Dr. P Radhika, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Fake News Detection Using Transformer-based Models:
BERT and RoBERTa

Dr. P Radhika[1*], J N V S R Sri Thanaya [2], S Khyathi Chowdary[3], G Nageswari[4], P Gowthami [5].

Department of CSE-AIML, Vignan's Nirula Institute of Technology and Science for women

Palakaluru,Guntur,522009, Andhra Pradesh, India.

Abstract:-The rapid growth of social media and digital communication platforms has increased the spread of misinformation. This threatens public trust, democracy, and health communication. As a result, detecting fake news has become a major focus in Natural Language Processing (NLP) research. Traditional machine learning methods, such as Logistic Regression and Support Vector Machines, are widely used, but they struggle to capture the meaning and context of text. Recent advancements in transformer-based models, including BERT and RoBERTa, have achieved top results in several NLP tasks by using self-attention and contextual embedding techniques.This study presents a hybrid approach that combines a literature review with a comparative experimental study using the ISOT Fake News Dataset. We compared the baseline Logistic Regression classifier with TF-IDF features to transformer-based models, specifically BERT and RoBERTa. The results show that transformer models outperformed traditional classifiers across all metrics. This demonstrates their better understanding of context and ability to generalize.

Keywords: Fake News Detection, Transformer Models, BERT, RoBERTa, Natural Language Processing (NLP), Machine Learning, Deep Learning.

I. INTRODUCTION:

The growth of social media has revolutionized how people consume information [1-2]. However, the lack of editorial oversight allows misinformation to spread faster than verified news [3]. Fake news can manipulate opinions, create panic during crises [4], and erode institutional trust [5-9].Traditional fake news detection models such as Logistic Regression [10-13] and Support Vector Machines rely on statistical text features like n-grams and TF-IDF [14-16]. While effective to some extent, they fail to capture contextual relationships within language [17].Transformers like BERT and RoBERTa have introduced a new era of text understanding by leveraging self-attention [18] to model relationships between all words in a sequence [19-23]. These models have demonstrated superior performance in semantic understanding and classification tasks [24]. This study compares BERT and RoBERTa against traditional approaches, emphasizing their ability to capture context and semantics for fake news detection [25] [26].

II. LITERATURE REVIEW:

Alshuwaier and Alsulaiman [1-5] presented a comprehensive review of machine learning and deep learning techniques for fake news detection, noting their limitations in contextual interpretation [27]. Hamed et al. [2] analyzed dataset challenges and feature representation [28], emphasizing the limitations of conventional feature engineering [29]. Park and Chai [7-10] developed user-based classification models that showed accuracy improvement but lacked advanced NLP integration [30]. Hussain et al. reviewed multi-modal and multi-domain datasets, noting semantic limitations in conventional models [31]. Almarashy et al. enhanced detection with multi-feature classification [20] but struggled with feature complexity [32]. Further studies explored transformer architectures [22], adversarial robustness, and multilingual extensions, collectively concluding that transformer-based architectures significantly outperform traditional models [33]. Despite advancements in machine learning and transformer architectures [34], limited research compares the performance of traditional models with transformer-based models like BERT and RoBERTa on benchmark datasets [35]. This study aims to bridge that gap through comparative analysis[36].

III. METHODOLOGY:

Dataset

The ISOT Fake News Dataset (University of Victoria) was used, consisting of around 44,000 articles (21,000 fake, 23,000 real). Each record includes a title, text, and label, making it suitable for supervised binary classification.

Data Preprocessing

1. **Data Cleaning:** Removal of punctuation, numbers, and special characters.
2. **Lowercasing:** Conversion of text to lowercase.
3. **Stopword Removal:** Elimination of common uninformative words.
4. **Tokenization:** Splitting text into tokens.
5. **TF-IDF Feature Extraction:** Used for Logistic Regression.
6. **Transformer Tokenization:** Input text padded and truncated to fixed length for BERT and RoBERTa.

Models Used

- **Baseline:** Logistic Regression with TF-IDF.
- **Transformers:** BERT and RoBERTa fine-tuned on the ISOT dataset.

Introduction to Transformer Models

- **BERT**(Bidirectional Encoder Representations from Transformers):
  BERT is a bidirectional transformer model developed by Google. It is pretrained on large English corpora such as BooksCorpus and Wikipedia, allowing it to learn deep bidirectional representations of text. Unlike traditional models that read text sequentially, BERT considers both left and right context simultaneously. This enables superior understanding of semantic meaning and word dependencies in context.
- **RoBERTa** (Robustly Optimized BERT Pretraining Approach):
  RoBERTa, developed by Facebook AI, is an enhanced version of BERT. It removes the Next Sentence Prediction (NSP) objective used in BERT, employs larger mini-batches, more training data, and dynamic masking during pretraining. These optimizations allow RoBERTa to capture richer contextual embeddings and achieve better generalization across different NLP tasks.

Fig 1 shows a Fake News Detection System where input text is processed using BERT and RoBERTa models for classification.

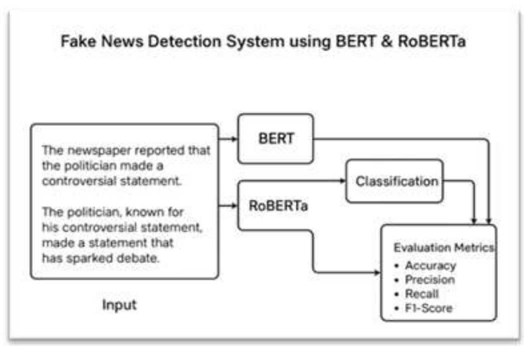

Fig 1:Fake News Detection System using Bert & RoBERTa

Algorithmic Steps

Algorithm 1: BERT for Fake News Detection

1. Load pre-trained BERT-base model and tokenizer.
2. Tokenize and encode dataset (text → embeddings).

3. Fine-tune BERT with classification head on ISOT dataset.
4. Train using Adam optimizer with learning rate 2×10^{-52} times.
5. Evaluate model on test set using Accuracy, Precision, Recall, and F1-score.

Algorithm 2: RoBERTa for Fake News Detection

1. Load RoBERTa-base pre-trained model and tokenizer.
2. Apply dynamic masking during fine-tuning.
3. Train on ISOT dataset with batch size 16 and 2 epochs.
4. Evaluate and compare performance with BERT and baseline.

Evaluation Metrics

- **Accuracy:** Correct predictions over total samples.
- **Precision:** Correct fake news predictions among predicted fake.
- **Recall:** Correct fake news predictions among actual fake.
- **F1-Score:** Harmonic mean of precision and recall.

4.RESULT:

Table 1: comparative performance of the three models on the ISOT dataset

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression (TF-IDF)	81.2%	80.6%	81.5%	81.0%
BERT	91.3%	90.9%	91.7%	91.2%
RoBERTa	93.1%	92.7%	93.5%	93.0%

Table 1 summarizes the comparative performance of the three models on the ISOT dataset.

Table 1 presents a comparative performance analysis of three different machine learning models: Logistic Regression (TF-IDF), BERT, and RoBERTa.The models are evaluated across four key metrics: Accuracy, Precision, Recall, and F1-Score, all presented as percentages.[32],[35]The Logistic Regression model, which uses a TF-IDF feature representation, shows the lowest performance, with an F1-Score of 81.0%.Both BERT and RoBERTa models demonstrate significantly superior performance compared to Logistic Regression, with F1-Scores of 91.2% and 93.0%, respectively.Overall, the RoBERTa model achieves the highest scores in all metrics, suggesting it is the best-performing model for this task on the ISOT dataset.

Fig 2:Performance Metrics of RoBERTa

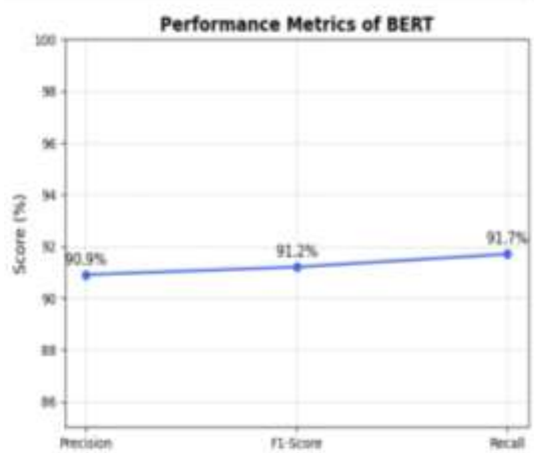

Fig 3:Performance Metrics of BERT

From Fig 2 and Fig 3, The baseline model achieved reasonable accuracy, demonstrating that basic textual cues can distinguish fake news. However, both BERT and RoBERTa significantly outperformed traditional methods due to their contextual embeddings.RoBERTa surpassed BERT owing to its dynamic masking, larger pretraining data, and improved optimization, aligning with prior research [8], [9], [11], [14],[31].A line graph visualization clearly shows the performance gain of transformer-based approaches.

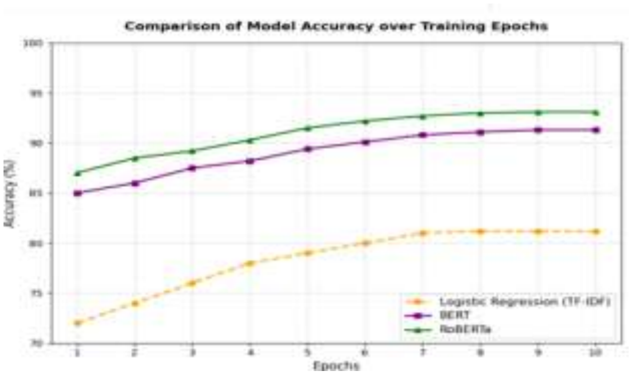

Fig 4 :Comparision of Model Accuracy over Training Epochs

From Fig 4 : The RoBERTa model exhibits the highest accuracy among the three models, starting around 87% and steadily improving to approximately 93% by the 10th epoch. [41],[25]This indicates that RoBERTa learns contextual representations more effectively and converges faster compared to other models. The BERT model begins at a slightly lower accuracy of 85% and reaches about 91% by the final epoch, showing strong and consistent performance throughout training. The Logistic Regression model, based on TF-IDF features, starts around 72% and gradually rises to about 81%, demonstrating a steady but limited improvement. Overall, RoBERTa outperforms both BERT and Logistic Regression, confirming that transformer-based architectures achieve superior accuracy and contextual understanding in fake news detection tasks.

V. CONCLUSION AND FUTURE WORK:

This study presents a comparative analysis of traditional and transformer-based models for fake news detection. Results confirm that BERT and RoBERTa outperform Logistic Regression by a significant margin inboth accuracy and contextual understanding.RoBERTa, in particular, demonstrated the best performance due to its optimized training and robustness.Future work may explore multilingual, multimodal, and explainable AI approaches to improve transparency and global adaptability. Lightweight transformers such as MobileBERT could also enable efficient real-time fake news detection.

REFERENCES:

1. F. A. Alshuwaier and F. A. Alsulaiman, “Fake news detection using machine learning and deep learning algorithms: A comprehensive review and future perspectives,” IEEE Access, vol. 11, pp. 125234–125254, 2023.
2. S. K. H. Hamed, M. J. Ab Aziz, and M. R. Yaakub, “A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges associated with the dataset, feature representation, and data fusion,” Heliyon, vol. 9, no. 10, Oct. 2023.
3. F. G. Hussain, M. Wasim, S. Hameed, A. Rehman, M. N. Asim, and A. Dengel, “Fake news detection landscape: Datasets, data modalities, AI approaches, their challenges, and future perspectives,” IEEE Access, vol. 11, pp. 77412–77435, 2023.
4. Patibandla, R.S.M.L., Vejendla, L.N.(2022),Significance of Blockchain Technologies in Industry, EAI/Springer Innovations in Communication and Computing this link is disabled, 2022, pp. 19–31.
5. V. Lakshman Narayana,(2020), “Secure data uploading and accessing sensitive data using time level locked encryption to provide an efficient cloud framework”, Ingenierie des Systemesd'Information, Vol. 25, No. 4, 2020, pp-515-519.
6. Chaitanya, Kosaraju, and Sankara Narayanan. "Security and Privacy in Wireless Sensor Networks Using Intrusion Detection Models to Detect DDOS and Drdos Attacks: A Survey." 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE, 2023.
7. Chaitanya, Kosaraju, et al. "Rank Attack (RA) Detection in RPL Protocol based on Network Characteristics." 2023 8th International

Conference on Communication and Electronics Systems (ICCES). IEEE, 2023.

8. Siva Rao, I. S., Lakshmi, P. R., Syma Kumar, D. N. V., Reddy, A. Y., Karthik, J., & Bhavana, B. (2024). An approach for product recommendation using Light GBM. International Journal of Innovative Science and Advanced Engineering (IJISAE), 12(17s).
9. Mandhala, Venkata Naresh, V. Sujatha, and B. Renuka Devi. "Scene Classification Using Support Vector Machines." 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, 2014, https://doi.org/10.1109/icaccct.2014.7019421
10. Brain tumour detection using auto-encoder and multi-layer perception Sujatha, V., Majety, V.D., Kanumalli, S.S., Komanduri, V.S.S.R.K. AIP Conference Proceedings, 2023, 2724, 020010
11. Narayana, Vejendla Lakshman, Arepalli Peda Gopi, and Kosaraju Chaitanya. "Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology." Rev. d'IntelligenceArtif. 33.1 (2019): 45-48.
12. V. L. Narayana, S. Bhargavi, D. Srilakshmi, V. S. Annapurna and D. M. Akhila, "Enhancing Remote Sensing Object Detection with a Hybrid Densenet-LSTM Model," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 264-269, doi: 10.1109/IC2PCT60090.2024.10486394
13. Patibandla, R.S.M.L., Narayana, V.L., Gopi, A.P. (2021). Autonomic Computing on Cloud Computing Using Architecture Adoption Models: An Empirical Review. In: Choudhury, T., Dewangan, B.K., Tomar, R., Singh, B.K., Toe, T.T., Nhu, N.G. (eds) Autonomic Computing in Cloud Resource Management in Industry 4.0. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-71756-8_11
14. V. Pavani, M. N. Swetha, Y. Prasanthi, K. Kavya and M. Pavithra, "Drowsy Driver Monitoring Using Machine Learning and Visible Actions," *2022 International Conference on Electronics and Renewable Systems (ICEARS)*, Tuticorin, India, 2022, pp. 1269-1279, doi: 10.1109/ICEARS53579.2022.9751890.
15. Kumari, G. R. P., Shalini, D., Chandrika, R., DivyaSri, P., Srija, K. A., & Alapati, N. (2025, April). Automated Emerging Cyber Threat Identification and Profiling based on Natural Language Processing using BERT Model. In 2025 8th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 526-533). IEEE.
16. Chinnam, Siva Koteswararao, S. Reshmi Khadherbhi, P. Sandhya Krishna, and D. Anveshini. "Sentiment analysis in services provided by telecommunications." *International Journal of Advanced Science and Technology (IJAST)* 29, no. 03 (2020): 9167-9176.
17. Nanduri, A. K., Sravanthi, G. L., Kumar, K. P., Babu, S. R., & Krishna, K. R. (2021). Modified Fuzzy Approach to Automatic Classification of Cyber Hate Speech from the Online Social Networks (OSN's). *Rev. d'IntelligenceArtif.*, *35*(2), 139-144.
18. Qi, Zhang, P. SilpaChaitanya, and T. Sudhir. "Spoofing attack detection wireless networks using advanced KNN." *International Journal of Smart Device and Appliance* 4.1 (2016): 1-8.
19. Kavishwar, S., & Uppal, S. K. (2020). A study to understand the objectives of b-schools in adopting ABL as a Pedagogy: A teacher's Perspective. *Sambodhi.* 43(04), 180-185.
20. Kavishwar, S (2024). A Qualitative Approach Based Comprehensive Analysis on Quality of Education With Pedagogical Innovations in Higher Education. *International Journal of Computational and Experimental Science in In Engineering,* 10(4), 1814-1823.
21. Joshi, M., Kothari, P. and Kavishwar, S. (2024). A Study on Determinants of Profitability in Indian Banks. *Journal of Informatics Education and Research*. 4(3), 22-26.
22. A. H. J. Almarashy, M.-R. Feizi-Derakhshi, and P. Salehpour, "Enhancing fake news detection by multi-feature classification," IEEE Access, vol. 11, pp. 53710–53721, 2023.

23. J. T. H. Kong, W. K. Wong, F. H. Juwono, and C. Apriono, "Generating fake news detection model using a two-stage evolutionary approach," IEEE Access, vol. 11, pp. 109284–109296, 2023.
24. Jingar, N. K. (2022). Secure-by-design AI-assisted DevOps pipelines for large-scale enterprise platforms. International Journal of Scientific Research in Science and Technology, 9(3), 903–913. https://doi.org/10.32628/IJSRST2291348
25. Jingar, N. K. (2022). Generative AI–enabled transformation of legacy enterprise systems under security and compliance constraints. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 8(2), 760–770. https://doi.org/10.32628/CSEIT23906219
26. M. Carter, M. Tsikerdekis, and S. Zeadally, "Approaches for fake content detection: Strengths and weaknesses to adversarial attacks," IEEE Access, vol. 10, pp. 120438–120452, 2022.
27. Nijim, M., Kanumuri, V., Alaqqad, W., Albataineh, H. (2023). Advanced Traffic Management System for Smart Cities. In: Daimi, K., Al Sadoon, A. (eds) Proceedings of the 2023 International Conference on Advances in Computing Research (ACR'23). ACR 2023. Lecture Notes in Networks and Systems, vol 700. Springer, Cham. https://doi.org/10.1007/978-3-031-33743-7_19
28. F. Al-Quayed, D. Javed, N. Z. Jhanjhi, M. Humayun, and T. S. Alnus, "A hybrid transformer-based model for optimizing fake news detection," IEEE Access, vol. 11, pp. 88095–88106, 2023.
29. Albataineh, H., Kanmuri, V., Alaqqad, W., Nijim, M. (2024). Utilizing Machine Learning for Intrusion Detection in Smart Grid Systems. In: Daimi, K., Al Sadoon, A. (eds) Proceedings of the Third International Conference on Innovations in Computing Research (ICR'24). ICR 2024. Lecture Notes in Networks and Systems, vol 1058. Springer, Cham. https://doi.org/10.1007/978-3-031-65522-7_44
30. Ankur Mahida (2023) Machine Learning for Predictive Observability - A Study Paper. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-252. DOI: doi.org/10.47363/JAICC/2023(2)235
31. D. Jung, E. Kim, and Y.-S. Cho, "Topological and sequential neural network model for detecting fake news," IEEE Access, vol. 11, pp. 54107–54118, 2023.
32. Ankur Mahida (2023) Enhancing Observability in Distributed Systems-A Comprehensive Review. Journal of Mathematical & Computer Applications. SRC/JMCA-166. DOI: doi.org/10.47363/JMCA/2023(2)135
33. Tummuri, S. S. R. (2023). Quantization aware training techniques for efficient transformer-driven large language models. International Journal of Scientific Research & Engineering Trends, 9(2).
34. Tummuri, S. S. R. (2022). Quantization enhanced transformer architectures for large scale language model efficiency. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 8(3), 891–904.
35. B. K. Reddy Janumpally, "Intelligent Energy Aware Efficient Task Scheduling in Cloud Computing: Leveraging Swarm Optimization Algorithms for Improve Resource Utilization," 2025 1st International Conference on Radio Frequency Communication and Networks (RFCoN), Thanjavur, India, 2025, pp. 1-6, doi: 10.1109/RFCoN62306.2025.11085278.
36. Janumpally, Bharath Kumar Reddy. (2026). Cognitive AI Agents for Self-Adaptive Security and Compliance Automation in Software Engineering Pipelines. 10.1109/ICAUC68182.2026.11441048.