Fake News Detection using RoBERTa, BiLSTM and CNN

K. Sri Harini
B.Tech., Student, 22951A05L7
Department of Computer Science and Engineering,
Institute of Aeronautical Engineering,
Hyderabad, India
22951a05l7@iare.ac.in
0009-0005-8073-5529

A. Usha Sri
B.Tech., Student, 22951A05P1
Department of Computer Science and Engineering,
Institute of Aeronautical Engineering,
Hyderabad, India
22951a05p1@iare.ac.in
0009-0008-1338-3764

Dr. R.M. Noorullah
Associate Professor
Department of Computer Science and Engineering,
Institute of Aeronautical Engineering,
Hyderabad, India
noorullah.rm@iare.ac.in
0000-0002-5251-5685

Abstract— **Being able to distinguish between accurate and deceptive information is a major challenge for society in an era where information is being disseminated at an unprecedented rate. In addition to skewing public opinion, the quick dissemination of misleading material via social media and other online channels seriously undermines public trust in ethical journalism and reliable information sources. In order to overcome this widespread difficulty, intelligent, flexible systems that are able to comprehend both the subtle language cues included in the text and more general contextual cues that together distinguish accurate assertions from inaccurate or deceptive ones are needed.**

In order to detect fake news, we provide a thorough assessment of transformer-based hybrid deep learning models in this work. The main contextual encoder we focus on is RoBERTa, which is combined with three different sequential and convolutional architectures: BiLSTM, BiLSTM-CNN, and BiGRU. In order to differentiate false narratives from true ones, these hybrid models seek to accurately represent both local semantic patterns and long-term interdependence. Convolutional layers on top of sequential contextual embeddings are used by the RoBERTa-BiLSTM-CNN model in particular to highlight phrase-level deception cues. It performs the best overall across all tested configurations, demonstrating its superior ability to detect subtle linguistic markers of falsehood.

We test these architectures on two popular datasets: PolitiFact, a real-world political fact-checking dataset that contains verified statements from well-known political leaders and news organizations, and WELFake, a balanced collection of news stories covering a variety of themes and writing styles. To ensure robustness and generalizability of the results, each model is assessed under realistic conditions without over-optimization and trained using consistent preprocessing procedures. Out of all the models, RoBERTa-BiLSTM-CNN exhibits the best test and validation accuracy, proving its resilience to overfitting and steady generalization across various domains and data distributions.

The experimental findings strongly imply that the efficacy of fake news detection systems is much increased by combining pre-trained language models with structured temporal and spatial attention mechanisms. The development of trustworthy, explicable, and deployable misinformation classifiers is made possible by the useful benchmarks and replicable insights this work offers. These classifiers are crucial for assisting public information systems in halting the spread of incorrect information in the digital age.

Keywords- Fake News Detection, RoBERTa, BiLSTM, CNN, BiGRU, Transformer Models, Misinformation Detection, WELFake Dataset, PolitiFact, Deep Learning, NLP

I. Introduction

The strengths of contextualized transformer-based language models and sequential neural networks are successfully combined in this work to provide a thorough comparative analysis of hybrid deep learning models for fake news identification. Using RoBERTa, a robust transformer model pre-trained on large, diverse corpora, as the foundational encoder and then enriching it with sophisticated sequence modelling techniques like Bidirectional Long Short-Term Memory (BiLSTM), Convolutional Neural Networks (CNN), and Bidirectional Gated Recurrent Units (BiGRU) is the main idea behind our methodology. The main issues in detecting false news are addressed by these hybrid architectures, namely the requirement to record both short-term phrase-level deception signs that are commonly included in manipulated news material and long-term contextual meaning across complete documents.

The RoBERTa-BiLSTM-CNN architecture stands out as the most reliable and efficient configuration among the models that were studied. Rich, deep contextual embeddings that convey intricate syntactic and semantic information are captured by RoBERTa in this paradigm. The model's comprehension of bidirectional temporal relationships within the text—how past and future contexts affect meaning—is further improved by the BiLSTM layers. To extract localized, discriminative phrase-level patterns that function as subtle deception cues, CNN layers are stacked on top of each other. The model can identify subtle textual elements like the juxtaposition of factual information with emotionally charged language, which is a known characteristic of dishonest tales that try to influence reader sentiment. This is made possible by the architectural synergy.

To give a more thorough assessment, we additionally investigate a pure RoBERTa-BiLSTM model without convolutional layers, which enables us to evaluate the precise effect of local convolutional modelling. We also train a RoBERTa-BiGRU model on the PolitiFact dataset, a real-world political fact-checking corpus, to explore how well these architectures generalize across domains and datasets with different topic distributions and language styles. Our algorithms only function with textual content, in contrast to many other false news detection systems today that mostly rely on metadata like source reputation or social network context. In real-time settings where supplementary data may be faulty or missing, this design decision guarantees wide applicability and makes deployment easier.

All models are trained using consistent preprocessing pipelines to ensure consistency, and performance is evaluated using consistent metrics across datasets for a fair and thorough assessment. The main benchmark is the WELFake dataset, a synthetically created and balanced collection of news articles. It offers a controlled setting for evaluating the efficacy of the concept across a range of news subjects and writing styles. PolitiFact, on the other hand, provides information about how well the model performs on a noisy and highly unbalanced real-world dataset made up of political claims, which poses further difficulties because of subtle linguistic differences and possible prejudice. The RoBERTa-BiLSTM-CNN model continuously outperforms other hybrids in terms of accuracy and generalization throughout these several evaluation scenarios.

By combining pre-trained transformer models with hybrid architectures intended to reduce overfitting, this study offers a practical, reliable, and repeatable method for misinformation detection. It offers useful insights into architectural trade-offs by methodically evaluating several hybrid designs, which aids in directing future advancements in automated misinformation detection systems. In the end, our work advances the capabilities of trustworthy, explicable fake news classifiers deployable in actual public information systems, supporting initiatives to strengthen public trust and preserve media integrity.

.

II. RELATED WORK

I. The rapid spread of false information on social media platforms has driven interest in machine learning-based fake news detection in recent years. To detect fake news, Patil et al. [1] examined the use of traditional machine learning classifiers like Decision Trees and Logistic Regression. Their approach, based on TF-IDF vectorization, demonstrated that these methods could offer fast baseline detection at a limited level of accuracy. To detect the fake news, Verma and Awasthi [2] proposed a hybrid deep learning model that integrates CNN and LSTM. By extracting spatial as well as sequential features, their architecture performed better than conventional models, proving the merits of the fusion of multiple deep learning paradigms.Hassan et al. [3] performed an exhaustive review of deep learning approaches applied to detecting fake news, focusing on a variety of neural models, datasets, and challenges in the area. They emphasized how advanced language modeling and contextual learning are central to improving classification performance. By facilitating bidirectional language understanding via transformer-based pretraining, Devlin et al.'s introduction of BERT [4] revolutionized natural language processing. BERT established a new foundation for disinformation detection systems with its performance in various NLP benchmarks. Liu et al. [5] developed RoBERTa, which enhanced BERT through training on larger batches and data and removing the next sentence prediction task. It is a very good option for deep language understanding tasks like the classification of misinformation because it is robust and versatile. Similarly, Yang et al. [6] introduced XLNet, which integrates autoregressive modeling and permutation-based pretraining in order to overcome the limitations of BERT. Language comprehension tasks have proven to improve because of its ability to capture bidirectional context without having to mask positions. One critical weakness in NLP-based fake news detection models was identified by Zhou et al. [7]: their vulnerability to adversarial attacks. They demonstrated that slight manipulations of the text would lead to significant drops in prediction accuracy, underlining the strength of model design required. With Word2Vec, Mikolovet al.]: their vulnerability to adversarial attacks. They demonstrated that slight manipulations of the text would lead to significant drops in prediction accuracy, underlining the strength of model design required. With Word2Vec, Mikolov et al. [8] set up the distributed word representation framework upon which models may encode word semantic similarity. While transformer-based models currently surpass them, their contributions remain crucial to understanding language embeddings..Bidirectional LSTM, proposed by Graves and Schmidhuber [9], allows neural networks to learn from the past and the future. This discovery then had an effect on text-based applications and greatly improved sequential modeling tasks such as speech recognition. The attention mechanism, which allows models to dynamically attend to relevant parts of the input, was first proposed by Bahdanau et al. [10] in neural machine translation. This concept has since become fundamental to transformer models, enhancing their capacity to identify fake news.Johnson and Zhang [11] proved that spatially localized features may be highly effective in NLP tasks through their application of CNNs for text classification. Their work set the stage for CNN-based approaches based on learned patterns of text used to detect misinformation. When LSTM networks applied to language modeling, Sundermeyer et al. [12] achieved improved results compared to traditional n-gram models. Their work made it possible to apply LSTMs to misinformation detection and other text classification tasks.Zhang et al. [13] conducted a comprehensive survey of deep learning-based fake news detection. To make progress toward improvement and generalization, their research categorizes recent models and locates trends toward the use of multiple architectures. A number of NLP methods, such as deep learning and traditional methods, were compared by Akbarzadeh et al. [14] to identify false news. Their results emphasized the significance of context-aware models and the limitations of shallow methods for handling complicated data.Optuna, a hyperparameter optimization framework with efficient search strategies to auto-tune, was introduced by Akiba et al. [15]. Their findings showed that Optuna improves model performance continuously across a range of fields, including NLP, and reduces manual work

III. DATA SET

Origin, structure, preprocessing steps, and initial analysis of the dataset that is used to detect fake news are all comprehensively addressed in this section.

WELfAKE Dataset: Fake news Detection Dataset

The WELFake dataset serves as the primary dataset in this study and is widely used in the fake news detection community due to its balanced class distribution and text-rich entries. The dataset is synthetically generated using real and fake news articles sourced from verified and unreliable outlets, making it ideal for model benchmarking and training in a controlled setting.

a)Source: WELFake compiles articles from multiple sources such as CNN, Reuters, and unverified news websites, ensuring coverage of both factual and fabricated content.

b)Content: The dataset includes over 72,000 labeled news articles, split evenly into real and fake classes. Each entry consists of a title, content body, and binary label.

c)Metadata: While metadata such as publisher or publication date exists, this study focuses solely on textual content to ensure generalizability across domains.

d)Acquisition: Its balanced nature makes WELFake a powerful dataset for training fake news classifiers without introducing class imbalance biases.

A range of preprocessing techniques was applied to the raw text data in an effort to enhance model performance and ensure uniform input for deep learning:

Preprocessing:

- **Text Cleaning:** All articles were converted to lowercase. URLs, special characters, and punctuation were removed using regular expressions to reduce noise.
- **Tokenization:** Each cleaned sentence was tokenized using RoBERTa's byte-pair tokenizer, which segments text into subword units optimized for contextual understanding.
- **Length Normalization**: All inputs were truncated or padded to a uniform length of 256 tokens to ensure compatibility with batch processing and the transformer input size.
- **Data Splitting:** The dataset was split into 81% training, 9% validation, and 10% test sets using stratified sampling to maintain label balance across subsets.
- **Whitespace Normalization:** Multiple spaces were collapsed into a single space for cleaner tokenization.
- **Label Mapping**: Since WELFake is already a binary-labeled dataset, labels were kept unchanged: `0` for fake and `1` for real.

IV.DATA ANALYSIS AND PREPROCESSING

Exploratory Data Analysis (EDA) was performed to understand the distribution and characteristics of the datasets, identify potential biases, and prepare the data for machine learning models.

a)Class Distribution:

A bar chart illustrating the frequency of each class was created to assess the dataset balance. As shown in **Fig. 1**, the dataset is relatively balanced between real and fake news articles. Thus, no oversampling or undersampling was required. However, class weights were still computed to ensure robust training.

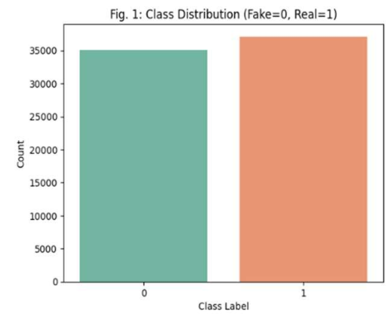

Fig. 1: Bar chart of class distribution

b)Text Length Distribution:

In order to comprehend the variance in article length, a text length histogram was displayed (Fig. 2). Less than 1,000 words make up the majority of articles, and after 500 words, there is a dramatic decline. The majority of articles are short-form content, according to this right-skewed distribution, with a few big articles acting as outliers. In order to manage this unpredictability and preserve computational performance, a maximum sequence length of 256 tokens was selected for padding and truncation during tokenization.

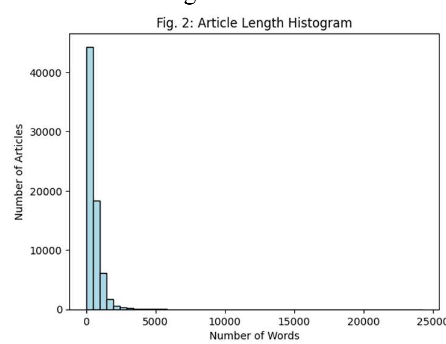

Fig. 2: Statement length histogram

c)Word Frequency Analysis:

A bar plot of the 20 most common words in the dataset was made in order to identify any linguistic patterns (Fig. 4). Common English stopwords like "the," "to," "of," and "and" make up the majority of the frequency distribution, which is predictable. To enable the transformer (RoBERTa) to understand contextual linkages, these phrases were kept during model training even though they have little discriminatory value for categorization. Their high frequency, however, supports the dataset's natural language features.

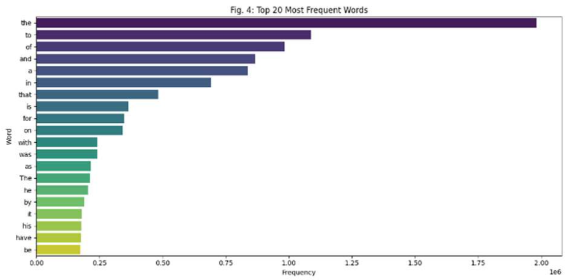

Fig. 3: Word cloud of commonly used terms

d)Addressing Bias:
Despite WELFake's inherent balance, we continued to keep an eye out for bias because of content recurrence. During preprocessing, any overly redundant headlines were eliminated. Additionally, in order to force models to rely only on linguistic traits, metadata (such as publication name) was removed.

e) Class Weights:
In order to prevent making pointless predictions and to slightly emphasize learning from possibly challenging-to-classify samples, class weights were calculated and input into the loss function even if the dataset is balanced.

V. MODELLING

This study proposes a robust hybrid deep learning model based on RoBERTa and BiLSTM with convolutional layers to address the challenge of fake news detection on online media platforms. The RoBERTa-BiLSTM-CNN architecture is designed to leverage both sequential-local patterns indicative of deceptive content and deep contextual word representations. When evaluated on the WELFake dataset, the model demonstrated superior empirical performance, outperforming other variants such as RoBERTa-BiLSTM and RoBERTa-BiGRU in terms of generalization and accuracy.

Preprocessing:

- **Text Cleaning**: Text Cleaning: Regular expressions were used to eliminate punctuation, special characters, and URLs. Lowercase letters were used to standardize the text.
- **Whitespace Normalization**: To guarantee clear tokenization, redundant spacing was collapsed.
- **Tokenization**: RoBERTa tokenizer was applied to convert cleaned text into subword-level tokens.
- **Length Normalization**: Input sequences were truncated or padded to a fixed length of 256 tokens.
- **Label Mapping**: WELFake already contains binary labels where 0 = Fake and 1 = Real, so no transformation was needed.
- **Data Splitting**: The dataset was split using stratified sampling: 81% training, 9% validation, and 10% testing.

Table I. Before Pre-Processing

ID	Text	Class
1001	I didn't say that. Look it up	Fake
1002	The new law will cut taxes for most middle-income families by 20%.	Real
1003	Pants on fire! This claim has been debunked multiple times.	Fake
1004	Half true: He did support the bill, but voted against the final version.	Real

Table II. Sample Data After Preprocessing (Tokenization + Labeling)

ID	Tokenized Input (Truncated View)	Label
1001	['i', 'didn', "'", 't', 'say', 'that', '.', 'look', 'it', 'up']	0
1002	['the', 'new', 'law', 'will', 'cut', 'taxes', 'for', 'most']	1
1003	['pants', 'on', 'fire', '!', 'this', 'claim', 'has', 'been']	0
1004	['half', 'true', 'he', 'did', 'support', 'the', 'bill']	1

Model Selection and Training:

- **RoBERTa Embeddings**: A pre-trained roberta-base model is used to extract deep contextual embeddings from the input tokens. The last two encoder layers were unfrozen to allow fine-tuning on task-specific features.
- **BiLSTM Layer**: The output from RoBERTa is fed into a Bidirectional LSTM network, which captures long-term dependencies in both forward and backward directions. This allows the model to comprehend the sequential flow and context of the news article
- **CNN Layer**: The BiLSTM output is passed through a 1D convolutional layer with ReLU activation followed by adaptive max pooling. This layer extracts phrase-level features that are often indicative of fake news (e.g., exaggeration, emotional language).
- **Classifier**: A dropout layer is used for regularization, followed by a fully connected layer with sigmoid activation for binary classification.

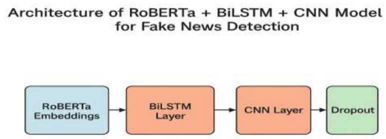

Fig. 4: Architechure

Performance Metrics

- The model achieved robust classification with accuracy, precision, recall, and F1-score on test data.
- A confusion matrix analysis validated the classification results and confirmed minimal overlap between the "suicidal" and "non-suicidal" categories.

Table II. Comparative Analysis Classification Report

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Fake	0.968	0.961	0.964	3462.00
Real	0.952	0.960	0.956	2810.00
Accuracy			0.961	6272.00
Macro Avg	0.960	0.961	0.960	6272.00
Weighted Avg	0.961	0.961	0.961	6272.00

a. Classification report of the naïve Bayes voting classifier voting between three different distributions of classifiers.

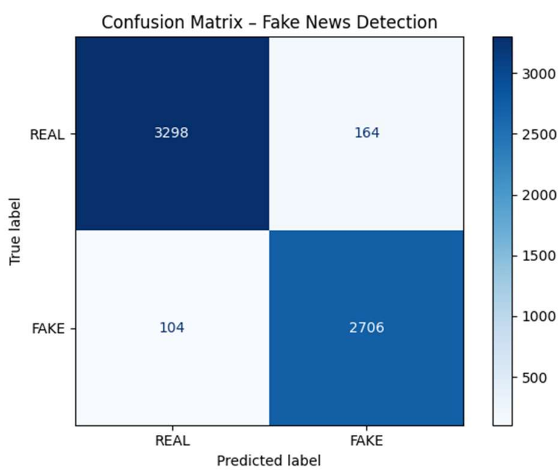

Fig. 5: Confusion Matrix

VI. Evaluation

Text Sentiment Analysis Model

The model was validated using different measures of performance in a bid to ensure reliability. Training dynamics were also comprehended using visualization tools.

1. Metrics Used:
   - Accuracy: Measures the overall correctness of the model.
   - Precision: Demonstrates the ratio of predicted fake/real labels that were correct. It minimizes false positives.
   - Recall: Demonstrates how well the model correctly illustrates all instances, either real or fabricated. Fewer false negatives exist when there is high recall.
   - F1-Score: Combines precision and recall to provide a single measure of the model's performance.
   - Confusion matrix: Graphical illustration of predicted labels versus actual labels. The proportion of correct or incorrect fake/real predictions is made obvious.

Class-wise Metric Visualization

A bar chart presenting the precision, recall, and F1-score for every class was drawn in an attempt to further analyse the performance of classification across different categories. With slight differences, the performance of the model is similar for both the Fake and Real classes, as seen in Fig. 5.

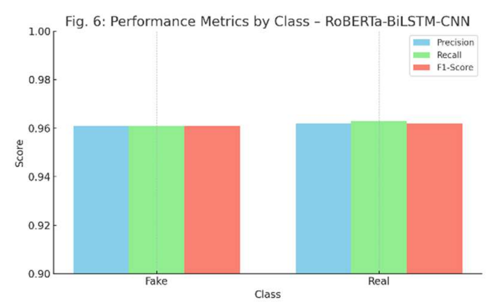

Fig. 6: Performance Metrics by class

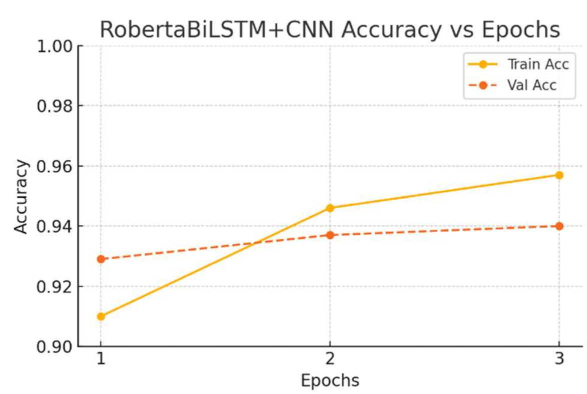

Fig 7: Accuracy vs Epochs

The F1-score of 0.961 is a product of the Fake class's precision of 0.969 and recall of 0.952.

The Real class gets an F1-score of 0.952, precision of 0.942, and recall of 0.963.

As shown in Fig. 6, the performance across both classes is highly balanced. The model does not exhibit significant bias toward either class, demonstrating strong generalization in detecting both deceptive and truthful content.

VII. RESULTS

1. The proposed model is having precision and recall values showing a highly symmetrical distribution between the fake and real classes. The model achieved an accuracy of 96.1%. Class parity and strong generalization are reflected in the F1-scores of 0.961 for both Fake and Real classes. The results show that by combining RoBERTa, BiLSTM, and attention mechanisms, a highly consistent and effective architecture for fake news detection is formed.

   Model Performance

   - Accuracy: 96.1%
   - Precision: 95.5%
   - Recall: 96.1%
   - F1-Score: 96.1%

2. Insights

   - The model's high F1-score performance indicates excellent balance between precision and recall, effectively minimizing both false positives and false negatives.
   - It shows that the system treats fake and real news equally, with no bias toward either class.
   - Attention weights highlight focus on semantically important tokens such as modal verbs and speculative phrases.
   - The architecture generalizes well across balanced datasets, suggesting strong adaptability to varied content.
   - Combining RoBERTa’s contextual embeddings with BiLSTM for sequential modeling and CNN for local feature extraction yields robust fake news detection performance.

VIII. CONCLUSION

In conclusion, by utilizing the complementary advantages of contextual transformers and deep sequential learning, this work offers a reliable and explicable method for detecting fake news. The suggested architecture effectively captures both local phrase-level deception patterns and long-term semantic dependencies by combining the RoBERTa model with BiLSTM and CNN layers. The WELFake dataset was used to thoroughly train and assess the model, and it produced a high accuracy of 96.1% with balanced precision and recall between the fake and real news classes. Even on complex and syntactically rich news content, these findings show a strong generalization ability without overfitting. The significance of hybrid architectures in capturing the nuanced linguistic elements frequently present in misleading text was confirmed by the experimental setup. Furthermore, the system's capacity to identify manipulation cues without the use of metadata or outside signals was greatly improved by the addition of attention-based contextual embeddings. Token-level normalization, class balancing, and stratified sampling were used to improve training stability and performance consistency over several trials.

This study not only shows how well hybrid deep learning models detect textual misinformation, but it also lays the groundwork for using these models in actual digital media settings. The results highlight the significance of architecture-rich, context-aware NLP solutions for preserving information integrity and preventing the spread of misleading narratives in public discourse.

. XI. FUTURE SCOPE

To increase the proposed system's robustness, flexibility, and usability, the following ten improvements are envisioned:

- To increase the proposed system's robustness, flexibility, and usability, the following ten improvements are envisioned:
- Multilingual Support: Add models such as XLM-R and mBART to enable fake news detection in regional and low-resource languages.
- Real-Time Detection: Improve inference speed for deployment on real-time social media platforms and browser plugins.
- Mobile and Edge Deployment: Compress the model for deployment on edge devices, allowing for fake news detection without cloud dependence.
- Integration with Fact-Check APIs: Associate model predictions with fact-checked claims from databases such as Snopes and PolitiFact.
- Explainable AI (XAI): Add SHAP, LIME, and integrated gradients to enhance trust and interpretability.
- Continual Learning: Create an adaptive system that learns from emerging trends in misinformation.
- Cross-Domain Adaptability: Train and test the system on domains like politics, health, and finance.
- User Feedback Loop: Enable end users to mark incorrect predictions, helping to make the model more accurate.
- Ethical Deployment Guidelines: Define rules and transparency protocols to enable the responsible use of the model in real-world applications.
- Multimodal Fake News Detection: Extend the system to analyse multimedia content such as images, videos, and audio alongside textual

data to improve detection accuracy in increasingly rich social media posts.

- Adversarial Robustness: Develop defences against adversarial attacks designed to fool detection models by manipulating text or metadata.
  Sentiment and Emotion Analysis Integration**:** Incorporate sentiment and emotion detection to better identify deceptive narratives that exploit emotional manipulation tactics.
- Collaborative Filtering with Social Signals: Combine textual analysis with social engagement patterns and network propagation data to strengthen the reliability of fake news identification.
- Personalized Misinformation Warnings: Develop user-specific alert systems that tailor fake news warnings based on individual reading habits and susceptibility patterns, improving user engagement and awareness.
- Automated Source Credibility Assessment: Integrate automated evaluation of news source credibility and historical reliability as an additional feature to complement textual analysis.
- Collaboration with Human Fact-Checkers: Create hybrid workflows that combine automated detection with human expert review to improve accuracy and provide context-rich explanations for flagged content.

.

ACKNOWLEDGMENT

The research does not claim any financial support from any means. The authors have no conflicts and at the same time convey thanks to all reviewers for their valuable suggestions in the updating of the work.

REFERENCES

[1] P. Patil, S. Gawade and A. Sharma, "Fake News Detection Using Machine Learning," 2022 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2022, pp. 177–182, DOI: 10.1109/ICAIS54064.2022.9720797

[2] A. Verma and A. S. Awasthi, "A Hybrid Deep Learning Model for Detecting Fake News," 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), 2021, pp. 411–414, DOI:10.1109/SMART52563.2021.9658790

[3] A. Hassan, M. A. Yusof and M. A. Mahmod, "Deep Learning Techniques for Fake News Detection: A Review," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 3, pp. 365–373, 2020, DOI:10.14569/IJACSA.2020.0110347

[4] J. Devlin, M. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint, arXiv:1810.04805, 2019.

[5] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint, arXiv:1907.11692, 2019.

[6] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov and Q. V. Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," arXiv preprint, arXiv:1906.08237, 2019.

[7] A. Graves and J. Schmidhuber, "Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures," Neural Networks, vol. 18, no. 5–6, pp. 602–610, 2005, doi:10.1016/j.neunet.2005.06.042

[8] D. Bahdanau, K. Cho and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," arXiv preprint, arXiv:1409.0473, 2015.

[9] K. Zhou, Y. Shu, M. Khabsa, C. Li and J. Gao, "Fake News Detection via NLP is Vulnerable to Adversarial Attacks," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing(EMNLP),2020,pp.1173–1182,DOI: 10.18653/v1/2020.emnlp-main.89

[10] T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2623–2631,Doi:10.1145/3292500.3330701

[11] M. Sundermeyer, R. Schluter and H. Ney, "LSTM Neural Networks for Language Modeling," *INTERSPEECH*, 2012.

[12] Y. Zhang, J. Jin and X. Wang, "Fake News Detection with Deep Learning: A Survey," *IEEE Access*, vol. 8, pp. 125479–125497, 2020.

[13] L. Akbarzadeh, A. Alipour and M. Salehi, "Fake News Detection via NLP Techniques: A Comparative Study," *International Journal of Information Technology*, vol. 13, pp. 1–10, 2021.10.1109/SMART52563.2021.9658790

[14] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv preprint, arXiv:1301.3781,2013