[1]Ahmed E. Eltoukhy
[2]Saad M. Darwish
[3]Mostafa R. Kaseb

# An Advanced Framework for Fake News Detection: Integrating Feature Extraction and Optimal Selection to Combat Misinformation

Journal of Electrical Systems

***Abstract:*** - The proliferation of fake news on digital platforms presents significant challenges to societal well-being, making robust detection mechanisms essential. This paper introduces an advanced fake news detection framework that integrates feature extraction and optimal feature selection to enhance model performance. The proposed approach systematically combines various features—content-based, style-based, source-based, and social context attributes—to create a nuanced identification process. To preprocess the textual data, rigorous cleaning procedures, including noise removal and normalization, are applied, followed by word embedding techniques to capture contextual understanding. Optimal feature selection is then performed using Recursive Feature Elimination methods (RFE). Statistical tests are utilized to identify the most relevant features from the extracted set. This dual-stage process of feature extraction followed by optimal feature selection not only reduces dimensionality but also enhances model robustness and mitigates overfitting. A neural network machine learning algorithm is then trained and evaluated on the refined feature set. Model performance is rigorously assessed using various metrics. Experimental results demonstrate the efficacy of the proposed model, highlighting the critical role of both feature extraction and optimal feature selection in developing reliable detection systems. This research contributes to ongoing efforts to combat misinformation, providing a foundation for more accurate and trustworthy news dissemination systems.

***Keywords:*** Fake News; Feature Extraction; Feature Selection; NLP; Machine Learning; Classification.

## I. INTRODUCTION

Fake news is not a recent phenomenon; it existed before Christ (BC). It is becoming widespread in newspaper articles, television shows, and some other traditional media [1]. Following that, the explosion of the World Wide Web in the mid-1990s significantly developed. Due to its convenience, quick speed, and low cost, social media has emerged as the primary medium for online human interaction and the transmission of information (see Fig. 1) [1]. Social media is flooded with millions of posts about news in different fields. The use of social media is increasing with the time. In 2024, there are over 5.17 billion users on social media (about 63% of the world population), and by 2028, it is expected that there will be around six billion users [2]. Social media has brought us many benefits like, faster and easier communication, brand promotions, customer feedback, etc.; however, it also has several drawbacks, and one of the prominent ones being fake news. Fake news is unarguably a threat to the society, it can cause damage to large corporations, stock markets, businesses, etc., and harm the people [3]. Also, it has become a challenging problem for social media users and researchers alike. A tremendous amount of fake news has been quickly and accurately disseminated over online social network, as shown in Fig. 2.

[1] *Corresponding author: Ahmed E. Eltoukhy, Computer Science Department, Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum, Egypt; Computer Science Department, Higher Institute of Management and Information Technology, KFS, Egypt.

[2] Information Technology Department, Institute of Graduate studies and Research, Alexandria University, Alexandria, Egypt;

[3] Computer Science Department, Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum, Egypt;

Emails: [1] ai1547@fayoum.edu.eg ; [2] saad.darwish@alexu.edu.eg ; [3] mrk00@fayoum.edu.eg

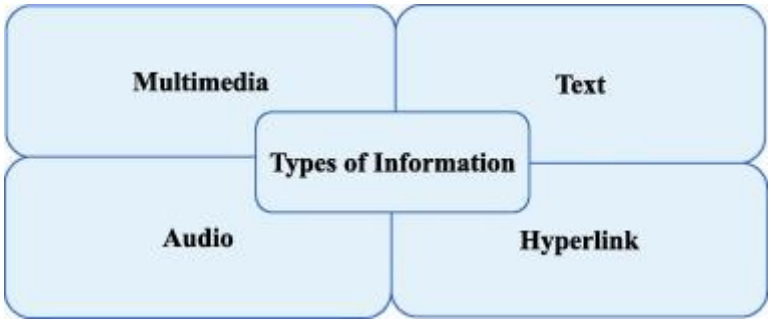


**Fig. 1.** Types of information on social media [1]

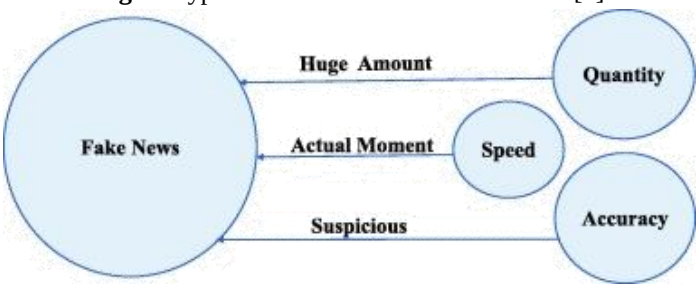


**Fig. 2.** Fake news quantity, accuracy, and speed [1]

Fake news detection involves analyzing various aspects of news articles, including textual content, social context, and source credibility, to differentiate between factual reporting and misleading information. Traditional approaches often rely on manual fact-checking or rule-based systems, which are limited in scalability and effectiveness. In contrast, the proposed fake news detection model leverages state-of-the-art feature extraction techniques to capture nuanced patterns and linguistic cues indicative of fake news. By extracting informative features such as linguistic, sentiment analysis, and social engagement metrics, the proposed system aims to discern subtle distinctions between authentic and fabricated news content. Furthermore, feature selection is pivotal in refining the model's predictive capabilities and reducing overfitting. Through rigorous statistical analyses such as Recursive Feature Elimination, the work prioritizes the most discriminative features while discarding irrelevant or redundant information. This ensures a streamlined and efficient model that generalizes well to unseen data.

The proposed approach embodies a data-driven and iterative methodology, continuously refining the feature extraction and selection process based on empirical insights and domain expertise. By integrating cutting-edge machine learning algorithms and domain-specific knowledge, the work strives to build a robust and scalable fake news detection system that empowers users to make informed judgments and combat misinformation effectively. In the subsequent sections, delve into the technical details of the proposed approach, including the feature extraction pipeline, feature selection strategies, model architecture, evaluation metrics, and experimental results. Through rigorous experimentation and validation, besides aiming to demonstrate the efficacy and practical utility of the proposed fake news detection approach.

The rest of the paper is arranged as follows. Section II discusses the most stat-of-the -art fake news detection approaches. The proposed model is presented and discussed in Section III. The experiment results and discussion are explained in Section IV, and the conclusion of the proposed work and the possible future work is presented in Section V.

## II. Related Work

The detection of fake news has been a focal point of research in recent years, driven by the rise of misinformation on social media and its potential impact on public opinion and societal trust. Various approaches and methodologies have been proposed and developed to tackle this challenge. This section reviews the related work in the field of fake news detection, highlighting significant contributions and methodologies. Content-based fake news detection systems primarily focus on analyzing the textual content of news articles. These methods leverage NLP techniques to extract meaningful features from the text. Fig. 3, illustrates the types of news content and its characterizations.

Linguistic and Stylistic Features: Early works explored the use of linguistic and stylistic features [4][5], such as part-of-speech tags (POS), syntactic structures, and readability scores, to identify deceptive language. These studies demonstrated that deceptive content often exhibits distinct linguistic patterns compared to truthful content. Machine Learning Classifiers: Traditional machine learning algorithms like Support Vector Machines (SVM), Naive Bayes, and Logistic Regression have been widely used for fake news detection. Authors in [6] employed a combination of Term Frequency-Inverse Document Frequency (TF-IDF) and n-gram features to train SVM classifiers, achieving notable performance in classifying fake news. Deep Learning Models: Recent advances in deep learning have further enhanced content-based detection. Authors in [7] introduced a hybrid model called CSI, which combines CNNs for text representation with RNNs for capturing temporal dynamics. Similarly, Authors in [8] utilized Long Short-Term Memory (LSTM) networks to analyze news content, showing improved detection accuracy.

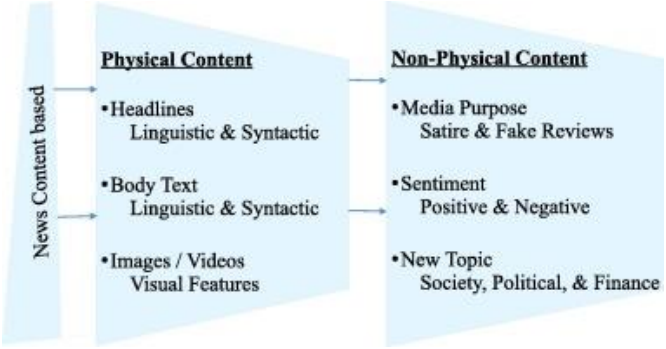


**Fig. 3.** Different types of news content based fake news on its characteristics [1]

Context-based approaches incorporate additional information beyond the textual content, such as social context, source credibility, and user behavior, to improve fake news detection. Fig. 4 illustrates the features of news and based on that the proposed model detect fake news on social media [1]. Social Network Analysis: authors explored the use of social network features, including user interactions, reposting patterns, and network topology, to identify fake news [9]. Their study highlighted the significance of examining how news propagates through social networks to detect misinformation. Source Credibility: Horne et al. focused on the credibility of news sources as a key factor in fake news detection [10]. The authors proposed features based on the historical reliability of sources, citation patterns, and the reputation of authors, demonstrating that credible sources are less likely to disseminate fake news. User Behavior: authors investigated user behavior on social media platforms to identify fake news [11]. besides analyzed engagement metrics, such as likes, shares, and comments, along with user credibility scores, showing that user behavior can provide valuable signals for detection.

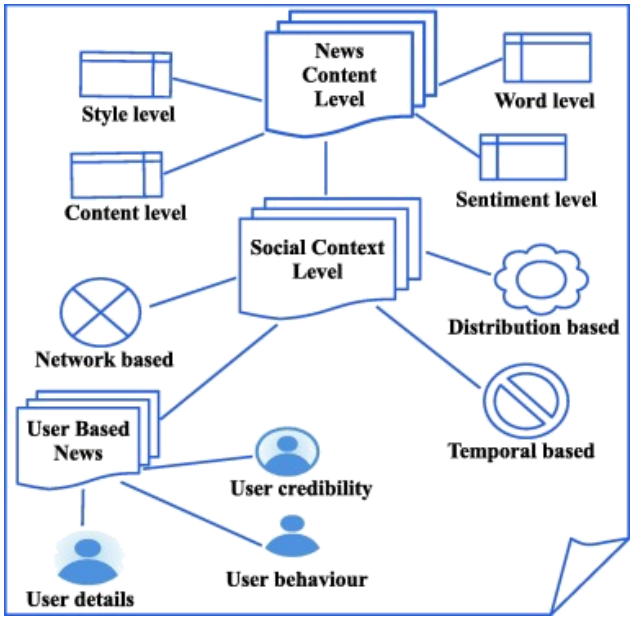


**Fig. 4.** Fake news detection based on its features [1]

Hybrid approaches combine content-based and context-based features to leverage the strengths of both methodologies, leading to more robust and accurate fake news detection systems. Feature Fusion; authors introduced a hybrid model that integrates linguistic features from the content with social context features from user interactions. By fusing these features, the model achieved improved performance compared to using content-based or context-based features alone [12]. Attention Mechanisms: authors utilized attention mechanisms in neural networks to dynamically weigh the importance of different features [13]. Their model combined textual features with user credibility scores, demonstrating that attention-based hybrid models can effectively capture the relevance of various features.

Effective feature selection is crucial for enhancing the performance of fake news detection systems by identifying the most informative features and reducing dimensionality. Recursive Feature Elimination (RFE): authors employed RFE to iteratively eliminate the least important features, refining the feature set and improving model performance [15]. This technique helped in selecting a subset of features that contribute the most to the prediction task. In conclusion, fake news detection has seen significant advancements through the development of content-based, context-based, and hybrid approaches. By leveraging sophisticated feature extraction, selection techniques, and robust machine learning models, researchers continue to improve the accuracy and reliability of fake news detection systems. However, despite significant advancements in fake news detection, existing systems still face several limitations that hinder their effectiveness and robustness. Especially, Hybrid approaches that combine content and context features face integration challenges as the diverse features from different domains (textual, social, and behavioral) can be complex and may not always lead to improved performance. Besides, the issues with interpretability, overfitting, and increased computational requirements.

## III. METHODOLOGY

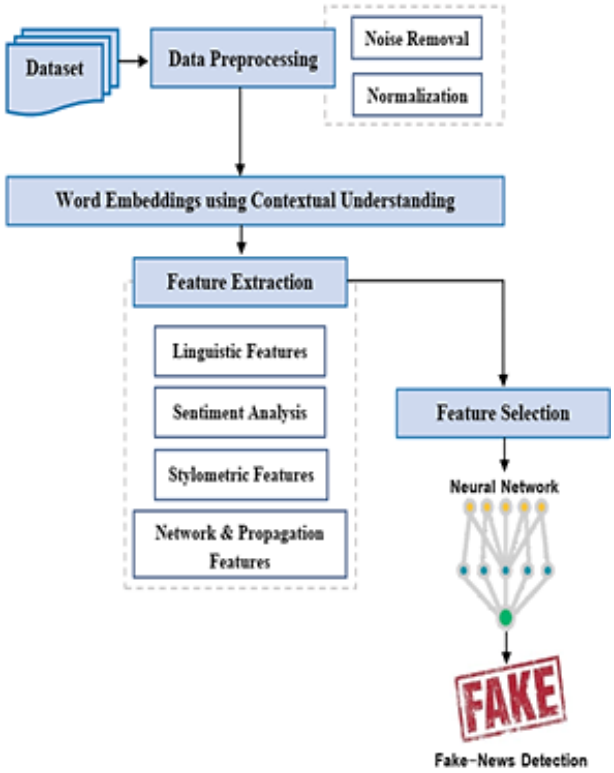


**Fig. 5.** The Proposed Fake New Detection Model

### *A. Dataset*

In the proposed model (see Fig. 5), the FakeNewsNet dataset has been used which is a comprehensive resource designed to facilitate the study and development of fake news detection algorithms [16]. This dataset is specifically curated to address the challenges posed by fake news on social media platforms and provides a rich set of features

encompassing content, social context, and network interactions. It is sourced from two well-known fake news datasets: Politifact, and GossipCop. Based on the available documentation, the approximate distribution is 432 for fake news articles, and 624 for real news articles in the Politifact. And is a 17,000 for both fake and real news articles in the GossipCop. So, the estimated total distribution is 17,432 for fake news articles, and 17,624 for real news articles. As a conclusion, the slight imbalance from the Politifact source is minor in comparison to the overall dataset size.

### *B. Preprocessing*

- *Noise Removal*

Noise Removal is an essential preprocessing step in the data collection phase that significantly enhances the quality and usability of the data for the fake news detection system. The primary goal of this step is to remove noise and irrelevant information from the raw text data, ensuring that the subsequent analysis is focused on meaningful and consistent content. The process begins with the removal of HTML tags, which are common in web-scraped data but do not contribute to the content's semantic value. URLs and special characters are also eliminated, as it often introduces unnecessary noise and do not add substantive information relevant to text analysis.

- *Normalization*

this step aiming to standardize the text data for consistent and accurate analysis. This process involves several techniques that transform the raw text into a more uniform format, making it easier for machine learning models to process and interpret the data. One of the first steps in normalization is tokenization, where the text is split into individual words or tokens. This helps in analyzing the text at a granular level and facilitates the extraction of meaningful features. Then lemmatization is applied to reduce words to their base or root forms.

Finally, preprocessing step not only improves the quality of the data but also ensures that the text is in a standardized form, which is crucial for accurate feature extraction and effective neural network model training in the fake news detection approaches. So, several tools and libraries are frequently used to implement these tasks such as NLTK (Natural Language Toolkit) which is a key library for data preprocessing in Python that provides a suite of libraries and tools for tokenization, normalization, and other NLP tasks [17][18].

### *C. Word Embeddings Using Contextual Understanding*

This step represents a significant advancement in NLP by capturing the meaning of words in their specific context. Unlike traditional embeddings that assign a fixed vector to each word, contextual embeddings generate dynamic representations that vary based on the surrounding text [19]. This is particularly powerful for fake news detection models, where the subtle nuances of language and context can play a crucial role in distinguishing between true and false information. In the proposed model, (BERT) Bidirectional Encoder Representations from Transformers is applied which is a state-of-the-art model for generating contextual word embeddings [20]. It captures the bidirectional context of words, meaning it considers both the preceding and succeeding words in a sentence to generate rich, context-aware embeddings. BERT has several advantages as it captures the context of each word, providing a deeper understanding of the text. Also, it allows it to consider the full context of a word within a sentence, improving the quality of embeddings. Besides, BERT is pre-trained on a large corpus of text, enabling it to capture general language patterns and nuances, which can be fine-tuned for specific tasks. By leveraging BERT's contextual embeddings, the fake news detection approach can achieve a higher level of accuracy and reliability, effectively identifying misleading information through nuanced text analysis.

### *D. Feature Extraction*

Feature extraction is a critical step in the pipeline of the fake news detection system. It involves transforming the cleaned and normalized text data into a set of measurable characteristics that can be used by machine learning models to distinguish between fake and real news [21]. This step leverages various techniques to capture the linguistic, semantic, Stylometric of the text, and social network and propagation, providing a rich set of features for accurate classification.

- *Linguistic Features*

Extract N-grams helps in capturing word sequences and co-occurrence patterns within the text. For instance, "breaking news" or "according to" might frequently appear in certain contexts and help in distinguishing the nature of the content. Perform Part-of-Speech (POS) Tagging analyzes the grammatical structure of sentences by tagging parts of speech (nouns, verbs, adjectives, etc.) which helps in understanding the syntactic roles of words. Fake news might exhibit specific syntactic patterns different from real news. Applying Named Entity Recognition (NER) classifies named entities (e.g., people, organizations, locations) which helps in analyzing the subjects of the news. The presence and frequency of certain entities can indicate potential biases or recurring fake news themes.

- *Semantic Feature*

Generating word embeddings captures semantic relationships between words, allowing the model to understand context and nuances in the text. Conduct sentiment analysis for assessing the sentiment of the text using tools like VADER or TextBlob to provide insights into the emotional tone (positive, negative, neutral). Fake news often employs sensationalism and emotional appeal, which can be detected through sentiment analysis. • Applying topic modeling using techniques like Latent Dirichlet Allocation (LDA) in order to identify the underlying topics within the text. This helps in understanding the broader themes and whether they align with typical fake news subjects.

- *Stylometric Features*

Calculate readability scores helps in assessing the complexity of the text. As the fake news may often be written in simpler language to appeal to a broader audience. Analyze writing style such as the use of punctuation, capitalization, and sentence length. Mostly, fake news articles might exhibit distinct stylistic patterns aimed at catching attention. Measuring the variety of vocabulary used in the text. Usually lower lexical diversity might indicate repetitive and simplistic language, which is often a characteristic of fake news.

- *Network and Propagation Features*

Conducting social network analysis by examining the propagation patterns of news articles on social media. Metrics like the speed of spread, the number of shares, likes, and comments can provide insights into the virality and potential manipulation of the content. Analyzing the credibility of the source based on historical accuracy, fact-checking scores, and domain reputation. Sources with a history of publishing fake news can be flagged and their content scrutinized more rigorously.

### *E. Feature Selection Module*

This paper introduces an advanced fake news detection framework that incorporates feature extraction and optimal feature selection using Recursive Feature Elimination (RFE) to enhance model performance. The proposed approach systematically combines various features—content-based, style-based, source-based, and social context attributes—to create a nuanced identification process. Optimal feature selection is performed using RFE, which iteratively removes the least important features based on a fake news detection model. RFE helps in identifying the most relevant features by recursively considering smaller sets of features and ranking them based on their importance. The main steps of RFE is:

1. Initial Model Training: RFE begins by training a fake news detection model on the entire feature set.
2. Importance Ranking: After training, the model assigns an importance score to each feature based on its contribution to the model's predictions.
3. Elimination Process: RFE eliminates the least important features (those with the lowest scores) and retrains the model on the remaining features.
4. Recursive Process: This process is repeated iteratively, recursively eliminating less important features until a specified number of features are left or a stopping criterion is met.

Fig. 6 is a visualization of the feature ranking obtained using Recursive Feature Elimination (RFE). The x-axis represents the index of each feature. The y-axis shows the ranking assigned to each feature by RFE. Lower rankings indicate more important features, with the most important features ranked as 1. Fig. 6 also helps to identify which features are considered most critical by the RFE process, aiding in understanding which features significantly

contribute to the fake news detection model. As shown, the feature selection step is critical for building an effective fake news detection system. By systematically identifying and selecting the most relevant features, it can enhance the model's performance, reduce overfitting, and improve interpretability.

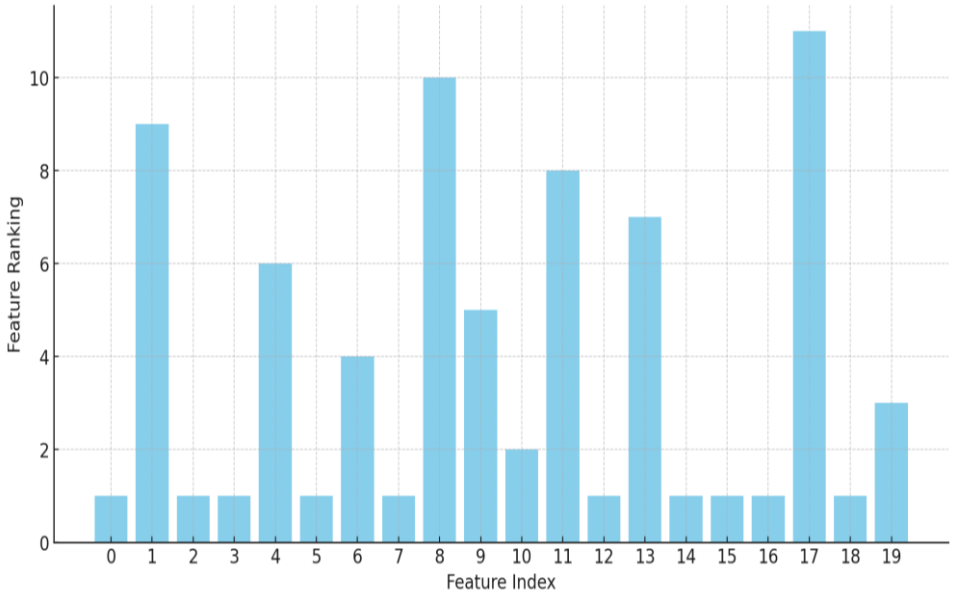


**Fig. 6** Feature Ranking using RFE

*F. Neural Networks Module*

RNN is a class of neural networks particularly suited for sequence data, such as text. In the context of fake news detection, RNN can capture temporal dependencies and context from sequential data, making them valuable for understanding the flow and context within news articles [24]. The RNN models can be built using various architectures. The Long Short-Term Memory (LSTM) is commonly used due to its ability to handle long-term dependencies and mitigate the vanishing gradient problem. In the proposed model, combining a simple RNN layer with an LSTM layer can leverage the strengths of both types of recurrent neural networks, making the model more robust for fake news detection. This stage includes:

1. Embedding layer: to transforms the input text data into dense vectors of fixed size, where each word is represented by a continuous vector in a high-dimensional space. This allows the model to work with numerical data instead of raw text, capturing semantic relationships between words.

2. Simple RNN Layer: to processes the embedded sequences to capture basic sequential patterns and dependencies within the text.

3. LSTM Layer: to capture long-term dependencies and more complex sequential patterns that the Simple RNN layer might miss.

4. Dense Layer: The final layer that performs binary classification to determine whether the input is fake news or real news. The sigmoid activation function in the Dense layer outputs a probability value between 0 and 1, suitable for binary classification tasks.

Combining a simple RNN layer with an LSTM layer, the fake news detection model benefits from both short-term and long-term pattern recognition. The RNN layer quickly processes simpler dependencies, allowing the LSTM layer to focus on more complex and long-term dependencies, which is crucial for the fake news detection where context and sequence matter significantly. This hybrid approach enhances the model's ability to accurately classify news as fake or real, leveraging the strengths of both RNN and LSTM architectures.

## IV. Experimental Results

The proposed model has been implemented and trained using Jupyter Notebook and pytorch. Jupyter Notebook is a powerful tool for data science, enabling users to create comprehensive, interactive documents that combine code execution, visualization, and narrative. It's widely used for data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more [25]. PyTorch framework is used for developing and training neural network models due to its flexibility and ease of use. It utilizes optimizations to backend operations contribute to overall performance improvements, making training and inference faster [26]. The system has been implemented using a laptop computer with Processor Intel(R) Core i7-6500U CPU @ 2.50 GHz 2.59 GHz with RAM: 12 GB.

### *A. Experiment 1: Accuracy Comparison for Different Classification Models*

To compare the accuracy of the proposed model with other classification models, it is necessary to implement and evaluate these models on the dataset (FakeNewsNet). The Common classification models that are used for fake news detection include logistic regression, Naive Bayes, support vector machines (SVM), and other deep learning models such as CNNs (see Fig. 7).

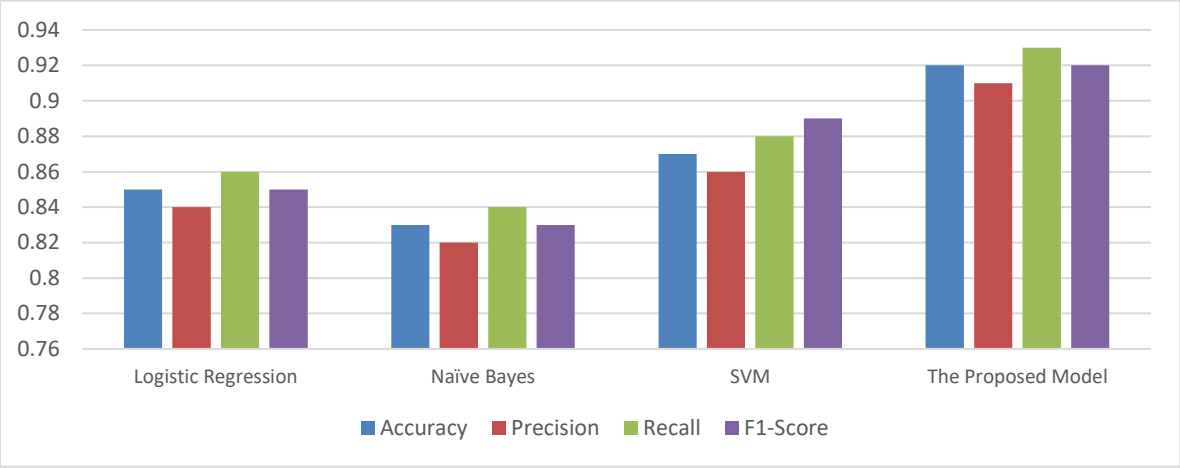


**Fig. 7.** Accuracy Comparison between Different Classification Models

As shown in the fig. 7, the proposed model shows the highest accuracy, close to 0.95. SVM has slightly lower accuracy, around 0.92. Logistic Regression and Naïve Bayes have similar accuracy, both around 0.88. As a conclusion, the proposed model outperforms the other classification models across all metrics (Accuracy, Precision, Recall, and F1-Score). This indicates that it provides the best balance between correctly identifying positive cases (high recall) and the precision of those positive identifications (high precision), leading to the highest F1-Score. Logistic Regression and Naïve Bayes, while commonly used for classification tasks, do not perform as well as SVM or the proposed model. This could be due to the inherent simplicity of these models, which might not capture complex patterns in the data as effectively as the proposed model. Accuracy alone can be misleading in imbalanced datasets, but the proposed model's superior performance in precision, recall, and F1-Score reinforces its effectiveness.

*Experiment 2: Accuracy comparison for different distributions of fake and real class samples of the trained dataset*

The aim of this experiment is to infer how the distribution of fake news in the dataset significantly impacts the accuracy of a fake news detection model. The results shown in Table 1 can be interpreted as follows:

- Extreme Imbalance (e.g., 10:90 Fake to Real):

✓ Model Bias: The model tends to become biased towards the majority class (real news) because it encounters many more examples of real news during training.

✓ Accuracy Paradox: High overall accuracy might be achieved simply because the model predicts the majority class most of the time. However, this doesn't reflect true performance as the model fails to detect fake news effectively.

- Moderate Balance (e.g., 30:70 or 40:60 Fake to Real):

✓ Improved Learning: With a more balanced distribution, the model gets sufficient examples of both classes, improving its ability to distinguish between fake and real news.
✓ Higher Overall Accuracy: Because the model is not overwhelmingly biased towards one class, it achieves higher and more meaningful accuracy.

- Equal Balance (50:50 Fake to Real):

✓ Optimal Learning: The model has equal opportunity to learn from both classes, which typically leads to the best performance metrics for both fake and real news detection.
✓ Balanced Performance: accuracy is usually high and well-balanced across both classes.

Table 1. Accuracy Comparisons for Different Distributions

| **Accuracy** | **Class Distribution** | |
|---|---|---|
| | ***Fake*** | ***Real*** |
| 0.85 | 10% | 90% |
| 0.87 | 20% | 80% |
| 0.88 | 30% | 70% |
| 0.86 | 40% | 60% |
| 0.84 | 50% | 50% |

In conclusion, the distribution of fake news in the dataset plays a crucial role in determining the model's accuracy and overall performance. While a balanced distribution (e.g., 50:50) often leads to optimal results, real-world scenarios may require handling class imbalance effectively to maintain high accuracy and generalizability. Techniques such as resampling, adjusting class weights, and using appropriate evaluation metrics can help mitigate the effects of class imbalance on model performance.

*Experiment 3: Accuracy Comparison on different Datasets*

Testing the validity of a fake news detection model across different types of datasets is crucial to ensure its robustness, generalizability, and effectiveness. This experiment is conducted to evaluate the robustness of the suggested fake news detection model on a diverse range of datasets to ensure it performs well across different contexts. The results shown in Figure 8 confirm the validity of the model across different dataset. For four different benchmark fake news datasets [27 - 30].

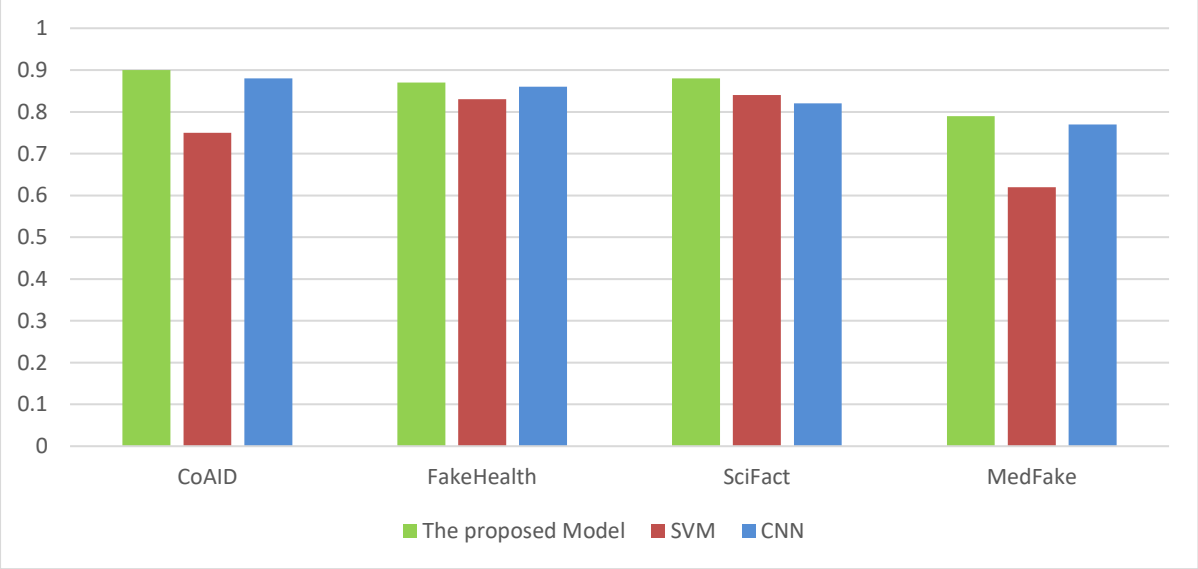


**Fig. 8.** Accuracy Comparison between different Datasets

In general, the accuracy of the fake news detection can vary significantly as the dataset changes due to several factors:

- Content Diversity: Different datasets may contain varying types of content, vocabulary, and writing styles.

- Topic Distribution: The prevalence of certain topics (e.g., specific diseases, treatments) can influence model performance. A model trained on COVID-19 related news may struggle with content about chronic diseases.
- Label Noise: inconsistent or incorrect labeling of fake and real news can impact the model's learning process and its ability to generalize. Different datasets may have varying levels of labeling accuracy.
- Insufficient Training Data: Smaller datasets may not provide enough examples for the model to learn effectively, especially for rare events or nuanced fake news patterns.
- Evolving Misinformation: Fake news trends and misinformation tactics evolve over time. A model trained on older data may not perform well on newer datasets if the nature of fake news has changed.
- Feature Drift: The underlying features and patterns in the data can change over time, affecting the model's accuracy on newer datasets.

## V. Conclusion

This work presents a novel methodology used in distinguishing news between true and fake. To evaluate the effectiveness of the proposed model, the FakeNewsNet dataset has been used. To ensure that the text is a standardized form, the preprocessing step is done. Besides, this step is useful to get the accurate features and more effective neural network model. To extract the best related-features. the model leverages various techniques to capture the linguistic, semantic, Stylometric of the text, and social network and propagation, providing a rich set of features for accurate classification. Moreover, the proposed model identifying and selecting the most relevant features using Recursive Feature Elimination selection technique (REF). As a key contribution, the integration of a Simple RNN layer with an LSTM layer in the fake news detection model leverages both short-term and long-term pattern recognition. The RNN layer efficiently handles simpler dependencies, enabling the LSTM layer to concentrate on more intricate and extended dependencies. This is particularly important for fake news detection, where context and sequence are critical. This hybrid approach enhances the model's accuracy in classifying news as either fake or real, capitalizing on the strengths of both RNN and LSTM architectures. On the other hand, using a hybrid RNN-LSTM model can bring several limitations: Combining RNN and LSTM layers results in a model with higher computational demands. The increased complexity can lead to longer training times and require more memory and processing power. With the increased complexity of the RNN-LSTM model, there is a greater risk of overfitting the training data.

In the future, addressing and dealing with the issues of feature drift is critical point, which is the ability of the system to adjust its detection mechanisms in response to changes in feature importance or relevance of fake news over time. In addition to explore multimodal approaches by integrating image and video analysis with textual data to provide a more comprehensive fake news detection system. Besides, implementing mechanisms to collect and incorporate user feedback into the model training process, enhancing the system's adaptability and accuracy.

## Acknowledgment

I would like to express my heartfelt gratitude to all the individuals and organizations who have contributed to the successful completion of my research paper on "Fake news detection" Their unwavering support, guidance, and encouragement have been invaluable throughout this academic journey. I extend my deepest appreciation to my supervisors who supported me during this research whose expertise and dedication have been instrumental in shaping the direction of this research. Their insightful feedback and constant motivation have greatly enriched the quality of this work. In conclusion, the successful completion of this research paper would not have been possible without the collective efforts and support of all. Thank you all for being an integral part of my academic journey.

## References

[1] M. Mahyoob, J. Al-Garaady, and M. Alrahaili, "Linguistic-based detection of fake news in social media", Forthcoming, International Journal of English Linguistics 11, no. 1, 2020.

[2] Feng, Song, Ritwik Banerjee, and Yejin Choi, "Characterizing stylistic elements in syntactic structure", Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, pp. 1522-1533. 2012.

[3] Zhou, Xinyi, Atishay Jain, Vir V. Phoha, and Reza Zafarani, "Fake news early detection: A theory-driven model." Digital Threats: Research and Practice 1, no. 2, pp. 1-25, 2020.

[4] Ruchansky, Natali, Sungyong Seo, and Yan Liu, "Csi: A hybrid deep model for fake news detection", Proceedings of the ACM on Conference on Information and Knowledge Management, pp. 797-806. 2017.
[5] Rashkin, Hannah, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking", Proceedings of the conference on empirical methods in natural language processing, pp. 2931-2937, 2017.
[6] Shu, Kai, Deepak Mahudeswaran, and Huan Liu, "FakeNewsTracker: a tool for fake news collection, detection, and visualization", Computational and Mathematical Organization Theory, vol. 25, pp. 60-71, 2019.
[7] Horne, Benjamin D., Jeppe Nørregaard, and Sibel Adali, "Robust fake news detection over time and attack", ACM Transactions on Intelligent Systems and Technology (TIST), vol. 11, no. 1, pp. 1-23, 2019.
[8] Monti, Federico, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M. Bronstein, "Fake news detection on social media using geometric deep learning", arXiv preprint arXiv:1902.06673, 2019.
[9] Wang, Haizhou, Sen Wang, and YuHu Han, "Detecting fake news on Chinese social media based on hybrid feature fusion method, "Expert Systems with Applications, vol. 208, pp. 118111, 2022.
[10] Karimi, Hamid, Proteek Roy, Sari Saba-Sadiya, and Jiliang Tang. "Multi-source multi-class fake news detection", Proceedings of the 27th international conference on computational linguistics, pp. 1546-1557, 2018.
[11] Jiang, Shengyi, Xiaoting Chen, Liming Zhang, Sutong Chen, and Haonan Liu, "User-characteristic enhanced model for fake news detection in social media, "In Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, Proceedings, Part I 8, pp. 634-646. Springer International Publishing, 2019.
[12] Zhang, Xichen, and Ali A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion", Information Processing & Management 57, no. 2, pp. 102025, 2020.
[13] Shu, Kai, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu, "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media", Big data, vol. 8, no. 3, pp. 171-188, 2020.
[14] Bauskar, Shubham, Vijay Badole, Prajal Jain, and Meenu Chawla, "Natural language processing based hybrid model for detecting fake news using content-based features and social features,"International Journal of Information Engineering and Electronic Business, vol. 11, no. 4, pp. 1-10, 2019.
[15] Lai, Chun-Ming, Mei-Hua Chen, Endah Kristiani, Vinod Kumar Verma, and Chao-Tung Yang, "Fake news classification based on content level features", Applied Sciences, vol. 12, no. 3, pp. 1116, 2022.
[16] Kapusta, Jozef, Dávid Držík, Kirsten Šteflovič, and Kitti Szabó Nagy. "Text Data Augmentation Techniques for Word Embeddings in Fake News Classification", IEEE Access 12, pp. 31538-31550, 2024.
[17] Choudhary, Anshika, and Anuja Arora, "Assessment of bidirectional transformer encoder model and attention based bidirectional LSTM language models for fake news detection", Journal of Retailing and Consumer Services, vol. 76, pp. 103545, 2024.
[18] Farhangian, Faramarz, Rafael MO Cruz, and George DC Cavalcanti. "Fake news detection: Taxonomy and comparative study", Information Fusion, vol. 103, 102140, 2024.
[19] Zaheer, Hamza, Saif Ur Rehman, Maryam Bashir, Mian Aziz Ahmad, and Faheem Ahmad, "A metaheuristic based filter-wrapper approach to feature selection for fake news detection", Multimedia Tools and Applications, pp. 1-30, 2024.
[20] Fayaz, Muhammad, Atif Khan, Muhammad Bilal, and Sana Ullah Khan. "Machine learning for fake news classification with optimal feature selection", Soft Computing, vol. 26, no. 16, pp. 7763-7771, 2022.
[21] Kaliyar, Rohit Kumar, "Fake news detection using a deep neural network", Proceedings In the 4th international conference on computing communication and automation (ICCCA), pp. 1-7. IEEE, 2018.
[22] Zaheer, Hamza, Saif Ur Rehman, Maryam Bashir, Mian Aziz Ahmad, and Faheem Ahmad, "A metaheuristic based filter-wrapper approach to feature selection for fake news detection", Multimedia Tools and Applications, pp. 1-30, 2024.
[23] Fayaz, Muhammad, Atif Khan, Muhammad Bilal, and Sana Ullah Khan. "Machine learning for fake news classification with optimal feature selection", Soft Computing, vol. 26, no. 16, pp. 7763-7771, 2022.
[24] Kaliyar, Rohit Kumar, "Fake news detection using a deep neural network", Proceedings In the 4th international conference on computing communication and automation (ICCCA), pp. 1-7. IEEE, 2018.
[25] https://docs.jupyter.org/en/latest/
[26] https://pytorch.org/docs/stable/index.html,
[27] L. Cui L, and D. Lee, Coaid: Covid-19 healthcare misinformation dataset. arXiv preprint arXiv:2006.00885. 2020.
[28] https://www.kdnuggets.com/2020/03/data-repository-fake-health-news.html
[29] https://allenai.org/data/scifact
[30] X. Ma, X. Chu, Y. Wang, H. Yu, L. Ma, W. Tang, and J. Zhao, "MedFACT: Modeling Medical Feature Correlations in Patient Health Representation Learning via Feature Clustering", arXiv preprint arXiv:2204.10011. 2022 Apr 21.