International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0121
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 790 - 795

Real-Time Fake News Detection using DeBERTa-V3 and Transformer

Mrs.T.Divya[1], B. K. Sandhiya[2], J.Kaviya[3], R.Ritika varthini[4]
[1]Assistant professor, Kamaraj College of Engineering and Technology Virudhunagar,Tamilnadu,India,
[2,3,4]Department of Computer Science Engineering,Kamaraj College of Engineering and Technology, Virudhunagar, Tamilnadu, India.
Assistant professor, Kamaraj College of Engineering and Technology Virudhunagar,Tamilnadu,India,
Emails: *sandhiyabk1723@gmail.com[1], jayakumarkaviya22@gmail.com[2], ritikavarthini90@gmail.com[3],* divyacse@kamarajengg.edu.in[4.]

Abstract
Fake news dissemination through digital and social media has grown into a real threat, as it distorts public opinion, decision-making, and confidence in institutions. Manual fact-checking is time-consuming and slow, hence inappropriate for a real-time environment where the misinformation spreads within minutes. This study proposes a real-time fake news detection framework that incorporates the pre-trained DeBERTa V3 transformer model for text classification, fact-checking based on Wikipedia, and sentiment analysis to provide explainable and transparent outputs. The system is trained and tested by the LIAR factchecking dataset of short political statements labeled, and its deployment is performed via an intuitive interface using Gradio so that users can easily input the news text and get classifications, sentiments, and evidence that support those sentiments in real-time. The experimental results indicated that the proposed system can effectively classify fake from real news, while providing interpretable justifications, hence improving users' trust and promoting responsible consumption of online information. The proposed approach is appropriate for embedding into media monitoring tools, browser extensions, and learning environments to mitigate the effects of misinformation.

Keywords:*Fake News Detection, DeBERTa-V3, Transformer, LIARDataset, FactVerification, Sentiment Analysis, Misinformation.*

1. Introduction
Although the dissemination of information through social networking platforms has increased the speed at which information is passed and is more accessible than in the past, it also allows for the promotion of. Misinformation has the potential to impact the election processes, the health department in dealing with the pandemic, markets, as well as promoting unity in society in ways through which the user does not have the means through which the information can be checked manually by journalists and specialized institutions, is true but delayed and cannot in any case sustain the tempo of online dynamics..Early automated approaches to fake news detection relied on classical machine learning models such as Support Vector Machines (SVM) and Naive Bayes using handcrafted lexical or syntactic features. While these provide a baseline performance, capturing deep contextual nuances, sarcasm, and subtle framing is difficult for them-features common in real-world misinformation. More recently, significant improvements in deep learning, particularly Transformer-based language models such as BERT, RoBERTa, and DeBERTa, have raised the bar on text classification tasks by modeling long-range dependencies and rich semantics. Yet, most of the fake news systems today still behave as black boxes, providing a label without explicit evidence, which can reduce trust by users.In this paper, we present an architecture of real-time fake news detection using the DeBERTa V3 transformer model that is strengthened by Wikipedia evidence fact-checking and sentiment analysis. Our proposed system accepts the news text input and identifies the type of news as true or false by finding the appropriate evidence from Wikipedia and performs an emotional bias analysis of the inputs automatically. The contributory aspects of our paper are listed below: (1) A DeBERTa V3 fake news

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0121
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 790 - 795

classification model that is fine-tuned using the LIAR dataset, (2) A Wikipedia fact-checking functionality that aggregates evidence in the Wikipedia style, (3) A sentiment analysis tool that analyzes the emotional bias of the input news in the content analysis phase of the proposed system, and (4) A Gradio GUI that enables the system for real-time use by common people without the requirement of an indepth understanding of the system's functionality.

2. Related Work

Early research on fake news detection has tended to focus on applying traditional machine learning algorithms combined with NLP features. Zhou et al. used TF–IDF, n-grams, and basic linguistic features with SVM and Naive Bayes classifiers for fake news classification, showing reasonable performance but not performing well in extracting deep semantics and contexts. Another challenge in these models is that they rely heavily on feature engineering and suffer from a lack of robust domain transfer. Since the proposal of the Transformer architecture, many different works have applied deep contextual models for fake news detection. Liu et al. showed that Transformer-based models outperform their classical counterparts in text classification tasks such as misinformation detection by far, thanks to their effective capturing of long-range dependencies and complicated language patterns. BERT and RoBERTa models are widely explored in fake news and rumor detection, showing substantial improvements in accuracy and robustness across datasets. DeBERTa, an upgraded version of BERT, was proposed by He et al., which enhanced BERT by using disentangled attention and more powerful decoding for better modeling of word content and position. This model has had state-of-the-art performances on several language understanding tasks and provides better contextual representations, thereby making the model a strong contender incomplex tasks such as the detection of fake news.Another interesting related task that has come to the forefront in the war against fake news is the verification of claims through evidence extracted from Wikipedia articles, and the FEVER dataset has become one of the largest benchmarks for this particular job, where most tasks involve the retrieval of sentences and the classification of claims as proven, disproven, and unproveable respectively. Evidence-based methods can further enhance the trust in these models. There are also methods that combine content-based characteristics and contextual information like profiles, propagation, and structure to improve the accuracy of these detectors. While these methods improve the performance, they either rely on platform-dependent metadata, potentially raising privacy or generalization issues. Our system goes beyond these previous lines of inquiry by joining a state-of-the-art DeBERTa V3 text classifier with Wikipedia-based evidence retrieval and sentiment analysis within a single pipeline and by emphasizing a practical real-time interface over an offline benchmark.

3. Problem Statement

The spreading of fake news through social media platforms is an issue that poses significant difficulties to individuals, groups, and governments. Social media platforms expose users to information that might be misleading or biased, with signs of authenticity that are not readily available through manual checking. Current automated solutions to the problem either depend on classical solutions that may lack contextual information or make use of more advanced solutions such as deep learning that do not make any attempt to explain their results[1]. It is proposed that an automatic approach should ideally be able to identify fake or real stories with speed while offering information to the user that is understandable[2]. The approach should have the capability of checking facts against authentic external knowledge bases while pointing to biased emotional expressions.The objective of this work is to design and implement a real-time fake news detection framework that[3]

- Improves classification accuracy using a DeBERTa-V3 Transformer-based model.
- Provides explainable outputs by integrating Wikipedia-based fact verification.
- Analyzes sentiment and potential bias in news content.
- Offers a simple and interactive user interface for practical deployment.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0121
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 790 - 795

4. Proposed Work

4.1.System Overview

The proposed system is an end-to-end system which takes news texts from the datasets or live sources and provides several levels of analysis. The system is designed using eight crucial modules, such as Data Collection, Preprocessing, Training of Models, Prediction, Explanation, Sentiment Analysis, Wikipedia Analysis, and User Interface. The news texts are initially gathered and preprocessed for further analysis using a fine-tuned DeBERTa V3 model for the purpose of classifying them as fake or real news. Concurrently, the identification of main claims is executed and analyzed for relevance using Wikipedia for supportive or opposing evidence. The final analysis involves the evaluation of sentimental polars of the news article. All these results are then displayed using a Gradio interface[4].

4.2.Collection Module

The LIAR dataset is used by this system because LIAR is a very popular benchmark dataset for research into fact-checking, with around 12K labeled statements with levels of label disagreement from "pants on fire" to "true." In this experiment, however, the system uses a simplified classification task with a binary classification of "fake" versus "true" labels. In this dataset, there are context variables of speakers, their parties, and their venues. The system proposed here, however, relies purely on the text representation of these statements. News APIs can be used to feed real-time data into this system to help detect whether statements are fake or true. Shows Figure 1 Architecture of the Proposed Fake News Detection and Verification System[5].

4.3.Preprocessing Module

The preprocessing step plays a vital role in enhancing the efficiency and efficacy of the model. The text is normalized by applying lowercasing, removing surplus whitespace characters, and optionally removing URLs and uninformative tokens and maintaining significant punctuation. The text is tokenized by applying the DeBERTa V3 tokenization technique. The tokenized text is divided into sequences of a fixed maximum length. The attention mask is then applied to differentiate the actual tokens and the paddings. The normalized and tokenized text is used to split it into training and validation sets.

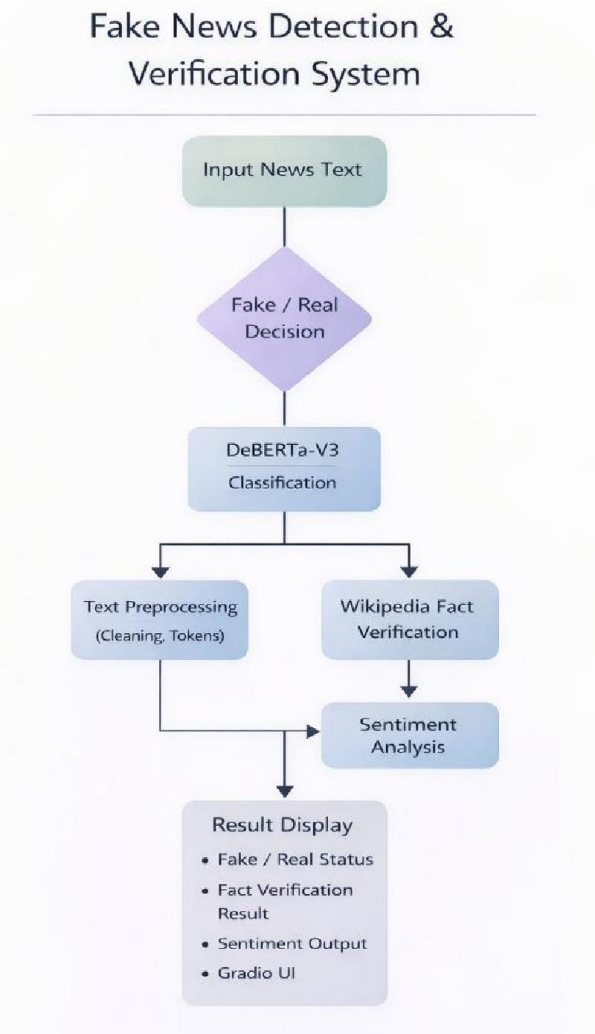

Figure 1 Architecture of the Proposed Fake News Detection and Verification System

4.4.Model Training Module

The prevalent classification architecture is based on DeBERTa V3, which is chosen for its disentangled attention mechanism that encodes content and position information separately, thus generating more informative context. To this architecture, a classification head is attached that comprises a pooling layer followed by a fully connected layer that calculates logits for fake or real. It is then fine-tuned on the LIAR dataset using the cross entropy loss function, where adaptation is done using Adam or AdamW. Early

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0121
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 790 - 795

stopping is used for preventing overfitting. It can be run on GPU devices such as NVIDIA GTX 1050 Ti to accelerate processing time[6].

4.5.Prediction Module

At the prediction phase, the new text containing unseen news is put through the same set of processes as in training. The DeBERTa V3 pre-trained model is finetuned, and the probability for each class is predicted. With the highest probability threshold value, the final result is determined. Alongside the result label, the confidence value is also displayed for the user in order to provide an idea about the predictions made by the model. This module runs in real-time mode and helps the user quickly assess the news or statement credibility.

4.6.Explanation Module (Wikipedia-Based Fact Verification)

In an effort to make these predictions more understandable and provide additional context-related information, the model has been designed with an explanation module based on fact verification systems such as those used on the FEVER dataset. The key entities and assertions are identified in the text by basic NLP methods or named entity recognition. The identified claims are then used to search the Wikipedia database and fetch possible articles containing evidence related to these claims. Statements in these articles are then shown as evidence both contradicting and supporting the identified claims and help users know how their assertions compare to the news content based on available knowledge. The model currently has not used classification of claim-verdict like FEVER but still offers contextual evidence along with the model's result

4.7.Sentiment Analysis Module

Sentiment analysis can be used to identify the emotional connotation and possible biased nature of the news article. Then, a classification model or the sentiment classification library can be used to classify the news article into categories like positive, negative, and neutral. By combining the result of sentiment analysis with the classification regarding the news article being true or fake, one can identify possible trends such as when highly negative news is emotionally manipulated in an attempt to deceive. Adding the sentiment component would be beneficial in informing the user not only on the possible accuracy of news but also on its intention to manipulate emotionally.

4.8.User Interface Model

The last aspect is the user-friendly interface developed employing Gradio or a light web framework like Flask. The interface will enable the user to enter or paste news, and the result will be fed into the system, which will generate visualizations of the outputs in the form of: the result (real or fake news classification), confidence, sentiment, and evidence from Wikipedia. The interface is developed with the intention of being user-friendly in order to reach the largest number of people possible, such as college students, teachers, or journalists. This aspect will allow the developed system to be used as an educational tool in the real world.

5. Experimental Setup and Results

For the experimental evaluation, the LIAR dataset is employed as the major comparison benchmark, and the typical training, validation, and test-split procedure is adopted to gauge model accuracy. Preprocessor functions and the DeBERTa V3 tokenizer are used to tokenize text statements. If the length of the statements exceeds the specified maximum length, truncation of the sentences takes place. The model has been trained on the machine, which has qualified specifications of consuming minimum 8GB of RAM, an Intel i5 processor, and an NVIDIA GTX 1050 Ti GPU, and the programming environments include Python, PyTorch, and the Transformers library, where the necessary hyperparameters like the number of epochs, batch size, and learning rate are varied to check accuracy on the validation set.For evaluation, common classification metrics such as accuracy, precision, recall, and F1-score are calculated on the held-out test set. Compared to traditional baselines, the DeBERTa V3 model greatly outperforms those reported in prior work on LIAR, reflecting the advantage of deep contextual language representations for this task.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0121
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 790 - 795

While the exact numerical values do depend on the chosen configuration and training regime, these results suggest that the proposed model should be able to reliably tell whether statements are fake or real on short textual inputs. Qualitative analysis further confirms the utility of the explanation and sentiment modules. In correctly classified fake news examples, the retrieved Wikipedia evidence often directly contradicts key claims in the text, giving users concrete references they can investigate. For real news statements, evidence often directly contradicts key claims in the text, giving users concrete references they can investigate. For real news statements, the retrieved evidence is often corroborative of the content of the statement. Sentiment analysis indicates that many fake or misleading statements contain strong negative or polarizing tone, which can be flagged to users as an additional cue. These observations suggest that a combination of classification, evidence retrieval, and sentiment analysis provides a richer understanding than one based on just classification. Which also includes development of browser extensions or mobile apps that let users integrate seamlessly into their daily news consumption and exploration of user studies for evaluation of how interface and explanations drive trust and decision-making.Shows Figure 2 User Interface of the Real-Time Fake News Detection System using DeBERTa-V3

Figure 2 User Interface of the Real-Time Fake News Detection System using DeBERTa-V3

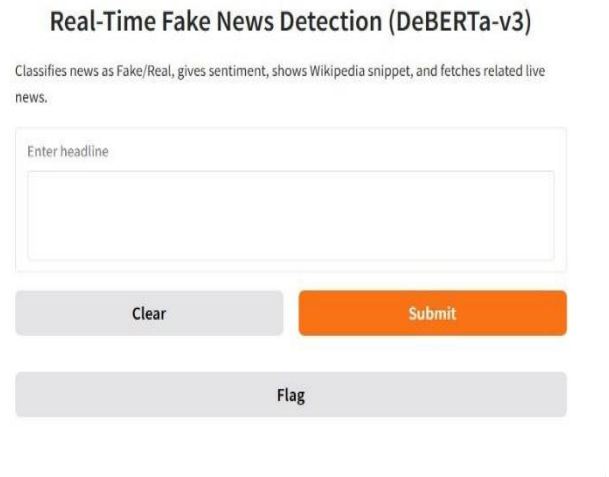

Figure 3 Testing the Proposed System with a Sample News Headline

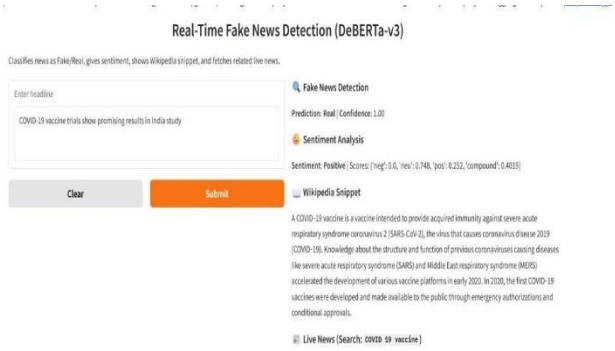

Figure 4 Output of the Real-Time Fake News Detection and Verification System

References

[1]. Z. Zhou et al., "Fake News Detection using NLP and Machine Learning," International Journal of Information Processing, 2020.
[2]. Y. Liu et al., "Transformer-Based Fake News Detection," Journal of Artificial Intelligence Research, 2021.
[3]. P. He et al., "DeBERTa: Decoding-Enhanced BERT with Disentangled Attention," Advances in Neural Information Processing Systems (NeurIPS), 2020.
[4]. J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL-

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0121
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 790 - 795

HLT, 2019.
[5]. W. Y. Wang, “Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection,” ACL, 2017.
[6]. J. Thorne et al., “FEVER: A Large-Scale Dataset for Fact Extraction and Verification,” NAACL-HLT, 2018.