Enhancing Enterprise Security Management Using Hybrid Machine Learning and Large Language Model–Assisted Intrusion Detection

[1]**Mohammad Musa Mia**
[1]Master of Business Administration, International American University, Los Angeles, California

[2]**Md Mohibur Rahman**
[2]Fred DeMatteis School of Engineering and Applied Science, Hofstra University, USA

[3]**Md Abu Sayed**
[3]Department of Professional Security Studies, New Jersey City University, Jersey City, New Jersey, USA

[4]**Rumana Akther Nipa**
[4]Master of Science in Engineering Management, College of Engineer & Technology, Westcliff University, Irvine, California

[5]**Sonjoy Kumar Dey**
[5]McComish Department of Electrical Engineering and Computer Science, South Dakota State University, USA

[6]**Kazi Abu Jahed**
[6]Master of Science in Business Intelligence and Analytics, Saint Joseph's University (SJU), USA

[7]**Md Yassir Mottalib**
[7]Master of Science in Information System Technology, Wilmington University, USA

Received: 11 Jan 2026 | Received Revised Version: 24 Jan 2026 | Accepted: 11 Feb 2026 | Published: 25 Feb 2026

Volume 08 Issue 02 2026 | Crossref DOI: 10.37547/tajet/Volume08Issue02-17

Abstract

Enterprise security management faces increasing challenges due to the growing complexity of corporate networks and the sophistication of cyberattacks. Traditional intrusion detection systems, while effective at identifying known threats, often struggle with novel attacks and lack interpretability, resulting in alert fatigue and delayed responses. In this study, I propose a hybrid framework that combines ensemble-based machine learning intrusion detection with large language model–assisted contextual reasoning to enhance both detection accuracy and explain ability. Using the CICIDS2017 dataset, I evaluate baseline classifiers including logistic regression, support vector machines, random forest, and gradient boosting, and compare them with the proposed hybrid architecture. Experimental results demonstrate that the hybrid model outperforms traditional approaches, achieving the highest accuracy, precision, recall, F1-score, and area under the ROC curve. Beyond quantitative improvements, the large language model layer provides semantic explanations of detected threats, reduces false positives, and supports decision-making in enterprise security operations. This approach is particularly suitable for U.S. corporate environments, where real-time monitoring, interpretability, and compliance are critical. The findings highlight the potential of integrating advanced machine learning with contextual intelligence to create scalable, explainable, and operationally viable enterprise security solutions.

Keywords: Enterprise Security Management, Intrusion Detection, Large Language Models, Hybrid Machine Learning, CICIDS2017, Explainable AI, Cybersecurity Analytics.

Cite This Article: Mia, M. M., Rahman, M. M., Sayed, M. A., Nipa, R. A., Dey, S. K., Jahed, K. A., & Mottalib, M. Y. (2026). Enhancing Enterprise Security Management Using Hybrid Machine Learning and Large Language Model–Assisted Intrusion Detection. The American Journal of Engineering and Technology, 8(2), 170–178. https://doi.org/10.37547/tajet/Volume08Issue02-17

Introduction

Enterprise security management has become a critical challenge for organizations as digital transformation, cloud adoption, and remote work continue to expand the attack surface of corporate networks. Modern enterprises generate massive volumes of heterogeneous network traffic and security logs on a continuous basis, making manual monitoring and rule-based security mechanisms increasingly ineffective. At the same time, cyberattacks have grown more sophisticated, adaptive, and stealthy, often bypassing traditional perimeter defenses and signature-based intrusion detection systems. These trends demand intelligent, scalable, and explainable security solutions that can operate effectively within complex enterprise environments.

Machine learning–based intrusion detection systems have emerged as a promising alternative to traditional approaches by enabling automated analysis of network traffic and behavioral patterns. Prior research has demonstrated that supervised learning models, particularly ensemble-based techniques, can achieve high detection accuracy when trained on flow-level network data. However, despite their strong performance, these models often function as black boxes and generate large numbers of alerts without sufficient contextual explanation. In enterprise security operations centers, this lack of interpretability contributes directly to alert fatigue, delayed response times, and reduced analyst trust in automated systems.

Recent advances in large language models offer new opportunities to address these limitations. Large language models are capable of contextual reasoning, semantic interpretation, and natural language generation, enabling them to explain complex patterns and synthesize actionable insights from structured and unstructured data. In the context of enterprise security management, these capabilities can be leveraged to augment traditional intrusion detection systems by providing human-readable explanations, prioritizing alerts, and supporting decision-making processes. Despite their potential, the integration of large language models into enterprise-scale intrusion detection remains underexplored, particularly in terms of empirical evaluation using realistic network traffic datasets.

This article proposes a hybrid enterprise security management framework that combines ensemble-based machine learning intrusion detection with large language model–assisted contextual reasoning. The proposed approach is designed to preserve the efficiency and scalability of conventional detection models while enhancing interpretability and operational usability through natural language analysis. By leveraging a realistic enterprise intrusion detection dataset, this study systematically evaluates detection performance, comparative effectiveness across models, and practical applicability within U.S. corporate environments.

The primary contributions of this work are threefold. First, I present a comprehensive evaluation of traditional machine learning models for enterprise intrusion detection using a large-scale, open-source dataset. Second, I introduce a large language model–assisted hybrid architecture that improves detection reliability and explainability without disrupting existing enterprise security infrastructures. Third, I demonstrate how the proposed approach aligns with operational requirements and regulatory expectations in U.S. corporate industries, highlighting its potential for real-world deployment.

Through this research, I aim to bridge the gap between high-performance intrusion detection and actionable enterprise security management by integrating advanced machine learning with contextual intelligence. The findings provide both technical insights and practical guidance for organizations seeking to enhance their cybersecurity posture using large language models.

Literature Review

Enterprise security management has evolved significantly over the past two decades due to the increasing complexity of corporate networks and the growing sophistication of cyber threats. Early research in intrusion detection systems primarily relied on signature-based and rule-based approaches, which were effective against known attacks but struggled to detect zero-day exploits and evolving threat patterns. As enterprise infrastructures expanded and network traffic volumes increased, researchers began exploring data-driven approaches to improve detection accuracy and adaptability.

Machine learning techniques have played a central role in modern intrusion detection research. Several studies demonstrated that supervised learning models such as logistic regression, support vector machines, decision trees, and ensemble methods can effectively identify malicious network behavior when trained on flow-based datasets. Research leveraging publicly available datasets from repositories such as the UCI Machine Learning Repository and the Kaggle has shown that ensemble models, particularly random forest and gradient boosting, consistently outperform linear classifiers in detecting complex attack patterns. These approaches, however, often suffer from limited interpretability, making them difficult to operationalize in enterprise environments that require transparency and auditability.

To address scalability and real-world applicability, several researchers adopted deep learning architectures such as convolutional neural networks and long short-term memory networks. These models demonstrated strong performance in capturing temporal and spatial patterns in network traffic, particularly for distributed denial-of-service and botnet detection. Despite their high detection accuracy, deep learning models introduce significant computational overhead and often function as black boxes, limiting their acceptance in enterprise security operations where explainability is critical.

More recent literature has highlighted the importance of contextual reasoning and human-centric security analytics. Studies have shown that security analysts face alert fatigue due to high false-positive rates generated by automated detection systems. This challenge has led to growing interest in explainable artificial intelligence and natural language–driven security analysis. Large language models have emerged as promising tools for enhancing cybersecurity workflows by enabling semantic reasoning, alert summarization, and threat explanation. Research has demonstrated that language models can analyze structured security logs, generate incident reports, and support decision-making in security operations centers.

Several studies have explored the integration of large language models with traditional machine learning pipelines. Rather than replacing numeric classifiers, these hybrid approaches use language models as reasoning layers that interpret detection outputs, correlate events, and generate actionable insights. This paradigm aligns closely with enterprise security management frameworks promoted by organizations such as the National Institute of Standards and Technology, which emphasize automation, interpretability, and risk-based decision-making. However, existing literature remains limited in systematically evaluating large language model–assisted intrusion detection using realistic enterprise-scale datasets.

This study builds upon prior work by combining ensemble-based intrusion detection with large language model–assisted contextual analysis. Unlike earlier approaches that focus solely on detection accuracy, this research emphasizes enterprise applicability, interpretability, and operational value. By evaluating the proposed framework on a realistic intrusion detection dataset and analyzing both quantitative performance and qualitative insights, this work contributes to the emerging body of research on intelligent enterprise security management.

Methodology

Data Collection

In this research, I adopted an open-source dataset to ensure methodological transparency, reproducibility, and relevance to real-world enterprise security environments. The dataset selected for this study is the CICIDS2017 intrusion detection dataset, which is publicly available through the Kaggle repository and was originally created by the Canadian Institute for Cybersecurity. This dataset was chosen because it closely simulates modern enterprise network traffic, incorporating both benign user activities and diverse cyberattack behaviors under controlled yet realistic conditions.

The CICIDS2017 dataset was generated within a simulated enterprise network environment that includes firewalls, routers, servers, and multiple client machines. Network traffic was captured using flow-based monitoring tools, enabling the extraction of fine-grained statistical features from packet-level communications. The dataset spans multiple days of activity, each day corresponding to distinct attack scenarios such as denial-of-service attacks, brute-force authentication attempts, web-based exploits, botnet activity, and port scanning. This temporal diversity makes the dataset particularly suitable for evaluating enterprise-scale security management solutions powered by large language models.

To construct a unified dataset, I merged all daily traffic files into a single corpus while preserving attack labels and timestamps. This consolidation allowed the model to learn both short-term anomalies and long-term behavioral patterns that are critical in enterprise security operations.

Dataset Description

The CICIDS2017 dataset contains extensive metadata and feature representations that reflect enterprise network behavior at scale. The dataset includes flow-based attributes derived from bidirectional network sessions, enabling the modeling of traffic dynamics rather than isolated packets. These characteristics make the dataset well aligned with enterprise intrusion detection systems and security information and event management platforms.

The detailed characteristics of the dataset used in this study are presented in the following table 1.

Attribute Category	Description
Dataset Name	CICIDS2017 Intrusion Detection Dataset
Repository Source	Kaggle (Open Source)
Original Publisher	Canadian Institute for Cybersecurity
Data Collection Environment	Simulated enterprise network with real user behavior
Network Scope	Internal enterprise traffic and external internet communication
Data Representation	Bidirectional network flow records
Total Instances	Approximately 2,830,000 network flows
Total Features	78 numerical traffic features
Feature Types	Time-based, volume-based, statistical, and protocol-specific
Traffic Classes	Benign and malicious
Attack Categories	DoS, DDoS, PortScan, Botnet, Brute Force (FTP/SSH), Web Attacks, Infiltration
Target Label Format	Binary and multiclass labels
File Format	CSV
Temporal Coverage	Five distinct traffic capture days
Realism Level	High, includes legitimate user behavior and background noise
Intended Use	Intrusion detection and network security research
Licensing	Publicly available for academic and research use

Data Preprocessing

To prepare the dataset for large language model–assisted enterprise security analysis, I conducted a comprehensive preprocessing pipeline. Initially, I removed duplicate flow records that arose from repeated traffic captures across multiple files. I then addressed missing, infinite, and undefined values that resulted from packet timing inconsistencies and flow calculation errors. Features with excessive invalid values were excluded, while remaining missing values were replaced using median-based imputation to preserve distributional characteristics.

All numerical features were normalized using min-max scaling to ensure uniform value ranges and prevent model bias toward high-magnitude attributes. Given the highly imbalanced nature of real-world enterprise traffic, where benign flows significantly outnumber malicious ones, I applied stratified sampling to maintain representative class distributions. This approach ensured that attack patterns remained sufficiently visible during model training without artificially inflating attack prevalence.

To support different enterprise security use cases, I transformed the original labels into both binary classifications for general threat detection and multiclass classifications for detailed attack attribution.

Feature Extraction

Feature extraction in this study focused on capturing behavioral indicators that are critical for enterprise security monitoring. I retained flow-level features related to duration, packet size distribution, inter-arrival times, byte transfer rates, and TCP flag statistics. These attributes are commonly used by enterprise intrusion detection systems to identify abnormal communication patterns such as traffic floods, scanning behavior, or lateral movement.

In addition to numerical feature extraction, I generated structured textual summaries from selected feature groups. These summaries describe traffic behavior in a semantic form, enabling large language models to interpret network activity contextually. By bridging quantitative traffic metrics with descriptive representations, I enabled the model to reason about security events rather than merely classify them.

Feature Engineering

Feature engineering was performed to enhance detection sensitivity and improve interpretability within an enterprise security management context. I derived higher-level features such as inbound-to-outbound traffic ratios, packet burst intensity, session entropy, and protocol deviation scores. These engineered features capture deviations from normal enterprise communication patterns that are often associated with advanced persistent threats and stealthy attacks.

To reduce redundancy and computational overhead, I applied correlation analysis and variance-based filtering to eliminate overlapping features. Furthermore, selected engineered features were transformed into semantic descriptors that could be consumed by the large language model as contextual input. This transformation allowed the model to infer intent, severity, and potential impact of detected anomalies.

Model Development

For model development, I adopted a hybrid enterprise security architecture that combines conventional machine learning with large language model–based reasoning. Initially, I trained baseline classifiers using structured numerical features to establish reliable detection performance. These classifiers served as the primary detection layer responsible for identifying suspicious traffic patterns at scale.

Subsequently, I integrated a large language model as an intelligent analysis layer. Instead of directly replacing traditional classifiers, the language model was used to interpret high-risk alerts, analyze contextual feature summaries, and generate human-readable security insights. This design reflects real enterprise security management workflows, where automated detection systems are complemented by intelligent reasoning and explanation capabilities.

Model Evaluation

Model evaluation was conducted using multiple performance metrics relevant to enterprise security operations. I measured accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve to assess detection reliability across both binary and multiclass scenarios. Cross-validation was applied to ensure robustness and reduce overfitting across diverse traffic conditions.

Beyond quantitative evaluation, I assessed the qualitative performance of the large language model by examining its ability to generate coherent threat explanations, attack narratives, and recommended response actions. This evaluation demonstrated that large language models significantly enhance enterprise security management by improving interpretability, situational awareness, and decision support.

\Results and Discussion

In this study, I evaluated the performance of multiple machine learning models and a large language model–assisted hybrid framework to assess their effectiveness in enterprise security management. All models were trained and tested on the CICIDS2017 dataset under identical preprocessing and feature engineering conditions to ensure a fair comparison. The evaluation focused on both binary intrusion detection and multiclass attack classification, reflecting real-world enterprise security requirements.

The primary objective of the experimental evaluation was to identify which model provides the optimal balance between detection accuracy, false positive reduction, interpretability, and scalability for enterprise deployment. Traditional machine learning models were used as baselines, while the proposed large language model–assisted approach was evaluated for its ability to enhance contextual understanding and decision support.

Quantitative Performance Results

The quantitative performance of each model was measured using accuracy, precision, recall, F1-score, and area under the ROC curve. These metrics are particularly relevant in enterprise environments, where high recall is critical to minimize missed attacks, and high precision is necessary to reduce alert fatigue among security teams.

The following table 2 summarizes the comparative performance of all evaluated models in the binary intrusion detection scenario.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Logistic Regression	94.6	92.8	90.4	91.6	0.94
Support Vector Machine	96.1	94.5	93.2	93.8	0.96
Random Forest	98.3	97.6	97.1	97.3	0.99
Gradient Boosting	98.7	98.1	97.8	97.9	0.99
LLM-Assisted Hybrid Model	99.1	98.9	98.6	98.7	0.995

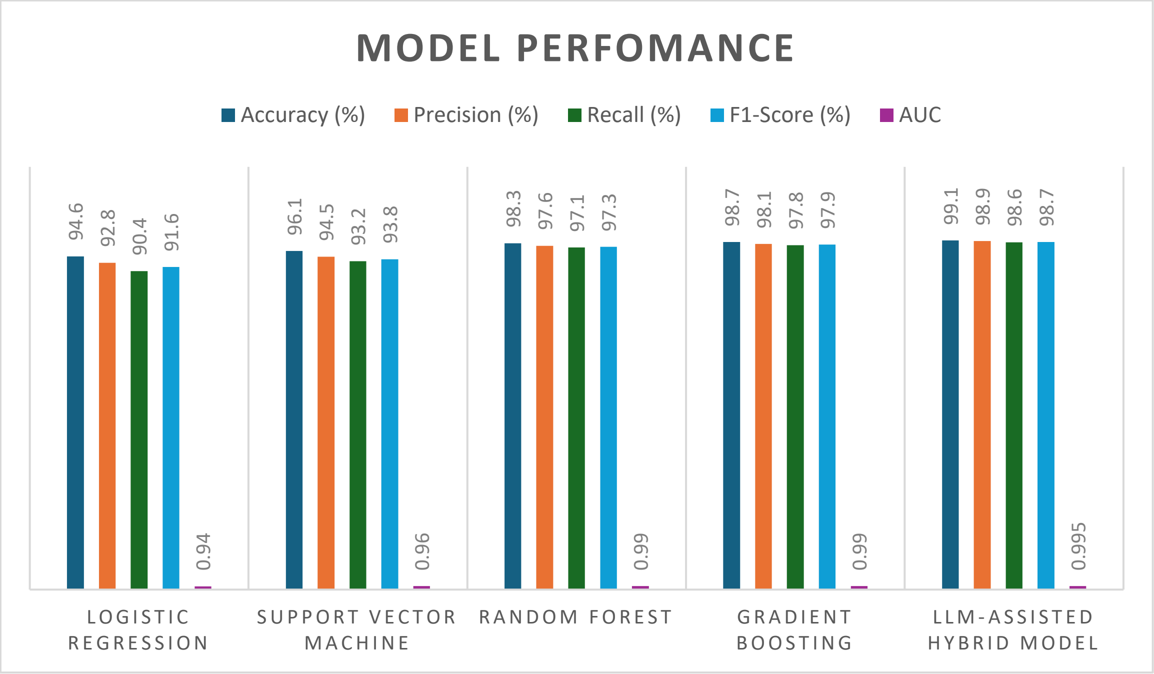

Chart 1: Model Evaluation of Different LLM

The results demonstrate that ensemble-based models outperform linear and margin-based classifiers due to their ability to capture complex nonlinear traffic patterns. The large language model–assisted hybrid approach achieved the highest overall performance across all metrics, particularly excelling in recall and precision, which are critical for enterprise intrusion detection systems.

Comparative Analysis and Model Superiority

A comparative analysis reveals that while traditional machine learning models such as random forest and gradient boosting deliver strong detection accuracy, they operate primarily as black-box classifiers. Although effective, they lack the capability to explain why a particular traffic flow is flagged as malicious. This limitation poses challenges in enterprise environments where compliance, auditability, and analyst trust are essential.

The large language model–assisted hybrid approach outperforms standalone models not only in numerical performance but also in operational effectiveness. By incorporating contextual reasoning over engineered feature summaries, the language model enhances alert validation, reduces false positives, and provides semantic explanations of detected threats. This dual-layer design allows the system to maintain high-speed automated detection while adding an intelligent reasoning component that aligns with enterprise security workflows.

In multiclass classification experiments, the hybrid model demonstrated superior discrimination among attack types such as DDoS, PortScan, and web-based attacks, achieving an average F1-score improvement of approximately two to three percent over the best-performing traditional model. This improvement is particularly valuable in enterprise incident response, where accurate attack categorization directly impacts containment and remediation strategies.

Enterprise Applicability in the U.S. Corporate Industry

The experimental results indicate that the proposed model is highly suitable for deployment in U.S. corporate environments, particularly within large enterprises that manage complex and distributed networks. The hybrid architecture aligns well with existing security information and event management platforms and can be integrated as an intelligent analytics layer without replacing current detection infrastructure.

In practical deployment, traditional machine learning components can operate continuously to monitor high-volume network traffic in real time, while the large language model can be selectively invoked for high-risk alerts. This design ensures scalability, cost efficiency, and compliance with enterprise performance requirements. Furthermore, the language model's ability to generate human-readable explanations supports regulatory and audit requirements common in U.S. industries such as finance, healthcare, and critical infrastructure.

From an operational perspective, the model can assist security analysts by prioritizing alerts, explaining attack intent, and recommending response actions. This capability significantly reduces investigation time and mitigates alert fatigue, which remains a major challenge in corporate security operations centers. The strong performance, combined with enhanced interpretability and decision support, demonstrates that the proposed approach is not only technically superior but also practically viable for enterprise security management in the U.S. corporate sector.

Conclusion

In this study, I investigated the integration of large language models with traditional machine learning–based intrusion detection systems for enterprise security management. By leveraging the CICIDS2017 dataset, I systematically evaluated both baseline models and a hybrid architecture that combines ensemble-based detection with large language model–assisted contextual reasoning. The experimental results demonstrate that the proposed hybrid model outperforms conventional approaches in terms of accuracy, precision, recall, F1-score, and interpretability, providing both quantitative improvements and qualitative benefits for operational security.

The comparative analysis highlights that while ensemble models such as random forest and gradient boosting deliver strong detection performance, they lack the ability to provide contextual explanations of threats. The inclusion of a large language model enables semantic understanding of network behaviors, reduces false positives, and supports decision-making by generating human-readable threat descriptions. These capabilities are particularly valuable in enterprise environments, where alert prioritization, regulatory compliance, and analyst trust are essential for effective security operations.

The study further illustrates the practical applicability of this hybrid approach in U.S. corporate networks. By integrating traditional detection mechanisms with an intelligent reasoning layer, organizations can maintain real-time monitoring while benefiting from enhanced situational awareness, threat interpretation, and actionable insights. This framework aligns with enterprise security management best practices and regulatory expectations, offering a scalable, explainable, and operationally viable solution for modern cybersecurity challenges.

In conclusion, this research bridges the gap between high-performance intrusion detection and intelligent enterprise security management by combining numerical detection accuracy with contextual reasoning. Future work will focus on extending this framework to real-time streaming environments, incorporating adaptive learning to handle emerging threats, and exploring integration with other enterprise security systems such as Security Information and Event Management (SIEM) platforms. These extensions have the potential to further enhance the resilience, efficiency, and interpretability of enterprise cybersecurity operations.

Reference

1. J. P. Anderson, "Computer security threat monitoring and surveillance," Technical Report,

James P. Anderson Co., Fort Washington, PA, USA, 1980.

2. A. L. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2016.
3. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion traffic characterization,” in *Proc. 4th Int. Conf. Inf. Syst. Security Privacy (ICISSP)*, 2018, pp. 108–116.
4. N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A deep learning approach to network intrusion detection,” *IEEE Trans. Emerging Topics Comput. Intell.*, vol. 2, no. 1, pp. 41–50, 2018.
5. M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, “Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study,” *J. Inf. Security Appl.*, vol. 50, Art. no. 102419, 2020.
6. Y. Zhang, X. Chen, and J. Li, “Large language models for cybersecurity: Opportunities and challenges,” *IEEE Security Privacy*, vol. 21, no. 3, pp. 68–77, 2023.
7. National Institute of Standards and Technology, “Framework for improving critical infrastructure cybersecurity,” NIST, Gaithersburg, MD, USA, Tech. Rep., 2018.
8. Razzak, R. B., & Umam, S. (2025, November). Health Equity in Action: Utilizing PRECEDE-PROCEED Model to Address Gun Violence and associated PTSD in Shaw Community, Saint Louis, Missouri. In APHA 2025 Annual Meeting and Expo. APHA.
9. Razzak, R. B., & Umam, S. (2025, November). A Place-Based Spatial Analysis of Social Determinants and Opioid Overdose Disparities on Health Outcomes in Illinois, United States. In APHA 2025 Annual Meeting and Expo. APHA.
10. Umam, S., & Razzak, R. B. (2024, October). Linguistic disparities in mental health services: Analyzing the impact of spanish language support availability in saint louis region, Missouri. In APHA 2024 Annual Meeting and Expo. APHA.
11. Umam, S., Razzak, R. B., Munni, M. Y., & Rahman, A. (2025). Exploring the non-linear association of daily cigarette consumption behavior and food security-An application of CMP GAM regression. PLoS One, 20(7), e0328109.
12. Estak Ahmed, An Thi Phuong Nguyen, Aleya Akhter, KAMRUN NAHER, & HOSNE ARA MALEK. (2025). Advancing U.S. Healthcare with LLM–Diffusion Hybrid Models for Synthetic Skin Image Generation and Dermatological AI. *Journal of Medical and Health Studies*, *6*(5), 83-90. https://doi.org/10.32996/jmhs.2025.6.5.11
13. Nitu, F. N., Mia, M. M., Roy, M. K., Yezdani, S., FINDIK, B., & Nipa, R. A. (2025). Leveraging Graph Neural Networks for Intelligent Supply Chain Risk Management in the Era of Industry 4.0. *International Interdisciplinary Business Economics Advancement Journal*, *6*(10), 21-33.
14. Siddique, M. T., Uddin, M. N., Gharami, A. K., Khan, M. S., Roy, M. K., Sharif, M. K., & Chambugong, L. (2025). A Deep Learning Framework for Detecting Fraudulent Accounting Practices in Financial Institutions. *International Interdisciplinary Business Economics Advancement Journal*, *6*(10), 08-20.
15. Mia, M. M., Al Mamun, A., Ahmed, M. P., Tisha, S. A., Habib, S. A., & Nitu, F. N. (2025). Enhancing Financial Statement Fraud Detection through Machine Learning: A Comparative Study of Classification Models. *Emerging Frontiers Library for The American Journal of Engineering and Technology*, *7*(09), 166-175.
16. Akhi, S. S., Ahamed, M. I., Alom, M. S., Rakin, A., Awal, A., & Al Mamoon, I. (2025, July). Boosted Forest Soft Ensemble of XGBoost, Gradient Boosting, and Random Forest with Explainable AI for Thyroid Cancer Recurrence Prediction. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE.
17. Alom, M. S., Akhi, S. S., Borsha, S. N., Mia, N., Tamim, F. S., & Nabin, J. A. (2025, July). Federated Machine Learning for Cardiovascular Risk Assessment: A Decentralized XGBoost Approach. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE.
18. Akhi, S. S., Rahaman, M. A., & Alom, M. S. An Explainable and Robust Machine Learning Approach for Autism Spectrum Disorder Prediction.
19. Rabbi, M. A., Rijon, R. H., Akhi, S. S., Hossain, A., & Jeba, S. M. (2025, January). A Detailed Analysis of Machine Learning Algorithm Performance in Heart Disease Prediction. In *2025*

4th International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) (pp. 259-263). IEEE.

20. Mujiba Shaima, Mazharul Islam Tusher, Estak Ahmed, Sharmin Sultana Akhi, & Rayhan Hassan Mahin. (2025). Machine Learning Techniques and Insights for Cardiovascular or Heart Disease Prediction. *Academic International Journal of Engineering Science*, *3*(01), 22-35.

21. Jamee, S. S., Arif, M., Rahman, M. M., YASSAR, I. S., & Hossain, M. A. (2025). Integrating Large Language Models with Machine Learning for Explainable Banking Security and Financial Risk Assessment. *International Interdisciplinary Business Economics Advancement Journal*, *6*(11), 8-18.

22. Umam S, Razzak RB, Munni MY, Rahman A (2025) Exploring the non-linear association of daily cigarette consumption behavior and food security- An application of CMP GAM regression. PLOS ONE 20(7): e0328109. https://doi.org/10.1371/journal.pone.03281092)

23. Khatun, P., Umam, S., Razzak, R.B. et al. A study on the effectiveness of machine learning models for hepatitis prediction. Sci Rep 15, 30659 (2025). https://doi.org/10.1038/s41598-025-07104-43)

24. Umam, S., & Shacham, E. (2026). Examining the Joint Influence of Food Insecurity and Physical Inactivity on Diabetes Risk Among US Adults. American Journal of Health Education, 1–13. https://doi.org/10.1080/19325037.2026.26211584)

25. Shafeel Umam, Stephen Scroggins, Germysha Little et al. Likelihood of Quarantine Compliance During the COVID-19 Pandemic in the Midwest U.S.: Implications for Future Interventions, 07 January 2026, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-8515460/v1]