International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(14s) 2026 pp. 770-781
https://doi.org/10.70917/ijcisim-2026-4282
Article

Adaptive Semantic Feature Refinement for Explainable Fake News Detection Using Pretrained Transformers

Sudha Patel[1], Shilpa Serasiya[2], Sachi Bhavsar[3], Maitri Rashmikant Sakarvadiya[4]

[1]Monark University, Naroda, Ahmedabad, Gujarat, India
sudhace@gmail.com
[2]Department of Information and Technology, Sankalchand Patel College of Engineering, Sankalchand Patel University, Visnagar, Gujarat, India
shilpapatel84@gmail.com
[3] SAL Engineering and Technical Institute, GTU, Ahmedabad, Gujarat, India
sachivbhavsar@gmail.com
[4]Monark University, Naroda, Ahmedabad, Gujarat, India
maitrisakarvadia208@gmail.com

Abstract: The online media has proliferated and become more accessible, so too has the ease with which misinformation can spread and be consumed and automated mechanisms to detect this form of online deception will be very important research targets moving forward. The implementation of new techniques from deep learning and transformer models pre-trained on large amounts of data has greatly improved the ability to detect misinformation, however many detectors are hindered by limitations on their ability to utilize semantic features and interpret the resulting predictions. In this study, we present our Adaptive Semantic Feature Refinement for Explainable Fake News Detection Utilizing Pre-Trained Transformers model, which uses a novel hybrid deep learning architecture that combines DeBERTa-v3, Bidirectional Gated Recurrent Unit (BiGRU), Multi-Head Self-Attention, and an Adaptive Semantic Refinement Module (ASRM) to produce higher-quality representations of the text used to classify fake news articles into binary categories. In addition, by applying SHAP (SHapley Additive exPlanations) values and Integrated Gradients to improve prediction transparency, we were able to produce both global and local explanations of the model's predictions. Our model was tested using a large corpus of fake news articles that included 682,661 articles, the resulting accuracy was 85.85%, with a ROC AUC statistic of .9294. Overall, results indicate that our proposed architecture successfully combines the process of refining the semantic features of text data while providing an explainable artificial intelligence solution for real-world applications of fake news detection.

Keywords: Fake News Detection, DeBERTa-v3, BiGRU, Multi-Head Self-Attention, Adaptive Semantic Refinement Module (ASRM), Explainable Artificial Intelligence, SHAP, Integrated Gradients, Natural Language Processing.

1. Introduction

The rise of the internet, the development of many different online news sites and social media pages has greatly changed how we receive, create, and share information. According to some, this has created a much more connected world and greater access to data. While this is true, the quick accessibility of information has also contributed to the rapid rise of false news and misinformation. It is widely accepted that the circulation of false and misleading information influences people's opinions and ultimately affects how they vote, the stock market, and their overall health. Therefore, developing algorithms that automate the detection of false news has become an increasingly important area of research. Given the sheer volume of digital content being produced daily, there is an urgent need for intelligent systems that can identify false or misleading information with a high degree of confidence, helping to ensure that the information being disseminated via the internet is credible and reliable.

Prior to the advent of pre-trained transformer models, fake news detection primarily relied on human-created linguistic features and traditional machine learning approaches to identify false stories however, these traditional methods did not adequately capture the complex semantic and contextual relationships found in text data [1]. The introduction of transformer architectures complemented with explainable AI (XAI) has provided an improved means for developing models to detect fake news by allowing for better contextual representation learning [2]. Finally, incorporating an attention mechanism within explainability techniques will both improve the reliability of predictions and make automated fake news detection systems more transparent/trustworthy overall [4].

In spite of these major changes, lots of current fake news detection strategies are still having difficulties to provide an adequate amount of comprehension of semantically structured representations while sustaining sufficient amounts of model interpretability [5]. Although there is now a way to have improved contextualization and accuracy in machine learning through recent advancements in transformers and explainable AI processes, the produced feature representations still lack adequate levels of refinement to be dependable when decision-making processes are conducted under complicated, real-world situations [7]. Thus, the effective balance between accurate classification performance with sufficient refinement of semantic features and transparent model explanation remains a major challenge problem of research in explainable fake news classification [9].

Considering the above, this research presents a method entitled Adaptive Semantic Refinement for Explainable Detection of Fake News using Pre-Trained Transformers. This approach combines DeBERTa-v3, BiGRU, multi-head self-attention, and an Adaptive Semantic Refinement Module (ASRM) aimed at the development of contextualized semantics that can be leveraged to produce binary classifications related to fake news. Additionally, SHAP and Integrated gradients are also included within this framework to provide both global and local explanations about model predictions. Finally, upon validation against a large-scale fake news data set using standard metrics for classification, the overall results indicate that the framework offers reliable, interpretable, and ultimately trustworthy methods for detecting fake news within real-world settings.

2. Literature Review

The Need to Improve Fake News Detection & Monitoring Systems, in order to produce accurate predictions, recent research has suggested ways in which transformer architectures may be improved by means of better understanding contextual data (i.e., multiple modalities, graph networks, and fusion of evidence) [14]. A lot of research has gone into creating multi-source data that contains complementary and/or overlapping semantics to achieve improved classification of news stories while still taking into account the ever-changing state of fake news [15]. Along with this research into contextual data, researchers have also developed large-sized language models that promote explainable AI, and have reported success with these methods in providing trustworthy, transparent predictions from contemporary fake-news detection systems [17],[19].

Research is building on hybrid transformer systems that provide semantic understanding for news articles through multimodal data sources and attention-based architectures to be more effective at detecting fake news [20],[21]. Hybrid transformers that integrate both textual data as well as social contextual sources have demonstrated increasingly better performance in capturing semantic relationships between words within a news story [22]. Additionally, transformer and multimodal attention architectures have shown promise in producing more rigorous and generalized results in classes of systems that automatically identify fake news across various datasets and contexts [23],[24], [25].

Nevertheless, despite these efforts, most of the existing work has focused exclusively on improving classification accuracy or explainability, with little consideration given to how to improve the quality of the semantic feature representations while providing transparent prediction mechanisms. Additionally, an additional challenge is to balance contextual understanding, semantic refinement and interpretable decision-making when detecting fake news in the real-world. To address these issues, the proposed framework consists of DeBERTa-v3, BiGRU, Multi-Head Self Attention and an Adaptive Semantic Refinement Module (ASRM) combined with SHAP and Integrated Gradients to improve the learning of semantic representation while providing the global and local explanations for the prediction results. A comparative summary of the most relevant transformer-based and explainable fake news detection approaches discussed in this review is presented in Table 1, highlighting their key methodologies, strengths, and research limitations.

Table 1. Comparative Analysis of Recent Fake News Detection Approaches

	Methodology / Dataset	**Key Contribution**	**Limitations**
Athira et al., 2025 [1]	Multimodal Transformer + SHAP / Multimodal fake news	SHAP-based explainable multimodal fake news detection.	High complexity; requires multimodal data.
Kumar et al., 2025 [3]	Graph-Augmented Transformer / Social media fake news	Improved contextual learning using graph-enhanced transformers.	High training cost and complexity.
Mohammed et al., 2025 [4]	AMGET / Social media fake news	Enhanced evolving fake news detection using graph learning.	Dependent on multimodal inputs.
Hashmi et al., 2024 [6]	FastText + DL + XAI / Fake news benchmarks	Improved prediction transparency using XAI.	Limited contextual understanding.
Liu et al., 2024 [8]	Teller Transformer / Fake news benchmarks	Improved explainability and trustworthy predictions.	No semantic feature refinement.
Thakur et al., 2026 [14]	Evidence Fusion / Multimodal fake news	Improved reasoning through multimodal evidence fusion.	Complex architecture; multimodal dependency.
Wang et al., 2025 [16]	Causal Transformer / Fake news datasets	Integrated causal reasoning for explainable detection.	High computational overhead.

Based on the reviewed literature, there remains a need for a framework that effectively combines contextual semantic learning with explainable artificial intelligence. Motivated by this research gap, the proposed methodology integrates semantic feature refinement and explainability to improve fake news detection performance and prediction transparency.

3. Proposed Methodology

This new framework aims to help detect fake news by utilizing techniques such as Contextual Language Understanding, Sequential Dependency Modeling, Semantic Feature Refinement, and Explainable Artificial Intelligence (XAI). It does this through a unified architecture that processes raw news articles using a structured pipeline. This includes Text Preprocessing (to clean up the raw news articles), Contextual Feature Extraction (with DeBERTa-v3), Sequential Learning (with Bidirectional Gated Recurrent Unit [BiGRU]), Multidirectional Attention (Multi head Self-Attention), and Adaptive Semantic Refinement Modules (ASRMs), before ultimately classifying it as either real or fake. The use of SHAP and Integrated Gradients will increase transparency, thus, helping to ensure trustworthy decisions by providing an analysis of the importance of each global feature and by enabling interpretation of how each token affects the model's prediction of class or score. This overall architecture should provide accurate, robust, and interpretable fake news detection while maintaining efficient semantic representation learning.

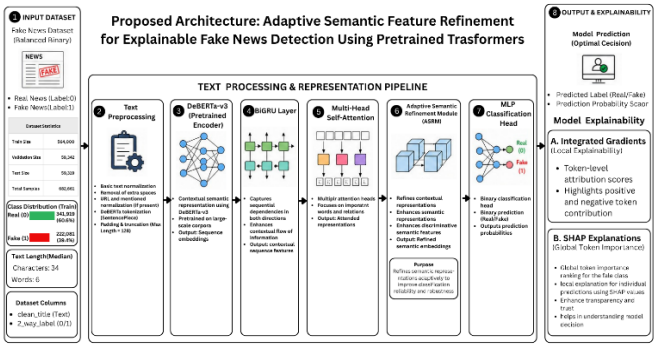

Figure 1. Proposed Architecture

The complete architecture of the proposed framework is presented in Figure 1, highlighting the sequential processing pipeline from text preprocessing and contextual feature extraction to semantic refinement, classification, and explainability for reliable fake news detection.

A. Dataset Description

To assess the performance of the proposed framework, an evaluation was performed on a large-scale binary fake news data collection composed of 682,661 news articles pulled from freely accessible online sources. The primary purpose of this data is to provide a source for supervised binary classification since the dataset is divided into two classes (referred to as Real or Fake). For each of the articles in this dataset, a news title is provided along with its corresponding class label to provide the necessary contextualized semantic information needed to build the model. This data split is also important for establishing a trustworthy basis for the development of the model, since it enables the creation of three distinct data subsets for training, validation, and testing. This approach permits the proposed framework to create valid semantic representations while minimizing instances of overfitting and improving the ability of the model to generalize from data that has not yet been utilized in the development of the model.

The large size and wide-ranging nature of the corpus of content within the dataset are the primary reasons for its selection to support the training of transformer-based language models for semantic understanding. The balanced classification of the samples in the various categories of news provides the proposed framework with the opportunity to separate true articles from/and false articles, while also providing a means to evaluate how well the proposed framework performs at classifying between these two classes of information.

B. Data Preprocessing

Developing models for deep learning requires having clean text data which is often not available prior to being created due to inconsistencies that introduce errors into the process of predicting values from training samples using prebuilt algorithms. As such, the first step in this processing technique will include cleaning or preparing the news articles so they can be used effectively when performing feature extraction based on transformer models.

Normalizing the input text was accomplished to address formatting differences but still capture the meaning within the input text to assist with determining which text contains fake news and non-fake news. After the input text was normalized, it was then transformed into tokens, this was done using the tokenizer associated with the pretrained Transformer Model DeBERTa-v3.

Once the text was in the form of tokens and had gone through the padding or truncation process to be converted into a fixed maximum sequence length for training purposes, the token IDs and attention mask were provided as input into the DeBERTa-v3 Encoder to extract the context from the tokens. This processor allows for the variable lengths of the news articles being processed and retains the semantically relevant information necessary for accurately determining whether an article is fake or not.

C. Proposed Hybrid Framework

A combination of pre-trained language modeling based on Transformers with sequential learning, attention mechanisms, and explainable AI is utilized for improving the detection of fake news. Each part of this pipeline provides a different operational role within the overall framework that is designed to allow the capture of contextual

semantics, emphasis on important text patterns, refinement of feature representations, and creation of interpretable prediction outputs. The next subsections that provide further details about these components are outlined in subsequent sections.

1) DeBERTa-v3 Encoder

The initial step in our proposed framework uses DeBERTa-v3 (Decoding-enhanced BERT with Disentangled Attention) as the main module for extracting features from data. This software uses a different approach to encoding semantic data and positional data through an innovative technique called disentangled attention. This allows for a better representation of contextual relationships between text.

DeBERTa-v3 also has an advanced architecture that allows for the ability to learn long-term dependencies, while preserving the meaning of the words in the article. This advantage will create better representations of the phrase, therefore allowing the developed framework to make accurate classifications of Breaking news.

In our developed framework, each news article will be pre-tokenized into individual tokens prior to being passed through the DeBERTa-v3 encoder. After completing this step, each token will be converted into an embedded high-dimensional vector with context-embedded meanings of syntax and semantic for classification. Rather than classifying the tokens immediately after producing their embeddings, they will be sent to later layers of processing to learn about sequential dependent structures within the converted embedded representation of tokens. This enables the created framework to learn richer, more discriminative textual features.

2) Bidirectional Gated Recurrent Unit (BiGRU)

DeBERTa-v3's rich contextual embeddings can be enhanced by modeling sequential relationships (dependencies) over the full length of the sequence and therefore, in this proposed framework, a Bidirectional Gated Recurrent Unit (BiGRU) will be added to the model to utilize both forward and backward contextual information from the embeddings generated by the transformer encoder. This will improve how the model represents the relationship between preceding and succeeding words (words that come before or after) and improve representations of semantic dependencies within a news article.

The BiGRU will take the contextual embeddings and process them in both directions (backwards and forwards) and produce one joint sequence representation that will retain the long ""range"" characteristics of the contextual information. The BiGRU will complement the contextual information learned by DeBERTa-v3 and, thus, strengthen the learning of sequential features and produce richer representations of the semantic content of the words that will be sent to the Multi-Head Self-Attention module to assist in finding the most informative patterns in the text.

3) Multi-Head Self-Attention

The output sequences produced by the BiGRU network undergo further processing with a multi-headed self-attention (MHSA) module, which allows us to extract the most significant semantic information contained within the news text. Although the BiGRU network provides bidirectional contextual dependencies for each token (word or other entity), not every token has the same influence on the ultimate prediction made by the model. The attention mechanism provides a method for the model to weight the relative importance of each token based upon its relationship to other tokens in the entirety of the input sequence, this enables the model to focus on relevant contextual information for producing an accurate classification.

By using multiple attention heads, the framework can learn from the different subspaces of token representations, thereby providing complementary semantic relationships between the tokens simultaneously. This improves the model’s capability to capture local and global contextual interactions necessary for differentiating between true and false news reports. The attention-based representations are passed to the Adaptive Semantic Refinement Module (ASRM) for further refinement prior to making the final classification.

4) Adaptive Semantic Refinement Module (ASRM)

The Adaptive Semantic Refinement Module (ASRM) was developed to improve how contextually representative feature representations are formed by earlier layers in preparation for the final layer of classification. The combination of DeBERTa-v3, BiGRU, and Multi-Head Self-Attention can capture both contextual and sequential information so it is conceivable that some of the resulting feature representations will contain redundant or less useful semantic information than others. The ASRM addresses this issue of redundancy by refining the previously produced attention-enhanced feature representations and enhancing those semantic features that have the greatest value for carrying out the fake news classification task.

Using the ASRM, our classification framework will adaptively process the feature representations at the end of the ASRM module to increase their discriminative strength but will maintain the useful/meaningful semantic relationships that have been established with prior layers. Refining feature representations using the ASRM will allow the classification model to be trained with improved feature representation and therefore increase prediction accuracy. Finally, the refined semantic feature representations created by the ASRM will be used as input to the multi-layer perceptron (MLP) classification head for binary classification.

5) Classification Head

The Multi-Layer Perceptron (MLP) classification head receives the refined semantic representations from the Adaptive Semantic Refinement Module (ASRM) and uses the fully connected layers of the MLP to predict scores based on the transformed and refined feature representations. The MLP predictions will allow the system to create a distinguishing boundary between Real or Fake news articles. The classification head learns this boundary based on the semantic information obtained and is refined during the previous layers.

The output of the MLP classification head is a probability score that will provide a news article with a classification corresponding to its two target classes. By using the refined semantic embeddings, the classification component will provide a more reliable and discriminative binary decision than if only the raw contextual features had been used and thus are compatible with the explainability methods developed in the proposed framework.

D. Explainability Module

Additional explanations will be provided through the use of integrated gradient techniques to offer local feature level explanations of the model's reasoning for detecting fake news across its predictions. Using the integrated gradient method will help explain how individual words, phrases, or phrases play a role in contributing to the decisions made by the model and provide explanations without changing or modifying how the model makes predictions.

1) SHAP (SHapley Additive exPlanations)

Each time that SHAP is used, it provides a global level of explanation for the overall fake news detection model. SHAP uses a cooperative model from the theory of cooperative games to assign each feature of the model (if it has been assigned a particular classification by the model) an importance value based on how much it influenced the model's ultimate classification. Through its methodology of measuring the contributions made to the decision process by the model, SHAP helps establish a better understanding of the main features of the overall model and assists with establishing the primary reasons for its predictions with some measure of significance and to maintain that the model centers around significant contextual influence rather than irrelevant or non-contextual influence within the text to be evaluated.

Figure 2. Global Token Importance using SHAP

The SHAP (SHapley Additive exPlanations)-based global token importance generated via the proposed framework is shown in Figure 2. As a result, the contribution of important text features to the overall prediction process is shown, producing a worldwide comprehension of the model's behavior.

2) Integrated Gradients

Integrated Gradients is a local interpretability method that measures how much each input token contributed to the overall prediction from a single article. Integrated Gradients also makes apparent the positive and negative contributions that each token has when providing predictions for all tokens in a particular document. As a result, you can see how each token contributed to the decision made by the AI model, improving the interpretability of an instance's prediction, while working in conjunction with global explanations provided by SHAP. These methods of explainability provide a complete understanding of the actions of the entire AI model, as well as how the AI model makes decisions on a particular instance.

Figure 3. Token-Level Attribution using Integrated Gradients

Here Figure 3 displays the token-level integrative gradient-based attribution. The positive and negative contributions of specific tokens to a prediction are illustrated, thereby enhancing clarity around individual classification deliberations.

D. Training Configuration

A new proposed framework using an end-to-end learning approach developed on the training data was used to validate against model performance during training with the validation dataset. The model was administered with three epochs of training, providing both the transformer encoder portion and subsequent neural network components an opportunity to learn how to make discriminative semantic representations of binary fake news for classification. After training, the optimized model was evaluated with the independent test dataset to determine the effectiveness of the model and ability for generalization and prediction accuracy.

E. Performance Evaluation Metrics

The effectiveness of the proposed framework was evaluated using commonly accepted classification measurements of accuracy, precision, recall, f1-score, confusion matrix and ROC-AUC. Accuracy indicates the overall accuracy of prediction, precision reflects upon the level of accuracy on the correct classifications only, and recall reflects upon the level of accuracy of all classifications regardless of whether they are correct. The f1-score is an overall measure of both precision and recall for binary classification. The confusion matrix gives an overview of prediction results in detail, and the ROC-AUC measures the effectiveness of the model in differentiating between real and fake news based on various classification thresholds as a complete assessment of the proposed framework utilized within this framework is provided through all metrics.

4. Result Analysis

The performance of the trained model using the proposed framework was assessed by observing how much accuracy and loss values for both training and validation improved through the duration of training across three epochs. Figure 4 illustrates that both training and validation accuracies improved continuously and that corresponding loss values decreased continuously during training.

Analysis of the model's learning curves provides evidence that the model learned semantic representations that were meaningful to it as it trained with no indications of being unstable during this time.

It is evident from the fact that the training curves for both the training and validation sets have very close agreement with one another and the proposed framework has produced a model that converged into a stable solution and did a good job generalizing previously unseen data. There is also no evidence of large amounts of divergence occurring between both training curves therefore, the model successfully reduced the amount of overfitting it experienced during the training period while maintaining a consistent rate of learning.

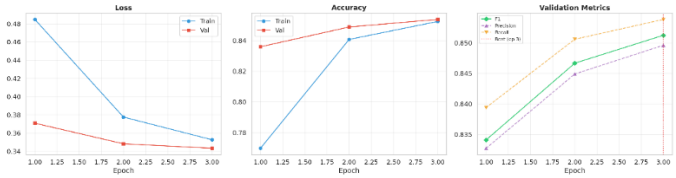

Figure 4. Training and Validation Accuracy and Loss Curves

Above Figure 4 illustrates the training and validation accuracy and loss obtained during model training. The curves demonstrate stable convergence and consistent learning behavior of the proposed framework across all training epochs.

A. Classification Performance

The proposed framework has been evaluated based on the performance of binary classification using a number of standard binary classification assessment metrics such as Accuracy, Precision, Recall, F1 Score and ROC-AUC to provide a complete view of the overall assessment of the ability for the model to provide predictions and the ability of the model to discriminate class-wise in addition to represent the balance between precision and recall using the independent test dataset to evaluate the ability for the proposed framework to generalize or predict external examples.

The overall classification performance of the proposed framework was 85.85% correct with 85.41% macro precision, 85.82% macro recall and 85.58% macro F1 score as summarized in Table 3. The ROC-AUC score of the proposed framework was 0.9294, indicating that it provides a strong ability to discriminate real news articles from fake news articles across all classification threshold levels as shown in the results. The table 2 provide evidence that the overall ability for the proposed framework to provide reliable and robust classification of fake news is made possible through the integration of contextual language representations, sequential learning, semantic refinement and explainable AI.

Table 2. Overall Performance of the Proposed Framework

Metric	**Value**	**Evaluation Type**
Accuracy	85.85%	Overall Performance
Precision	85.41%	Classification Metric
Recall	85.82%	Classification Metric
F1-score	85.58%	Classification Metric
ROC-AUC	0.9294	Threshold-independent Metric

B. Confusion Matrix Analysis

A confusion matrix is a representation of the classification results and can help us assess performance by showing how many articles are correctly or incorrectly classified. In this experiment (as shown in figure 5), there are

27,700 true articles and 20,073 fabricated articles classified correctly. The misclassified articles include 4,506 true articles that have been labelled as false and 3,370 false articles that have been labelled as true.

These classification results for both true and false articles demonstrate that the proposed system is capable of effectively separating true and false content. The majority of articles, for both classes, were correctly classified so the classification results indicate that there were consistent performance and reliable prediction among the entire testing data. These findings are verified through the accuracy, F1-score and ROC-AUC results that were obtained during the evaluation process.

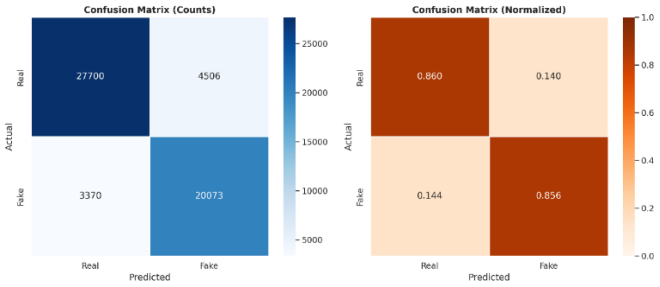

Figure 5. Confusion Matrix of the Proposed Framework

Figure 5 shows that the proposed framework achieves a high degree of accuracy and a significantly greater degree of accuracy for distinguishing between true (27,700) and false (20,073) articles than for all other articles combined (less than 4,730).

C. ROC Curve Analysis

An ROC Curve was created to assess the effectiveness of the proposed system for various cutoff values. Figure 6 presents an ROC Curve, which has demonstrated a global average area under the curve (ROC AUC) (0.9294), illustrating that real and fake news articles can be well distinguished from one another from a statistical perspective using the proposed methodology. The ROC curve was found to be close to the upper-left corner, indicating that for each true positive classification there were a minimum number of false positive classifications.

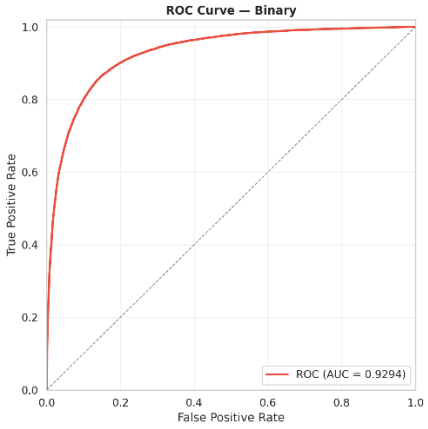

Figure 6. ROC Curve of the Proposed Framework

The ROC curve shown in Figure 6 is the ROC Curve generated for the test set using the proposed method. The ROC curve had an area underneath the curve (ROC AUC) of 0.9294, illustrating that the proposed method can classify news articles as either ‘fake’ or ‘real’ with a high degree of accuracy.

5. Conclusion

An Adaptive Semantic Feature Refinement for Explainable Fake News Detection Using Pretrained Transformers was proposed in this paper. The hybrid framework, leveraging the strengths of DeBERTa-v3, Bidirectional Gated Recurrent Units (BiGRU), Multi-Head Self-Attention and Adaptive Semantic Refinement Modules, significantly enhances the ability of semantic representations to learn from binary class (fake news) data. To

provide both global and local explanations of predictions made by the model and aid transparency and trustworthy decision making, Integrated Gradients and SHAP methods were used. Using a large-scale dataset of fake news articles, the framework demonstrated an overall accuracy of 85.85% and ROC-AUC score of 0.9294, proving reliable classification performance while providing interpretability of the model. With a unified architecture that combines contextual understanding through language, refinement of semantic features, and an explainable AI methodology, the proposed framework provides a viable solution to automating the detection of fake news. Future work may consider extending the framework to multimodal detection, multilingual datasets, and developing more efficient transformer architecture to improve scalability, robustness, and application in the real world.

References

1. Athira, A.B., Kumar, S.M. and Chacko, A.M., 2025. Pretrained transformers for multimodal fake news detection: Explainability using SHapley Additive exPlanations for contributions from text, image, and image captions. Engineering Applications of Artificial Intelligence, 162, p.112590.
2. Anand Babu, G.L., Sekhar Reddy, G., Sravan Kumar, G., Prashanth Rao, A., Nagalakshmi, N. and Vijay Kumar, S., 2026. A hybrid framework for fake news detection using transformer models with MASHAP. International Journal of Data Science and Analytics, 21(1), p.41.
3. Kumar, C., Bansal, M., Khan, M.A., Kaushik, V., Arquam, M. and Alabdultif, A., 2025. Graph-augmented transformer ensemble framework for robust and scalable fake news detection in social media ecosystems. Scientific Reports.
4. Mohammed, B.K., Nuiaa Alogaili, R.R., Hasan, B.M., Dashoor, Z.A., Alshami, H.G., Manickam, S. and Arjuman, N.C., 2025. AMGET: An Adaptive Multimodal Graph-Enhanced Transformer for Cybersecure Detection of Evolving Fake News on Social Media. International Journal of Intelligent Engineering & Systems, 18(10).
5. LekshmiAmmal, H.R. and Madasamy, A.K., 2025. A reasoning based explainable multimodal fake news detection for low resource language using large language models and transformers. Journal of Big Data, 12(1), p.46.
6. Hashmi, E., Yayilgan, S.Y., Yamin, M.M., Ali, S. and Abomhara, M., 2024. Advancing fake news detection: Hybrid deep learning with fasttext and explainable ai. IEEE access, 12, pp.44462-44480.
7. Szczepański, Mateusz, Marek Pawlicki, Rafał Kozik, and Michał Choraś. "New explainability method for BERT-based model in fake news detection." Scientific reports 11, no. 1 (2021): 23705.
8. Liu, H., Wang, W., Li, H. and Li, H., 2024, August. Teller: A trustworthy framework for explainable, generalizable and controllable fake news detection. In Findings of the Association for Computational Linguistics: ACL 2024 (pp. 15556-15583).
9. Bashaddadh, Omar, Nazlia Omar, Masnizah Mohd, and Mohd Nor Akmal Khalid. "Machine learning and deep learning approaches for fake news detection: A systematic review of techniques, challenges, and advancements." IEEE Access (2025).
10. Alex, M.G. and Peter, S.J., 2024. A Hybrid Approach for Integrating Deep Learning and Explainable AI for augmented Fake News Detection. Journal of Computational Analysis & Applications, 33(6).
11. Dolati, Mona, Sayvan Soleymanbaigi, Fatemeh Daneshfar, and Mourad Oussalah. "IMGA-Net: inconsistency-aware multiview graph attention network for explainable fake news detection." International Journal of Machine Learning and Cybernetics 17, no. 5 (2026): 202.
12. Likha, O. V., S. Sachin Kumar, Neethu Mohan, and O. K. Sikha. "Fake News and Offensive Content Detection in Malayalam Using Machine Learning, Deep Learning and Transformer based method with XAI." IEEE Access (2025).
13. Singh, S.K., Assi, S., Ginige, T., Mohammed, A.H., B. Wahit, F. and Al-Jumeily OBE, D., 2024, December. Fake News Detection Using Explainable Artificial Intelligence. In The International Conference on Data Science and Emerging Technologies (pp. 477-488). Singapore: Springer Nature Singapore.
14. Thakur, R., Kavindu Jayasinghe, J.P. and Mehta, P., 2026. Interpretable and robust multimodal fake news detection with evidence fusion and reliable reasoning. International Journal of Data Science and Analytics, 22(1), p.4.
15. Liu, Bingyi, Anqi Wang, and Chengqian Xia. "Interpretable Chinese Fake News Detection with Chain-of-Thought and In-context Learning." IEEE Access (2025).
16. Wang, J., Qian, S., Hu, J., Dong, W., Huang, X. and Hong, R., 2025. End-to-end explainable fake news detection via evidence-claim variational causal inference. ACM Transactions on Information Systems, 43(4), pp.1-26.
17. Bhuiyan, Maniruzzaman, Farzana Sultana, and Aha Mudur Rahman. "Fake news classifier: Advancements in natural language processing for automated fact-checking." Strategic Data Management and Innovation 2, no. 01 (2025): 181-201.
18. Rama Moorthy, H., Avinash, N.J., Krishnaraj Rao, N.S., Raghunandan, K.R., Dodmane, R., Blum, J.J. and Gabralla, L.A., 2025. Dual stream graph augmented transformer model integrating BERT and GNNs for context aware fake news detection. Scientific reports, 15(1), p.25436.
19. Patel, Jainee, Chintan M. Bhatt, Himani Trivedi, and Thanh Thi Nguyen. "Misinformation detection using large language models with explainability." In 2025 8th International Conference on Algorithms, Computing and Artificial Intelligence (ACAI), pp. 1-7. IEEE, 2025.
20. Al-Quayed, F., Javed, D., Jhanjhi, N.Z., Humayun, M. and Alnusairi, T.S., 2024. A hybrid transformer-based model for optimizing fake news detection. IEEE Access, 12, pp.160822-160834.

21. Raza, S. and Ding, C., 2022. Fake news detection based on news content and social contexts: a transformer-based approach. International Journal of Data Science and Analytics, 13(4), pp.335-362.
22. Ma, J., Ma, J., Zhang, W., Liu, Y. and Han, M., Explainable Fake News Detection with Large Language Models Fine-Tuned by Motivation Psychology. Available at SSRN 5354827.
23. Lu, Y. and Yao, N., 2025. A fake news detection model using the integration of multimodal attention mechanism and residual convolutional network. Scientific Reports, 15(1), p.20544.
24. Uddin, R., Basit, A., Khan, Y., Shazib, M.S.J. and Hossain, S., 2025. BERT-Based Fake News Detection: A Transformer-Driven Approach for Misinformation Classification on Twitter. IJSAT-International Journal on Science and Technology, 16(1).
25. Zuin Gigli, L.R., 2025. Enhancing Fake News and Rumor Detection Using Metadata, Context, and User Interaction with a Hybrid GCN-Transformer.
26. Author Biographies
27. First Author The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author's educational background is listed. The degrees should be listed with type of degree in what field, which institution, city, state or country, and year degree was earned. The author's major field of study should be lower-cased. If a photograph is provided, the biography will be indented around it. The photograph is placed at the top left of the biography.
28. Second Author The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author's educational background is listed. The degrees should be listed with type of degree in what field, which institution, city, state or country, and year degree was earned. The author's major field of study should be lower-cased. If a photograph is provided, the biography will be indented around it. The photograph is placed at the top left of the biography.
29. Third Author The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author's educational background is listed. The degrees should be listed with type of degree in what field, which institution, city, state or country, and year degree was earned. The author's major field of study should be lower-cased. If a photograph is provided, the biography will be indented around it. The photograph is placed at the top left of the biography.