DETECTING FAKE POSTS IN SOCIAL MEDIA USING MACHINE LEARNING AND DEEP LEARNING

Damarasingi Sai Teja[1]

[1]Computer Science and Engineering, GMR Institute of Technology Rajam,532127, India.

ABSTRACT

Online reviews are crucial for consumer decision-making, yet the prevalence of fake reviews undermines their reliability. This study explores the application of machine learning techniques to detect fake reviews, focusing on Natural Language Processing (NLP) and deep learning models like Long Short Term Memory (LSTM) networks. The study evaluates these methods based on accuracy, efficiency, and scalability, and compares the trade-offs between simpler and more complex models. Challenges such as obtaining quality training data and adapting to evolving fake review tactics are also examined. To address

these issues, the study proposes and tests hybrid detection systems that integrate multiple techniques. The goal is to enhance the integrity and trustworthiness of online review platforms, making them more reliable resources for consumers.

Keywords: Fake news detection, Machine learning, Natural Language Processing (NLP), Deep learning, LSTM, Hybrid models, Online review integrity

1. INTRODUCTION

Social media has become a major platform for information sharing, but it is also plagued by the widespread dissemination of fake posts, which can mislead users and cause significant societal harm. Detecting these fake posts is a complex challenge, as they often mimic legitimate content, making traditional moderation methods insufficient. As a result, the need for automated and intelligent systems to identify and filter out false information is more urgent than ever. This research explores the application of machine learning techniques to detect fake posts on social media. By utilizing Natural Language Processing (NLP) and deep learning models, such as Long Short-Term Memory (LSTM) networks, the study focuses on improving detection accuracy and scalability. Additionally, the study addresses challenges like the quality of training data and the evolving nature of misinformation tactics. The proposed system integrates multiple machine learning approaches to enhance detection capabilities, ultimately aiming to curb the spread of misinformation and maintain the reliability of social media platforms. The goal of this research is to develop an effective machine learning system for accurately detecting fake posts on social media, enhancing the reliability of information shared online.

2. RELATED WORK

2.1 Alnabhan, Mohammad Q., and Paula Branco. "Fake News Detection Using Deep Learning: A Systematic Literature Review." IEEE Access (2024).

The objectives of this research involve using deep learning algorithms to analyze and classify fake news, with a focus on evaluating their effectiveness.

Publicly available datasets for fake news detection are also analyzed, with particular attention to their characteristics and suitability. Additionally, the study highlights key challenges and research gaps in transfer learning and class imbalance. The methodology includes deep learning algorithms, transfer learning, class imbalance handling, and data extraction and analysis. Advantages of this approach include identifying gaps in current research, applying transfer learning, and providing a detailed analysis of deep learning methods. However, there are disadvantages, such as the limited scope, dependence on available datasets, and the complexity of implementation.

2.2 Sudhakar, M., and K. P. Kaliyamurthie. "Detection of fake news from social media using support vector machine learning algorithms." Measurement: Sensors, vol. 32, 2024, p. 101028.

The objectives of this study are to evaluate the effectiveness of various machine learning and deep learning classifiers in detecting fake news, investigate the spread and impact of misinformation during the COVID-19 pandemic using a large dataset of tweets, and improve prediction rates for fake news classification by employing advanced algorithms. The methodology involves implementing machine learning and deep learning models for training and testing using a comprehensive dataset of 1,375,592 tweets related to COVID-19. Advantages include high accuracy, wide applicability, diverse techniques, and a robust framework, while disadvantages involve computational demands, the risk of false positives, and dependence on data quality.

2.3 Khan, Uroosa, and Mohd Nazim. "Fake News Detection over the Social Media by using Machine Learning Techniques: A Systematic Literature Review." Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing. 2023.

The objectives of this study are to identify machine learning techniques for detecting fake news on social media, evaluate their effectiveness, and determine the most common social media platforms where fake news spreads. Additionally, the study aims to highlight research gaps for future studies. The methodology involves conducting a systematic literature review, data collection, quality assessment, and data synthesis. Advantages include providing a comprehensive overview of detection techniques, identifying gaps in the existing literature, offering practical insights, and raising awareness. Disadvantages include high computational requirements, limited adaptability to new tactics, and interpretability issues in complex models**.**

2.4 Cheng, Li-Chen, et al. "Detecting fake reviewers from the social context with a graph neural network method." Decision Support Systems 179 (2024): 114150.

The objectives of this study are to develop a framework for detecting spammers, utilize a social context network to capture user interactions for detecting fake reviews, and introduce a novel two-stage architecture for data classification to address the issue of imbalance. The methodology involves the use of Graph Neural Networks (GNN), subgraphs, and a two-stage architecture with Focal Loss. Advantages include combining social interaction data with review content, enhancements through Focal Loss, and evaluation using real-world data. Disadvantages include the complexity of the model, high computational resource requirements for GNN, and dependence on the quality of training data**.**

2.5 Chang, Qian, Xia Li, and Zhao Duan. "Graph global attention network with memory: A deep learning approach for fake news detection." Neural Networks 172 (2024): 106115.

The objectives of this study are to develop a novel GANM to enhance fake news detection on social media, leverage deep learning and NLP techniques to analyze news context and user content, and improve the identification of fake news by integrating structural and contextual information from social networks. The methodology involves a Graph Global Attention Module with Memory, Partial Key Message Learning, and the use of three graph convolutional networks (GCNs) to extract features. Advantages include combining global and local information, efficiently processing multimodal data, and demonstrating promising performance on real-world datasets. Disadvantages include the need for significant computational resources, difficulties in adapting to new fake news tactics, and reliance on high-quality data for effective model training.

2.6 Visweswaran, M., et al. "Synergistic Detection of Multimodal Fake News Leveraging TextGCN and Vision Transformer." Procedia Computer Science 235 (2024): 142-151.

The objectives of this study are to develop a multimodal fake news detection system to classify Reddit posts as fake or true using both text and image content, improve classification performance despite class imbalances in the dataset, and analyze the propagation patterns of misinformation to understand how fake news spreads. The methodology involves utilizing the Multimodal Fakeddit Dataset, feature extraction, a fusion model, and testing models like SVC, RF, ANN, and dimensionality mapping. Advantages include achieving high accuracy of 94.17% for binary classification, effective multimodal integration, and comprehensive misinformation analysis. Disadvantages include the potential impact of class imbalance on performance, increased model complexity, and risks of overfitting.

2.7 Hashmi, Ehtesham, et al. "Advancing fake news detection: hybrid deep learning with fasttext and explainable AI." IEEE Access (2024).

The objectives of this study are to improve fake news detection accuracy and adaptability using machine learning (ML), deep learning (DL), and transformer-based models, utilize supervised and unsupervised FastText embeddings to efficiently handle subword and out-of-vocabulary (OOV) words, enhance model transparency and interpretability through Explainable AI, and leverage multiple datasets to improve model performance across different types of fake news. Advantages include improved accuracy, efficient handling of OOV words, scalability, enhanced explainability, and the use of diverse datasets. Disadvantages include increased model complexity, challenges with generalizability, potential bias and subjectivity, and difficulties in contextual understanding.

2.8 Luqman, Muhammad, et al. "Utilizing ensemble learning for detecting multi-modal fake news." IEEE Access (2024).

The objectives of this study are to accurately classify news articles as fake or true by leveraging multimodal data, utilize sentiment analysis to assess the influence of sentiment in identifying fake news, and demonstrate the significance of preprocessing in enhancing classification performance. The study also aims to compare the proposed method with existing models. The methodology involves data collection, preprocessing with NLP, sentiment analysis, feature extraction, ensemble learning, and evaluation metrics. Advantages include high accuracy, effective multimodal

INTERNATIONAL JOURNAL OF PROGRESSIVE
RESEARCH IN ENGINEERING MANAGEMENT
AND SCIENCE (IJPREMS)

(Int Peer Reviewed Journal)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 11, November 2024, pp : 2827-2833

e-ISSN :
2583-1062

Impact
Factor :
7.001

integration, the impact of sentiment analysis, robust preprocessing, and the benefits of an ensemble approach. Disadvantages include model complexity, dependency on data quality, the risk of overfitting, limitations of sentiment analysis, and challenges related to class imbalance.

2.9 Goyal, Bharti, et al. "Detection of fake accounts on social media using multimodal data with deep learning." IEEE Transactions on Computational Social Systems (2023).

The objectives of this study are to accurately detect fake accounts by reliably identifying fake or manipulated social media profiles, enhance multimodal feature extraction to improve the accuracy of account classification and authenticity assessment, and develop a flexible architecture capable of adapting to different social media contexts while maintaining performance. The methodology involves utilizing deep learning approaches such as CNN, LSTM, and ANN, along with feature fusion and normalization techniques. Advantages include comprehensive data analysis, an adaptive and flexible architecture, increased classification accuracy, and scalability. Disadvantages include high computational demand, dependency on data quality, the risk of overfitting, and ethical and privacy concerns.

2.10 Zhang, Dong, et al. "A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information." Decision Support Systems 166 (2023): 113911.

The objectives of this study are to improve the detection of fake reviewers by using a behavior-sensitive feature extractor and contextualized text representations, and to evaluate the proposed method's performance against state-of-the-art benchmarks using the YelpZIP dataset. The methodology includes creating a data testbed, data preprocessing, utilizing a behavior-sensitive feature extractor, feature assessment, and evaluating metrics such as precision, recall, F1-score, and AUC. Advantages include superior performance with an F1-score of 0.7664 and AUC of 0.8358, behavior-sensitive detection, and context-aware attention mechanisms. Disadvantages include the potential for bias introduction, the impact of data quality, longer training times, and variability in performance across different platforms.

2.11 Truică, Ciprian-Octavian, Elena-Simona Apostol, and Panagiotis Karras. "DANES: Deep neural network ensemble architecture for social and textual context-aware fake news detection." Knowledge-Based Systems 294 (2024): 111715.

The objectives of this study are to develop a context-aware model that combines both social and textual data for detecting fake news and to investigate the effectiveness of these features in enhancing fake news detection, particularly in low-data settings. The methodology involves conducting experiments on the BuzzFace, Twitter15, and Twitter16 datasets using GRU, LSTM, and CNN architectures, along with ablation testing to evaluate performance. Advantages include the combination of social and textual content, making the model practical for real-world applications, and its flexibility. However, the study faces disadvantages such as high computational complexity and an insufficient amount of social and user interaction data.

2.12 Altheneyan, Alaa, and Aseel Alhadlaq. "Big data ML-based fake news detection using distributed learning." IEEE Access 11 (2023): 29447-29463.

The objectives of this study are to address the detection of fake news on Twitter, evaluate state-of-the-art methods using big data technology (Spark) and machine learning, and propose a stacked ensemble classification model to improve accuracy with the FNC-1 dataset. The methodology uses the FNC-1 dataset with classes like agree, disagree, discuss, and unrelated, employing feature extraction techniques such as N-grams, Hashing TF-IDF, and count vectorizer. A stacked ensemble model was implemented on a decentralized Spark cluster using Random Forest, Logistic Regression, and Decision Trees, with performance evaluated using precision, recall, accuracy, and F1 score. Advantages include an F1 score of 92.45%, a 9.35% improvement over the baseline, scalability through Spark, and enhanced performance via ensemble learning. Disadvantages involve high computational costs from feature extraction integration, slower training times due to Spark's setup, and a focus on text-based detection that overlooks multimedia forms of fake news.

2.13 Devarajan, Ganesh Gopal, et al. "AI-assisted deep NLP-based approach for prediction of fake news from social media users." IEEE Transactions on Computational Social Systems (2023).

The objectives of this study are to develop an AI-assisted deep NLP approach for detecting fake news across social media, propose a four-layered architecture using deep learning techniques (CNN, Bi-LSTM), and verify user and publisher credibility through multiheaded attention mechanisms. The methodology involves a four-layer architecture (Publisher, Social Media, Edge, Cloud) utilizing a Deep CNN Bi-LSTM model (N-DCBL) for detection and a multiheaded attention network (DCBMHA) for credibility verification. Datasets include Buzzface, FakeNewsNet, and Twitter, with preprocessing steps such as data cleaning, stop-word removal, word segmentation, and word embedding (TF-IDF, GloVe). Advantages include achieving up to 99.72% accuracy and a 98.33% F1-score, credibility checks for users and publishers, and a distributed architecture for efficient detection and validation. Disadvantages include increased computational costs due to model complexity, potential overfitting from high-dimensional sparse matrices,

and a limited focus on text-based detection with minimal multimedia integration.

2.14 Kondamudi, Medeswara Rao, et al. "A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches." Journal of King Saud University-Computer and Information Sciences 35.6 (2023): 101571.

This study provides a comprehensive survey of fake news on social networks, analyzing its attributes, features, and propagation mechanisms. It reviews various detection techniques, emphasizing machine learning and deep learning methods while categorizing fake news based on content, social context, and creators. The methodology explores multiple classification techniques, including machine learning, deep learning, and geometric deep learning, and examines datasets like FakeNewsNet and LIAR for effective detection. Key advantages of this survey include its thorough coverage of linguistic and social features, the presentation of diverse detection methods-ranging from manual to automated fact-checking-and insights into contemporary techniques. However, it also highlights certain limitations, such as a lack of practical implementation of new detection models, scalability issues with manual fact-checking in large datasets, and minimal focus on real-time detection of rapidly spreading fake news.

2.15 Uppada, Santosh Kumar, and Parth Patel. "An image and text-based multimodal model for detecting fake news in OSN's." Journal of Intelligent Information Systems 61.2 (2023): 367-393.

The objectives of this study are A novel framework that integrates both image and text data to enhance fake news detection on social media platforms, utilizing the extensive Fakeddit dataset comprising over one million samples. The methodology employs advanced architectures, featuring a BERT-based text model and various deep learning image models like Xception and ResNet50, alongside Error Level Analysis (ELA) for identifying manipulated images. This multimodal approach achieves an impressive accuracy of 91.94%, significantly outperforming unimodal models, while precision, recall, and F1-score metrics also reflect strong performance at 93.43%, 93.07%, and 93%, respectively. However, the model faces challenges in visual sentiment analysis due to a limited dataset for sentiment detection, and its reliance on human-annotated data raises scalability concerns. Additionally, the framework may struggle to generalize across different domains and languages, and it does not leverage metadata or user engagement data, highlighting areas for further research.

3. METHODOLOGY
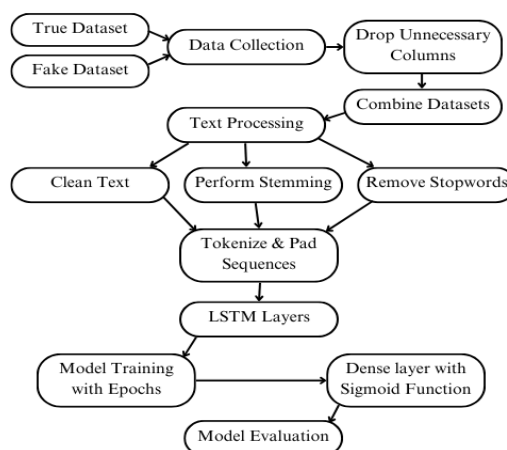
1. Data Preparation

- Merged fake.csv and true.csv datasets, assigning labels (0 for fake, 1 for true).

- Removed irrelevant columns (date, subject) and handled missing values.

2. Text Preprocessing

- Cleaned text by removing non-alphabetical characters, converting to lowercase, and tokenizing.

- Applied stopword removal and stemming using the Porter Stemmer.

3. Feature Engineering

- Encoded text into numerical format using one-hot encoding.

- Padded sequences to ensure uniform input length.

4. Model Design

- Built a keras-based sequential model with:

- Embedding Layer for vector representation.

- LSTM Layer for sequence learning.

- Dropout Layer to prevent overfitting.

- Dense Layer with sigmoid activation for classification.

5. Training and Testing

- Split data into training and testing sets.

- Used metrics like accuracy, precision, recall, and F1-score for evaluation.

4. RESULTS

4.1 Classification Accuracy

• **LSTM Model**: Achieved an accuracy of **93.62%** in classifying fake and real posts, showing reliable performance for text-based fake news detection.

• **Multimodal Model (Reference 1)**: Reached a higher accuracy of **92.17%** by integrating both text and image data, leveraging a TextGCN and Vision Transformer. However, this approach demands significantly more computational resources.

• **SVM (Reference 2)**: Achieved high accuracy with a Support Vector Machine classifier on a COVID-19 Twitter dataset, using TF-IDF for text representation. Although effective, this model lacks the sequential context understanding that LSTM provides.

• **Explainable AI Model (Reference 3)**: Achieved high accuracy with a hybrid deep learning model and explainable AI but at the cost of increased complexity, making it less efficient for real-time applications.

4.2 Computational Efficiency

• **LSTM Model**: Lower computational demands compared to multimodal and ensemble models, making it suitable for real-time social media monitoring. Dropout layers in the LSTM model reduce overfitting, supporting reliable performance on unseen data.

• **Multimodal Model**: The distributed learning model employed by Altheneyan & Alhadlaq (2023) leverages Spark, which increases scalability but at the expense of higher computational costs.

• **Traditional SVM and TF-IDF**: Although SVM and TF-IDF are computationally efficient, they may miss the deeper contextual dependencies that LSTM models capture, particularly in nuanced fake news content.

4.3 Adaptability and Robustness

• **LSTM Model**: Effective in adapting to varied text data and evolving linguistic patterns, making it robust for general-purpose fake news detection. However, being limited to text-only input restricts its potential in multimodal contexts.

• **Multimodal Model**: Highly adaptable due to its use of multimodal inputs, allowing it to capture more features from various content types. Yet, the complexity limits its real-time applicability.

• **Explainable AI Model**: Improved interpretability with explainable AI, but added model complexity and potential bias reduce its adaptability for broad social media content.

Confusion Matrix:

Breakdown of prediction results to display the counts of True Positives, False Positives, True Negatives, and False Negatives.
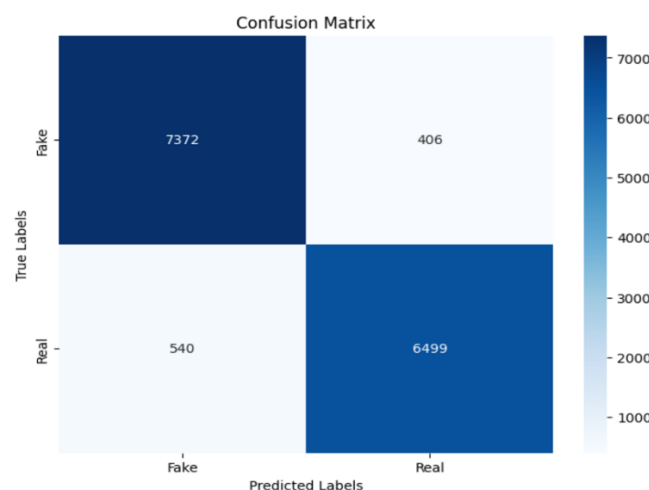

Fig. 4.1

ROC Curve for True Positive and False Positive Rates

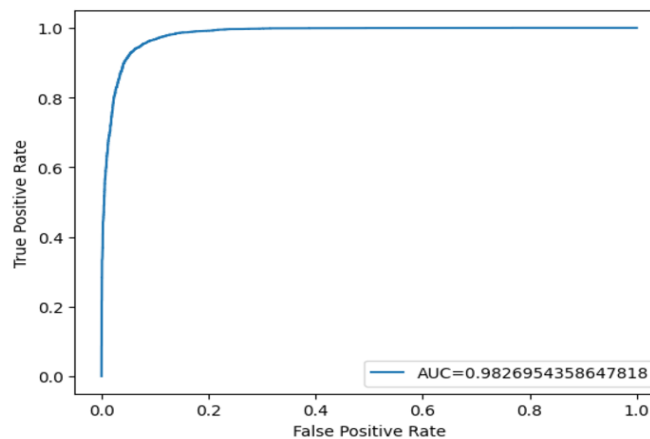Evaluates the model's ability to distinguish between classes (fake vs. real)

Fig. 4.2

Model Accuracy vs. Epochs:

This plot helps evaluate how well the model is learning over the specified number of epochs.
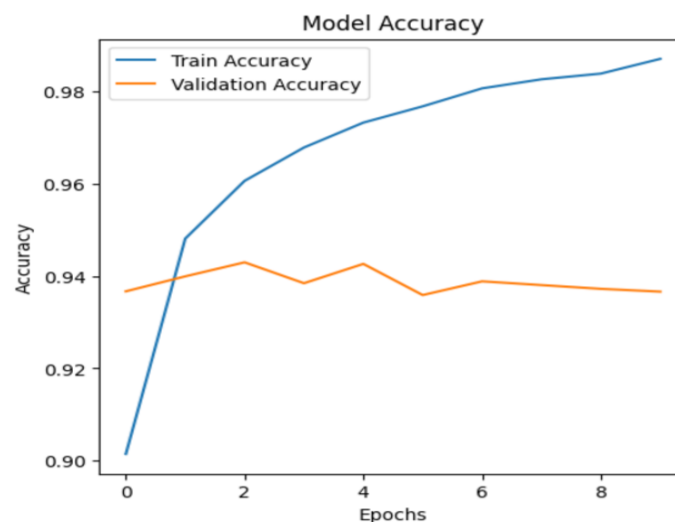

Fig. 4.3

5. CONCLUSION

The study highlights the potential of deep learning techniques, particularly LSTM networks, in addressing the challenge of fake news detection with a high degree of accuracy. By utilizing a preprocessed dataset with tokenization, stopword removal, and padding, the model effectively captures the underlying patterns in the text data. The integration of LSTM layers allows the model to learn sequential dependencies, which is crucial for understanding the context of news articles. The results underscore the importance of effective preprocessing and sequential modeling for capturing textual nuances. The strong accuracy achieved on the test data demonstrates the model's reliability in distinguishing between genuine and fake news. Furthermore, the approach demonstrates scalability, suggesting its applicability to diverse datasets and languages. Future work could focus on improving accuracy further by expanding the dataset, optimizing the model, and incorporating multimodal data sources to enhance robustness and generalization. Future work could also explore real-time detection systems to further strengthen its practical impact.

6. REFERENCES

[1] Alnabhan, Mohammad Q., and Paula Branco. "Fake News Detection Using Deep Learning: A Systematic Literature Review." IEEE Access (2024).

[2] Sudhakar, M., and K. P. Kaliyamurthie. "Detection of fake news from social media using support vector machine learning algorithms." Measurement: Sensors, vol. 32, 2024, p. 101028.

[3] Khan, Uroosa, and Mohd Nazim. "Fake News Detection over the Social Media by using Machine Learning Techniques: A Systematic Literature Review." Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing. 2023.

[4] Cheng, Li-Chen, et al. "Detecting fake reviewers from the social context with a graph neural network method." Decision Support Systems 179 (2024): 114150.

[5] Chang, Qian, Xia Li, and Zhao Duan. "Graph global attention network with memory: A deep learning approach for fake news detection." Neural Networks 172 (2024): 106115.

[6] [6] Visweswaran, M., et al. "Synergistic Detection of Multimodal Fake News Leveraging TextGCN and Vision Transformer." Procedia Computer Science 235 (2024): 142-151.

[7] Hashmi, Ehtesham, et al. "Advancing fake news detection: hybrid deep learning with fasttext and explainable AI." IEEE Access (2024).

[8] Luqman, Muhammad, et al. "Utilizing ensemble learning for detecting multi-modal fake news." IEEE Access (2024).

[9] Goyal, Bharti, et al. "Detection of fake accounts on social media using multimodal data with deep learning." IEEE Transactions on Computational Social Systems (2023).

[10] Zhang, Dong, et al. "A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information." Decision Support Systems 166 (2023): 113911.

[11] Zhang, Dong, et al. "A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information." Decision Support Systems 166 (2023): 113911.

[12] Altheneyan, Alaa, and Aseel Alhadlaq. "Big data ML-based fake news detection using distributed learning." IEEE Access 11 (2023): 29447-29463.

[13] Devarajan, Ganesh Gopal, et al. "AI-assisted deep NLP-based approach for prediction of fake news from social media users." IEEE Transactions on Computational Social Systems (2023).

[14] Kondamudi, Medeswara Rao, et al. "A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches." Journal of King Saud University-Computer and Information Sciences 35.6 (2023): 101571.

[15] Uppada, Santosh Kumar, and Parth Patel. "An image and text-based multimodal model for detecting fake news in OSN's." Journal of Intelligent Information Systems 61.2 (2023): 367-393.