Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

Deep Learning Multimodal Methods to Detect Fake News

S. Saraswathi[1], S. Jeevithaa[2], K. Vishwabharathy[3], D. Eyuvaraj[4]

[1,2,3,4]Department of Information Technology, Puducherry Technological University, Puducherry, India

Email: [1]s.saraswathi@ptuniv.edu.in, [2]jeevithaa.s@pec.edu, [3]vishwabharathy.k@pec.edu, [4]eyuvaraj.d@pec.edu

Abstract

Fake news, characterized by false information disseminated intentionally with malicious intent, has become a critical societal issue. Its impact spans political, economic, and social domains, fueled by the rapid proliferation of digital communication channels, particularly social media. To combat this menace, researchers have turned to automated mechanisms for detection, leveraging machine learning algorithms and curated datasets. In this exploratory research, the landscape of machine learning algorithms is employed in identifying fake news. Notably, the research focus on algorithms such as the Bidirectional Encoder Representations from Transformers (BERT) and Convolutional Neural Network (CNN) respectively. However, most of these studies rely on controlled datasets lacking real-time information from social networks—the very platforms where disinformation thrives. The findings underscore the need for research in social network environments, where fake news spreads most prolifically. Additionally, future investigations should extend beyond political news, considering hybrid methods that combine NLP and deep learning techniques. This study serves as a valuable resource for researchers, practitioners, and policymakers seeking insights into the evolving landscape of the ability to combat fake news effectively.

Keywords: Deep Learning, Fake News Detection, Natural Language Processing, Pre-Trained Embedding, Text Classification, Multimodal Datasets, Evaluation Metrics.

DOI: https://doi.org/10.36548/jtcsst.2024.2.004
Received: 08.04.2024, received in revised form: 30.04.2024, accepted: 13.05.2024, published: 22.05.2024

1. Introduction

Our focus is on advanced algorithms such as the BERT and CNN. This approach not only enhances the accuracy of fake news detection but also improves the system's generalizability across political news [8-12]. However, a notable limitation in current research is the reliance on controlled datasets that lack real-time information from social networks—the very platforms where disinformation thrives. It underscores the need for future investigations to consider hybrid methods that combine NLP and deep learning techniques. we aim to contribute significantly to the existing body of knowledge in the field of fake news detection and inspire future research in this critical area [13-15].

2. Related Works

This section offers a summary of the literature on fake news detection. In paper [1], the principles behind word embeddings are discussed, focusing on the GloVe methodology and the BERT model. Deep learning techniques such as CNN, LSTM, and FakeBERT, which employs three parallel 1D-CNN blocks with convolutional, max-pooling, and dense layers, are also covered. Paper [2] handles multimodal data, using SBERT-Mpnet for textual content and Vision Transformers (ViTs) for visual content. Various machine learning and deep learning models are employed as classification models, analyzing the fused features and assigning each sample to predefined classes like 'Support,' 'No-Evidence,' 'Refute,' or 'Satirical,' effectively determining the claim's veracity and providing valuable insights into its legitimacy. The Paper [3] combines Transformer-based encoding and decoding, consisting of two key parts: encoder aims to learn meaningful representations from the provided fake news data. decoder focuses on predicting future behavior based on its understanding of past observations. The Paper [4] uses convolutions and non-linear activations. Generalized autoregressive model (XLNet) predicting the next word based on both preceding and succeeding context. EANN Combines Text-CNN and VGG19 for textual and visual feature extraction, MVAE is used as a variational autoencoder to learn correlations between textual and visual features for reconstruction. The Paper [5] the HyproBert models input Layer prepares the data by tokenizing, indexing, and padding. Embedding Layer converts tokens into numerical representations. convolutional Layer extracts spatial features from the embeddings. BiGRU Layer captures deep features. The Paper [6] techniques recurrent neural networks and convolutional neural networks automatically learn powerful representations from text data. As news content diversified beyond text, research shifted to harnessing visual features as well. The Paper [7] proposes convolutional neural network (CNN), learning network (FLN), TF-IDF, with methods such as classification and feature extraction using ReLU for activation function.

Table 1. Summary on the Studies

Year	Authors	Contribution	Dataset	Methodology	Limitation
2021	Rohit Kumar Kaliyar, et al.[1]	"FakeBERT to improve fake new performance	"U.S. General Presidential Election-2016"	Word embedding, GloVe, BERT	only focuses on social media, and does not consider other sources or domains of misinformation
2022	Suryavardan, et al. [2]	Curated multimodal fact verification dataset	Factify 2	SBERT-Mpnet, Vision Transformers (ViT)	multi-class problem, which may not account for the nuances and uncertainties
2022	Balasubramanian Palani , et al. [3]	model can capture the most textual and visual achieve better classification accuracy	Politifact and Gossipcop	BERT CapsNet	does not handle news articles containing multiple images or videos, which may have different levels of relevance and credibility
2022	S.Raza, et al.[4]	object detection, research background, challenges faced by traditional algorithms	NELA-GT Fakeddit	transformer-based encoding and decoding	by tuning hyperparameters, we can improve model performance
2023	Muhammad Imran Nadeem, et al.[5]	CapsNet with 3-routing and	FA-KES ISOT	HyproBERT	content-based features, ignoring the

		4D five capsules			significance of social context and multimedia content in fake news detection.
2023	Jing Jing, et al.[6]	feature-based) and transformer-based ML and Dl respectively	Twitter Weibo	Multimodal progressive fusion network (MPFN)	Dataset used is confidential and cannot be extracted

3. Proposed Work

The proposed work represents a comprehensive research initiative aimed at advancing the field of fake news detection. At its core, it recognizes the multifaceted nature of deceptive content in the digital landscape and seeks to address this complexity through a multi-modal approach. Fake news can take various forms, including text, images and the proposed research is dedicated to developing a framework that can effectively handle all these modalities. By taking this holistic approach, the work strives to ensure that no facet of deceptive content remains unaddressed, providing a more complete and versatile solution.

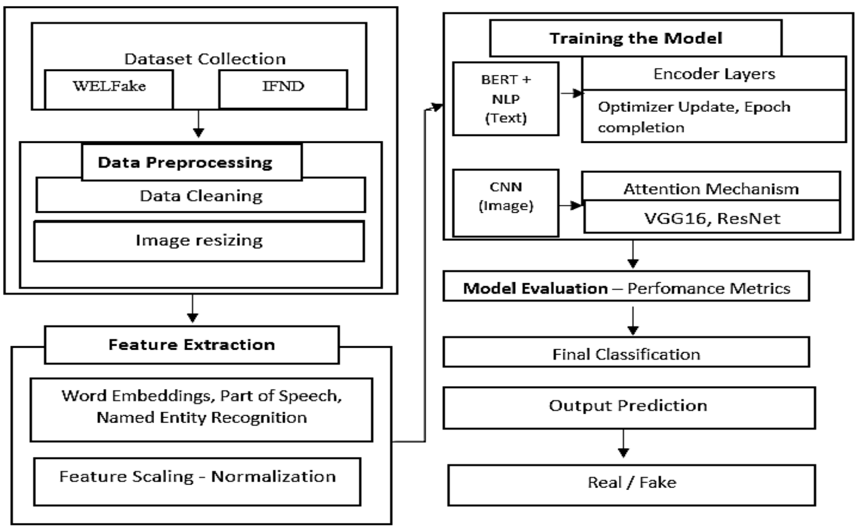

Figure 1. Block Diagram of Proposed Work Multi-Modal Fake News Detection Using Natural Language Processing and Deep Learning

We have 3 modules namely data preprocessing, model development on algorithms such as the BERT and CNN respectively and model evaluation.

a. Data Pre-Processing

The data preprocessing does clean such as removing punctuation, special characters, stop words, converting text to lowercase, handling typos and abbreviations. Normalizing technique such as reducing words to their root which is lemmatization. Label creation, column dropping from the dataset and handles empty metadata by assigning null value and combines the metadata. Dropna function remove any rows that contains missing values. Resizing the images to a uniform size for efficient processing by models, normalization converting the pixel values to a common range.

b. Dataset Used

The WELFake dataset comprises 72,134 news articles, including 35,028 real and 37,106 fake news items. It merges four popular datasets (Kaggle, McIntire, Weibo, BuzzFeed Political) to prevent overfitting and provide ample text data. Specifically focused on Indian news, the IFND dataset was created by scraping Indian fact-checking websites. It features both real and fake news, collected from reliable sources, with a focus on political topics. Each of the 46,755 news articles has a unique identifier, title, image URL, topic, and date. News articles are labeled as '1' for true news and '0' for fake news.

c. Model Development

Model developed with NLP techniques to capture the contextual information from the text. Pretrained language models such as optimized BERT are utilized for feature extraction from the text. BERT captures contextual information by considering the surrounding words and their relationships, resulting in rich representations of the text. For the image pretrained CNN models, such as, ResNet for tasks that require very deep architectures or demand high accuracy, such as image recognition in complex scenes, object detection, or semantic segmentationare employed to extract visual features from images. These CNN models are pretrained on large image datasets like ImageNet and are capable of extracting high-level features from images through hierarchical convolutional layers.

4. Evaluation

The training process involves several stages to evaluate the performance of a model. Typically, TensorFlow is employed for training deep learning models, offering functionalities to monitor training metrics like loss and accuracy across each epoch. Throughout training, a training history object is utilized to log these metrics for subsequent analysis. After each training epoch, the training history object records changes in the model's loss and accuracy. This information is then visualized using Matplotlib. Once the training phase is complete, the trained model can be used to predict labels, such as "real" or "fake," for unseen data, the test set. To evaluate the model's accuracy on this test data, confusion matrices can be generated using scikit-learn. These matrices depict the frequency of correct and incorrect predictions made by the model. Finally, from the confusion matrix, various metrics such as precision, recall, F1-score, and overall accuracy can be calculated. These metrics provide a comprehensive understanding of the model's performance, aiding in further analysis and refinement.

5. Results and Discussion

The metrics accuracy, precision, recall, and F1-score are selected to assess the model's performance across textual and image modalities. These metrics provide a holistic view of the model's effectiveness, considering its ability to correctly classify fake news instances and minimize false positives and false negatives. Accuracy measures the overall proportion of correctly classified instances. It simply calculates the percentage of samples your model predicted correctly regardless of the class.

$$\text{Accuracy} = ((\text{True Positives} + \text{True Negatives}))/ (\text{Total Samples}) \quad (1)$$

Precision focuses on the model for avoiding false positives. It is the proportion of predicted positive cases that are positive.

$$\text{Precision} = (\text{True Positives})/ (\text{True Positives} + \text{False Positives}) \quad (2)$$

Recall focuses on the model is for avoiding false negatives. It is the proportion of actual positive cases that are correctly identified.

$$\text{Recall} = (\text{True Positives})/ (\text{True Positives} + \text{False Negatives}) \quad (3)$$

F1-score is a harmonic mean of precision and recall, it takes the average of their inverses. This offers a combined measure that considers both precision and recall.

$$\text{F1-Score}=2*((\text{Precision}*\text{Recall}))/(\text{Precision}+\text{Recall}) \quad (4)$$

Table 2. Classification Report for the Proposed Model Text- BERT

	precision	recall	f1-score	support
0	0.99	0.99	0.99	7006
1	0.99	0.99	0.99	7302
accuracy			0.99	14308
macro avg	0.99	0.99	0.99	14308
weighted avg	0.99	0.99	0.99	14308

The confusion matrix shows how often the model's predictions were correct or incorrect. In this case, it seems the model performed well at identifying both real and fake news articles. The value under "0" and "0" is 0.99, indicating that 99% of class 0 articles were correctly classified. The value under "1" and "1" is 0.99, indicating that 99% of class 1 articles were correctly classified. The model shows a good performance overall, with an accuracy of 99% and an F1-score of 0.99 for both classes. This suggests that the model can effectively distinguish between real and fake news articles. The F1-score indicates a good balance between precision and recall. Macro average and weighted average provide average precision, recall, and F1-score scores across all classes, with macro avg giving equal weight to each class and weighted avg considering the number of data points in each class.

Table 3. Classification Report for the Proposed Model Image-CNN

	precision	recall	f1-score
Fake	0.50	1.00	0.67
Fake_news	0.00	0.00	0.00
Real	1.00	1.00	1.00
accuracy			0.80
macro avg	0.50	0.67	0.56
weighted avg	0.70	0.80	0.73

The confusion matrix visualizes how often the model's predictions were correct or incorrect. In this case, it seems the model performed well at identifying real news articles. The value under "Real" and "Real" is 1.00, indicating all real news articles were correctly classified.

On the other hand, the model seems to have struggled more with fake news articles. The value under "Fake" and "Real" is 1.00, which means the model correctly classified all fake news articles as real news.

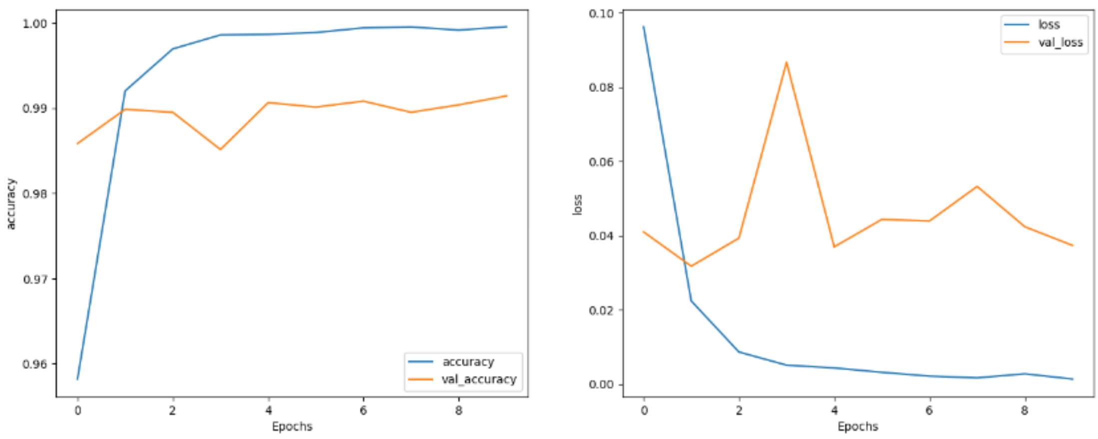

Figure 2. Accuracy and Loss for the Proposed Model Text

The graph analysis gives the accuracy and loss being plotted for both the training and validation sets. Ideally, the accuracy increases over time (epochs) during training which suggests the model is learning the patterns in the data and improving its ability to make correct predictions. A rapid increase in accuracy or decreases in loss might suggest overfitting. The low learning rate causes slow convergence (reaching optimal performance).

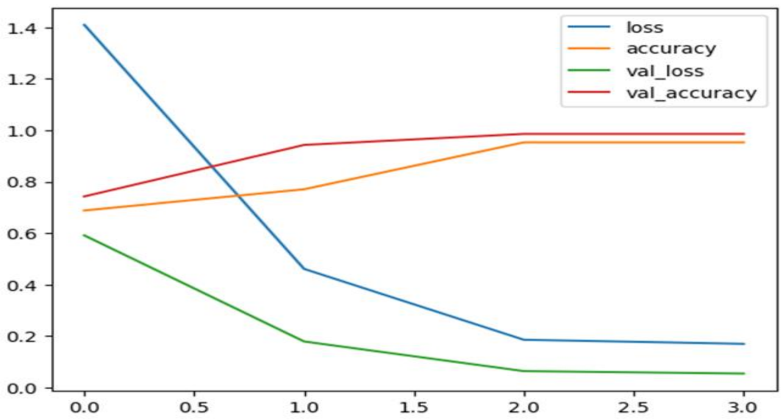

Figure 3. Accuracy and Loss for the Proposed Model Fake Image

The gap between the training and validation curves, a small and persistent gap is considerable else it is overfitting. Validation data is not used to train the model it’s used for evaluation only. Validation data can be used for hyperparameter tuning. Hyperparameters are settings that control the training process of a model learning rate, number of epochs.

Table 4. Performance Analysis of Existing Models with Proposed Model Using Text News Corpus

Parameters	CB-FAKE	MPFN	TEXT-GCN	BERT (Proposed Work)
Accuracy	0.93	0.83	0.52	0.91
Precision	0.87	0.85	0.48	0.86
Recall	0.81	0.88	0.49	0.89
F1 score	0.84	0.89	0.49	0.92

CB-Fake - CapsNet BERT – Fake, MPFN- modal progressive fusion network, TEXT-GCN - Text Graph Convolutional Network

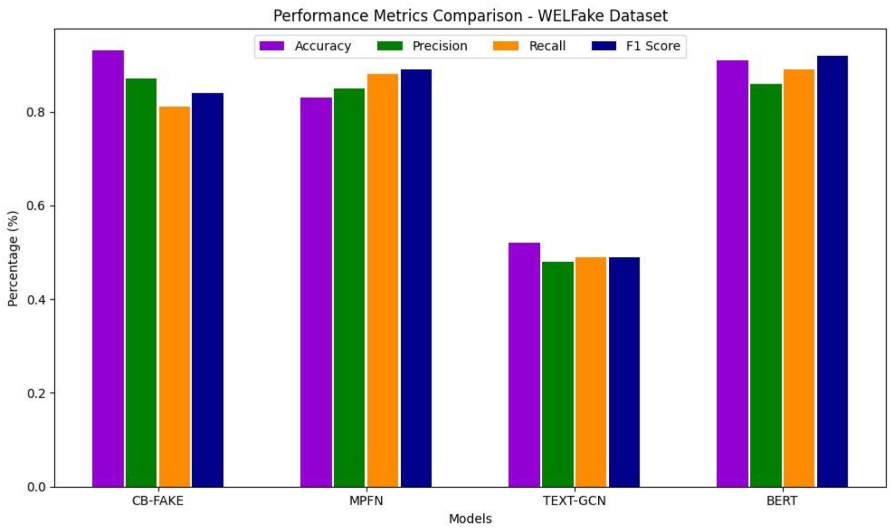

Figure. 4 Performance Metrics Comparison – WELFake Dataset

BERT achieves the highest average score across all metrics (Accuracy, Precision, Recall, F1-score), suggesting a more balanced performance in terms of correctly identifying relevant data points (Recall) while minimizing false positives (Precision). Based on Accuracy, CB-

FAKE achieves the highest score (0.93), indicating it correctly classifies the highest proportion of text data points.

Table 5. Performance Analysis of Existing Models with Proposed Model Using Image News Corpus

Parameters	GRU	MVAE	EANN	CNN (Proposed)
Accuracy	0.63	0.74	0.64	0.80
Precision	0.58	0.80	0.81	0.76
Recall	0.81	0.71	0.49	0.83
F1 score	0.67	0.75	0.61	0.79

EANN- event adversarial neural network, MVAE-Multimodal variational autoencoder, GRU-Gated Recurrent Unit

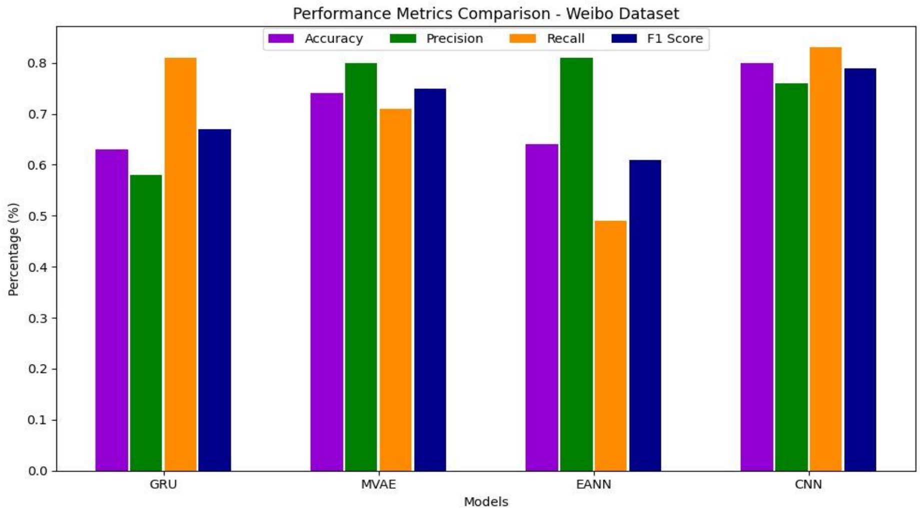

Figure 5. Performance Metrics Comparison – Weibo Dataset

CNN achieves the highest Accuracy (0.80) and F1-score (0.79), indicating it correctly classifies the highest proportion of data points and offers a good balance between precision and recall. CNN emerges as the strongest performer in this comparison based on Accuracy, F1-score, and a more balanced performance across precision and recall.

6. Limitations

Limited Multimodal Fusion exploring diverse deep learning models for multimodal fusion. utilization of synthetic fake news data for more effective multi-modal solutions. Lack of System excellence no single system excelled across all categories, indicating the multifaceted challenges in multi-modal fact verification. The scarcity of labeled data poses a significant challenge in training and evaluating fake news detection models. This limitation can lead to biased models and hinder the generalizability of the system. The lack of transparency and interpretability in existing models can undermine trust in the system and limit its adoption. Understanding how the model makes decisions is crucial for building trust and ensuring transparency in the domain of fake news detection**.**

7. Conclusion

The development of a multi-modal fake news detection system is a complex task that involves integrating deep learning, natural language processing. The project's main modules include data collection and preprocessing, multi-modal model development, adversarial attack simulation, and model evaluation and validation. These modules collectively aim to develop a robust system for detecting fake news across different modalities. The project addresses the pervasive issue of fake news by leveraging cutting-edge technologies to uphold the integrity of digital discourse and fortify information authenticity. The integration of user context analysis, adversarial attack simulation, and model transparency and interpretability are crucial aspects, ensuring the system's adaptability, robustness, and trustworthiness in combating the proliferation of fake news.

8. Future Enhancement

In the future, different deep learning models for fusing textual and visual features can be investigated to better understand the relationship between different modalities to recognize fake news. The proposed model analyzes English political fake news datasets for identifying if the news is fake or not, but it can be extended to include other popular languages fake news datasets. Fusion of algorithms can also be done in future for much better results.

References

[1]Kaliyar, Rohit Kumar, Anurag Goswami, and Pratik Narang. "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach." Multimedia tools and applications 80, no. 8 (2021): 11765-11788.

[2]Suryavardan, S., Shreyash Mishra, Megha Chakraborty, Parth Patwa, Anku Rani, Aman Chadha, Aishwarya Reganti et al. "Findings of factify 2: multimodal fake news detection." arXiv preprint arXiv:2307.10475 (2023).

[3]Balasubramanian Palani, Sivasankar Elango, Vignesh Viswanathan, “CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT”, Multimedia Tools and Applications 2022 Volume 81 pp 5587–5620.

[4]Raza, Shaina, and Chen Ding. "Fake news detection based on news content and social contexts: a transformer-based approach." International Journal of Data Science and Analytics 13, no. 4 (2022): 335-362.

[5]Nadeem, Muhammad Imran, Syed Agha Hassnain Mohsan, Kanwal Ahmed, Dun Li, Zhiyun Zheng, Muhammad Shafiq, Faten Khalid Karim, and Samih M. Mostafa. "Hyprobert: A fake news detection model based on deep hypercontext." Symmetry 15, no. 2 (2023): 296.

[6]Jing Jing, Hongchen Wu, Jie Sun, Xiaochang Fang, Huaxiang Zhang “Multimodal fake news detection via progressive fusion networks” Information Processing and Management 60 (2023) 103120.

[7]Ahmed Hashim Jawad Almarashy, Mohammad-Reza Feizi-Derakhshi (2023), “Enhancing Fake News Detection by Multi-Feature Classification” 2023 IEEE Access.DOI: 10.1109/ACCESS.2023.3339621.

[8]Guobiao Zhang, Anastasia Giachanou, Paolo Rosso, “SceneFND: Multimodal fake news detection by modelling scene context information”, Journal of Information Science 2022 1–13.

[9]Mohit Mayank Outplay, Shakshi Sharma, Rajesh Sharma, “DEAP-FAKED: Knowledge Graph based Approach for Fake News Detection”. Journal on Interactive Systems, 2022 arXiv:2107.10648v3 [cs.CL]

[10] Zhang, X. Y., & Ghorbani, A. A. "An Overview of Online Fake News: Characterization, Detection, and Discussion." Information Processing & Management, 2020 57(2), 102025.

[11] Zellers, R., Holtzman, A., Bisk, Y., Farhadi, A., & Choi, Y. "Defending Against Neural Fake News." Advances in Neural Information Processing Systems 2019, 32, 9054-9065.

[12] Essa, Ehab, Karima Omar, and Ali Alqahtani. "Fake news detection based on a hybrid BERT and LightGBM models." Complex & Intelligent Systems 9, no. 6 (2023): 6581-6592.

[13] Balshetwar, Sarita V., and Abilash Rs. "Fake news detection in social media based on sentiment analysis using classifier techniques." Multimedia tools and applications 82, no. 23 (2023): 35781-35811.

[14] Sharma, Uma, Sidarth Saran, and Shankar M. Patil. "Fake news detection using machine learning algorithms." International Journal of creative research thoughts (IJCRT) 8, no. 6 (2020): 509-518.

[15] Jamal Abdul Nasir, Osama Subhani Khan, Iraklis Varlamis, “ Fake news detection: A hybrid CNN-RNN based deep learning approach” International Journal of Information Management Data Insights Volume 1, Issue 1, April 2021, 100007.

Author’s Biography

Dr. S. Saraswathi is Professor in the Department of Information Technology, Puducherry Technological University. She completed her PhD, in the area of speech recognition for Tamil language. Her areas of interest are Natural Language Processing, Design of Intelligent Systems and Speech Processing.

Email ID: s.saraswathi@ptuniv.edu.in

S. Jeevithaa, a B.Tech (IT) student at Puducherry Technological University, passionate on data science. Her areas of interest is in Python, database management, and various machine learning tools, all geared towards tackling complex problems and crafting innovative solutions.

Email ID: jeevithaa.s@pec.edu

K Vishwabharathy, a B.Tech (IT) student at Puducherry Technological University, is deeply involved in exploring machine learning and deep learning. Proficient in Python, Java, database management, and machine learning tools, his adept at implementing forward-thinking solutions, particularly in machine learning projects, showcasing exceptional problem-solving skills in crafting intricate ML models.

Email ID: vishwabharathy.k@pec.edu

D. Eyuvaraj, a B.Tech (IT) student at Puducherry Technological University, is deeply involved in his exploration of machine learning. Proficient in Python, Flask, Fast API, database management, and Machine Learning tools. His endeavors are directed towards implementing forward-thinking solutions, particularly in the realm of machine learning projects. His expertise lies in crafting ML models.

Email ID: eyuvaraj.d@pec.edu