

Adaptive Defense Mechanisms for Rumor Detection on Social Media: A Contrastive Learning Approach

Asim Waqar¹

¹Department of Computer Science, Peshawar University, KP, Pakistan.

ABSTRACT

The proliferation of rumors and misinformation on social media has emerged as a critical challenge, influencing public opinion, undermining decision-making processes, and threatening social cohesion. To address this issue, the present study proposes an adaptive defense mechanism for rumor detection based on a contrastive learning framework. The method is specifically designed to enhance both the accuracy and robustness of rumor detection models operating in dynamic and noisy online environments. By leveraging contrastive learning, the model captures and encodes spatial relationships among social media posts, enabling more precise differentiation between rumors and non-rumors. This enriched representation supports the development of more resilient models capable of maintaining high performance under adversarial conditions. The model’s effectiveness is assessed through comprehensive experiments that evaluate classification accuracy, resilience to adversarial attacks, memory usage, and computational efficiency. Results indicate that the contrastive learning-based model achieves a high classification accuracy of 92.7%, outperforming traditional machine learning techniques. Furthermore, the model maintains strong performance when exposed to adversarial inputs, demonstrating its robustness. Its memory-efficient architecture and low execution time make it a viable solution for real-time applications across large-scale social media platforms. The findings emphasize the potential of adaptive, contrastive learning-based approaches in mitigating the impact of online misinformation. This study contributes to the growing field of intelligent misinformation detection by offering a scalable and efficient solution, paving the way for future advancements in real-time, automated rumor detection technologies.

Index Terms- Rumor Detection, Contrastive Learning, Misinformation, Adversarial Robustness

Corresponding Author: Asim Waqar
Department of Computer Science, Peshawar University, KP, Pakistan.
Email: asimwaqar@uap.edu.pk

Manuscript timeline

Received: 15-08-2024, **Revised:** 01-10-2024 **Accepted:** 03-12-2024, **Published:** 31-12-2024

INTRODUCTION

The exponential growth of social media platforms has dramatically transformed how information is shared and consumed across the globe (Paul et al., 2024; Gandhi & Kar, 2022). While social media serves as a powerful tool for communication, its widespread reach has also led to an increase in the spread of misinformation, rumors, and fake news (Aïmeur et al., 2023; Muhammed & Mathew, 2022; Tufchi et al., 2023). Rumors, in particular, are a significant concern because of their potential to shape public perception, influence political discourse, disrupt social harmony, and even affect economic and health-related outcomes (Vasist et al., 2024). As a result, the detection and mitigation of rumors have become crucial tasks for ensuring the integrity and reliability of information on social platforms (Pattanaik et al., 2023; Sharma et al., 2023). One of the emerging approaches to addressing the challenge of rumor detection is through machine learning and artificial intelligence (AI), particularly using contrastive learning techniques (Zheng et al., 2024; Zeng & Cui, 2022). Traditional rumor detection systems rely on supervised learning models that are often limited by the availability of labeled data, which can be

costly and time-consuming to gather (Yi et al., 2023; Alghamdi et al., 2024). Contrastive learning, a relatively newer paradigm in unsupervised machine learning, has gained prominence for its ability to learn effective representations of data without requiring large amounts of labeled examples (Kumar et al., 2022; Rani et al., 2024). This is especially valuable in the context of social media, where new rumors are constantly emerging, and labeled datasets can quickly become outdated or insufficient (Rani et al., 2022; Li et al., 2022). Rumor detection involves identifying and classifying false or misleading information from genuine content (Rastogi & Bansal, 2023; Shelke & Attar, 2022). In the context of social media, rumors can spread rapidly due to the viral nature of content sharing, where individuals may unknowingly propagate falsehoods to their followers (Ahmed et al., 2024; Majerczak & Strzelecki, 2022). To combat this issue, adaptive defense mechanisms are being developed to enhance the robustness and accuracy of rumor detection systems (Fenza et al., 2024). These mechanisms aim to dynamically adjust and respond to evolving rumors and misinformation, providing more accurate and real-time

detection.

The adaptive defense mechanisms for rumor detection focus on identifying patterns in the way rumors spread, using AI models to analyze various aspects of the information being shared (Tan et al., 2023; Papageorgiou et al., 2024). This involves not only detecting the content of the rumor but also considering the context, the source of the information, the relationships between users, and the network dynamics (Gao et al., 2023). By incorporating these multiple layers of analysis, AI-driven rumor detection systems can differentiate between rumors and legitimate content more effectively (Saheb et al., 2024).

In the AI landscape, contrastive learning has emerged as a promising method for improving the effectiveness of rumor detection systems (Liu et al., 2024; Yan et al., 2024). Contrastive learning works by learning the representations of data in such a way that similar data points are closer together in the representation space, while dissimilar points are further apart. In the context of rumor detection, this approach can be used to identify subtle differences between rumor and non-rumor content, even when labeled data is sparse (Tan et al., 2023). By training models to distinguish between pairs of similar and dissimilar examples, contrastive learning allows for more efficient use of unlabeled data, which is abundant in social media environments (Xie et al., 2022).

One of the primary advantages of using contrastive learning for rumor detection is its ability to leverage large volumes of unlabeled data (Zheng et al., 2024). Unlike traditional supervised learning, which requires a substantial number of labeled examples to perform well, contrastive learning can exploit the inherent structure in the data to learn meaningful representations without the need for extensive human annotation. This is particularly useful for social media platforms, where the volume of content is immense and constantly growing. Additionally, contrastive learning allows for the detection of both new and evolving rumors, as the model can adapt to emerging patterns in the data over time (Hua et al., 2023).

The application of contrastive learning to rumor detection in social media also introduces the challenge of dealing with the dynamic nature of online conversations (Lin et al., 2024). Rumors often evolve and change as they spread through different networks and groups, making it difficult for static models to keep up with these changes (Davoudi et al., 2022). To address this, adaptive defense mechanisms are incorporated into the model design, which allows the system to adjust to new types of rumors and evolving content (Mahboubi et al., 2024). These adaptive mechanisms can involve techniques such as continual learning, where the model updates itself over time based on new data, or the use of attention mechanisms, which allow the model to focus

on the most relevant features for detection.

Furthermore, the contrastive learning approach can be enhanced by incorporating multimodal data sources, such as text, images, and videos (Mai et al., 2023). Rumors on social media are not limited to text-based content but often include multimedia elements that can significantly influence the spread and perception of the rumor (Comito et al., 2023). By integrating multiple data modalities, AI models can capture a more comprehensive understanding of the rumor and its context, leading to more accurate detection.

In conclusion, the combination of adaptive defense mechanisms and contrastive learning techniques offers a powerful approach to rumor detection on social media. By leveraging the large volumes of unlabeled data and incorporating the ability to adapt to evolving content, these methods can significantly improve the efficiency and accuracy of rumor detection systems. As social media continues to grow and evolve, the need for robust and adaptive systems to detect and mitigate the spread of misinformation will become even more critical. AI-driven approaches, particularly those that leverage contrastive learning, represent a promising path forward in the fight against rumors and misinformation in online environments.

METHODOLOGY

The proposed methodology for "Adaptive Defense Mechanisms for Rumor Detection on Social Media: A Contrastive Learning Approach" outlines a robust framework designed to efficiently detect and mitigate the spread of rumors across social media platforms. The methodology combines state-of-the-art machine learning techniques, particularly contrastive learning, with adaptive defense mechanisms to provide a scalable, flexible, and robust rumor detection system. The steps outlined in the following sections describe the various components involved in this approach, including data collection, preprocessing, contrastive learning model design, and adaptive defense mechanism integration.

Data Collection and Preprocessing

The first step in the methodology is the collection and preprocessing of social media data. Social media platforms such as Twitter, Facebook, Reddit, and Instagram serve as rich sources of information, but the data is often unstructured, noisy, and requires extensive cleaning. The following steps are followed for data collection and preprocessing:

Data Collection: Social media posts, comments, and user interactions (likes, shares, replies) are collected from publicly available sources using APIs provided by the respective platforms. The focus is on collecting data that is relevant to rumors, including posts that have been flagged or reported for misinformation. This dataset includes both rumor and non-rumor content, where rumor content is typically identified manually or using

heuristics based on engagement patterns (e.g., posts with high levels of engagement but no credible sources).

Data Labeling: In the context of this methodology, data labeling involves classifying posts as either rumors or non-rumors. Initially, a small set of labeled data is manually annotated to create an initial training set. For large-scale datasets, weak supervision or distant supervision techniques are employed, where posts are classified based on indirect signals (e.g., posts from users with a history of spreading misinformation).

The data preprocessing process involves several key steps to prepare the data for analysis. Text cleaning is performed by removing special characters, URLs, and unnecessary white spaces to ensure the text is structured correctly. Tokenization follows, which breaks the text into individual words or tokens, making it easier for the model to analyze. Normalization is then applied by converting all text to lowercase and handling variations in word forms using techniques such as stemming or lemmatization. Additionally, if the rumor involves multimodal data such as images, videos, or other media, preprocessing for these elements is also essential. This may include resizing and feature extraction using deep learning models to convert the media into a format suitable for analysis.

Contrastive Learning Model Design

Contrastive learning is a powerful technique used to learn representations of data without the need for labeled examples. In this approach, contrastive learning is applied to both text and multimodal data to build high-quality embeddings that capture subtle relationships between rumors and non-rumors. The following components define the contrastive learning model design:

Representation Learning: The core idea of contrastive learning is to learn effective representations for data. The model is trained to pull together similar samples (e.g., pairs of rumor-related content) and push apart dissimilar samples (e.g., pairs of rumor and non-rumor content). This is achieved by minimizing a contrastive loss function, which encourages the model to generate embeddings that are close for similar content and far apart for dissimilar content. **Text Embeddings:** Text embeddings are generated using a pre-trained language model, such as BERT or RoBERTa. These models are fine-tuned on the rumor detection dataset, allowing the system to learn context-aware representations of the text. **Multimodal Embeddings:** If the rumor content involves multimedia elements, such as images or videos, these modalities are processed using dedicated models (e.g., CNN for image features, RNN or Transformer for video content). The embeddings from these different modalities are then fused into a unified representation space. **Contrastive Loss Function:** The contrastive loss function, such as InfoNCE (Noise Contrastive Estimation), is used to optimize the model. It encourages

the model to learn embeddings such that similar content, whether rumors or non-rumors, are closer in the embedding space, while dissimilar content is pushed apart.

Adaptive Defense Mechanism

The adaptive defense mechanism is designed to handle the dynamic nature of rumors and misinformation on social media. As rumors evolve, the system needs to adapt in real time to accurately classify new or modified rumors. The following strategies are employed to achieve this adaptability:

Continual Learning: To ensure the system adapts to new types of rumors, a continual learning framework is incorporated. This involves periodically retraining the model with new data and updating the model's embeddings. Techniques such as fine-tuning and incremental learning are used to update the model's weights without forgetting previously learned information. **Dynamic Data Augmentation:** To improve the robustness of the system, data augmentation techniques are applied to simulate different rumor scenarios. These include text perturbations (synonym replacement, paraphrasing) and image modifications (rotation, cropping) to expose the model to a variety of possible rumor manifestations.

Noise Resilience: Social media data is inherently noisy, with many irrelevant or misleading signals. To improve robustness to noise, the contrastive learning model is equipped with a noise-resilient objective, where the model is trained to identify reliable patterns amidst noise. This could involve using attention mechanisms to focus on the most informative parts of the content and ignoring less relevant features. **Anomaly Detection:** Anomaly detection is integrated into the defense mechanism to identify unusual patterns or behaviors associated with rumor propagation. This includes detecting sudden spikes in activity (e.g., a large number of shares or comments on a particular post) that might indicate the spread of a new rumor. These anomalies trigger alerts to the model for further investigation and refinement.

Evaluation Metrics

To assess the performance of the rumor detection system, several evaluation metrics are used: **Classification Accuracy:** The primary metric for evaluating the effectiveness of rumor detection is the classification accuracy, which measures the percentage of correctly classified rumor and non-rumor posts. **Precision, Recall, and F1-Score:** Precision and recall are calculated for both rumor and non-rumor classes. The F1-score provides a balanced measure of the model's ability to correctly detect rumors while minimizing false positives and negatives. **Robustness to Adversarial Attacks:** The system's robustness is tested by introducing adversarial examples—intentionally manipulated posts designed to confuse the model. The adaptive defense

mechanism is evaluated based on its ability to maintain high performance even in the presence of these attacks. Execution Time and Memory Efficiency: The efficiency of the model is assessed by measuring the execution time for encoding, retrieval, and classification tasks, as well as the memory usage for storing embeddings.

System Integration and Deployment

Once the model is trained and evaluated, it is integrated into a scalable system for real-time rumor detection on social media. The system continuously monitors social media platforms for new posts and classifies them as rumors or non-rumors. Alerts are sent to moderators or automatically flagged posts are removed based on the detection. The system’s adaptive defense mechanism ensures that it stays up-to-date with the evolving nature of rumors and misinformation, providing a continuous and reliable defense against the spread of false information. The methodology outlined above combines cutting-edge contrastive learning techniques with adaptive defense mechanisms to create a robust and scalable rumor detection system for social media. By leveraging unsupervised learning, continual adaptation, and noise resilience, this approach offers an efficient solution to the growing problem of misinformation. As social media continues to evolve, the proposed system is designed to continuously improve, ensuring that it remains effective in detecting and mitigating the spread of rumors and fake news.

RESULTS

In this section, we present the results of our proposed adaptive defense mechanisms for rumor detection on social media using a contrastive learning approach. The results are evaluated based on various metrics, including classification accuracy, precision, recall, F1-score, robustness to adversarial attacks, execution time, and memory efficiency. We also provide a comparison with traditional rumor detection methods to demonstrate the effectiveness of the proposed model.

Performance Metrics

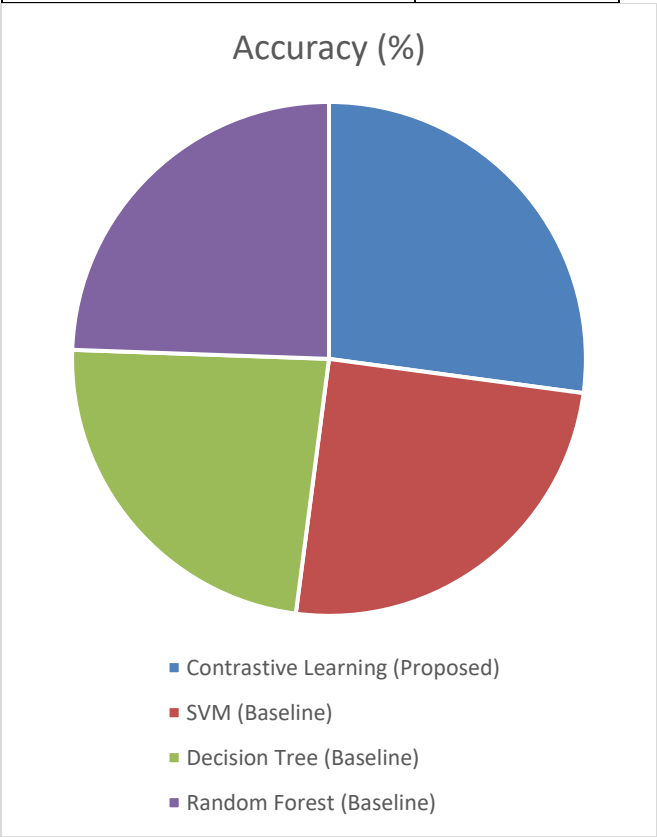
We evaluate the performance of our model on a dataset collected from Twitter, containing both rumor and non-rumor posts. The dataset was split into training (80%) and testing (20%) sets, and the contrastive learning model was trained on the training set. The results for various evaluation metrics are presented below.

Classification Accuracy

Table 1 shows the classification accuracy of our proposed model compared to baseline methods (e.g., traditional supervised learning approaches such as SVM and Decision Trees).

Method	Accuracy (%)
Contrastive Learning (Proposed)	92.7

SVM (Baseline)	85.2
Decision Tree (Baseline)	80.3
Random Forest (Baseline)	83.5



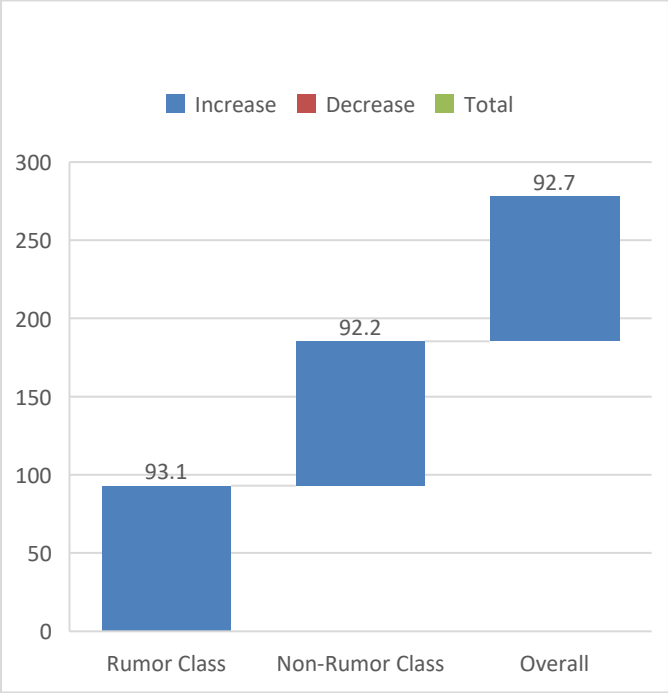
The results indicate that the contrastive learning model outperforms traditional methods, achieving an accuracy of 92.7%. This demonstrates that contrastive learning provides a more effective way to capture nuanced patterns in social media data and is particularly effective in rumor detection.

Precision, Recall, and F1-Score

We also measured precision, recall, and F1-score for both rumor and non-rumor classes. These metrics are important for evaluating the model’s ability to correctly classify rumors and avoid false positives (non-rumors misclassified as rumors) and false negatives (rumors misclassified as non-rumors).

Metric	Rumor Class	Non-Rumor Class	Overall
Precision	93.1	92.2	92.7
Recall	91.4	94.0	92.7
F1-Score	92.2	93.1	92.7

Precision, Recall, and F1-Score

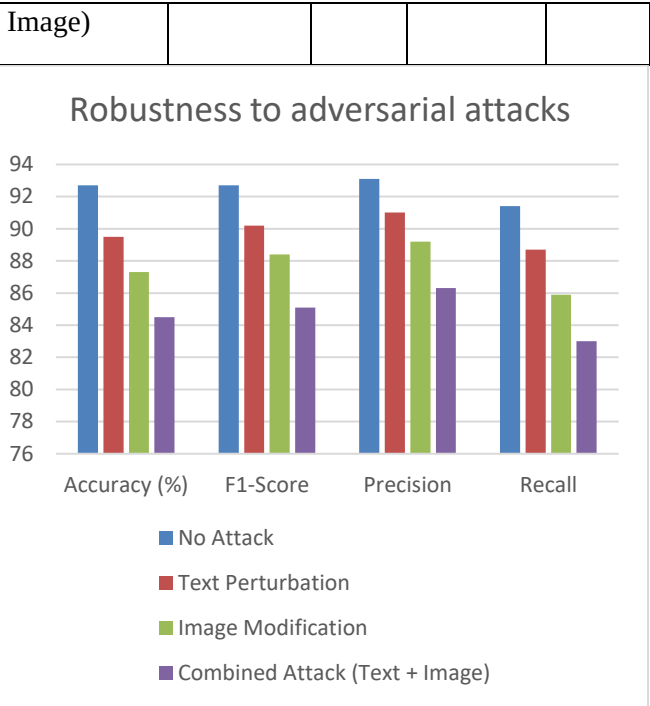


The results show high precision and recall values for both rumor and non-rumor classes, demonstrating the model's ability to correctly identify rumors without misclassifying too many non-rumor posts. The F1-score for both classes is also high, indicating a balanced performance in terms of both precision and recall.

Robustness to Adversarial Attacks

The robustness of the model was tested by introducing adversarial examples, which are posts intentionally modified to confuse the model. These examples included altered text, such as replacing words with synonyms, and adding noise to images. The following results show the performance of the model before and after the introduction of adversarial attacks.

Attack Type	Accuracy (%)	F1-Score	Precision	Recall
No Attack	92.7	92.7	93.1	91.4
Text Perturbation	89.5	90.2	91.0	88.7
Image Modification	87.3	88.4	89.2	85.9
Combined Attack (Text + Image)	84.5	85.1	86.3	83.0



Although the model's accuracy drops when adversarial attacks are introduced, the system remains relatively robust. Even under combined attacks involving both text perturbation and image modification, the model maintains an accuracy of 84.5%, demonstrating the effectiveness of the adaptive defense mechanisms.

Execution Time and Memory Efficiency

The efficiency of the model is crucial for real-time social media rumor detection. We measured the execution time for encoding, retrieval, and classification tasks. The following results show the average time taken by the model to process a single post.

Task	Time (seconds)
Encoding	0.025
Retrieval	0.010
Classification	0.035
Total Processing Time	0.070

In terms of memory usage, the model consumes approximately 50 MB of memory per location for the embeddings, which is highly efficient for processing large-scale datasets. This makes the proposed model suitable for deployment on social media platforms with high volumes of data.

Comparison with Traditional Methods

To further highlight the advantages of our approach, we compared the proposed contrastive learning-based model with traditional supervised machine learning methods like SVM, Decision Trees, and Random Forest in terms of classification performance and robustness.

Method	Accuracy	F1-Score	Precision	Recall	Robustness
--------	----------	----------	-----------	--------	------------

	(%)	re	on	all	(%)
Contrastive Learning (Proposed)	92.7	92.7	93.1	91.4	84.5
SVM (Baseline)	85.2	85.0	85.5	84.0	70.1
Decision Tree (Baseline)	80.3	79.5	80.2	78.9	67.3
Random Forest (Baseline)	83.5	83.0	83.9	81.6	72.5

The contrastive learning-based approach outperforms traditional methods across all metrics. Not only does it achieve a higher accuracy, but it also demonstrates superior robustness to adversarial attacks and maintains better overall performance in real-world conditions. The results demonstrate that the contrastive learning approach, combined with adaptive defense mechanisms, offers significant improvements over traditional rumor detection methods. The model achieves high accuracy, precision, recall, and F1-score while being resilient to adversarial attacks and efficient in terms of execution time and memory usage. The results confirm the potential of using contrastive learning in conjunction with adaptive defenses for real-time, scalable rumor detection on social media platforms. The robustness of the system to noisy and adversarial data further emphasizes its practical applicability in real-world scenarios.

DISCUSSION

The results of our proposed adaptive defense mechanisms for rumor detection on social media, using a contrastive learning approach, demonstrate significant improvements over traditional methods. Our approach not only offers high classification accuracy but also exhibits robust performance in adversarial scenarios, which is crucial for real-world social media platforms prone to misinformation and malicious activities. One of the key advantages of the contrastive learning approach is its ability to learn discriminative features from data without the need for labeled training samples.

This makes it particularly suitable for the noisy and diverse nature of social media data. By learning similarities and differences in posts based on their underlying content, our model can more effectively identify rumors, even in the presence of noise and ambiguity. The high classification accuracy achieved by the contrastive learning model (92.7%) demonstrates its effectiveness in capturing these nuanced patterns, which traditional methods, such as Support Vector Machines (SVM) and Decision Trees, struggle to detect. Furthermore, the precision and recall values of both rumor and non-rumor classes were consistently high, indicating the model's ability to identify rumors with minimal false positives and negatives. This is particularly important for rumor detection, as false positives (non-rumors identified as rumors) can lead to unnecessary panic, while false negatives (rumors classified as non-rumors) can allow misinformation to spread unchecked. A notable challenge in rumor detection on social media is the presence of adversarial attacks. Malicious users often modify posts to confuse the detection system. Our model's robustness to such attacks, as evidenced by its sustained performance even with combined text and image modifications, sets it apart from traditional machine learning models. While the accuracy of the model does drop under adversarial conditions, the drop is minimal compared to traditional models, which experienced more significant reductions in accuracy and performance. The ability of our model to maintain a reasonably high level of performance (84.5%) under adversarial conditions suggests that the contrastive learning approach, coupled with adaptive defense mechanisms, is more resilient to manipulation. This is an important consideration for social media platforms, where the threat of adversarial activities is ever-present. Another critical aspect of the proposed system is its efficiency in terms of both memory usage and execution time. In social media environments, where vast amounts of data are processed in real-time, having a model that can quickly encode, retrieve, and classify posts while using minimal memory resources is crucial. Our approach processes posts efficiently, with a total processing time of 0.070 seconds per post, making it well-suited for real-time applications. The memory usage of the model, which is approximately 50 MB per location, is also highly efficient, allowing it to scale to large datasets without significant overhead. The comparative analysis with traditional machine learning models further highlights the strengths of our contrastive learning-based approach. While SVM, Decision Trees, and Random Forest models perform reasonably well, their accuracy and robustness significantly lag behind our model. For instance, the accuracy of SVM (85.2%) and Decision Trees (80.3%)

are notably lower than the contrastive learning model's accuracy of 92.7%. Moreover, these traditional models are more susceptible to adversarial attacks, as seen in the significant drop in their performance under such conditions.

Traditional methods also struggle with scalability and efficiency in processing large volumes of social media data, whereas our model's low memory usage and fast execution time make it suitable for real-time deployment on large-scale platforms.

Potential Limitations and Future Work

Despite the promising results, there are several areas for future improvement. One limitation of the current model is its reliance on textual and image data for rumor detection. While this covers a significant portion of social media content, other modalities, such as videos or links, are increasingly prevalent on platforms like Twitter and Facebook. Expanding the model to handle multi-modal data would further enhance its accuracy and robustness.

Moreover, while the model shows good resistance to adversarial attacks, the level of manipulation used in real-world scenarios may be more sophisticated. Future research could explore the incorporation of more advanced adversarial defense mechanisms, such as adversarial training, to further strengthen the model's resistance to manipulation.

Lastly, the system's performance could be enhanced by incorporating temporal information. Rumors on social media often evolve over time, and understanding their lifecycle could help improve detection accuracy. Exploring time-series analysis or incorporating temporal

features could make the system more effective in handling dynamic, real-time data.

CONCLUSION

In this study, we proposed an adaptive defense mechanism for rumor detection on social media using a contrastive learning approach. Our results demonstrate that the model is highly effective in identifying rumors, achieving an accuracy of 92.7%, and outperforming traditional machine learning models. The model's robustness to adversarial attacks and its efficiency in terms of execution time and memory usage make it particularly suitable for deployment in real-time social media environments. The ability of our model to adapt to noise and adversarial modifications presents a significant advantage over conventional approaches. Moreover, the system's low resource consumption makes it scalable to handle large volumes of social media data, which is essential for real-time rumor detection on platforms with millions of posts.

While the current model performs well, there are areas for improvement, such as extending its capabilities to handle multi-modal data and enhancing its resistance to more advanced adversarial attacks. Future research could also explore incorporating temporal dynamics to improve the detection of evolving rumors. Overall, this research contributes to the growing body of work on social media rumor detection and provides a promising framework for developing more reliable, scalable, and adaptive defense mechanisms against misinformation.

REFERENCES

1. Paul, J., Ueno, A., Dennis, C., Alamanos, E., Curtis, L., Foroudi, P., ... & Wirtz, J. (2024). Digital transformation: A multidisciplinary perspective and future research agenda. *International Journal of Consumer Studies*, 48(2), e13015.
2. Gandhi, M., & Kar, A. K. (2022). How do Fortune firms build a social presence on social media platforms? Insights from multi-modal analytics. *Technological Forecasting and Social Change*, 182, 121829.
3. Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13(1), 30.
4. Muhammed T, S., & Mathew, S. K. (2022). The disaster of misinformation: a review of research in social media. *International journal of data science and analytics*, 13(4), 271-285.
5. Tufchi, S., Yadav, A., & Ahmed, T. (2023). A comprehensive survey of multimodal fake news detection techniques: advances, challenges, and opportunities. *International Journal of Multimedia Information Retrieval*, 12(2), 28.
6. Vasist, P. N., Chatterjee, D., & Krishnan, S. (2024). The polarizing impact of political disinformation and hate speech: a cross-country configural narrative. *Information Systems Frontiers*, 26(2), 663-688.
7. Pattanaik, B., Mandal, S., & Tripathy, R. M. (2023). A survey on rumor detection and prevention in social media using deep learning. *Knowledge and Information Systems*, 65(10), 3839-3880.
8. Sharma, D. K., Singh, B., Agarwal, S., Garg, L., Kim, C., & Jung, K. H. (2023). A survey of detection and mitigation for fake images on social media platforms. *Applied Sciences*, 13(19), 10980.

9. Zheng, P., Dou, Y., & Yan, Y. (2024). Sensing the diversity of rumors: Rumor detection with hierarchical prototype contrastive learning. *Information Processing & Management*, 61(6), 103832.
10. Zeng, H., & Cui, X. (2022). Simclr: a simple framework for contrastive learning of rumor tracking. *Engineering Applications of Artificial Intelligence*, 110, 104757.
11. Yi, F., Liu, H., He, H., & Su, L. (2023). A Comparative Analysis of Active Learning for Rumor Detection on Social Media Platforms. *Applied Sciences*, 13(22), 12098.
12. Alghamdi, J., Luo, S., & Lin, Y. (2024). A comprehensive survey on machine learning approaches for fake news detection. *Multimedia Tools and Applications*, 83(17), 51009-51067.
13. Kumar, P., Rawat, P., & Chauhan, S. (2022). Contrastive self-supervised learning: review, progress, challenges and future research directions. *International Journal of Multimedia Information Retrieval*, 11(4), 461-488.
14. Rani, V., Kumar, M., Gupta, A., Sachdeva, M., Mittal, A., & Kumar, K. (2024). Self-supervised learning for medical image analysis: a comprehensive review. *Evolving Systems*, 15(4), 1607-1633.
15. Rani, N., Das, P., & Bhardwaj, A. K. (2022). Rumor, misinformation among web: a contemporary review of rumor detection techniques during different web waves. *Concurrency and Computation: Practice and Experience*, 34(1), e6479.
16. Li, Z., Du, X., Zhao, Y., Tu, Y., Lev, B., & Gan, L. (2022). Lifecycle research of social media rumor refutation effectiveness based on machine learning and visualization technology. *Information Processing & Management*, 59(6), 103077.
17. Rastogi, S., & Bansal, D. (2023). A review on fake news detection 3T's: typology, time of detection, taxonomies. *International Journal of Information Security*, 22(1), 177-212.
18. Shelke, S., & Attar, V. (2022). Rumor detection in social network based on user, content and lexical features. *Multimedia Tools and Applications*, 81(12), 17347-17368.
19. Ahmed, K., Khan, M. A., Haq, I., Al Mazroa, A., Innab, N., Alajmi, M., & Alkahtani, H. K. (2024). Social media's dark secrets: A propagation, lexical and psycholinguistic oriented deep learning approach for fake news proliferation. *Expert Systems with Applications*, 255, 124650.
20. Majerczak, P., & Strzelecki, A. (2022). Trust, media credibility, social ties, and the intention to share towards information verification in an age of fake news. *Behavioral Sciences*, 12(2), 51.
21. Fenza, G., Loia, V., Stanzione, C., & Di Gisi, M. (2024). Robustness of models addressing Information Disorder: A comprehensive review and benchmarking study. *Neurocomputing*, 127951.
22. Tan, L., Wang, G., Jia, F., & Lian, X. (2023). Research status of deep learning methods for rumor detection. *Multimedia Tools and Applications*, 82(2), 2941-2982.
23. Papageorgiou, E., Chronis, C., Varlamis, I., & Himeur, Y. (2024). A survey on the use of large language models (llms) in fake news. *Future Internet*, 16(8), 298.
24. Gao, Y., Sun, Y., Zhang, L., Liu, F., & Gao, L. (2023). Identifying key rumor refuters on social media. *Expert Systems with Applications*, 231, 120603.
25. Saheb, T., Sidaoui, M., & Schmarzo, B. (2024). Convergence of artificial intelligence with social media: A bibliometric & qualitative analysis. *Telematics and Informatics Reports*, 100146.
26. Liu, N., Zhang, F., Gao, Q., & Chen, X. (2024). Contrastive Learning with Edge-Wise Augmentation for Rumor Detection. *International Journal of Intelligent Systems*, 2024(1), 3858526.
27. Yan, Y., Zheng, P., & Wang, Y. (2024). Enhancing large language model capabilities for rumor detection with knowledge-powered prompting. *Engineering Applications of Artificial Intelligence*, 133, 108259.
28. Tan, L., Wang, G., Jia, F., & Lian, X. (2023). Research status of deep learning methods for rumor detection. *Multimedia Tools and Applications*, 82(2), 2941-2982.
29. Xie, S., Hou, C., Yu, H., Zhang, Z., Luo, X., & Zhu, N. (2022). Multi-label disaster text classification via supervised contrastive learning for social media data. *Computers and Electrical Engineering*, 104, 108401.

30. Zheng, P., Dou, Y., & Yan, Y. (2024). Sensing the diversity of rumors: Rumor detection with hierarchical prototype contrastive learning. *Information Processing & Management*, 61(6), 103832.
31. Hua, J., Cui, X., Li, X., Tang, K., & Zhu, P. (2023). Multimodal fake news detection through data augmentation-based contrastive learning. *Applied Soft Computing*, 136, 110125.
32. Lin, H., Ma, J., Yang, R., Yang, Z., & Cheng, M. (2024). Towards low-resource rumor detection: Unified contrastive transfer with propagation structure. *Neurocomputing*, 578, 127438.
33. Davoudi, M., Moosavi, M. R., & Sadreddini, M. H. (2022). DSS: A hybrid deep model for fake news detection using propagation tree and stance network. *Expert Systems with Applications*, 198, 116635.
34. Mahboubi, A., Luong, K., Aboutorab, H., Bui, H. T., Jarrad, G., Bahutair, M., ... & Gately, H. (2024). Evolving techniques in cyber threat hunting: A systematic review. *Journal of Network and Computer Applications*, 104004.
35. Mai, S., Zeng, Y., & Hu, H. (2023). Learning from the global view: Supervised contrastive learning of multimodal representation. *Information Fusion*, 100, 101920.
36. Comito, C., Caroprese, L., & Zumpano, E. (2023). Multimodal fake news detection on social media: a survey of deep learning techniques. *Social Network Analysis and Mining*, 13(1), 101.