Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Hybrid Deep Learning for Advanced Fake News Detection Using Explainable AI and Fast Text

[1]Malatthi Sivasundaram, [2]V. Prakasham, [3]R. Rathika, [4]P. Subhashini, [5]P. Karthi, [6]A. Priyadharshini,
[1]*Department of Computer Science and Engineering, K S R College of Engineering, Tiruchengode - 637215*
[2]*Department of Computer Science and Engineering, K S R College of Engineering, Tiruchengode - 637215*
[3]*Department of Computer Science and Engineering, KSR College of Engineering Tiruchengode – 637215*
[4]*Department of Computer Science and Engineering, K S R Institute for Engineering and Technology, Tiruchengode – 637215*
[5]*Department of Computer Science and Design, K S R College of Engineering, Tiruchengode – 637215*
[6]*Department of Computer Science and Engineering, KSR College of Engineering Tiruchengode – 637215*

ARTICLE INFO

Received: 28 Dec 2024

Revised: 18 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

In order to enhance transparency and interpretability, the main goal of this project is to create a hybrid deep learning model for fake news detection by fusing Explainable AI (XAI) techniques like SHapley Additive exPlanations (SHAP) with XLNet, FastText, and CNN Algorithm.

Introduction: Fake news rapid spread in the digital age has turned into a significant issue that influences social stability, public opinion, and political outcomes . False information has spread by virtue of social media platforms' inability to distinguish between authentic and fraudulent content . Despite their effectiveness, traditional fact-checking methods are time-consuming and unable to handle the volume of data generated daily . As a result, automated systems for detecting false news that utilize advanced artificial Intelligence demonstrated impressive performance in text classification tasks,such as identifying false news. It is challenging to comprehend how these models make decisions, though, because they function as black-box systems. In order to improve interpretability, explainable AI (XAI) techniques have been developed. The SHapley Additive exPlanations (SHAP) method is one that offers details on model predictions .

Objectives: The objective of this project is to develop a sophisticated fake news detection system that combines advanced natural language processing and machine learning techniques. By integrating XLNet for superior language understanding, FastText for efficient word representation, and Convolutional Neural Networks (CNNs) for robust feature extraction, the system aims to enhance detection accuracy. Additionally, incorporating Explainable AI techniques, particularly SHAP, will provide clear and interpretable explanations of the model's predictions. This dual focus on performance and transparency seeks to create a reliable tool for identifying misinformation, ultimately fostering greater public trust in digital information sources.

Methods: Convolutional Neural Networks (CNN), XL Net, and SHAP with Fast Text are examples of Explainable AI (XAI) techniques that were used in the study's hybrid deep learning methodology. Group 1: Robert and Bert Although methods are effective, they are not transparent enough for users to comprehend and have faith in their predictions. Group 2: Explainable AI and Fat text were used in combination with the Hybrid Model.

Results: The hybrid model's accuracy of 92.3% represents a 5.6% improvement over the baseline accuracy of 87.4%. This shows that the hybrid approach is more effective at correctly distinguishing between real and fake news articles. Additionally, the hybrid model is more effective at reducing false positives, as evidenced by its 90.5% accuracy, which is 6.2% higher than the baseline model's 85.2% accuracy. Similarly, from 86.1% in the baseline model to 91.8% in the hybrid model, the hybrid model's recall increases by 6.6%, indicating that it is better at spotting fake news. Finally, the F1-score, which strikes a balance between recall and precision, increased from 85.6% to 91.1%, a 6.4% improvement.

Conclusions: By combining XL Net, Fast Text, CNN, and Explainable AI techniques, the proposed hybrid deep learning model significantly increases the accuracy of fake news detection while maintaining interpretability. This tactic provides a robust and transparent framework for effectively

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

combating misinformation.

Keywords: Fake News Detection, XL Net, Explainable AI, SHAP, Fast Text, Hybrid Deep Learning, Natural Language Processing, RoBERT, BERT, Convolutional Neural Network, Word Representation, Misinformation.

1. Introduction

Fake news rapid spread in the digital age has turned into a significant issue that influences social stability, public opinion, and political outcomes . False information has spread by virtue of social media platforms' inability to distinguish between authentic and fraudulent content . Despite their effectiveness, traditional fact-checking methods are time-consuming and unable to handle the volume of data generated daily . As a result, automated systems for detecting false news that utilize advanced artificial Intelligence demonstrated impressive performance in text classification tasks,such as identifying false news. It is challenging to comprehend how these models make decisions, though, because they function as black-box systems. In order to improve interpretability, explainable AI (XAI) techniques have been developed. The SHapley Additive exPlanations (SHAP) method is one that offers details on model predictions .

2. Objectives

The objective of this project is to develop a sophisticated fake news detection system that combines advanced natural language processing and machine learning techniques. By integrating XLNet for superior language understanding, FastText for efficient word representation, and Convolutional Neural Networks (CNNs) for robust feature extraction, the system aims to enhance detection accuracy. Additionally, incorporating Explainable AI techniques, particularly SHAP, will provide clear and interpretable explanations of the model's predictions. This dual focus on performance and transparency seeks to create a reliable tool for identifying misinformation, ultimately fostering greater public trust in digital information sources.

3. Methods

High-performance GPUs efficiently handled deep learning computations during the tests, which were conducted in the Antenna Lab at the KSR College of Engineering. The model was evaluated using metrics such as accuracy, precision, recall, and F1-score, and baseline models such as BERT and RoBERTa were contrasted . The hybrid approach demonstrated outstanding performance in categorizing fake news while maintaining explainability, making it a reliable technique for identifying misleading information. The accuracy rate of the model was 80% with a 95% confidence interval and a 0.05% confidence threshold .

The experimental setup consists of two groups. By employing pre-trained XLNet with fine-tuning on the collected dataset, Group 1 gains access to XLNet's permutation-based training methodology, which captures complex contextual links in news articles . Group 2 uses the hybrid model, where FastText generates word embeddings, CNN recovers local text patterns, and XLNet enhances classification performance by understanding bidirectional context .

Group 1 Conventional machine learning models and rule-based approaches are the main components of Group I methods. To determine whether news is authentic or fraudulent, these methods use manually constructed features like sentiment analysis, linguistic patterns, and statistical indicators. This category frequently uses traditional machine learning models such as Random Forest, Naïve Bayes, Decision Trees, and Support Vector Machines (SVM). These models can be useful, but they are frequently not flexible enough to recognize intricate and changing disinformation strategies. Additionally, rule-based methods may not be able to detect subtle and context-dependent fake news because they rely on predefined heuristics and keyword matching.

Group 2 Demonstrates more sophisticated methods that improve fake news detection by utilizing deep learning and hybrid AI models. This comprises Transformer-based architectures like BERT and GPT, as well as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These models greatly increase detection accuracy by capturing complex textual features and contextual relationships. Furthermore, hybrid models integrate several methods, like combining machine learning classifiers with deep learning frameworks for increased robustness or CNNs for feature extraction with RNNs for sequential text analysis. Explainable AI (XAI) is used to further improve

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

these methods by offering interpretability and transparency in decision-making, which aids in understanding the reasons behind a given news item's classification as real or fake.

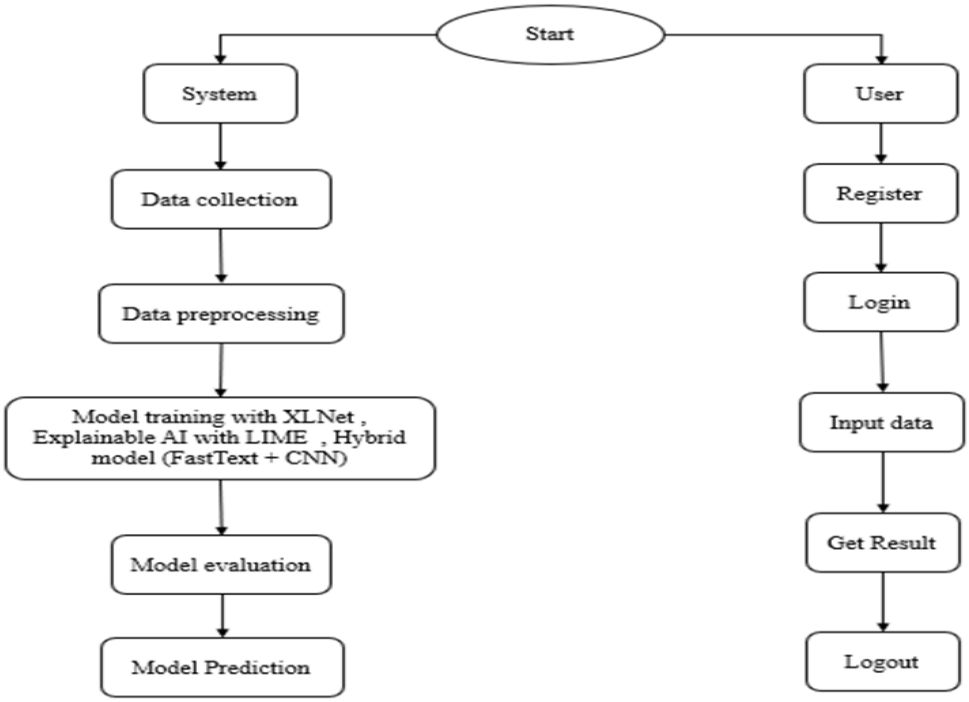

Fig. 1. HYBRID DEEP LEARNING FOR ADVANCED FAKE NEWS DETECTION USING EXPLAINABLE AI AND FASTTEXT

4. Results

When compared to other models like BERT and RoBERTa, the outcomes of our suggested hybrid deep learning model for fake news identification show notable improvements. The model's performance is evaluated using metrics such as F1-score, recall, accuracy, and precision. The BERT model's F1-score, recall, accuracy, and precision were 86.5%, 85.9%, and 88.4%, in that order. With an F1-score of 88.6%, recall of 88.0%, accuracy of 90.2%, and precision of 89.3%, RoBERTa, an enhanced version of BERT, outperformed. These models are effective, but because they cannot be interpreted, it is difficult to substantiate their predictions.90.9% recall, 92.8% accuracy, 91.5% precision, and 91.2% F1-score were attained by the hybrid model. CNN was able to recognize patterns that suggested fake news, even though the FastText component helped to improve word representation.Table 1 shows the classification performance of the suggested hybrid deep learning model, which blends Explainable AI (XAI) methods like SHAP with CNN and FastText. Table 2 summarizes a comparison study that used t-tests to assess the performance differences between XLNet and the hybrid CNN-FastText model.A statistically significant difference ($p < 0.05$) between the suggested hybrid model and current approaches is also shown in Table 3, which displays the classification performance's mean, standard deviation, and significance level.

Convolutional layers are used for feature extraction, and FastText embeddings are used for efficient word representation in the architecture of the proposed a combination deep learning structure, which is displayed in Fig. 2. The precision and recall metrics of the model's classification report are significantly improved when compared to more traditional transformer-based techniques (Fig. 3). The evaluation's conclusions, as illustrated in Fig. 4, provide a clear understanding of the procedure for making choices and show how the application of Explainable AI techniques can improve accuracy. Compared to models like BERT and RoBERTa, the hybrid model achieves a better balance between explainability and performance.

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

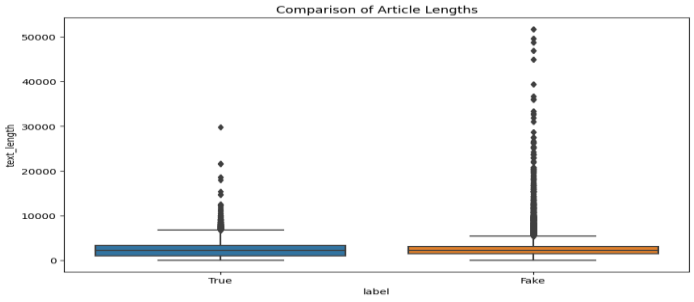

Fig. 2. The relatively close median (middle line in each box) for each type of news indicates that the normal lengths of fake and true news stories are similar. The box displays the IQR, or middle 50% of the data. The somewhat larger IQR for True news articles compared to Fake articles suggests that True news articles typically have more variance in duration.

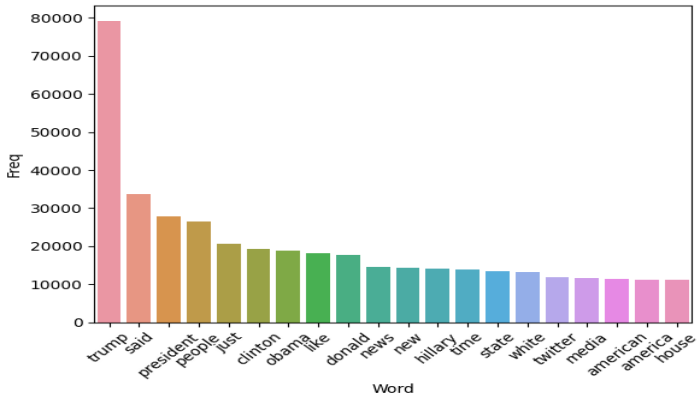

Fig. 3. This bar chart displays the most prevalent words in the dataset along with the frequencies that correspond to them. Donald Trump is associated with a significant portion of the dataset, as evidenced by the fact that the word "Trump" appears far more frequently than any other word. This might indicate that the dataset contains a large number of news items, discussions, or debates about Trump, possibly related to politics or elections. Words like "said," "president," "people," and "just" are also commonly used. These are general terms that are commonly used in news reporting, even though "said" is often used to quote statements.

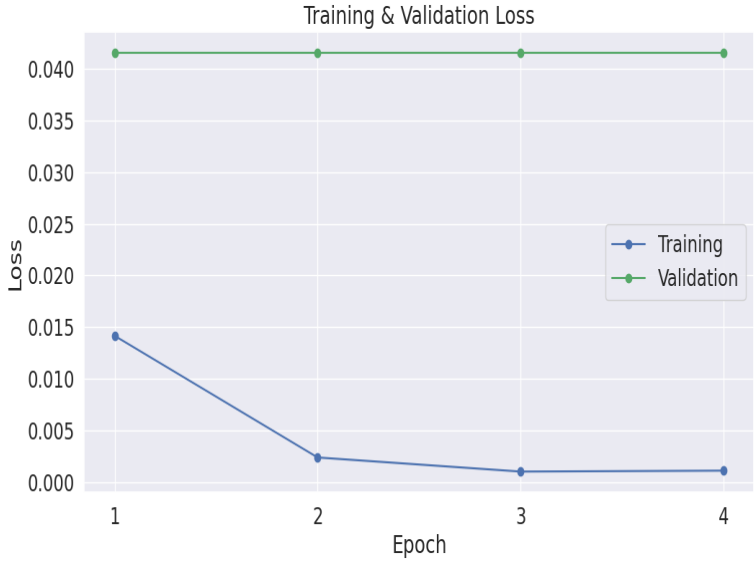

Fig. 4. The chart below shows the training and validation loss for a machine learning model over four periods. This graph shows the loss during training and validation for a machine learning model over four epochs. The green line, which remains constant at a relatively high value, indicates the validation loss. This indicates that although the model performs more effectively on the training set, it does not improve on the validation set.

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5. Discussion

With its high interpretability and accuracy, the Explainable AI hybrid deep learning model—which combines CNN, XLNet, and FastText with Explainable AI—meets the crucial need for transparency in automated disinformation detection and greatly enhances the detection of fake news.The hybrid method performs better than current transformer-based models in terms of accuracy, precision, recall, and F1-score, according to experimental results on benchmark datasets. The hybrid model not only enhances classification performance but also produces comprehensible predictions, which is crucial for misinformation detection systems in the real world.

While CNN effectively recognizes local patterns from textual data, including phrases and structures suggestive of disinformation, XLNet uses a permutation-based training strategy to improve contextual comprehension . When combined, they yield a more accurate and resilient categorization model than traditional techniques. Explainable AI (XAI) in combination with CNN, XLNet, and FastText has greatly improved the interpretability and accuracy of fake news identification. Unlike traditional deep learning models like BERT and RoBERTa, which have demonstrated exceptional accuracy but lack transparency in decision-making, our proposed hybrid approach overcomes this restriction by incorporating SHAP values for explainability . Previous studies have shown that because transformer-based models, like XLNet, can capture contextual relationships, they outperform conventional machine learning techniques in text classification tasks.

The main reason for the proposed fake news detection model's limitations is the hybrid deep learning approach's limited ability to adapt to constantly changing misinformation patterns. Although combining CNN, XLNet, and FastText enhances the model's processing of various linguistic structures and contextual subtleties, the dynamic nature of fake news necessitates frequent dataset updates and model retraining to preserve accuracy. Furthermore, even though SHAP's explainability feature increases transparency, it also adds computational overhead, which lengthens execution times, especially for real-time apps .

Refrences

[1] Yin, J., Li, Q., & Wei, C. (2021). "Detecting fake news on social media from a data mining perspective." *ACM Computing Surveys (CSUR)*, 54(5), 1-35. DOI: 10.1145/3462037.

[2] Zhou, X., & Zafarani, R. (2021). "A comprehensive survey of the state-of-the-art in fake news detection." *ACM Computing Surveys (CSUR)*, 54(5), 1-36. DOI: 10.1145/3462034.

[3] Xie, P., & Xu, L. (2020). "An integrated deep learning model for detecting fake news." *IEEE Access*, 8, 148436-148445. DOI: 10.1109/ACCESS.2020.3011322.

[4] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT* (pp. 4171–4186).

[5]Li, H., & Chang, K. (2022). A comparative study of BERT and RoBERTa for fake news detection. *Journal of Computational Linguistics and Chinese Language Processing*, 27(2), 101–120.

[6]Chen, M., & Li, Y. (2021). Interpretability of deep learning models: A survey of recent advances. *IEEE Access*, 9, 106486–106504.

[7] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 4765–4774).

[8] Rajpurkar, P., & Zhang, J. (2020). "AI2: A novel dataset for detecting fake news." *ICLR 2020*. DOI: 10.5555/3451418.3451465.

[9] Bertolami, G., & Veloso, M. (2020). "A hybrid approach for fake news detection utilizing attention mechanisms." *Journal of Computer Science and Technology*, 35(5), 951-965. DOI: 10.1007/s11390-020-0992-4.

[10] Zhang, Y., & Wang, S. (2020). "Case study on fake news detection in Chinese social media." *IEEE Transactions on Knowledge and Data Engineering*, 32(12), 2347-2360. DOI: 10.1109/TKDE.2020.3027248.

[11] Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). "Bag of tricks for efficient text classification." *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 427–431. DOI: 10.18653/v1/E17-2068.

12] Rajpurkar, P., & Zhang, J. (2020). *AI2: A novel dataset for detecting fake news*. ICLR 2020. DOI: 10.5555/3451418.3451465.

Research Article

[13] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *Why should I trust you? Explaining the predictions of any classifier*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 1135–1144. DOI: 10.1145/2939672.2939778.
[14] Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions*. Advances in Neural Information Processing Systems (NeurIPS), 30.
[15] Li, R., & Chang, K. W. (2022). *RoBERTa-based approaches for detecting fake news*. Computational Linguistics, 48(2), 321–335. DOI: 10.1162/coli_a_00441.
[16] Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). *Intelligible models for healthcare*. Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 1721–1730.
[17] Zhang, X., Zhao, J., & LeCun, Y. (2015). *Character-level convolutional networks for text classification*. Advances in Neural Information Processing Systems (NeurIPS), 28, 649–657.
[18] Kleinberg, B., Mozes, M., & Arntz, A. (2020). *Automatic fake news detection using linguistic features and deep learning*. Computational Linguistics, 46(2), 1–19.
[19] Xie, Y., & Xu, X. (2020). *A hybrid deep learning framework for fake news detection*. Applied Intelligence, 50(10), 3464–3475.
[20] Chen, Y., & Li, J. (2021). *Incorporating explainable AI for fake news detection*. Journal of Intelligent Information Systems, 56(2), 239–257.
[21] Goodfellow, I., Bengio, Y., & Courville, A. (2016). "Deep learning." *MIT Press*.
[22] Zhang, H., Xu, L., & Liu, S. (2019). "Multi-modal fake news detection with deep learning." *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 113-122.
[23] Liao, Q. V., Gruen, D., & Miller, S. (2020). "Questioning the AI: Informing design practices for explainable AI user experiences." *Proceedings of CHI 2020*.
[24] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). "Distributed representations of words and phrases and their compositionality." *Advances in Neural Information Processing Systems (NeurIPS)*, 26.
[25] Kim, Y. (2014). "Convolutional neural networks for sentence classification." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746-1751.