IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 6, Issue 1, February 2026

Impact Factor: 8.2

A Hybrid Multimodal Machine Learning Framework for Fake News Detection

Suraj Pal[1] and Dr. Devender Kumar[2]

Research Scholar, Department of Computer Science & Application[1]
Associate Professor, Department of Computer Science & Application[2]
BMU, Asthal Bohar, Rohtak

Abstract: *Fake News Detection (FND) has emerged as a critical challenge due to the rapid dissemination of deceptive information across social media and online platforms. The increasing use of multimedia content, particularly the combination of textual narratives with misleading visuals, has further complicated the detection process. This chapter addresses the FND problem by proposing a robust multimodal approach that jointly analyzes textual and visual information to improve detection accuracy. The study utilizes three widely adopted multimodal benchmark datasets Fakeddit, TwitterMediaEval2015, and the FakenewsNet repository to ensure comprehensive evaluation and generalizability of the proposed framework. The proposed multimodal architecture is designed to identify deceptive patterns by capturing superficial and semantic dependencies between text and images. Transformer-based techniques are employed to enhance word sequence modeling and internal feature representations, enabling effective cross-modal interaction and contextual understanding. By integrating visual and textual cues, the framework overcomes the limitations of unimodal approaches that rely solely on either text or images. Experimental results demonstrate the effectiveness of the proposed method. The multimodal model achieves an accuracy of 91.35% on the PolitiFact dataset and 98.59% on the Gossip Cop dataset, significantly outperforming traditional unimodal and baseline multimodal models. Furthermore, a comparative evaluation with an established multimodal framework, FakeRevealer, shows an accuracy of 80.00%, surpassing the state-of-the-art performance on the TwitterMediaEval2015 dataset by 2.23%. These findings highlight the superiority of multimodal learning in detecting fake news by leveraging complementary information from multiple data sources. Overall, the results confirm that multimodal approaches provide more reliable and accurate solutions for fake news detection in modern social media environments..*

Keywords*:* Fake News Detection, Fakeddi , TwitterMediaEval2015, Accuracy

I. INTRODUCTION

Online community like Facebook, Twitter, etc. are at the center of the current wave of digitalization in society. They provide a means for people to communicate, share ideas, and keep up with current events as a result, they have become a fundamental aspect of numerous individuals everyday practices [1,2]. However, it has also resulted in a rapid increase in the number of 'fake news' articles, which are news articles that contain intentionally false information. Typically, these news articles are produced through the manipulation of images, text, audio, and video. According to several news reports, it has been alleged that Russia has engaged in the creation of numerous fake accounts and social media bots to disseminate misleading information during electoral periods [3]. The dissemination of false information is pervasive, negatively impacting society and individuals. Disseminating false information can significantly disrupt the integrity of truthfulness within the news ecosystem. False news can hurt people, so it's important to make a system that can automatically spot it when it shows up on social media. However, there are certain difficult research problems about how to identify fake news on different social media platforms to determine the origin of the particular news or data uploaded to the social network. Understand the true meaning or intention behind the content that is posted. Determine the degree of authenticity and validity of the post. Researchers have investigated several techniques, including transfer

DOI: 10.48175/568

learning, cross-lingual learning, and zero-shot learning, to solve the fake news problem [4]. Transfer learning entails pre-training a model on an extensive dataset in one language and subsequently fine-tuning it with a smaller dataset in a different language. Crosslingual learning uses language embeddings to transfer text from several languages to a standard representation space that can be utilized to detect fake news in various languages [5]. Prior research has focused on detecting fake news using either handcrafted features or unimodal (text) detection. The issue lies in the failure to consider multiple media types within tweets. An empirical observation suggests that tweets incorporating visual components, such as images and videos (including GIFs), are more likely to capture user attention than tweets that rely only on text-based content. The objective of this chapter is to discuss the FND problem in a multimodal context. Section I represent introduction of fake news detection. Section II represented previous algorithm. Section III, IV represented proposed methodology and result & discussion in detail. In the last section conclude the paper with future scope.

II. RELATED WORK

Tianyi Huang et al. (2025) present an optimized Transformer-based model that integrates Bayesian algorithms with a Bidirectional Gated Recurrent Unit (BiGRU) for fake news classification, marking the first application of this combined framework in misinformation detection. The study begins by extracting textual features using the TF-IDF method, converting news content into numerical representations suitable for machine learning. Two experimental models were developed: a BiGRU-enhanced Transformer and a Bayesian-assisted BiGRU-Transformer. Results indicate that the BiGRU-optimized Transformer achieves 100% accuracy on the training set and 99.67% on the test set, while the addition of Bayesian optimization further improves test accuracy to 99.73%, demonstrating a 0.06% performance gain. Training analysis shows rapid convergence around the 10th epoch with accuracy nearing 100%, reflecting both the efficiency and effectiveness of the proposed approach. The combined Bayesian–BiGRU–Transformer architecture exhibits robust learning ability, high precision, and fast classification, making it highly suitable for real-time fake news detection. Overall, the study validates the strong potential of integrating Bayesian optimization and BiGRU structures within Transformer models, offering a promising and accurate solution to address the growing challenge of misinformation in the era of information overload. These findings provide valuable insights for future advancements in automated fake news identification systems [6].

Jingyuan Yi et al. (2025) highlight the growing threat posed by the widespread dissemination of fake news across social media, which undermines public trust, societal stability, and democratic systems. Addressing this challenge requires advanced detection methods capable of handling the dynamic, multi-modal nature of misinformation. The review discusses recent progress driven by large language models (LLMs), multimodal frameworks, graph-based techniques, and adversarial training. Enhanced LLMs significantly improve detection accuracy through deeper semantic understanding and cross-modal fusion. However, key gaps persist, including limited adaptability to rapidly changing online content, insufficient real-time detection, and weak cross-platform generalization. Promising future directions include style-agnostic models, cross-lingual detection systems, and stronger policies to counter LLM-generated misinformation. The survey highlights successful innovations such as MiLk-FD and FNDLLM, which combine LLMs with multimodal and graph-based methods, as well as approaches like SheepDog and DAFND that address stylistic variability and data scarcity through style-agnostic and few-shot learning techniques. [7].

Xiaochuan Xu et al. (2025) address the growing challenge posed by the rapid spread of fake news in an increasingly information-dense online environment. The study introduces a novel Large Language Model (LLM)-based detection framework that integrates both statistical textual features and deep semantic representations to improve fake news identification. By leveraging the contextual understanding strengths of LLMs and incorporating a hybrid attention mechanism, the model dynamically emphasizes the most informative feature combinations. Experiments conducted on the WELFake dataset demonstrate a significant performance gain, achieving an F1 score of 0.945—an improvement of 1.5% over the best existing method. The approach maintains high recall while keeping false alarm rates low, making it suitable for real-world applications. Additionally, interpretability is enhanced through attention heat maps and SHAP-based analyses, offering clear insights into the model's decision-making process and supporting content moderation strategies. The study's contributions lie in its effective fusion of statistical and semantic features, adaptive hybrid attention design, and interpretable analysis. However, the authors acknowledge limitations, including potential feature bias when confronting emerging misinformation trends, the need to validate performance across multilingual datasets, and computational complexity concerns that may impact large-scale, real-time deployment. Nonetheless, the proposed framework offers a scalable and robust solution for strengthening online information reliability [7].

Jumana Jouhar et al. (2024) highlight the growing threat posed by the widespread circulation of false information in today's digital environment. Motivated by the need for effective countermeasures, the study employs machine learning techniques to detect fake news and strengthen information reliability. By examining multiple machine learning models and evaluating them using metrics such as precision and accuracy, the research identifies the most effective algorithm for classifying news articles as real or fake. The methodology encompasses a detailed literature review, dataset selection, preprocessing and cleaning, vectorization, model selection, training, optimization, and performance assessment. Beyond technical advancements, the study contributes to enhancing public trust, supporting media authenticity, and assisting stakeholders such as news agencies, social media platforms, and government systems. Despite its promising results, several challenges persist. Fake news evolves rapidly, spreads across diverse platforms, and increasingly involves multimedia content, making precise classification difficult. The authors emphasize the need for future work focusing on improved data cleaning techniques, alternative vectorization approaches like Word2Vec and BERT, exploration of additional algorithms, and optimization methods such as grid and random search. They further stress the importance of real-time monitoring and automated fact-checking to strengthen model applicability and effectiveness in real-world scenarios [8].

Pummy Dhiman et al. (2024) address the growing challenge of fake news, a consequence of expanded global connectivity and widespread internet access. As false information increasingly threatens social harmony, politics, the economy, and public opinion, the need for reliable detection methods has become critical. To tackle this issue, the authors propose a novel hybrid framework named Generative Bidirectional Encoder Representations from Transformers (GBERT), which integrates the strengths of GPT's generative capabilities with BERT's deep contextual understanding. This combination enables a richer, more comprehensive representation of textual content for fake news classification. The model is fine-tuned on two benchmark datasets and achieves strong results, including 95.30% accuracy, 95.13% precision, 97.35% recall, and a 96.23% F1-score. Comparative analysis shows that GBERT outperforms traditional machine learning techniques such as XGBoost, Naive Bayes, and hybrid CNN–LSTM models, as well as individual transformer models like BERT and GPT-2. The study highlights that the integration of BERT and GPT for fake news detection remains relatively unexplored, making GBERT a notable contribution. Despite its promising performance, the authors acknowledge certain limitations. Statistical significance may not always translate to real-world effectiveness, where factors such as data diversity, robustness, and computational cost influence performance. Thus, further refinement is essential for deployment in dynamic, real-world environments [9].

Mahabuba Akhter et al. (2024) highlight how the widespread use of mobile and networked devices has accelerated information sharing, but has also amplified the circulation of misinformation. During the COVID-19 pandemic, this problem intensified, leading the World Health Organization (WHO) to label the surge of false COVID-19 content as an "infodemic." Such misinformation created significant challenges for governments and public health efforts, underscoring the need for reliable detection systems. To address this, the authors propose an effective CNN-based deep learning model using word embeddings for detecting COVID-19-related fake news. They optimize the CNN architecture through grid search to determine the best hyperparameters. To validate its performance, they compare the model against several machine learning algorithms, including Multinomial Naive Bayes, Logistic Regression, SVM, Decision Tree, and Random Forest—each also fine-tuned using grid search. The custom CNN achieves the highest performance with 96.19% accuracy, a 95% F1-score, and an AUC of 0.985, outperforming all baseline classifiers [10].

Mutaz A. B. Al-Tarawneh et al. (2024) investigate the effectiveness of different word embedding techniques—TF-IDF, Word2Vec, and FastText—applied to both machine learning (ML) and deep learning (DL) models for fake news detection. Using the Truth Seeker dataset, which includes labeled news articles and social media posts spanning over a decade, the study evaluates classifiers such as Support Vector Machines (SVMs), Multilayer Perceptrons (MLPs), and Convolutional Neural Networks (CNNs). Results indicate that SVMs and CNNs using TF-IDF embeddings achieve the highest overall performance in accuracy, precision, recall, and F1 score. TF-IDF effectively highlights discriminative features in text, benefiting models like SVMs that handle sparse data well, while CNNs leverage TF-IDF to capture localized textual patterns. Word2Vec and FastText embeddings, though capable of capturing semantic and syntactic nuances, introduce complexity that may not always enhance traditional ML models. The study emphasizes the importance of aligning embedding techniques with model architecture to maximize detection performance. Findings suggest that DL models excel with complex embeddings for capturing contextual information, whereas simpler embeddings like TF-IDF suffice for ML models. These insights provide valuable guidance for developing robust fake news detection systems, promoting content authenticity and mitigating misinformation on social media platforms [11].

Azka Kishwar et al. (2023) address the growing challenge of fake news in Pakistan and develop the first comprehensive Pakistani fake news detection dataset using multiple fact-checked news APIs. The study evaluates the dataset with five machine learning models—Naive Bayes, KNN, Logistic Regression, SVM, and Decision Trees—and two deep learning models, CNN and LSTM, using GloVe and BERT embeddings. Results show that LSTM with GloVe embeddings performs best, achieving an F1-score of 0.94, outperforming BERT embeddings. Analysis of misclassified samples indicates that even human judgments can be inaccurate, highlighting the complexity of fake news detection and the effectiveness of the proposed dataset and models [12].

Minjung Park and Sangmi Chai (2023) examine fake news detection on social media by incorporating user characteristics, news content, and social network features based on social capital, moving beyond traditional methods that focus solely on linguistic traits. The study applies XGBoost to evaluate feature importance and identifies four key variables—word sentiment, in-degree centrality, word similarity, and total number of tweets—that significantly influence fake news detection. Using these features, five machine learning models (SVM, RF, LR, CART, NNET) are constructed and evaluated through cross-validation and oversampling to address data imbalance. The Random Forest (RF) model achieved the highest accuracy of 94.1%, while NNET showed the lowest at 92.1%. The research highlights that word sentiment, particularly negative or overly positive tones, plays a critical role in identifying misleading content, reflecting how fake news manipulates readers' emotions. The study also emphasizes the importance of feature prioritization for developing robust detection systems and suggests that ensemble models like RF generally outperform single models. Future research should explore combining ensemble methods with Explainable AI (XAI) to enhance both accuracy and interpretability. These findings provide a foundation for designing effective, reliable, and transparent fake news detection systems for social media platforms [13].

III. RESEARCH METHODOLOGY

For the Fakeddit dataset, we used various text based architectures such as XLNet, distilled version of RoBERTa, DistiBERT for text classification. To extract image features, ResNet50 is used. We utilize DistilBERT for text feature extraction and VGG16 for image feature extraction. The extracted features are then processed through several dense layers to perform the multimodal FND task.

DistilBERT was chosen for text feature extraction due to its strong balance between efficiency and performance—it retains 97% of BERT's accuracy while using 40% fewer parameters and running 60% faster, making it ideal for resource-constrained environments. Despite being a distilled model, it effectively captures deep semantic relationships in text, producing high-quality embeddings with lower computational cost compared to models like full BERT or RoBERTa. In contrast, traditional models like LSTMs and CNNs fall short in handling long-range dependencies and global context, areas where transformer-based models like DistilBERT excel.

VGG16 was selected for image feature extraction due to its proven effectiveness and simplicity. it's a widely used CNN architecture known for extracting rich visual features. Its availability with pretrained weights on ImageNet enables efficient transfer learning, reducing the need for extensive training. While models like ResNet and EfficientNet offer improvements in accuracy and efficiency, VGG16 strikes a practical balance between performance and computational demands, making it suitable for multimodal applications. Alternatives like ResNet can be more resource-intensive, and EfficientNet, though efficient, often requires complex scaling and fine-tuning.

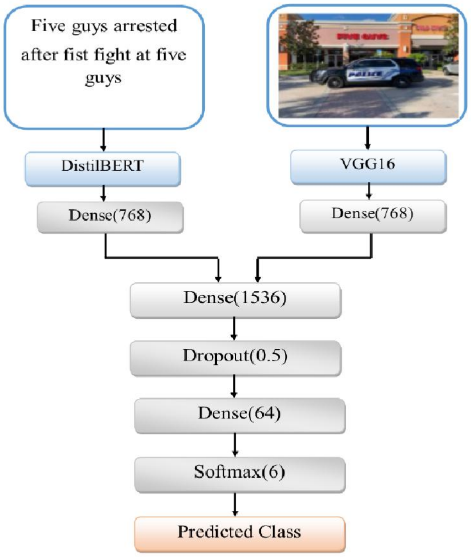

Figure 1: Multimodal Architecture Used for Fakeddit Dataset

For the Fakeddit dataset, we experimented with three text-based architectures, including DistilBERT and DistilRoBERTa, and compared the results with those from the BERT architecture used in [14], as indicated in Table 1. The findings show that the suggested textual architectures, such as DistilBERT and DistilRoBERTa, outperform the results reported in [64]. We achieved the highest accuracy of 88.60% using the DistilRoBERTa architecture for the 2-way classification task. DistilRoBERTa is a streamlined variant of RoBERTa, engineered to be more compact and efficient than BERT while preserving around 97% of RoBERTa's efficacy. With fewer parameters, it uses less memory and delivers quicker inference, making it ideal for environments with limited computational resources. Unlike BERT, DistilRoBERTa incorporates improved pretraining methods and omits next-sentence prediction, enhancing its effectiveness on specific tasks. Overall, it provides a balanced compromise between accuracy and efficiency. Table 2 demonstrates the model's performance with the ResNet50 architecture on image data.

Table 1: Comparative Analysis of Various Textual Based Architectures

Model/way	**2 way Test**	**3 way Test**	**6 way Test**
Fakeddit(BERT)	0.8644	0.8580	0.7677
DistilBERT	0.8798	0.8633	0.7982
DistilRoBERTa	0.8860	0.8747	0.8032

IV. RESULT ANALYSIS

The proposed multimodal architecture named FakeRevealer is a combination of DistilRoBERTa and ViT models, that surpass the existing leading model SpotFake [15] regarding accuracy. In Table 4.6, we show the comparison of our multimodal architecture to the current state-of-the-art multimodal architectures. The proposed FakeRevealer multimodal does better performance on the TwitterMediaEval2015 dataset than EANN, MVAE, and SpotFake [15]. The extracted features from both modalities are fused using a variety of fusion techniques, including concatenating, multiplying, and taking the maximum of both features. The cosine function performs better on our multimodal architecture. This is because it aids in capturing cross-modal relationships, performs well with high-dimensional data, manages different scales effectively, allows flexibility in weighting various modalities, enhances robustness in decision-making, handles imbalanced modalities efficiently, and can improve generalization, among other benefits. Concatenation, which is simply concatenating both feature lists; add, which is adding the values of the features; maximum, i.e., selecting the maximum out of both; minimum, which is selecting the minimum out of both; average, which is taking the average of both feature lists; dot, which is getting the by-product of both feature lists; and cosine fusion technique, which is selecting the maximum out of both. In Table 4.7, a comparative analysis of the outcomes of various fusion methods is provided.

Table 3: Comparative Analysis of Fake Revealer vs. Other Multimodal Architectures

Model	**Accuracy**
EANN	64.8%
VQA	63.1%
Neural Talk	61%
att-RNN	66.4%
MVAE	74.5%
SpotFake	77.7%
Multimodal Naive Bayes / SVM	65–75%
Simple Dual-Stream CNN	70–78%
FakeRevealer	80%

The table 3 illustrates a clear performance trend: models employing deep multimodal fusion and latent representation learning consistently outperform traditional and single-modality approaches. Advanced architectures such as MVAE, SpotFake, and FakeRevealer demonstrate that capturing semantic alignment and inconsistencies across modalities is crucial for effective fake news detection in social media environments. The table compares multimodal fake news detection models based on classification accuracy. Early approaches such as Neural Talk, VQA, EANN, and att-RNN show moderate performance (61–66%) due to limited cross-modal alignment. Advanced models like MVAE, SpotFake, and dual-stream CNNs achieve higher accuracy (70–78%) through improved multimodal fusion. FakeRevealer performs best (80%), demonstrating the effectiveness of robust cross-modal inconsistency modeling.

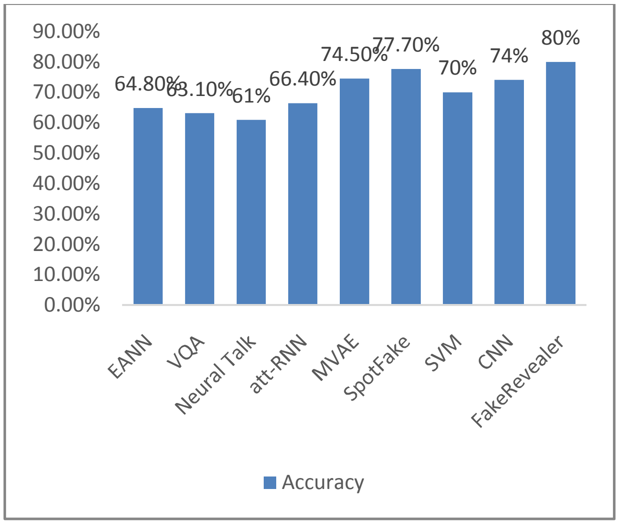

Fig 2 Comparative Analysis of Fake Revealer vs. Other Multimodal Architectures

Table 4: Fusion Techniques (FakeRevealer)

Modality	**Fusion Technique**	**Accuracy**
Text + Image	Concatenation	67.20%
Text + Image	Maximum	74.55%
Text + Image	Minimum	67.27%
Text + Image	Add	56.35%
Text + Image	Dot	65.45%

Text + Image	Cosine	80.00 %
Text + Image	Average	74.55%

The table 4 analyses the impact of different fusion techniques for combining text and image modalities on classification accuracy. Simple concatenation and minimum-based fusion yield moderate performance, while maximum and average fusion improve accuracy by emphasizing dominant features. Additive fusion performs poorly due to information dilution. Dot-product fusion captures limited interactions. Cosine similarity achieves the highest accuracy (80%), indicating that measuring semantic alignment between text and image features is most effective for multimodal fake news detection.

V. CONCLUSION

This paper examines the FND challenge through the application of a multimodal approach. The work utilizes three distinct multimodal datasets, namely Fakeddit, TwitterMediaEval2015, and FakenewsNet repository. The proposed chapter aims to examine the deceptive visuals in multimodal systems, such as social media and other networks that combine text and images. This chapter also aims to develop reliable models for the FND system. A multimodal approach is designed to identify fake news effectively. This approach captures the superficial dependencies between visual and textual content and improves word sequences and internal representations using transformer-based techniques. The multimodal architecture achieves a accuracy of 91.35% on PolitiFact and 98.59% on GossipCop. Examining a distinct multimodal framework, FakeRevealer gives a multimodal accuracy of 80.00%, which exceeds the current leading Twitter- MediaEval2015 dataset accuracy by 2.23%. The results suggest that the utilisation of multimodal approaches yields better outcomes than unimodal approaches.

REFERENCES

[1]. Y.-J. Lu and C.-T. Li, “Gcan: Graph-aware co-attention networks for explainable fake news detection on social media,” arXiv preprint arXiv:2004.11648, 2020.

[2]. C. Grimme, M. Preuss, L. Adam, and H. Trautmann, “Social bots: Humanlike by means of human control?” Big data, vol. 5, no. 4, pp. 279–293, 2017.

[3]. S. Lewandowsky, U. K. Ecker, and J. Cook, “Beyond misinformation: Understanding and coping with the “post-truth” era,” Journal of applied research in memory and cognition, vol. 6, no. 4, pp. 353–369, 2017.

[4]. D. Kar, M. Bhardwaj, S. Samanta, and A. P. Azad, “No rumours please! A multi-indic-lingual approach for covid fake-tweet detection,” in 2021 grace hopper celebration India (GHCI), IEEE, 2021, pp. 1–5.

[5]. G. K. Shahi and D. Nandini, “Fakecovid–a multilingual cross-domain fact check news dataset for covid-19,” arXiv preprint arXiv:2006.11343, 2020. Huang, Tianyi, Zeqiu Xu, Peiyang Yu, Jingyuan Yi, and Xiaochuan Xu. "A hybrid transformer model for fake news detection: Leveraging Bayesian optimization and bidirectional recurrent unit." *arXiv preprint arXiv:2502.09097* (2025).

[6]. Yi, Jingyuan, Zeqiu Xu, Tianyi Huang, and Peiyang Yu. "Challenges and innovations in LLM-Powered fake news detection: A synthesis of approaches and future directions." In *Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security*, pp. 87-93. 2025.

[7]. Xu, Xiaochuan, Peiyang Yu, Zeqiu Xu, and Jiani Wang. "A hybrid attention framework for fake news detection with large language models." In *2025 5th International Conference on Neural Networks, Information and Communication Engineering (NNICE)*, pp. 587-590. IEEE, 2025.

[8]. Jouhar, Jumana, Anju Pratap, Neharin Tijo, and Meenakshi Mony. "Fake news detection using python and machine learning." *Procedia Computer Science* 233 (2024): 763-771.

[9]. Dhiman, Pummy, Amandeep Kaur, Deepali Gupta, Sapna Juneja, Ali Nauman, and Ghulam Muhammad. "GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection." *Heliyon* 10, no. 16 (2024).

[10]. Akhter, Mahabuba, Syed Md Minhaz Hossain, Rizma Sijana Nigar, Srabanti Paul, Khaleque Md Aashiq Kamal, Anik Sen, and Iqbal H. Sarker. "COVID-19 fake news detection using deep learning model." *Annals of Data Science* 11, no. 6 (2024): 2167-2198.

[11]. Al-Tarawneh, Mutaz AB, Omar Al-irr, Khaled S. Al-Maaitah, Hassan Kanj, and Wael Hosny Fouad Aly. "Enhancing fake news detection with word embedding: A machine learning and deep learning approach." *Computers* 13, no. 9 (2024): 239.

[12]. Kishwar, Azka, and Adeel Zafar. "Fake news detection on Pakistani news using machine learning and deep learning." *Expert Systems with Applications* 211 (2023): 118558.

[13]. Park, Minjung, and Sangmi Chai. "Constructing a user-centered fake news detection model by using classification algorithms in machine learning techniques." *IEEE Access* 11 (2023): 71517-71527.

[14]. K. Nakamura, S. Levy, andW. Y.Wang, R/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection, 2019.

[15]. Álvaro Ibrain Rodríguez Department of Computer Science (2019). Fake News Identification Using Deep Learning. arXiv:1910.03496v2