IRJAEH

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 153-158
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0021

Leveraging DistilBERT-Multilingual for Robust and Efficient AI-Based Fake News Detection

Arpita Kulkarni[1]*, Vaishnavi Pawade*[2]*, Shilpa Mangshetty*[3]
[1,2]*UG – Computer Science and Design, PDA College of Engineering, Kalaburagi 585102, Karnataka, India.*
[3]*Associate Professor, Computer Science and Engineering, PDA College of Engineering, Kalaburagi 585102, Karnataka, India.*
Emails: *arpitakulkarnii09@gmail.com*[1]*, pawadevaishnavi571@gmail.com*[2]*, shilpamangshetty@pdaengg.com*[3]

Abstract

In this research, a machine learning–based framework for fake news detection using the DistilBERT-base-multilingual-cased Transformer model is proposed. The rapid growth of social media platforms has significantly increased the spread of misinformation, making accurate and scalable fake news detection a critical challenge. To address this issue, the proposed system focuses on multilingual text classification while maintaining low computational complexity, making it suitable for real-world applications. A large and diverse training dataset is created by merging multiple open-source datasets, enabling the model to effectively learn linguistic patterns across different languages and sources. Prior to training, the textual data is carefully pre-processed through cleaning, normalization, and tokenization to improve classification accuracy and reduce unnecessary computational overhead. The DistilBERT model is then fine-tuned to classify news content as real or fake based on semantic and contextual information. Extensive experiments conducted on social media–based text data demonstrate that the proposed approach achieves an accuracy of 88.97% and a precision of 0.99 in detecting fake misinformation. These results highlight the effectiveness of lightweight Transformer-based architectures in identifying misinformation while maintaining efficiency and scalability. The study confirms that artificial intelligence–driven and deep learning–based models can play a significant role in mitigating the spread of fake news and improving the reliability of online information ecosystems.

Keywords: *DistilBERT; Fake news detection; Misinformation; Multilingual text classification; Transformer model.*

1. Introduction

The rapid expansion of digital communication platforms has transformed how information is produced, shared, and consumed. Social media networks, online news portals, and instant messaging applications enable content to reach millions of users within seconds. While this accessibility has improved information flow, it has also accelerated the spread of fake news and misinformation. Misleading content has the potential to influence public opinion, disrupt social harmony, and erode trust in institutions, making fake news detection a pressing challenge in today's digital ecosystem (Shu et al., 2017; Zhou and Zafarani, 2018). Early approaches to fake news detection relied on manual verification, rule-based systems, and classical machine learning models such as Naïve Bayes, Support Vector Machines, and Logistic Regression. These methods typically depend on handcrafted textual features like bag-of-words and TF-IDF representations. Although effective to a limited extent, such approaches struggle to capture contextual meaning, semantic relationships, and long-range dependencies in text. As misinformation becomes more sophisticated, these limitations reduce the effectiveness of traditional detection systems (Zhang et al., 2021). Recent advances in natural language processing have led to the adoption of deep learning and transformer-based architectures for fake news detection. Models such as BERT introduced bidirectional attention mechanisms that significantly improved contextual understanding and classification performance (Devlin et al., 2019). Several studies have demonstrated that transformer-based models outperform earlier neural architectures in detecting deceptive content across different domains (Hanselowski et al., 2018; Shu et al., 2017). However, the large size and high computational requirements of full-scale transformer models limit their suitability for real-time or resource-constrained applications. To address these challenges, lightweight transformer models such as DistilBERT were introduced. DistilBERT achieves substantial reductions in model size and inference time while preserving most of BERT's language understanding capability (Sanh et al., 2019). Its multilingual variant further extends this advantage by enabling effective processing of content across multiple languages, which is particularly important in linguistically diverse regions. Motivated by these developments, this research proposes a multilingual fake news detection framework based on the DistilBERT-base-multilingual-cased model, aiming to achieve high detection accuracy while maintaining computational efficiency [1].

1.1. Background and Related Work

Research on fake news detection has evolved from simple content-based analysis to more advanced deep learning approaches. Early studies primarily focused on identifying linguistic patterns and statistical features associated with misinformation (Shu et al., 2017). While these methods provided valuable insights into fake news characteristics, they lacked robust contextual understanding. The introduction of transformer architectures marked a significant shift in this field. BERT-based models demonstrated strong performance by capturing bidirectional context within text (Devlin et al., 2019). Subsequent research highlighted the effectiveness of transformer-based representations for fake news detection, particularly when compared to CNN- and LSTM-based models (Zhang et al., 2021). However, the computational cost of these models remained a key limitation. DistilBERT addressed this issue by applying knowledge distillation to create a smaller and faster model without substantial loss in accuracy (Sanh et al., 2019). Recent studies have shown that DistilBERT performs competitively in misinformation detection tasks while offering improved efficiency, making it suitable for large-scale and multilingual applications [2].

1.2. Motivation and Contribution

Despite advances in fake news detection, many existing systems remain limited by high computational requirements and narrow language coverage. Most models are trained primarily on English datasets, reducing their effectiveness in multilingual environments. Additionally, heavyweight transformer architectures are difficult to deploy in real-time scenarios. This study is motivated by the need for a practical and scalable fake news detection system that balances accuracy, efficiency, and multilingual capability. The primary contribution of this work is the development of a lightweight multilingual framework based on DistilBERT-base-multilingual-cased. By combining multiple open-source datasets and fine-tuning a compact transformer model, this research demonstrates that efficient architectures can achieve reliable fake news detection across languages while remaining suitable for real-world deployment [3].

2. Method

This section describes the methodology adopted for developing the multilingual fake news detection system using the DistilBERT-base-multilingual-cased model. The methods are presented concisely while providing sufficient technical detail to allow reproducibility by a qualified reader. Established preprocessing and transformer-based modeling procedures are referenced from prior work, while only the task-specific implementation details are described in depth [4].

2.1. Dataset Description

To build a robust multilingual fake news detection system, multiple publicly available datasets were combined to form a unified training corpus. The datasets include the Fake and Real News Dataset from Kaggle, Bharat Fake News Kosh, and the Zenodo Fake News Corpus. These datasets contain labeled news articles classified as Fake or Real and cover content written in English, Hindi, and Kannada. Before merging, all datasets were standardized into a common structure by aligning column names, label formats, and text fields. Duplicate entries and incomplete records were removed to prevent bias during training. The final dataset provides a diverse linguistic and topical representation, enabling the model to learn misinformation patterns across multiple languages and sources Shown in Table 1 [5].

Table 1 Summary of Datasets Used for Training

Dataset Name	Language (s)	Number of Samples	Source
Fake and Real News Dataset	English	~45,000	Kaggle
Bharat Fake News Kosh	Hindi, English	~25,000	Kaggle
Zenodo Fake News Corpus	Kannada, English	~20,000	Zenodo

2.2. Text Preprocessing

Raw news articles collected from online sources often contain noise such as URLs, HTML tags, emojis, special characters, and inconsistent spacing. To address this, a preprocessing pipeline was applied to clean and normalize the text. The process involved removing URLs, non-textual symbols, repeated punctuation, and irrelevant metadata. Since the selected model is DistilBERT-base-multilingual-cased, original word casing was preserved to retain semantic information. Classical NLP techniques such as stop-word removal, stemming, and lemmatization were intentionally avoided, as transformer-based models rely on contextual word representations rather than reduced lexical forms. This preprocessing approach ensures clean and semantically rich input for tokenization and model training [6].

2.3. Tokenization and Model Architecture

Tokenization was performed using the DistilBERT multilingual tokenizer, which converts text into subword tokens suitable for transformer-based processing. Special tokens such as [CLS] and [SEP] were added automatically, and all input sequences were padded or truncated to a fixed length to ensure uniformity. Attention masks were generated to distinguish actual tokens from padded values. The DistilBERT-base-multilingual-cased model consists of six transformer encoder layers and employs multi-head self-attention mechanisms to capture contextual relationships across words and sentences. The contextual embedding corresponding to the [CLS] token was extracted and passed to a dense classification layer for binary prediction, classifying each news article as Fake or Real. This architecture enables efficient learning while significantly reducing computational overhead compared to full-scale transformer models [2] Shown in Figure 1 [7].

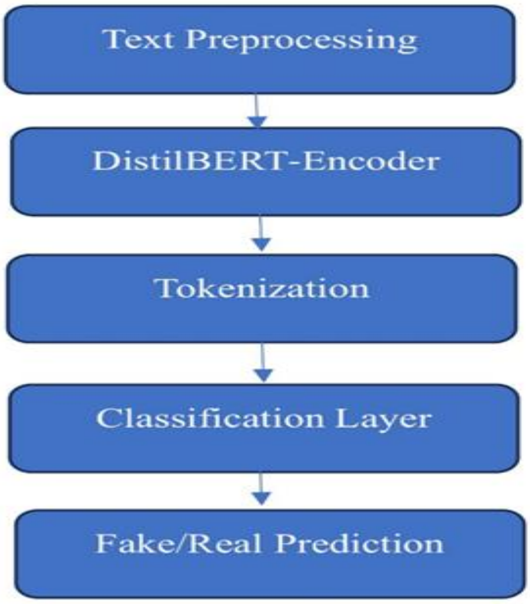

Figure 1 Workflow of the Proposed Multilingual Fake News Detection System

2.4. Model Training and Evaluation Setup

The model was fine-tuned using the AdamW optimizer with a low learning rate to preserve the pre-trained semantic knowledge of the transformer. The dataset was divided into training and testing sets using an 80:20 split. Class weighting was applied during training to handle class imbalance and improve sensitivity toward fake news samples. The model performance was evaluated using standard metrics, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive understanding of classification effectiveness and reliability. All experiments were conducted in a GPU-enabled environment to ensure efficient training and evaluation Shown in Table 2 [8].

Table 2 **DistilBERT Model Training Parameters**

Parameter	Value
Model	DistilBERT-base-multilingual-cased
Optimizer	AdamW
Learning Rate	1e-5
Batch Size	16
Epochs	3
Loss Function	Binary Cross Entropy
Evaluation Metrics	Accuracy, Precision, Recall, F1-score

3. Results and Discussion

This section presents the experimental results obtained from the proposed multilingual fake news detection system and provides an interpretation of the model's performance. The evaluation focuses on classification accuracy and class-wise performance to assess the effectiveness and reliability of the DistilBERT-based approach [9].

3.1. Results

The experiments were conducted to evaluate the performance of the fine-tuned DistilBERT-base-multilingual-cased model on unseen news articles. After preprocessing and tokenization, the dataset was divided into training and testing sets using an 80:20 split. The trained model was then evaluated on the test set to measure its ability to correctly classify news articles as Fake or Real Shown in Table 3.

Table 3 **Performance Metrics of the Proposed Model**

Class	Precision	Recall	F1-Score	Accuracy
Fake	0.99	0.78	0.87	—
Real	0.83	0.99	0.90	—
Overall	—	—	—	0.8897

The model achieved an overall classification accuracy of 88.97%, demonstrating strong generalization capability across multilingual inputs. Performance was further analyzed using precision, recall, and F1-score to obtain a detailed understanding of class-wise behavior. These metrics are particularly important in fake news detection, where minimizing false positives is critical. The results show that the model achieves very high precision for the Fake class, indicating that when the system identifies content as fake, it is highly likely to be correct. This behavior is desirable in real-world applications, as incorrect labeling of genuine news as fake can negatively impact credibility and trust Shown in Figure 2 Process of the Dataset Used for Training and Evaluation [10].

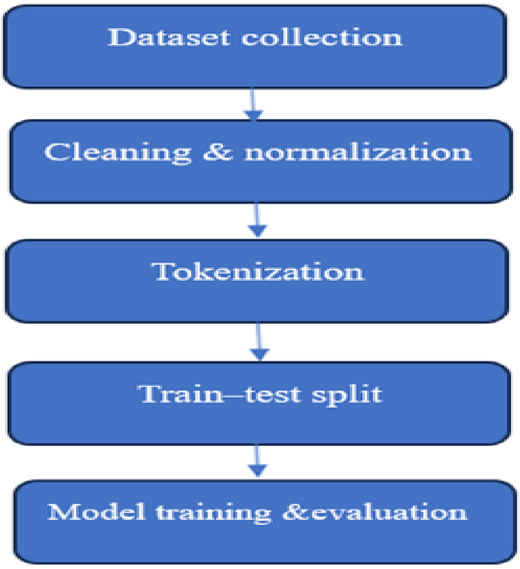

Figure 2 **Process of the Dataset Used for Training and Evaluation**

3.2. Discussion

The experimental results indicate that the proposed DistilBERT-based multilingual fake news detection system performs effectively across diverse linguistic inputs. The high precision achieved for the Fake news class highlights the model's strong capability to identify misleading content with minimal false positives. This is particularly important in misinformation detection, where wrongly flagging real news can have serious social and ethical implications. The recall value for fake news suggests that while the majority of misleading articles are successfully detected, a small portion may still remain undetected. This trade-off reflects the conservative nature of the model, which prioritizes precision over recall to ensure reliability. The strong recall observed for the Real news class further confirms the model's ability to correctly recognize authentic information. The effectiveness of the system can be attributed to the contextual encoding capability of the DistilBERT architecture. Unlike traditional machine learning models that rely on surface-level features, the transformer-based approach captures semantic relationships and linguistic patterns within text. Additionally, the use of multilingual datasets enables the model to generalize across different languages, addressing a key limitation of many existing fake news detection systems. Overall, the results demonstrate that a lightweight transformer model such as DistilBERT can achieve a favorable balance between accuracy, efficiency, and multilingual support. The findings confirm that computationally efficient architectures are well-suited for large-scale and real-time misinformation detection without compromising performance.

Conclusion

This study addressed the growing challenge of fake news and misinformation by proposing a multilingual fake news detection system based on the DistilBERT-base-multilingual-cased model. The system was designed to accurately classify news articles as fake or real while maintaining low computational complexity, making it suitable for practical deployment. By combining multiple publicly available datasets and applying effective preprocessing and fine-tuning strategies, the proposed model achieved an overall accuracy of 88.97% with a high precision of 0.99 for detecting fake news. These results confirm that lightweight transformer-based architectures can deliver reliable performance comparable to larger models while offering improved efficiency. The findings from the results and discussion validate the effectiveness of the proposed approach in addressing multilingual misinformation detection. The system demonstrates strong potential for real-world applications, particularly in environments where linguistic diversity and scalability are critical. Overall, this work confirms that efficient transformer models can play a significant role in mitigating the spread of fake news and improving the reliability of digital information ecosystems.

Acknowledgements

The authors would like to express their sincere gratitude to Dr. Shilpa Mangshetty, Associate Professor, Department of Computer Science and Engineering, PDA College of Engineering, Kalaburagi, for her continuous guidance, valuable suggestions, and constructive feedback throughout the course of this research. Her expertise and encouragement played a crucial role in shaping the methodology and improving the overall quality of this work. The authors also acknowledge the support provided by the faculty members of the Department of Computer Science and Design, PDA College of Engineering, for offering a conducive academic environment and necessary resources for carrying out this study. Additionally, sincere appreciation is extended to friends and peers for their helpful discussions and cooperation. The authors are grateful to their families for their constant motivation and support during the completion of this research.

References

[1]. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the NAACL-HLT Conference.

[2]. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

[3]. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu,

H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36.

[4]. Hanselowski, A., et al. (2018). A benchmark dataset for fake news detection. Proceedings of the Workshop on Fact Extraction and VERification (FEVER).

[5]. Jin, Z., Cao, J., Zhang, Y., & Luo, J. (2017). Multimodal fusion with recurrent neural networks for fake news detection. Proceedings of the ACM Multimedia Conference, 795–816.

[6]. Kaggle. (2018). Fake and real news dataset. Available at: https:/ /www. Kaggle .com/ datasets/clmentbisaillon/fake-and-real-news-dataset

[7]. Kaggle. (2020). Bharat fake news kosh dataset. Available at: https:// www. kaggle. com/datasets/man2191989/bharatfakenewskosh

[8]. Zenodo. (2023). Fake news corpus. Available at: https://zenodo.org/records/11408513

[9]. Zhou, Y., & Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315.

[10]. Zhang, C., et al. (2021). A survey on deep learning for fake news detection. IEEE Transactions on Knowledge and Data Engineering.