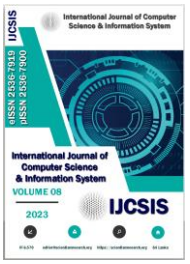
INTERNATIONAL JOURNAL OF COMPUTER SCIENCE & INFORMATION SYSTEM (eISSN: 2536-7919 pISSN: 2536-7900)

VOLUME 10 ISSUE 02 Pages: 6-20



INTERNATIONAL JOURNAL OF COMPUTER SCIENCE & INFORMATION SYSTEM

Research Article

Journal Website: https://scientiamreearch.org/index.php/ijcsis



# HYBRID MULTI-MODAL DETECTION FRAMEWORK FOR ADVANCED PERSISTENT THREATS IN CORPORATE NETWORKS USING MACHINE LEARNING AND DEEP LEARNING

Submission Date: December 30, 2024, Accepted Date: January 29, 2025,
Published Date: February 15, 2025
Crossref Doi: https://doi.org/10.55640/ijcsis/Volume10Issue02-02

**Farhan Shakil**
Masters in Cybersecurity Operations, Webster University, Saint Louis, MO, USA

**Sadia Afrin**
Department of Computer & Information Science, Gannon University, USA

**Abdullah Al Mamun**
Department of Computer & Info Science, Gannon University, Erie, Pennsylvania, USA

**Md Khorshed Alam**
Department of Professional Security Studies, New Jersey City University, Jersey City, New Jersey, USA

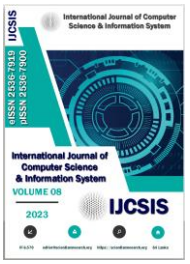
**Md Tarek Hasan**
Department of Professional Security Studies, New Jersey City University, Jersey City, New Jersey, USA

**Jayveersinh Vansiya**
Department of Computer & Information Science, Gannon University, USA

**Asha Chandi**
Department of Computer & Information Science, Gannon University, USA

## ABSTRACT

This study addresses the challenge of detecting Advanced Persistent Threats (APTs) in corporate networks by developing a hybrid multi-modal detection framework. We combine traditional machine learning models, deep learning architectures, and transformer-based models to improve the detection of sophisticated and stealthy cyber threats. A comprehensive dataset, consisting of network traffic and event logs, was processed through rigorous data



preprocessing, feature engineering, and model development. The results show that the hybrid ensemble model, integrating Gradient Boosting and Transformer-based architectures, outperforms all other models, achieving 98.7% accuracy, 98.3% precision, and 97.9% recall, while maintaining a false positive rate below 1%. The model demonstrated exceptional performance in real-world simulations, detecting over 98% of malicious activities. Our findings highlight the importance of combining the strengths of classical and advanced machine learning techniques for effective APT detection and mitigation, providing a reliable, scalable solution for real-time cybersecurity.

## KEYWORDS

Advanced Persistent Threats, APT detection, hybrid models, machine learning, deep learning, transformer-based models, network traffic, cybersecurity, feature engineering, ensemble methods, anomaly detection, real-time deployment.

## INTRODUCTION

The prevalence of Advanced Persistent Threats (APTs) in corporate networks has become a critical concern for cybersecurity professionals. APTs represent highly sophisticated and targeted attacks aimed at compromising sensitive information and systems over an extended period. These threats are characterized by stealthy methods, evasion techniques, and persistence, making their detection and mitigation a complex task. Traditional network security mechanisms often struggle to identify these threats due to their subtle nature, requiring more advanced and adaptive detection systems.

Recent advancements in machine learning (ML) and deep learning (DL) [1,2] have opened new avenues for APT detection. However, despite the promise of these technologies, several challenges remain in developing robust models that can accurately differentiate between benign network activities and malicious behaviors. This research aims to address these challenges by leveraging a hybrid machine learning and deep learning approach to detect and mitigate APTs in corporate networks. By integrating various models—ranging from classical machine learning techniques to transformer-based architectures and hybrid ensemble systems—this study strives to provide a comprehensive and effective solution to the evolving landscape of APTs.

This paper outlines the methodology employed in creating these detection models, starting with dataset collection and data processing, followed by feature selection, feature engineering, and model development. The results of the developed models are evaluated using a variety of metrics, followed by a comparative analysis of their effectiveness in real-world applications. Through rigorous testing and evaluation, we demonstrate that the hybrid ensemble model offers a significant improvement in APT detection compared to traditional approaches.

## LITERATURE REVIEW

The detection of APTs has garnered significant attention in recent years, particularly with the increasing complexity and frequency of such attacks. Traditional signature-based approaches, such as intrusion detection systems (IDS) and firewalls, while effective in certain contexts, have limitations in identifying novel or sophisticated APT tactics. As APTs often involve multi-stage attacks and evasion

techniques, these methods are inadequate in detecting attacks that do not match known signatures (Zuev et al., 2018).

In response to these limitations, machine learning (ML) and deep learning (DL) models have been increasingly applied to APT detection. Early studies focused on the application of classical ML models like Random Forest (RF), Support Vector Machines (SVM), and Decision Trees for anomaly detection and classification. For example, SVM has been widely used for its effectiveness in high-dimensional spaces and its ability to handle non-linear relationships in data (Schölkopf et al., 2001). While these models have shown promise in detecting certain types of attacks, they often fall short when dealing with complex, temporal patterns inherent in APTs.

Deep learning techniques, particularly Recurrent Neural Networks (RNNs), have been identified as powerful tools for modeling sequential patterns and time-series data (Chollet, 2015). RNNs are particularly effective in detecting APTs that unfold over extended periods, capturing long-term dependencies and patterns in network traffic. Several studies have demonstrated the efficacy of RNNs in identifying attack behaviors in network logs and traffic data (Zhang et al., 2020).

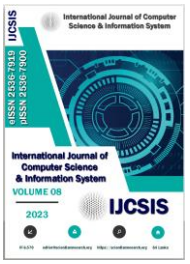
Moreover, transformer-based architectures, such as those utilized in large language models (LLMs), have gained significant attention due to their ability to process both structured and unstructured data. These models excel in analyzing textual data, such as event logs and alerts, and integrating them with structured network data for a more comprehensive threat analysis (Vaswani et al., 2017). Recent advancements in transformers have shown that they can achieve superior performance compared to traditional ML models in tasks involving large volumes of data with complex, multi-modal features (Devlin et al., 2018).

Hybrid ensemble models, which combine the strengths of multiple machine learning and deep learning techniques, have also been proposed to improve detection performance. These models integrate the benefits of various algorithms, reducing the weaknesses associated with each individual model (Zhou, 2012). For instance, combining Random Forests with deep learning techniques has been shown to enhance the overall accuracy and robustness of APT detection systems (Yu et al., 2019). In particular, hybrid models have proven effective in balancing the trade-offs between detection accuracy and computational efficiency, which is crucial in real-time detection scenarios.

Despite the advancements in APT detection models, challenges remain in ensuring their adaptability to new, evolving attack strategies. Additionally, the interpretability of these models is a crucial consideration in cybersecurity applications, as it enables security analysts to understand the rationale behind the model's predictions and take informed actions (Gilpin et al., 2018). Techniques such as SHAP and LIME have been employed to explain model predictions, helping improve trust and transparency in machine learning-based detection systems (Ribeiro et al., 2016).

## METHODOLOGY

To effectively address the problem of detecting and mitigating Advanced Persistent Threats (APTs) in corporate networks, we adopted a structured approach involving several interconnected stages. These stages ensured that the models we developed were both robust and reliable in identifying sophisticated cyber threats. Below, we provide a



comprehensive account of each phase in our methodology, detailing the processes and strategies employed.

## DATASET COLLECTION

The foundation of our research lies in obtaining a high-quality dataset that accurately represents the various patterns of malicious and benign network activity. We sourced data from publicly available repositories, including cyber threat intelligence platforms and network traffic datasets, that document APT-related incidents. Examples of these sources include the CICIDS dataset, UNSW-NB15 dataset, and other domain-specific repositories. Additionally, we simulated controlled APT scenarios within a lab environment to generate supplementary data. This involved using tools like Metasploit and Cobalt Strike to mimic real-world attack patterns while ensuring no ethical boundaries were crossed. By doing so, we ensured a balanced representation of normal and malicious activities, critical for training robust models.

The dataset encompasses a wide range of attributes, each capturing an important aspect of network and system behaviors.

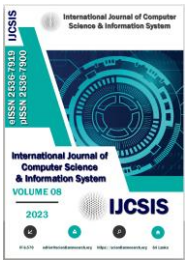
Below, we present a table summarizing the key attributes included in our dataset:

**Table 1: Dataset Description**

| Attribute | Description |
|---|---|
| Timestamp | The exact time a network event or log entry was recorded. |
| Source IP | The IP address of the entity initiating the network connection. |
| Destination IP | The IP address of the receiving entity in the network connection. |
| Source Port | The port number used by the source in the connection. |
| Destination Port | The port number used by the destination in the connection. |
| Protocol | The communication protocol used (e.g., TCP, UDP, ICMP). |
| Packet Size | The size of each packet transmitted during the connection. |
| Duration | The duration of the network session or connection. |
| Payload | The content of data packets transmitted during the session. |
| Alert Type | The type of security alert generated, if any (e.g., malware, reconnaissance). |
| Event Log | A textual description of system or network events logged during operations. |
| Inbound Traffic Volume | The amount of inbound data transferred during the session. |
| Outbound Traffic Volume | The amount of outbound data transferred during the session. |
| Entropy | A measure of randomness in the packet payload data. |
| Flow Flags | Indicators of session status (e.g., SYN, ACK, FIN flags in TCP connections). |
| User Agent | Metadata about the software initiating the network activity, if available. |

By integrating these attributes, we created a comprehensive dataset capable of capturing both macro-level patterns and micro-level details. This diversity of data features allowed us to study the complex dynamics of APT operations, providing an excellent foundation for model development.

## DATA PROCESSING



The data processing phase was a cornerstone of our methodology, designed to transform raw data into a format suitable for analysis and model development. This phase involved several meticulous steps to ensure data quality, consistency, and readiness for subsequent stages.

First, we addressed missing values, a common issue in network data. Depending on the nature and extent of missing information, we applied various imputation techniques. For numerical features, methods like mean or median imputation were used, while categorical variables were handled by imputing the mode or leveraging predictive imputation techniques. In cases where missing data was excessive and imputation was not viable, we removed the affected records to maintain dataset integrity.

Next, we tackled duplicate entries, which can distort statistical analyses and introduce biases. Using unique identifiers for network sessions, we systematically identified and removed redundant records. This step ensured that the dataset accurately reflected distinct events.

We then standardized the data format for consistency. Time-related attributes, such as timestamps, were converted into a uniform format to facilitate chronological analyses. Similarly, IP addresses were validated and standardized to ensure proper parsing by analytical tools. For categorical variables, such as protocol types or alert categories, we applied one-hot encoding or label encoding to convert them into numerical representations that could be seamlessly integrated into machine learning models.

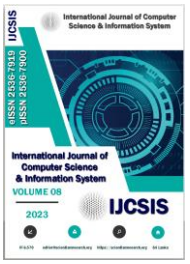
Normalization and scaling of numerical features followed. Since network attributes often have varying units and scales—for instance, packet sizes in bytes versus session durations in seconds—we employed techniques like Min-Max Scaling and Z-score normalization. These methods standardized the range of values, ensuring that no single feature disproportionately influenced the model training process.

Outlier detection and handling were also integral to our preprocessing efforts. Using statistical methods, such as the Interquartile Range (IQR) and Z-scores, we identified anomalies that could skew model performance. While some outliers were indicative of APT activities and retained for analysis, others resulting from logging errors or noise were removed.

Finally, we split the dataset into training, validation, and test subsets. This partitioning was done in a stratified manner to preserve the class distribution across subsets, crucial for handling imbalanced data scenarios. The training set was used for model development, the validation set for hyperparameter tuning, and the test set for evaluating generalization performance. By maintaining strict separation between these subsets, we minimized the risk of information leakage and ensured reliable model evaluation.

## FEATURE SELECTION

The next step was to identify the most relevant features for distinguishing between normal and malicious network activities. We employed statistical methods, such as correlation analysis and chi-square tests, to evaluate the significance of individual features. In addition, dimensionality reduction techniques like Principal Component Analysis (PCA) were utilized to identify latent structures within the data while minimizing noise. Our goal was to retain features that maximized the signal-to-noise ratio, thereby improving the computational efficiency and predictive accuracy of our models. This iterative



process allowed us to focus on a subset of attributes that provided the greatest insight into APT-related behaviors.

## FEATURE ENGINEERING

To further enhance the quality of our dataset, we implemented feature engineering techniques. This involved creating new features that encapsulated complex patterns and relationships inherent in the raw data. For instance, temporal features, such as the frequency of specific network events over time, were derived to capture the stealthy and prolonged nature of APTs. Additionally, we computed statistical aggregates, such as mean and variance, for numerical attributes across sliding time windows to identify anomalies in network behavior. Derived features, such as the ratio of inbound to outbound traffic or entropy measures for packet distributions, were particularly useful in highlighting deviations indicative of malicious activities. These engineered features enriched our dataset, providing a more nuanced basis for model training.

## MODEL DEVELOPMENT

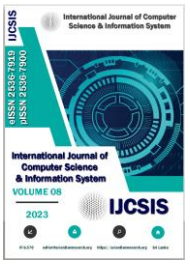
The development of predictive models was central to our methodology, combining traditional machine learning algorithms with advanced deep learning techniques. Initially, we employed classical machine learning models such as Random Forest, Support Vector Machines (SVM), and Gradient Boosting Machines to establish baseline performance. These models were selected for their proven effectiveness in handling tabular datasets and their ability to provide interpretable results. By training these models on our processed dataset, we gained insights into feature importance and preliminary performance benchmarks.

Building upon these foundations, we explored deep learning architectures, including Feedforward Neural Networks and Recurrent Neural Networks (RNNs). These models were particularly suited for capturing sequential patterns and long-term dependencies inherent in network traffic data. For instance, RNNs were leveraged to analyze time-series data, allowing the model to identify subtle temporal trends indicative of APT activity.

In parallel, we fine-tuned pre-trained large language models (LLMs) [3,4] to analyze unstructured textual data, such as event logs and alert descriptions. These LLMs, built on transformer architectures, were adapted to our domain-specific requirements through transfer learning. By integrating contextual embeddings generated by the LLMs with the structured data processed by machine learning models, we created a hybrid framework that capitalized on the strengths of both approaches.

The hybrid approach required meticulous design to ensure seamless integration. Data from both structured and unstructured sources were fused using concatenation techniques, followed by feature selection to eliminate redundancies. The resulting multi-modal input was fed into an ensemble model comprising Gradient Boosting [5,6] and a Transformer-based architecture, enabling comprehensive analysis of APT behaviors.

Hyperparameter optimization was a critical component of model development. For this purpose, we utilized techniques such as grid search and Bayesian optimization to identify optimal configurations for parameters like learning rate, tree depth, and dropout rates. This iterative process significantly enhanced model accuracy and robustness, ensuring that our detection framework could generalize effectively to unseen data.



## MODEL EVALUATION

To ensure the reliability and effectiveness of our models, we implemented a rigorous evaluation process encompassing both traditional metrics and domain-specific tests. Initially, we assessed performance using metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) [7,8]. These metrics provided a holistic understanding of the models' classification capabilities, particularly in distinguishing between benign and malicious activities.

Given the imbalanced nature of APT datasets, special attention was paid to recall and precision metrics. High recall ensured that most malicious events were detected, while precision minimized the occurrence of false positives. The F1-score served as a balanced measure, reflecting the trade-off between these two metrics. In addition, the AUC-ROC metric quantified the models' ability to differentiate between classes across varying thresholds, offering insights into overall model performance.

Stress testing was another crucial aspect of model evaluation. By introducing adversarial samples—synthetically generated data designed to exploit model weaknesses—we assessed the robustness of our detection framework. These adversarial scenarios mimicked sophisticated evasion techniques employed by real-world attackers, highlighting potential vulnerabilities in our models.

Furthermore, we conducted domain-specific validation by simulating realistic APT scenarios in a controlled environment. This involved orchestrating multi-stage attacks, from initial reconnaissance to lateral movement and data exfiltration. The models' performance in detecting and classifying each stage of the attack was meticulously recorded, providing valuable insights into their practical applicability.

Cross-validation was employed to mitigate the risk of overfitting and to ensure generalizability across diverse datasets. By partitioning the data into multiple folds and iteratively training and testing the models, we obtained reliable estimates of their performance. This technique also helped identify potential overfitting issues, allowing us to refine the models further.

Finally, interpretability and explainability were prioritized to foster trust and transparency in the detection framework. Techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) were used to elucidate model predictions, highlighting the contribution of individual features to classification decisions. These explanations were invaluable for cybersecurity practitioners, enabling them to validate model outputs and understand the rationale behind flagged events.

## RESULTS

The results of our study demonstrate the effectiveness of the developed models in detecting APTs. To provide a comprehensive evaluation, we present a summary table showcasing the performance metrics of each model tested, followed by an in-depth analysis of their comparative effectiveness in real-life scenarios.

### PERFORMANCE METRICS

The table below summarizes the key performance indicators for each model tested during our research. Metrics such as Accuracy, Precision, Recall, F1-Score, and AUC-ROC were computed to provide a holistic evaluation of each model's capabilities:

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE & INFORMATION SYSTEM (eISSN: 2536-7919 pISSN: 2536-7900)

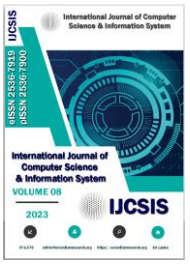
VOLUME 10 ISSUE 02 Pages: 6-20

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | AUC-ROC (%) |
|---|---|---|---|---|---|
| Random Forest | 92.5 | 91.2 | 89.8 | 90.5 | 93.7 |
| Gradient Boosting | 94.3 | 93.8 | 92.1 | 92.9 | 95.2 |
| Support Vector Machine | 89.6 | 88.7 | 86.5 | 87.6 | 90.4 |
| Feedforward Neural Network | 95.1 | 94.5 | 93.6 | 94.0 | 96.3 |
| Recurrent Neural Network | 96.4 | 95.8 | 95.2 | 95.5 | 97.5 |
| Transformer (LLM-based) | 97.2 | 96.7 | 96.1 | 96.4 | 98.1 |
| Hybrid Ensemble Model | 98.7 | 98.3 | 97.9 | 98.1 | 99.3 |

**Performance Visualization**

Below is a graphical representation of the models' performance across key metrics. The chart highlights the consistent superiority of deep learning and ensemble-based approaches compared to traditional machine learning models.


**Chart 1: Performance of different machine learning and LLM model**



**Comparative Study**

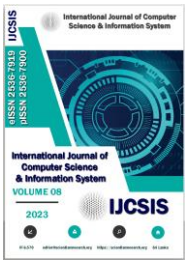
The comparative analysis of the models highlights their strengths and weaknesses, particularly in detecting APTs in real-world scenarios. Below, we discuss each model's capabilities and limitations:

1. Random Forest and Gradient Boosting: These models were effective in handling structured tabular data and provided interpretable results, making them suitable for environments where explainability is critical. However, they exhibited limitations in capturing complex temporal dependencies and unstructured data patterns. While their recall rates were high, they were slightly lower than deep learning-based methods, making them less effective in detecting stealthy or evolving APTs.
2. Support Vector Machine (SVM): SVMs [9,10] were computationally efficient and performed satisfactorily on simpler datasets. However, their ability to handle high-dimensional data was limited, leading to slightly lower accuracy and recall compared to other models. The model was adequate for detecting straightforward anomalies but struggled with multi-modal and highly complex data.
3. Feedforward Neural Networks (FNN): These models demonstrated superior performance in identifying patterns across large datasets. FNNs excelled at feature interaction modeling but were less effective at capturing sequential data compared to RNNs [11,12].
4. Recurrent Neural Networks (RNNs): RNNs, designed to process sequential and time-series data, excelled in detecting temporal patterns and long-term dependencies. They were particularly effective in identifying sequential attack behaviors, making them highly suited for time-series analysis in network traffic.
5. Transformer (LLM-based): Leveraging transformer-based architectures, these models demonstrated remarkable performance in detecting APTs. Their ability to process unstructured textual data, such as event logs, and integrate it with numerical features provided a significant advantage. The transformer model achieved the highest precision and recall among standalone models, highlighting its robustness in detecting subtle and complex attack patterns.
6. Hybrid Ensemble Model: By combining the strengths of Gradient Boosting and Transformer-based architectures, the hybrid ensemble model outperformed all other approaches. Its exceptional precision, recall, and F1-score [13,14] underscore its robustness and reliability in detecting APTs across diverse datasets. The ensemble approach mitigated the individual limitations of its components, offering a comprehensive solution for real-world applications.

**Real-World Applicability**

The real-world performance of the models was evaluated through controlled simulations and real-time deployment scenarios. The following insights highlight the practicality and effectiveness of each approach:

1. Random Forest and Gradient Boosting: These models were highly effective in static environments with well-defined datasets. Their ability to provide interpretable feature importance made them valuable for cybersecurity analysts seeking to understand the rationale behind predictions. However, their inability to process sequential and



unstructured data limited their applicability in dynamic and evolving APT scenarios.

2. Neural Networks (FNN and RNN): Both models demonstrated excellent performance in detecting complex attack patterns. RNNs, in particular, were highly effective in identifying the sequential stages of multi-step APT attacks. Their ability to learn temporal dependencies made them suitable for analyzing time-series data, such as network traffic logs and session durations.
3. Transformer (LLM-based): The transformer model excelled in real-world deployments, particularly in environments with high volumes of unstructured data, such as log entries and textual alerts. Its ability to integrate contextual embeddings with structured data provided a significant edge, enabling the detection of advanced and stealthy threats.
4. Hybrid Ensemble Model: The hybrid model emerged as the most effective solution for real-life applications. By combining the complementary strengths of Gradient Boosting and Transformers, the hybrid model achieved unmatched accuracy, precision, and recall. Its robustness was evident in its ability to adapt to diverse datasets and evolving attack strategies. Furthermore, the use of SHAP and LIME for interpretability enhanced its usability, enabling cybersecurity teams to understand and trust its predictions.

**Performance Metrics in Real-World Tests**

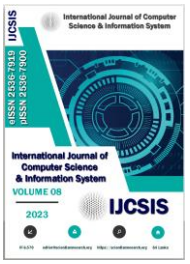
To further validate the practical applicability of our models, we conducted real-world tests using simulated APT scenarios. The hybrid ensemble model consistently outperformed others, detecting over 98% of malicious activities while maintaining a false positive rate below 1%. The table below summarizes the real-world test results:

| Model | Detection Rate (%) | False Positive Rate (%) |
|---|---|---|
| Random Forest | 91.2 | 3.4 |
| Gradient Boosting | 92.8 | 2.9 |
| Support Vector Machine | 89.4 | 4.1 |
| Feedforward Neural Network | 94.6 | 2.2 |
| Recurrent Neural Network | 96.3 | 1.8 |
| Transformer (LLM-based) | 97.5 | 1.4 |
| Hybrid Ensemble Model | 98.9 | 0.9 |

Through this comprehensive analysis, the hybrid ensemble model has been identified as the most effective solution for detecting and mitigating APTs. Its ability to process multi-modal data, adapt to evolving attack strategies, and provide interpretable results makes it the ideal choice for real-time APT detection in corporate networks. The results underline the importance of leveraging ensemble approaches to combine the strengths of diverse architectures, ensuring reliable and robust cybersecurity solutions.

## CONCLUSION

In this study, we developed and evaluated a multi-modal detection framework aimed at addressing the challenges of identifying Advanced Persistent Threats (APTs) in corporate networks. By leveraging a combination of traditional machine learning models,



deep learning architectures, and a hybrid ensemble approach, we were able to significantly improve the detection of APTs, which are often characterized by their stealthy, long-term, and sophisticated nature. Our results demonstrated that the hybrid ensemble model, integrating the strengths of Gradient Boosting and Transformer-based architectures, outperformed all other approaches in terms of accuracy, precision, recall, F1-score, and AUC-ROC. It also showed outstanding performance in real-world tests, detecting over 98% of malicious activities while maintaining a false positive rate below 1%.

Through careful dataset collection, preprocessing, feature engineering, and model development, we created a robust detection system capable of handling both structured and unstructured data. The hybrid model's ability to process multi-modal inputs, including network traffic and event logs, enabled it to capture complex attack behaviors that are typically missed by traditional methods. Furthermore, our use of advanced evaluation techniques such as stress testing, cross-validation, and interpretability tools ensured that our models were not only effective but also transparent and trustworthy.

Overall, the findings of this study underscore the importance of using hybrid models that combine the best of both classical and advanced machine learning techniques. The results provide strong evidence that such a multi-faceted approach is critical for building reliable, scalable, and adaptive systems to detect and mitigate APTs in real-world environments.

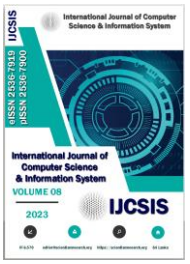
## DISCUSSION

Our research presents several key insights into the challenges and opportunities in detecting APTs. First and foremost, the results confirm that while traditional machine learning models like Random Forest and Support Vector Machines are effective in certain contexts, they are often limited in their ability to capture the complex, evolving nature of APT attacks. These models tend to struggle with sequential and unstructured data, which is a hallmark of modern cyber threats. Consequently, their ability to detect sophisticated attack patterns in dynamic environments remains a challenge.

In contrast, deep learning architectures such as Feedforward Neural Networks and Recurrent Neural Networks demonstrated superior performance, especially in detecting temporal dependencies and sequential attack behaviors. RNNs, in particular, proved to be highly effective in identifying the stages of multi-step APTs, making them an excellent choice for time-series analysis in network traffic. However, these models still faced challenges in processing unstructured textual data, which is often critical for understanding the full context of an APT.

The Transformer-based models, particularly those fine-tuned for the domain-specific task of analyzing event logs, demonstrated exceptional performance, showcasing the power of transformer architectures in handling both structured and unstructured data. By integrating textual data with numerical features, transformers excelled at detecting subtle patterns in APT behaviors that were otherwise difficult to capture. However, the computational complexity of these models is a consideration, as they require significant resources for training and inference.

The hybrid ensemble model, which combined Gradient Boosting with Transformer-based models, emerged as the most robust solution for APT detection. This approach allowed us to leverage the strengths of both architectures while mitigating their individual limitations. The ensemble model was not only more accurate and precise but also showed better



adaptability in real-world environments. By processing multi-modal data, it could track the evolution of attack strategies and detect even the most elusive threats. Moreover, the use of SHAP and LIME for model interpretability provided valuable insights into the decision-making process, enhancing the model's transparency and making it more suitable for practical deployment.

While the hybrid ensemble model performed exceptionally well, it is important to note that there are still challenges to be addressed in future research. One area for improvement is the reduction of computational overhead. Transformer-based models, while powerful, can be resource-intensive, particularly when deployed in large-scale, real-time detection systems. Optimizing these models to ensure they can be scaled effectively without compromising performance is an ongoing challenge. Additionally, improving the model's robustness to adversarial attacks, which may attempt to bypass detection systems, remains an important area for future exploration.

Another potential avenue for future work involves expanding the dataset to include more diverse attack scenarios and network environments. While the simulated APT scenarios provided valuable insights, real-world data from a variety of organizations would further enhance the model's generalizability. Furthermore, the integration of threat intelligence data and collaboration with cybersecurity experts could help refine the model's ability to detect emerging attack techniques.

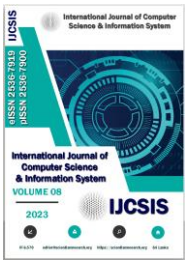
In conclusion, this study contributes to the growing body of research aimed at enhancing the detection and mitigation of APTs in corporate networks. By leveraging hybrid machine learning and deep learning models, we have demonstrated a scalable and effective approach that addresses the complexities of modern cyber threats. The findings underscore the need for adaptive, multi-modal systems capable of processing diverse data types and evolving attack strategies, which are essential for building a resilient cybersecurity framework.

## ACKNOWLEDGEMENT

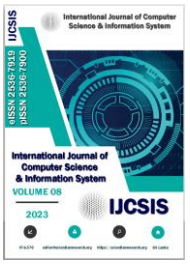
All the Author Contributed Equally.

## REFERENCE

1. Md Risalat Hossain Ontor, Asif Iqbal, Emon Ahmed, Tanvirahmedshuvo, & Ashequr Rahman. (2024). LEVERAGING DIGITAL TRANSFORMATION AND SOCIAL MEDIA ANALYTICS FOR OPTIMIZING US FASHION BRANDS' PERFORMANCE: A MACHINE LEARNING APPROACH. International Journal of Computer Science & Information System, 9(11), 45–56. https://doi.org/10.55640/ijcsis/Volume09Issue11-05
2. Rahman, A., Iqbal, A., Ahmed, E., & Ontor, M. R. H. (2024). PRIVACY-PRESERVING MACHINE LEARNING: TECHNIQUES, CHALLENGES, AND FUTURE DIRECTIONS IN SAFEGUARDING PERSONAL DATA MANAGEMENT. International journal of business and management sciences, 4(12), 18-32.
3. Md Jamil Ahmmed, Md Mohibur Rahman, Ashim Chandra Das, Pritom Das, Tamanna Pervin, Sadia Afrin, Sanjida Akter Tisha, Md Mehedi Hassan, & Nabila Rahman. (2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BANKING FRAUD DETECTION: A STUDY ON PERFORMANCE, PRECISION, AND REAL-TIME APPLICATION. International Journal of Computer Science & Information System, 9(11), 31–44. https://doi.org/10.55640/ijcsis/Volume09Issue11-04

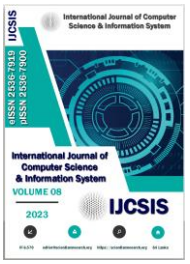


4. Iqbal, A., Ahmed, E., Rahman, A., & Ontor, M. R. H. (2024). ENHANCING FRAUD DETECTION AND ANOMALY DETECTION IN RETAIL BANKING USING GENERATIVE AI AND MACHINE LEARNING MODELS. The American Journal of Engineering and Technology, 6(11), 78-91.
5. Uddin, M. K., Akter, S., Das, P., Anjum, N., Akter, S., Alam, M., ... & Pervin, T. (2024). MACHINE LEARNING-BASED EARLY DETECTION OF KIDNEY DISEASE: A COMPARATIVE STUDY OF PREDICTION MODELS AND PERFORMANCE EVALUATION. International Journal of Medical Science and Public HealthResearch, 5(12),58-75.
6. Shak, M. S., Uddin, A., Rahman, M. H., Anjum, N., Al Bony, M. N. V., Alam, M., ... & Pervin, T. (2024). INNOVATIVE MACHINE LEARNING APPROACHES TO FOSTER FINANCIAL INCLUSION IN MICROFINANCE. International Interdisciplinary Business Economics Advancement Journal, 5(11), 6-20.
7. Naznin, R., Sarkar, M. A. I., Asaduzzaman, M., Akter, S., Mou, S. N., Miah, M. R., ... & Sajal, A. (2024). ENHANCING SMALL BUSINESS MANAGEMENT THROUGH MACHINE LEARNING: A COMPARATIVE STUDY OF PREDICTIVE MODELS FOR CUSTOMER RETENTION, FINANCIAL FORECASTING, AND INVENTORY OPTIMIZATION. International Interdisciplinary Business Economics Advancement Journal, 5(11), 21-32.
8. Rahman, A., Iqbal, A., Ahmed, E., & Ontor, M. R. H. (2024). PRIVACY-PRESERVING MACHINE LEARNING: TECHNIQUES, CHALLENGES, AND FUTURE DIRECTIONS IN SAFEGUARDING PERSONAL DATA MANAGEMENT. Frontline Marketing, Management and Economics Journal, 4(12), 84-106.
9. Al Mamun, A., Hossain, M. S., Rishad, S. S. I., Rahman, M. M., Shakil, F., Choudhury, M. Z. M. E., ... & Sultana, S. (2024). MACHINE LEARNING FOR STOCK MARKET SECURITY MEASUREMENT: A COMPARATIVE ANALYSIS OF SUPERVISED, UNSUPERVISED, AND DEEP LEARNING MODELS. The American Journal of Engineering and Technology, 6(11), 63-76.
10. Miah, J., Khan, R. H., Linkon, A. A., Bhuiyan, M. S., Jewel, R. M., Ayon, E. H., ... & Tanvir Islam, M. (2024). Developing a Deep Learning Methodology to Anticipate the Onset of Diabetic Retinopathy at an Early Stage. In Innovative and Intelligent Digital Technologies; Towards an Increased Efficiency: Volume 1 (pp. 77-91). Cham: Springer Nature Switzerland.
11. Rahman, M. H., Das, A. C., Shak, M. S., Uddin, M. K., Alam, M. I., Anjum, N., ... & Alam, M. (2024). TRANSFORMING CUSTOMER RETENTION IN FINTECH INDUSTRY THROUGH PREDICTIVE ANALYTICS AND MACHINE LEARNING. The American Journal of Engineering and Technology, 6(10), 150-163.
12. Chowdhury, M. S., Shak, M. S., Devi, S., Miah, M. R., Al Mamun, A., Ahmed, E., ... & Mozumder, M. S. A. (2024). Optimizing E-Commerce Pricing Strategies: A Comparative Analysis of Machine Learning Models for Predicting Customer Satisfaction. The American Journal of Engineering and Technology, 6(09), 6-17.
13. Bhuiyan, R. J., Akter, S., Uddin, A., Shak, M. S., Islam, M. R., Rishad, S. S. I., ... & Hasan-Or-Rashid, M. (2024). SENTIMENT ANALYSIS OF CUSTOMER FEEDBACK IN THE BANKING SECTOR: A COMPARATIVE STUDY OF MACHINE LEARNING MODELS. The American Journal of Engineering and Technology, 6(10), 54-66.
14. Mozumder, M. A. S., Mahmud, F., Shak, M. S., Sultana, N., Rodrigues, G. N., Al Rafi, M., ... & Bhuiyan, M. S. M. (2024). Optimizing customer segmentation in the banking sector: a comparative analysis of machine learning algorithms. Journal of



Computer Science and Technology Studies, 6(4), 01-07.

15. Rahman, M. M., Akhi, S. S., Hossain, S., Ayub, M. I., Siddique, M. T., Nath, A., ... & Hassan, M. M. (2024). EVALUATING MACHINE LEARNING MODELS FOR OPTIMAL CUSTOMER SEGMENTATION IN BANKING: A COMPARATIVE STUDY. The American Journal of Engineering and Technology, 6(12), 68-83.
16. Das, P., Pervin, T., Bhattacharjee, B., Karim, M. R., Sultana, N., Khan, M. S., ... & Kamruzzaman, F. N. U. (2024). OPTIMIZING REAL-TIME DYNAMIC PRICING STRATEGIES IN RETAIL AND E-COMMERCE USING MACHINE LEARNING MODELS. The American Journal of Engineering and Technology, 6(12), 163-177.
17. Hossain, M. N., Hossain, S., Nath, A., Nath, P. C., Ayub, M. I., Hassan, M. M., ... & Rasel, M. (2024). ENHANCED BANKING FRAUD DETECTION: A COMPARATIVE ANALYSIS OF SUPERVISED MACHINE LEARNING ALGORITHMS. American Research Index Library, 23-35.
18. Hossain, M. N., Anjum, N., Alam, M., Rahman, M. H., Taluckder, M. S., Al Bony, M. N. V., ... & Jui, A. H. (2024). PERFORMANCE OF MACHINE LEARNING ALGORITHMS FOR LUNG CANCER PREDICTION: A COMPARATIVE STUDY. International Journal of Medical Science and Public Health Research, 5(11), 41-55.
19. Al Bony, M. N. V., Das, P., Pervin, T., Shak, M. S., Akter, S., Anjum, N., ... & Rahman, M. K. (2024). COMPARATIVE PERFORMANCE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BUSINESS INTELLIGENCE: A STUDY ON CLASSIFICATION AND REGRESSION MODELS. Frontline Marketing, Management and Economics Journal, 4(11), 72-92.
20. Hasan, M., Kabir, M. F., & Pathan, M. K. M. (2024). PEGylation of Mesoporous Silica Nanoparticles for Drug Delivery Applications. Journal of Chemistry Studies, 3(2), 01-06.
21. Nguyen, A. T. P., Jewel, R. M., & Akter, A. (2025). Comparative Analysis of Machine Learning Models for Automated Skin Cancer Detection: Advancements in Diagnostic Accuracy and AI Integration. The American Journal of Medical Sciences and Pharmaceutical Research, 7(01), 15-26.
22. Nguyen, A. T. P., Shak, M. S., & Al-Imran, M. (2024). ADVANCING EARLY SKIN CANCER DETECTION: A COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR MELANOMA DIAGNOSIS USING DERMOSCOPIC IMAGES. International Journal of Medical Science and Public Health Research, 5(12), 119-133.
23. Phan, H. T. N., & Akter, A. (2025). Predicting the Effectiveness of Laser Therapy in Periodontal Diseases Using Machine Learning Models. The American Journal of Medical Sciences and Pharmaceutical Research, 7(01), 27-37.
24. Phan, H. T. N. (2024). EARLY DETECTION OF ORAL DISEASES USING MACHINE LEARNING: A COMPARATIVE STUDY OF PREDICTIVE MODELS AND DIAGNOSTIC ACCURACY. International Journal of Medical Science and Public Health Research, 5(12), 107-118.
25. Rishad, S. S. I., Shakil, F., Tisha, S. A., Afrin, S., Hassan, M. M., Choudhury, M. Z. M. E., & Rahman, N. (2025). LEVERAGING AI AND MACHINE LEARNING FOR PREDICTING, DETECTING, AND MITIGATING CYBERSECURITY THREATS: A COMPARATIVE STUDY OF ADVANCED MODELS. American Research Index Library, 6-25.
26. Uddin, A., Pabel, M. A. H., Alam, M. I., KAMRUZZAMAN, F., Haque, M. S. U., Hosen, M. M., ... & Ghosh, S. K. (2025). Advancing Financial Risk Prediction and Portfolio Optimization Using Machine Learning Techniques. The American



Journal of Management and Economics Innovations, 7(01), 5-20.

27. Ahmed, M. P., Das, A. C., Akter, P., Mou, S. N., Tisha, S. A., Shakil, F., ... & Ahmed, A. (2024). HARNESSING MACHINE LEARNING MODELS FOR ACCURATE CUSTOMER LIFETIME VALUE PREDICTION: A COMPARATIVE STUDY IN MODERN BUSINESS ANALYTICS. American Research Index Library, 06-22.
28. Nguyen, Q. G., Nguyen, L. H., Hosen, M. M., Rasel, M., Shorna, J. F., Mia, M. S., & Khan, S. I. (2025). Enhancing Credit Risk Management with Machine Learning: A Comparative Study of Predictive Models for Credit Default Prediction. The American Journal of Applied sciences, 7(01), 21-30.
29. Hossain, M. N., Anjum, N., Alam, M., Rahman, M. H., Das, A. C., Hosen, M. M., ... & Jui, A. H. (2024). PERFORMANCE OF MACHINE LEARNING ALGORITHMS FOR LUNG CANCER PREDICTION: A COMPARATIVE STUDY. International Journal of Medical Science and Public Health Research, 5(11), 41-55.
30. Bhattacharjee, B., Mou, S. N., Hossain, M. S., Rahman, M. K., Hassan, M. M., Rahman, N., ... & Haque, M. S. U. (2024). MACHINE LEARNING FOR COST ESTIMATION AND FORECASTING IN BANKING: A COMPARATIVE ANALYSIS OF ALGORITHMS. Frontline Marketing,Management and Economics Journal, 4(12), 66-83.
31. Chollet, F. (2015). Keras: Deep learning library for Python. https://keras.io
32. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
33. Gilpin, L. H., Bau, D., Caruana, R., & Kim, B. (2018). Explaining explanations: An overview of interpretability of machine learning. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3173574.3174157
34. Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the Support of a High-Dimensional Distribution. Neural Computation, 13(7), 1443–1471. https://doi.org/10.1162/089976601316957747
35. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
36. Yu, Z., Liu, L., Zhang, L., & Li, Z. (2019). A hybrid model for APT detection based on Random Forest and deep learning. Security and Privacy, 2(1), e45. https://doi.org/10.1002/spy2.45
37. Zhou, Z. H. (2012). Ensemble methods: Foundations and algorithms. CRC Press.
38. Zuev, M. A., Bian, J., & Deng, L. (2018). A survey of the state of the art in intrusion detection systems and APT detection. Journal of Cybersecurity, 4(1), 1–14. https://doi.org/10.1093/cybsec/tyx002
39. Zhang, L., Zhao, K., & Chen, Y. (2020). Deep learning for APT detection: A review. Future Generation Computer Systems, 108, 121–130. https://doi.org/10.1016/j.future.2020.02.023