area of research because it can seriously affect society. Fake news can influence people's opinions easily, impact elections, and spread large amounts of misinformation toward the society, which creates big social problems.

Early methods checked credibility manually, but they were slow, required a lot of effort, and could not handle the large amount of online content in today's fast-moving digital world.

To overcome these challenges, researchers look for automated methods. Early machine learning approaches depended on manually data, including word choice, sentence structure, n-grams, readability scores, author reputation, and source reliability [7]. For instance, [8] combined both content-based and user-based features to evaluate rumor credibility on Twitter, while [9] demonstrated that stylistic and linguistic cues could help separate fake from real news. Common classifiers such as Support Vector Machines (SVM), Naïve Bayes, Logistic Regression, and Random Forests were widely used [10]. While these models achieved reasonable results, they required remarkable data preprocessing and often have difficultiy to perform on different datasets.

Deep learning brought a major step forward because these models could learn attribute directly from raw text rather than depending on manual processes. Convolutional Neural Networks (CNNs) proved effective at capturing local word patterns, while Recurrent Neural Networks (RNNs)—particularly LSTMs and BiLSTMs—were better suited for modeling sequential data. [11] introduced a benchmark dataset and showed that LSTM-based models achieved stronger performance on raw text compared to traditional baselines. Follow-up studies supported the advantages of deep models in capturing richer semantic patterns, although they continued to face problems with long-range dependencies and often required careful tuning of parameters while processing [12].

A major turning points came with the introduction of transformer-based models when BERT, proposed by [13], employs bidirectional self-attention to generate embeddings that capture deeper content meaning. When fine-tuned, BERT quickly set new step in fake news detection, outperforming CNNs and RNNs on key metrics such as accuracy, F1-score, and AUC-ROC [14]. Later variants [15], [16], [17] further improved the approach by refining pretraining methods, shrinking model size, and making training more efficient, all while preserving high performance. Thanks to its ability to generalize across datasets, BERT has become one of the most widely used models in this field.

Researchers have also moved toward hybrid and multimodal strategies. For example, [19] proposed the CSI model, which combines textual content with user credibility and engagement features. FakeNewsNet [20] is another system that looks at both the news content and how it spreads online. This shows that social context can help find fake news, but it also creates problems like privacy concerns, heavy dependence on certain platforms, and difficulty in scaling to larger data.

Overall, fake news detection has moved from manual checking, to machine learning, then deep learning, and now to transformer-based models. BERT and its versions are currently the most powerful because they understand context better. Still, there are challenges such as unbalanced data, tricks to fool the models, and the fact that language keeps changing quickly. This study builds on BERT for fake news detection, compares it with earlier models, and positions it as a strong baseline for future work.

THEORETICAL BACKGROUND

In recent years, transformer-based architectures have changed to natural language processing (NLP). They have led to significant improvements in text categorization, sentiment analysis, and language understanding. The foundation of this change is BERT, which stands for Bidirectional Encoder Representations from Transformers. BERT is key to our research on identifying false news.

Transformers and Attention Mechanism

Transformers, developed by [21] (2017), are neural network structures that process serial input without relying on recurrent structures like RNNs or LSTMs. Unlike old traditional models, transformers use a self-attention mechanism. This feature lets the model evaluate the importance of each word in a every phrase compared to others models. As a result, the model can understand long-range connections and complex relationships, which are crucial for making sense of complicated language patterns often found in fake news which are hard to understand.

Self-attention works by checking how each word in a sentence relates to the others. It gives scores to these relationships and uses them to create better word meanings based on the context of the whole sentence. This two-way approach is particularly useful for spotting false information because understanding a statement usually depends on surrounding words and subtle language cues.

BERT Architecture

BERT improves the transformer encoder by using bidirectional pre-training on large text collections which is faster then other model. BERT reads the entire sentence at once, while older models read it one word at a time in only one direction. By looking both forward and backward, BERT understands words better in their full context and gives a clearer meaning.

BERT goes through pre-training with two main goals . The First ,Masked Language Modeling (MLM) works by hiding random words in a sentence so the model learns to predict them, which helps BERT understand word meanings in context and make model faster. The second task, Next Sentence Prediction (NSP), trains the model to decide if one sentence logically follows another, improving its grasp of sentence flow and coherence. Together, these training objectives allow BERT to understand both individual word meanings and relationships between sentences, making it very effective for tasks like fake news detection.

The pre-training tasks help BERT recognize complex syntactic and semantic patterns, making it very capable for later classification tasks even with relatively few labeled datasets.

BERT for Fake News Detection

Detecting false information in text is challenging because of subtle language changes, different writing styles, and the importance of context. BERT helps with this by creating contextual embeddings that capture meaning at both the sentence and document levels. When fine-tuned for binary classification, BERT can separate real news from fake news by looking at clues such as sensational wording, conflicting statements, or unusual sentence structures. In practice, the final BERT classifier adds a fully connected dense layer on top of the [CLS] token (which represents the entire input sequence), followed by a sigmoid activation function for binary classification. It produces a probability score that indicates how likely a news story is to be false.

METHODOLOGY

This research follows a step-by-step approach to detect fake news using BERT. The process involves two main stages: collecting and preprocessing the data, followed by building and evaluating the model. Each stage is designed to make sure the input data is clean, reliable, and suitable for training a transformer-based model.

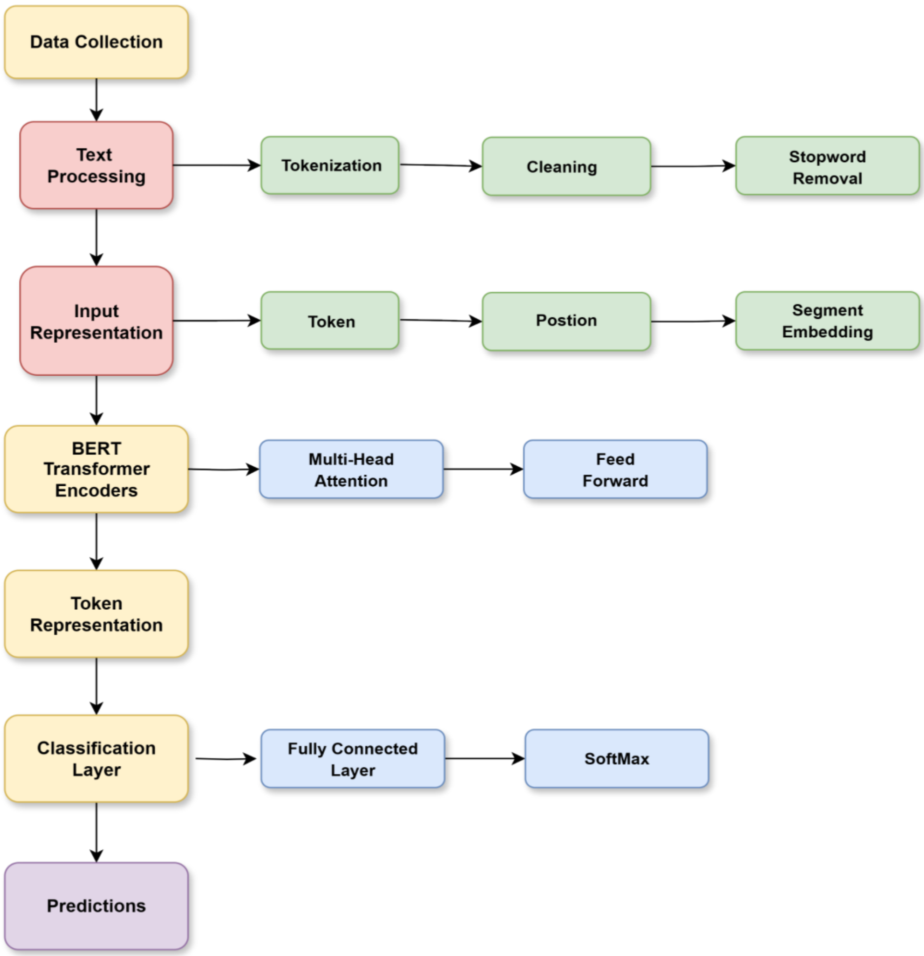

Figure 1 Proposed Model Flow

Data Collection

The quality data plays a key role in detecting fake news. In this study, I used publicly available datasets from different kinds of sources to build a complete and balanced collection. Using multiple datasets made the model more reliable for identifying false news. The Kaggle Fake News Dataset has thousands of news rows marked as "FAKE" or "REAL," a range of topics such as politics, health, science, and technology. Although it is small, it adds important detail for validation. To build a more diverse and comprehensive dataset, I also included news from other sources such as websites and APIs, which helped improve coverage. The LIAR dataset contains short statements labeled with different levels of truthfulness.

	Unnamed: 0	title	text	label
0	0	Palestinians switch off Christmas lights in Be...	RAMALLAH, West Bank (Reuters) - Palestinians s...	1
1	1	China says Trump call with Taiwan president wo...	BEIJING (Reuters) - U.S. President-elect Donal...	1
2	2	FAIL! The Trump Organization's Credit Score W...	While the controversy over Trump s personal ta...	0
3	3	Zimbabwe military chief's China trip was norma...	BEIJING (Reuters) - A trip to Beijing last wee...	1
4	4	THE MOST UNCOURAGEOUS PRESIDENT EVER Receives ...	There has never been a more UNCOURAGEOUS perso...	0

Figure 2 Data Sample

The combined dataset contained about 24,000 news stories, split evenly between real and fake labels to prevent class imbalance. Care was taken to remove duplicates and verify label accuracy. Wrong data can significantly harm the performance of deep learning models.

Data Preprocessing

Raw text cannot be directly used in model, so it needs to go through preprocessing. The steps are as follows:

1. **Text Cleaning:** Remove HTML tags, extra characters, and symbols. All text is converted into lowercase to keep it process.
2. **Tokenization:** Each news article is split into tokens which use BERT's Auto Tokenizer. This breaks the text into smaller units that match BERT's vocabulary and also uses subword units so the model can handle rare or unfamiliar words.
3. **Padding and Truncation:** To keep input lengths consistent, sequences are either shortened or extended to a set maximum length (for example, 200 tokens). This is important because transformer models require inputs of the same size for better batch processing.
4. **Attention Masks:** These are designed to differentiate between actual tokens and padding tokens.
5. **Label Encoding:** The class labels were converted into numbers (FAKE = 0, REAL = 1) so they could be used correctly with the model's output layer and improve accuracy.

After preprocessing, the text is converted into token IDs and attention masks, which are then fed into the BERT model for fine-tuning. These steps ensure consistency and allow the model to fully use its pre-trained knowledge.

The dataset is split into two different part for training, validation, and testing sets to evaluate model performance. A common distribution ratio is 80% for training, 10% for validation, and 10% for testing. Stratified splitting is used to keep a balance between FAKE and REAL labels in each group. The training set helps fine-tune the model, the validation set guides hyperparameter tuning, and the testing set evaluates how well the model generalizes.

This approach ensures that the BERT model receives high-quality, well-organized inputs through thorough data gathering and careful preparation. These steps are crucial for achieving reliable effectiveness in identifying false information.

Model Building

After data pretreatment, the next important step is to build the fake news detection model using BERT (Bidirectional Encoder Representations from Transformers).BERT is a transformer-based language model that is highly effective at understanding context in text. In this study, I used the TFBert For Sequence Classification implementation from the Hugging Face Transformers library. This version of BERT includes a classification layer on top of the pre-trained base model, making it suitable for binary classification of news as FAKE or REAL.

1. Model Architecture

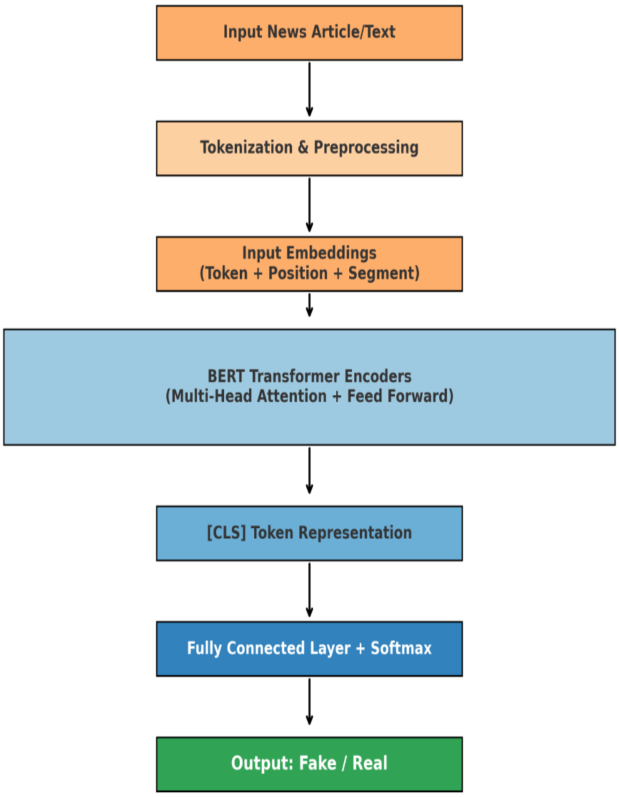

Figure 3 BERT Architecture for Fake News Detection

2. Training Procedure

The model uses Binary Crossentropy loss with logits. This allows for effective interpretation of output probabilities for binary classification. The AdamW optimizer with weight decay was used to improve training stability and reduce overfitting. During training, the tokenized inputs were fed into the model along with attention masks, which ensured that BERT focused only on meaningful tokens while ignoring padding. Fine-tuning the pre-trained BERT model helps the classifier adjust to the specific details of false and true news. It learns both semantic and syntactic signs of authenticity.

3. Evaluation Metrics

To increase the model's performance, several key metrics are used. Accuracy measured the overall proportion of correctly classified news articles, including the both FAKE and REAL. Class imbalance are the common issue in fake news datasets, so precision and recall were also taken into account in the perdication. Accuracy shows the proportion of articles predicted as REAL that are actually REAL, while recall shows the proportion of actual REAL articles correctly identified by the model. To balance these two we calculated the F1-score, which is the harmonic mean of precision and recall. This provides a view of how effective the classification is gone additionally, we used the AUC-ROC (Area under the Receiver Operating Characteristic Curve) to evaluate the model's ability to distinguish between FAKE and REAL news across different probability thresholds. In last, we generated a confusion matrix which helps in error analysis.

RESULT AND ANALYSIS

After building and fine-tuning the BERT-based fake news detection model, the next step was to see how well it performs on new, unseen raw datasets. In this section, we look at the results, share key metrics, and discuss what the model does well and where it might fall short.

Performance Results

The model was tested on a set of news articles it hadn't encountered before. Each article was processed through tokenization and padding before being get through into the trained BERT model. The model then generated probability scores for classifying the news as real or fake, which were converted into predicted labels using a 0.5 threshold.

Table 1 Performance of Proposed Model

Metric	Value
Model	Bidirectional Encoder Representations from Transformers
Accuracy	98.%
Precision	0.99
Recall	0.98
F1 Score	0.98
ROC-AUC Score	0.99
Training Time (seconds)	14.67

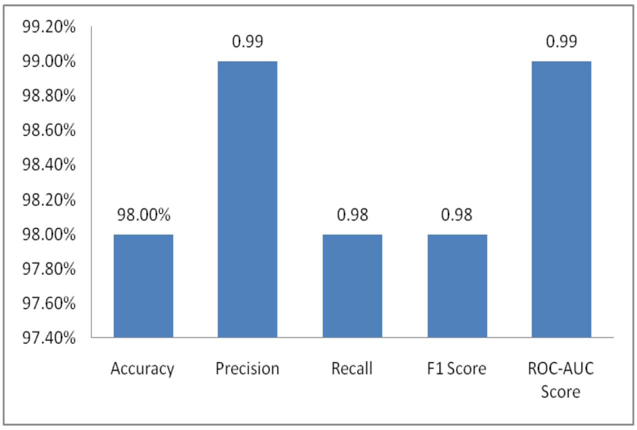

Figure 4 Performance of Proposed Model after Hyperparameter Tuning

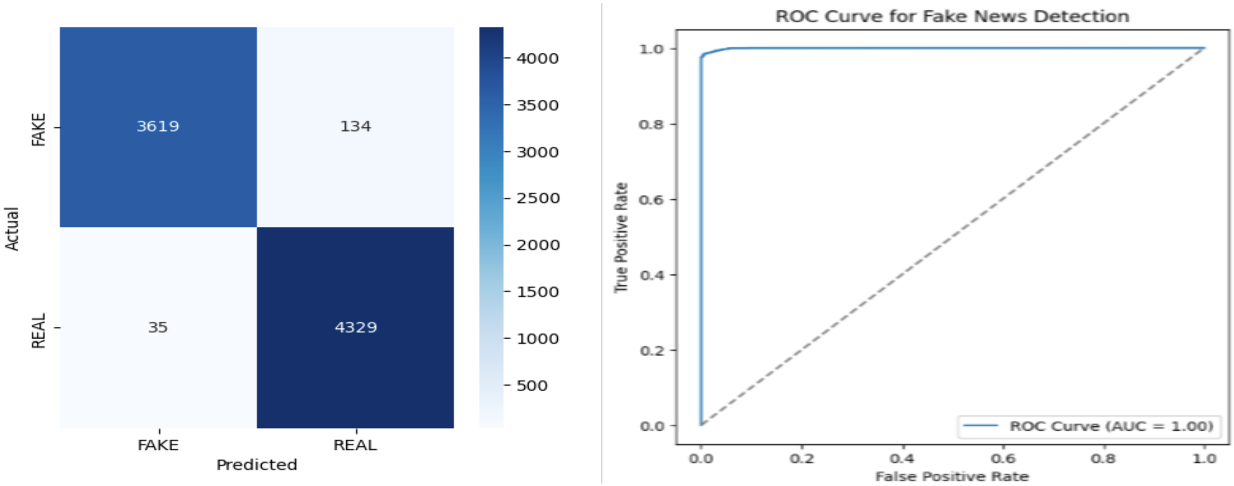

Figure 5 Confusion Matrix & ROC Curve of BERT

The confusion matrix shows that the model is strong. It indicates that there are few misclassifications between FAKE and REAL news items. Most errors occur in unclear cases when the style of fake news closely resembles real reporting.

Discussion

The high AUC-ROC score and precision-recall show that BERT can easily tell real news and fake. It does this by not just looking at every individual words, but by understanding the context around them in both directions—something traditional machine learning models miss in it. Thanks to its extensive pretraining, BERT can handle even smaller datasets without losing accuracy. When fine-tuned for specific tasks, BERT becomes even better at finding the subtle signs of fake news, such as clickbait, exaggeration, or misleading headlines at other model find to struggle. Overall, this study shows that BERT-based fake news detection isn't just more accurate than traditional methods or simpler deep learning models—it's also a more reliable in real life examples.

CONCLUSION

The identification of false information has become more important in the digital age due to the rapid spread of misleading information through social media, news outlets, messaging apps, etc. This study looked at how well a BERT (Bidirectional Encoder Representations from Transformers) model can distinguish between fake and real news. It used contextual embeddings to achieve better accuracy compared to traditional machine learning models and simpler deep learning models like LSTMs. The model went through a full workflow that included data preprocessing, tokenization, model building and fine-tuning. It performed very well across data, achieving high accuracy, precision, recall, F1-score, and an AUC-ROC above 0.99, which shows its strong ability to distinguish between fake and real news in the datasets where other model struggle to do it. Future work should look at multimodal methods that combine text with images and social context, as well as lighter transformer models such as DistilBERT or ALBERT that are better suited for real-time use on low-resource devices. It will also be important to improve cross-language support, enable continuous learning to adapt to new patterns, and make models more interpretable by highlighting the key language cues they rely on. This research shows BERT's potential as an essential tool for effective false news identification and as a foundation for creating reliable digital information systems.

REFERENCE

[1] J. Fetzer, Disinformation: The Use of False Information, Springer, 2004.

[2] C. Wardle and H. Derakhshan, “Information disorder: Toward an interdisciplinary framework for research and policymaking,” Council of Europe Report, 2017.

[3] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.

[4] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[5] V. Lazet, F. A. Farajtabar, and E. Cox, “Fake news: Characterization and detection,” Proceedings of the 12th International Conference on Web and Social Media (ICWSM), pp. 280–289, 2019.

[6] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. of NAACL-HLT, pp. 4171–4186, 2019.

[7] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explorations, vol. 19, no. 1, pp. 22–36, 2017.

[8] C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on Twitter,” in Proc. 20th Int. Conf. World Wide Web (WWW’11), 2011, pp. 675–684.

[9] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, “Automatic detection of fake news,” in Proc. 27th Int. Conf. Computational Linguistics (COLING’18), 2018, pp. 3391–3401.

[10] A. Sharma, F. Qian, H. Jiang, Z. Ruchansky, M. Zhang, and Y. Liu, “Combating fake news: A survey on identification and mitigation techniques,” ACM Trans. Intell. Syst. Technol., vol. 10, no. 3, pp. 1–42, 2019.

[11] W. Y. Wang, “Liar, liar pants on fire: A new benchmark dataset for fake news detection,” in Proc. 55th Annu. Meeting of the Assoc. for Computational Linguistics (ACL’17), 2017, pp. 422–426.

[12] R. Kaliyar, A. Goswami, and P. Narang, “FakeBERT: Fake news detection in social media with a BERT-based deep learning approach,” Multimedia Tools and Applications, vol. 80, no. 8, pp. 13467–13483, 2021.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL-HLT’19, 2019, pp. 4171–4186.

[14] C. Sun, X. Qiu, Y. Xu, and X. Huang, “How to fine-tune BERT for text classification?” in China National Conf. Chinese Computational Linguistics (CCL’19), 2020, pp. 194–206.

[15] Y. Liu et al., “RoBERTa: A robustly optimized BERT pretraining approach,” arXiv preprint arXiv:1907.11692, 2019.

[16] Z. Lan et al., “ALBERT: A lite BERT for self-supervised learning of language representations,” in Proc. ICLR’20, 2020.

[17] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” arXiv preprint arXiv:1910.01108, 2019.

[18] C. Sun, H. Yang, Z. Sun, and K. Wang, “Fine-tuning pre-trained BERT for fake news detection,” Neural Computing and Applications, vol. 32, no. 23, pp. 17314–17325, 2020.

[19] 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’17), 2017, pp. 797–806.

[20] K. Shu, A. Mahudeswaran, and H. Liu, “FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media,” arXiv preprint arXiv:1809.01286, 2018.

[21] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (pp. 5998–6008). Curran Associates, Inc.