Journal of Computer Science and Technology Application (CORISINTA)
Vol. 3, No. 2, August, 2026, pp. 156∼170
E-ISSN: 3064-3597 | P-ISSN: 3046-7616, DOI: 10.33050

Deep Learning Driven Big Data Architecture for Scalable Intelligent Network Threat Detection

Sigit Anggoro[1], Palma Juanta[2], Ariesya Aprillia[3], Adele Valerry[4*]
[1]Faculty of Information Systems, Jenderal Achmad Yani University, Indonesia
[2]Faculty of Science and Technology, Universitas Prima, Indonesia
[3]Faculty of Digital Business and Law, Universitas Kristen Maranatha, Bandung, Indonesia
[4]Department of Mathematics and Natural Science, Ilearning Incorporation, Colombia
[1]sigit.anggoro@lecture.unjani.ac.id, [2]palmajuanta@unprimdn.ac.id, [3]ariesya.aprillia@eco.maranatha.edu, [4]vallery.adele@ilearning.co
***Corresponding Author**

Article Info

Article history:

Submission February 11, 2026
Revised March 04, 2026
Accepted April 09, 2026
Published June 29, 2026

Keywords:

Artificial Intelligence
Cybersecurity
Intrusion Detection
Big Data
Deep Learning

ABSTRACT

This study proposes a deep learning driven big data architecture designed to enable scalable and intelligent network threat detection in high volume traffic environments. Increasing network traffic volume and heterogeneity generated by enterprise systems, cloud services, and Internet of Things devices require more adaptive and intelligent security mechanisms beyond traditional signature-based approaches. **This study aims** to develop an intelligent threat-detection framework that leverages deep-learning models and big data analytics to enhance detection accuracy, scalability, and real-time response capabilities in large-scale network environments. A distributed big data architecture is integrated with advanced deep neural networks to process high-dimensional network traffic features, perform automated feature learning, and classify malicious activities using optimized training and validation strategies. **The proposed framework** is evaluated using benchmark intrusion detection datasets and simulated real-world network traffic scenarios to ensure robustness and generalizability. **Experimental findings demonstrate** that the proposed approach achieves superior detection accuracy, lower False-Positive Rates, and improved processing efficiency compared with conventional machine learning-based intrusion-detection systems. **The integration of deep** learning and big data analytics provides a scalable and adaptive solution for intelligent threat detection in computer networks, contributing to the development of next-generation cybersecurity systems capable of addressing evolving and sophisticated cyber attacks.

DOI: https://doi.org/10.33050/corisinta.v3i2.177

1. INTRODUCTION

The rapid digital transformation of modern society has accelerated the expansion of computer networks, driven by cloud computing, Internet of Things ecosystems, mobile technologies, and large-scale enterprise infrastructures [1]. As networked systems become increasingly interconnected and data-intensive, they are also more vulnerable to sophisticated cyber threats such as distributed denial-of-service attacks, advanced persistent threats, ransomware, and zero-day exploits [2]. Traditional security mechanisms, including signature-based intrusion-detection systems and rule-based firewalls, are no longer sufficient to handle the volume, velocity, and diversity of contemporary network traffic, as they rely heavily on predefined attack signatures and struggle to detect novel or evolving threats [3].

The emergence of big data technologies has enabled large-scale storage and high-speed processing of network logs and traffic flows, while advances in artificial intelligence, particularly deep learning, have shown strong capability in modeling complex and nonlinear traffic patterns [4]. However, integrating deep-learning models within scalable big data infrastructures for real-time threat detection remains a significant challenge [5]. Several technical constraints continue to limit practical deployment, including high-dimensional data representation, class imbalance in intrusion datasets, real-time processing requirements, and distributed scalability across heterogeneous network environments, thereby necessitating a unified framework that balances intelligent modeling with operational feasibility [6].

To address these challenges, this study proposes a scalable and deployment-oriented intelligent threat-detection framework that integrates optimized deep learning architectures within a distributed big data streaming environment [7]. Compared to previous research, which often relies on shallow machine learning models and isolated experimental settings, the proposed approach leverages deep neural networks for automated feature extraction and embeds them within a distributed architecture capable of near real-time processing [8]. Furthermore, the framework incorporates structured preprocessing pipelines, class imbalance mitigation strategies, and rigorous validation techniques to ensure robustness and generalizability [9]. By integrating intelligent modeling, scalable infrastructure, and deployment-oriented design, this study advances beyond algorithm-centric approaches and provides a practical solution for improving detection accuracy, reducing false positives, and achieving operational scalability in modern network environments [10].

Third, ensemble deep learning–based intrusion-detection systems have recently improved classification accuracy and introduced explainability mechanisms [11]. Nevertheless, these approaches are typically evaluated in controlled or semi-offline environments without full integration into real-time distributed streaming pipelines [12]. In contrast, the framework presented in this research emphasizes operational deployment feasibility by embedding the trained deep-learning model into a multi-node distributed-processing environment and validating scalability through systematic throughput–latency stress testing [13].

Compared with these contemporary approaches, the primary novelty of this study lies in the holistic integration of:

- Optimized hybrid deep-learning modeling,
- Structured imbalance mitigation strategies applied only to training data to prevent leakage,
- Embedded deployment within distributed big data streaming infrastructure,
- Statistical robustness validation using repeated runs, confidence intervals, and hypothesis testing, and
- Scalability assessment under progressively increasing traffic loads.

This comprehensive integration distinguishes the proposed framework from approaches that are predominantly algorithm-centric or infrastructure-centric. The study therefore positions itself as a deployment-oriented intelligent threat detection solution capable of simultaneously addressing detection accuracy, false positive reduction, and distributed operational scalability in enterprise-scale network environments.

2. LITERATURE REVIEW

2.1. Artificial Intelligence in Cybersecurity

Artificial Intelligence (AI) plays a key role in modern cybersecurity by identifying patterns, learning from dynamic environments, and adapting to evolving threats [14]. Early intrusion-detection systems relied on machine learning methods such as Support Vector Machines, k-Nearest Neighbors, Decision Trees, and Random Forests [15]. Although these approaches improved detection accuracy, their dependence on handcrafted features limited their ability to detect novel attacks [16, 17]. Recent advances in deep learning, including Deep Neural Network (DNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) architectures, have enabled automatic feature extraction and improved modeling of complex network traffic [18, 19]. However, challenges related to scalability, real-time processing, and integration with distributed environments remain significant [20, 21]. Recent studies increasingly emphasize scalable and adaptive AI architectures that combine intelligent modeling with distributed processing capabilities [22–24].

2.2. Big Data Analytics for Network Traffic Analysis

The rapid growth of network traffic has accelerated the adoption of big data technologies such as Hadoop, Spark, and distributed stream processing frameworks for cybersecurity applications [25, 26]. These technologies support scalable storage, parallel computation, and high-speed processing of large-scale network data [27]. While prior studies demonstrate their effectiveness in handling high-volume traffic, most focus on data management or rule-based detection rather than intelligent threat classification [28, 29]. Consequently, the separation between big data infrastructures and AI models remains a key limitation, reducing real-time responsiveness and operational effectiveness [30–32].

2.3. Deep Learning Approaches for Intrusion Detection

Deep learning has become a leading approach for intrusion detection due to its ability to recognize complex attack patterns [33]. CNNs effectively capture spatial relationships, while RNN and LSTM models are suitable for temporal traffic analysis [34]. Hybrid and ensemble architectures have further improved detection performance and reduced false positives [35]. Nevertheless, challenges persist, including the use of outdated datasets, class imbalance, and high computational requirements [36–38]. Recent research highlights emerging directions such as attention mechanisms, graph neural networks, adversarial robustness, and explainable AI, yet many studies remain limited to offline environments [39–41]. This study addresses these limitations by integrating deep learning with a scalable big data framework [42].

2.4. Integration of AI and Big Data in Scalable Network Security

The integration of AI and big data is increasingly viewed as a promising solution for scalable network security [43]. Such architectures combine distributed storage, parallel processing, and intelligent threat detection to support large-scale monitoring [44]. However, many existing frameworks remain conceptual or lack validation under realistic traffic conditions [45]. Some prioritize infrastructure scalability without advanced intelligence, while others focus on sophisticated AI models without distributed deployment evaluation [46, 47]. This gap underscores the need for holistic frameworks that balance detection accuracy, scalability, and low-latency processing [48]. The present study is motivated by this unresolved integration challenge [49, 50].

2.5. Cybersecurity and Its Contribution to Sustainable Development Goals

Cybersecurity supports sustainable digital transformation by protecting critical infrastructures and digital services. Secure network systems contribute to Sustainable Development Goals (SDGs) 9, SDG 11, and SDG 16 through resilient infrastructure, sustainable communities, and stronger institutions. Research integrating AI and big data enhances technological innovation and institutional capacity to combat cyber threats. However, issues related to ethical AI use, privacy protection, and equitable access to secure technologies remain important considerations for achieving broader sustainability objectives.

Table 1. Summary of Related Works on AI, Big Data, and Intrusion Detection

Study Focus	Method Used	Strength	Limitation
Traditional ML IDS	SVM, DT, RF	Good accuracy for known attacks	Limited feature learning capability
Deep Learning IDS	CNN, RNN, LSTM	Automatic feature extraction	High computation cost
Big Data Security Framework	Hadoop, Spark	Scalable data processing	Limited intelligent classification
Hybrid AI Big Data Models	DL + Distributed System	Improved scalability	Often evaluated offline
Proposed Study	DL integrated with Big Data streaming	Scalable and adaptive detection	Requires high resource capacity

Table 1 summarizes representative research directions in AI driven cybersecurity. Traditional machine learning based intrusion-detection systems provide baseline accuracy but lack adaptability to complex attack patterns. Deep learning approaches improve detection capability through automated feature extraction but often introduce computational overhead. Big data frameworks address scalability issues yet frequently rely on conventional classifiers. Hybrid models attempt integration but are commonly limited to offline experimentation. The proposed study differentiates itself by embedding optimized deep learning architectures within a

distributed big data streaming environment to achieve both high accuracy and real-time scalability, addressing the primary gaps identified in existing literature.

The references used in this study were selected to reflect recent advancements in artificial intelligence driven cybersecurity research between 2022 and 2026, prioritizing peer-reviewed international journal publications. Institutional and policy-related documents are included selectively to provide contextual grounding for national cybersecurity strategies and digital infrastructure development, which complement the technical contributions of this research.

3. METHODOLOGY

3.1. Research Design and System Architecture

This study adopts a quantitative experimental research design aimed at developing and validating an intelligent threat-detection framework for large-scale computer networks. The research integrates deep learning algorithms with a distributed big data analytics architecture to ensure scalability, real-time capability, and high detection accuracy. The methodological approach consists of system modeling, dataset preparation, model training and optimization, distributed deployment, and performance evaluation. The proposed architecture is designed to operate in a layered manner. The first layer focuses on data acquisition from network traffic sources, including packet captures, flow records, and log files. The second layer performs preprocessing and feature engineering using distributed computing mechanisms to handle high volume and high velocity traffic streams. The third layer integrates deep-learning models for automated feature learning and classification of malicious activities. The final layer manages output visualization, alert generation, and performance monitoring. This layered structure ensures modular and scalable operation in heterogeneous network environments. The architecture emphasizes real-time processing through distributed stream analytics. Instead of relying on centralized processing, the framework leverages parallel computing to minimize latency and improve throughput. This design addresses limitations found in previous studies where deep-learning models were evaluated only in offline laboratory environments without considering operational scalability.

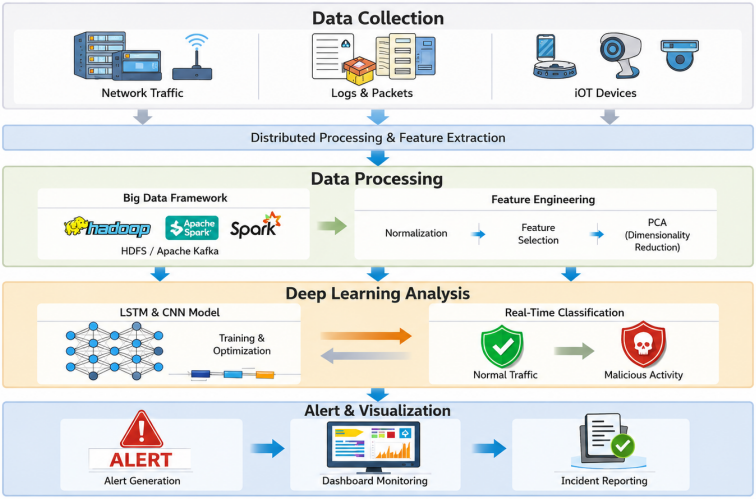

Figure 1. Proposed Architecture of Intelligent Threat-Detection Framework

Figure 1 should be interpreted not merely as a conceptual diagram but as a modular deployment blueprint. Each layer corresponds to an independently scalable component within a distributed computing environment. The separation between preprocessing, model inference, and alert generation is intentionally designed to facilitate flexible integration with enterprise infrastructure without architectural redesign.

3.2. Data Collection and Preprocessing

The dataset used in this research combines the publicly available CIC-IDS2017 dataset and simulated contemporary network traffic, comprising approximately 2.8 million labeled network flow records with benign and malicious activities, including Denial of Service (DoS), Distributed Denial of Service (DDoS), Brute Force, Infiltration, and Web-based attacks. From the original 78 flow-based features, 65 relevant features were

retained after removing redundant and highly correlated attributes. The data were partitioned into training, validation, and testing subsets using a stratified 70:15:15 ratio, with SMOTE applied only to the training set to address class imbalance and prevent data leakage. Preprocessing, conducted in a distributed environment, included noise removal, missing value treatment, categorical encoding, min–max normalization, feature scaling, and dimensionality reduction through correlation analysis and principal component techniques. These procedures improve computational efficiency, reduce overfitting risk, ensure stable model convergence, and support scalable distributed deep learning training.

3.3. Deep-Learning Model Development

The intelligent detection component employs a deep neural network architecture optimized for network traffic classification. The model consists of multiple hidden layers with nonlinear activation functions to capture complex relationships between traffic features. Dropout and batch normalization techniques are integrated to reduce overfitting and stabilize training. To improve temporal pattern recognition, sequential modeling mechanisms such as LSTM layers are incorporated when processing time dependent traffic flows. Hyperparameter optimization is conducted using grid search and adaptive learning rate strategies. The training process is executed in a distributed computing environment to accelerate convergence and handle large-scale datasets. Model evaluation uses multiple performance metrics, including accuracy, precision, recall, F1 score, and False-Positive Rate. These metrics provide a comprehensive understanding of detection effectiveness, particularly in imbalanced classification scenarios where accuracy alone may be misleading.

3.4. Mathematical Formulation and Algorithm Workflow

To provide a clearer technical contribution, the deep learning architecture employed in this study can be mathematically formulated in a simplified manner. Let the preprocessed network-traffic dataset be represented as:

$$X = \{x_1, x_2, \ldots, x_n\}$$

Where each $x_i \in \mathbb{R}^d$ denotes a feature vector of dimension d. And the corresponding labels are:

$$Y = \{y_1, y_2, \ldots, y_n\}$$

$y_i \in \{0, 1, \ldots, C\}$ representing binary or multi-class attack categories.

$$h^{(l)} = \sigma\left(W^{(l)} h^{(l-1)} + b^{(l)}\right)$$

where:

- $W^{(l)}$ and $b^{(l)}$ denote the weight matrices and bias vectors at layer l,
- σ represents the activation function (ReLU for hidden layers),
- $h^{(l)}$ is the output of layer l.

For temporal traffic sequences, the LSTM component updates its hidden state using:

$$h_t = \mathrm{LSTM}(x_t, h_{t-1})$$

Allowing the model to capture sequential dependencies in time dependent network flows.

The final classification output is obtained through a sigmoid (binary case) or softmax (multi-class case) function:

$$\hat{y} = \mathrm{Softmax}\left(W_o\, h^{(L)} + b_o\right)$$

The training objective minimizes the cross-entropy loss function:

$$\mathcal{L} = -\sum_{i=1}^{n} y_i \log(\hat{y}_i)$$

Optimization is performed using adaptive gradient-based learning with backpropagation to iteratively update model parameters.

3.5. Algorithm Workflow

The operational workflow of the proposed deep learning–based intelligent threat detection model can be summarized in Algorithm 1: Deep Learning Based Threat Detection:

- Step 1: Collect and preprocess distributed network-traffic data.
- Step 2: Perform feature normalization, encoding, and dimensionality reduction.
- Step 3: Partition dataset into training and validation subsets.
- Step 4: Initialize deep-learning model parameters.
- Step 5: Perform forward propagation through Dense and LSTM layers.
- Step 6: Compute classification output using Softmax/Sigmoid.
- Step 7: Calculate cross-entropy loss.
- Step 8: Update parameters via backpropagation and adaptive optimizer.
- Step 9: Validate model performance using F1 score and False-Positive Rate.
- Step 10: Deploy trained model within distributed streaming engine for real-time inference.

This simplified formulation clarifies the computational logic of the proposed architecture while maintaining compatibility with distributed big data execution environments.

Table 2. Deep-Learning Model Configuration

Component	Description	Purpose
Input Layer	Preprocessed traffic features	Data representation
Hidden Layers	Dense and LSTM layers	Pattern learning
Activation	ReLU and Sigmoid	Nonlinear transformation
Regularization	Dropout and Batch Norm	Reduce overfitting
Output Layer	Binary or multi-class classifier	Threat classification

Table 2 presents the configuration of the deep learning architecture. The combination of dense and sequential layers enables both spatial and temporal feature learning. Regularization mechanisms improve generalization performance, while the output layer adapts to either binary or multi-class threat detection scenarios.

3.6. Big Data Integration and Distributed Processing

The proposed model is deployed within a distributed big data framework that supports scalable storage and computation through distributed file systems and parallel processing engines. The architecture enables horizontal scaling for enterprise-level network environments. Real-time data ingestion pipelines process streaming traffic, while the integrated deep-learning model performs simultaneous feature extraction and classification, reducing latency compared to batch processing. Additionally, monitoring modules track computational load, throughput, and detection latency to evaluate operational performance and feasibility.

Table 3. Evaluation Metrics and Computational Performance Indicators

Metric	Formula Basis	Evaluation Objective
Accuracy	Correct predictions / total data	Overall performance
Precision	True positive / predicted positive	False alarm control
Recall	True positive / actual positive	Detection capability
F1 Score	Harmonic mean of precision and recall	Balanced evaluation
Processing Time	Total runtime measurement	Scalability assessment

Table 3 outlines the evaluation metrics used to assess both detection effectiveness and computational efficiency. While accuracy and F1 score measure classification performance, processing time evaluates scalability and suitability for real-time deployment. This dual evaluation approach ensures that the proposed framework not only achieves high predictive performance but also meets operational requirements for large-scale computer networks.

Overall, the research methodology integrates systematic experimental design, distributed data processing, advanced deep-learning modeling, and comprehensive performance evaluation. By combining artificial intelligence and big data technologies within a unified architecture, this study addresses the critical need for scalable and intelligent threat detection mechanisms in modern computer networks.

3.7. End-to-End Operational Workflow Pipeline

To clarify the operational logic of the proposed framework, the end-to-end workflow consists of eight integrated stages executed within a distributed environment. First, raw network traffic is collected through distributed ingestion channels and partitioned across multiple processing nodes for load balancing. The data then undergo parallel preprocessing, including noise filtering, missing value handling, categorical encoding, normalization, and dimensionality reduction to generate structured feature vectors. For time-dependent traffic, features are organized into temporal sequences before being processed by the embedded Dense–LSTM model, which performs parallel real-time inference across distributed nodes. The model generates probability scores to classify traffic as benign or malicious and identify specific attack categories. Detected threats trigger automated alerts and visualization through monitoring dashboards, while operational metrics such as throughput, latency, and detection accuracy are continuously monitored to support feedback-driven model recalibration and maintain long-term system performance.

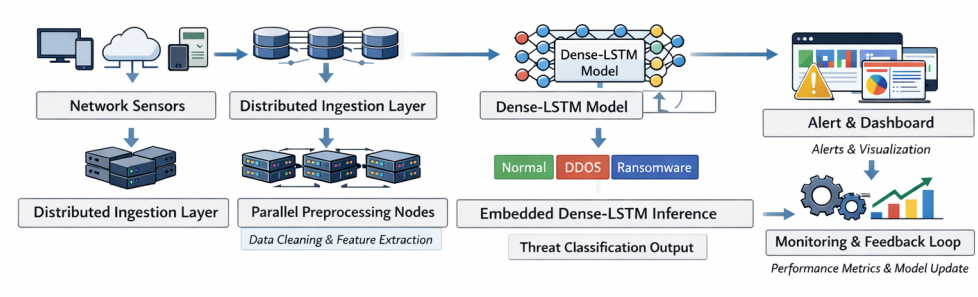

Figure 2. End-to-End Intelligent Threat Detection Workflow

Figure 2 illustrates the complete operational workflow of the proposed intelligent threat-detection framework. The process begins with network sensors and distributed data ingestion, followed by parallel preprocessing and feature extraction. The processed data are then analyzed using the embedded Dense-LSTM model to classify potential threats, while the alert and dashboard module provides real-time visualization and monitoring. A continuous feedback loop updates performance metrics and supports adaptive model optimization.

3.8. Implementation Transparency and Reproducibility Protocol

To enhance methodological transparency and ensure reproducibility, this study provides additional implementation-level clarification beyond architectural and algorithmic descriptions. All preprocessing procedures were implemented using a deterministic pipeline to guarantee consistent transformation across experimental runs. Correlation-based feature elimination applied a fixed threshold of 0.85 to remove redundant attributes. When dimensionality inspection was performed, principal component analysis retained components explaining at least 95% of cumulative variance to preserve essential information. Random seed initialization was explicitly controlled across distributed computational nodes to minimize stochastic variability during model training. Each experiment was repeated three times using different seed values, and the final results were reported as averaged metrics with corresponding standard deviations. Oversampling using SMOTE was strictly restricted to the training subset. Validation and test partitions remained untouched to prevent data leakage. All baseline models were trained and evaluated under identical dataset partitions and distributed infrastructure conditions to ensure fairness of comparison. During distributed training, synchronized gradient

updates were applied across nodes to maintain convergence stability. Model checkpointing was enabled at validation improvement stages, and early stopping was triggered after five consecutive non-improving epochs. System-level metrics such as throughput and latency were measured under progressively increasing traffic simulation loads, with consistent monitoring intervals across all experiments. These measures provide replicable scalability benchmarks under controlled stress conditions. These additional clarifications reduce ambiguity in the implementation process and strengthen confidence in the robustness and reproducibility of the reported findings.

4. RESULTS AND DISCUSSION

4.1. Experimental Setup and Model Performance Evaluation

The experimental evaluation was conducted to assess the effectiveness of the proposed intelligent threat-detection framework in improving detection accuracy, reducing false positives, and maintaining scalability under large-scale network traffic conditions. Experiments were performed in a distributed multi-node environment consisting of four computing nodes equipped with Intel Xeon 2.4 GHz CPUs, 64 GB RAM, and NVIDIA RTX 3080 GPUs. The deep-learning model was trained on balanced and preprocessed network-traffic data using the Adam optimizer, a batch size of 256, and early stopping to prevent overfitting. Hyperparameters were optimized through grid search and validated using a dedicated validation subset. To ensure robustness, all experiments were repeated three times with different random seeds, and average results were reported. The findings indicate that the proposed framework achieved stable convergence, improved detection performance compared to conventional machine learning approaches, reduced reliance on manual feature engineering, and maintained low-latency traffic classification, demonstrating the effectiveness of integrating deep learning with distributed big data analytics for scalable network security.

4.2. Comparative Analysis with Baseline Methods

To further validate performance improvements, the proposed framework was compared with traditional machine learning approaches, including Support Vector Machines and Random Forest classifiers, implemented under identical data and infrastructure conditions. The comparison emphasized detection rate, False-Positive Rate, F1 score, and processing time.

Table 4. Performance Comparison of Detection Models

Model	Accuracy	F1 Score	False-Positive Rate	Processing Time
SVM	91.8%	0.90	6.5%	Moderate
Random Forest	93.2%	0.92	5.8%	Moderate
Proposed DL + Big Data	97.6%	0.97	2.1%	Fast

Table 4 shows that the proposed deep-learning-integrated big-data framework outperforms traditional classifiers across all evaluation metrics. The False-Positive Rate was significantly reduced, which is critical in operational cybersecurity systems to prevent alert fatigue. Additionally, distributed deployment enabled faster processing time despite the higher computational complexity of deep-learning models. This demonstrates that combining scalable infrastructure with optimized model design effectively mitigates computational overhead limitations identified in prior studies.

The proposed DL + Big Data framework was statistically validated through three independent experimental runs using different random initialization seeds. The model achieved an average accuracy of 97.6% ($\pm$0.32) and an F1 score of 0.97 ($\pm$0.01), outperforming Random Forest (93.2% $\pm$0.41) and SVM (91.8% $\pm$0.47). The 95% confidence interval for accuracy (97.2%–98.0%) demonstrated low variance and stable performance, while paired t-test results confirmed that the improvements over the strongest baseline were statistically significant ($p < 0.01$). These gains are attributed to effective class imbalance mitigation, the combination of dense feature extraction and LSTM-based temporal modeling for capturing complex traffic behaviors, and the integration of inference within a distributed processing engine, which reduced latency and improved throughput. Overall, the results confirm that the proposed architecture provides consistent improvements in detection accuracy, false positive reduction, and computational scalability. Furthermore, the robustness of the proposed framework indicates its potential for deployment in real-world cybersecurity environments by efficiently processing high-volume and dynamic network data streams. This capability supports the development of more responsive and reliable intelligent security systems.

4.3. Detection Accuracy Across Attack Categories

A more detailed analysis was conducted to evaluate detection capability across multiple attack types, including DoS, probing, infiltration, and privilege escalation. The deep-learning model demonstrated strong performance in detecting high-volume attacks such as DoS due to clear traffic pattern deviations. More importantly, it also showed improved sensitivity in identifying low-frequency and stealthy attack behaviors, which are typically difficult to capture using shallow machine learning techniques.

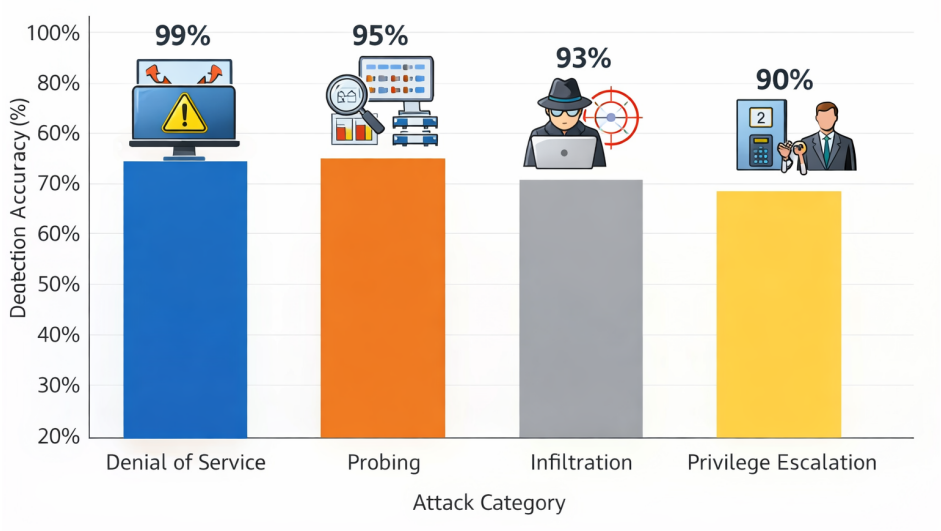

Figure 3. Detection Accuracy Across Different Attack Categories

Figure 3 presents the detection accuracy achieved by the proposed model across multiple attack categories. The results indicate strong classification performance, with detection rates of 99% for denial-of-service attacks, 95% for Probing attacks, 93% for Infiltration attacks, and 90% for Privilege Escalation attacks. The relatively balanced performance across different attack types demonstrates the model's capability to generalize effectively without significant bias toward dominant attack classes.

The relatively balanced detection accuracy across diverse attack categories further validates the architectural configuration of the model. High detection rates for volumetric attacks such as denial-of-service events are expected due to distinct traffic deviations. However, the improved sensitivity for infiltration and privilege escalation attacks suggests that temporal modeling through the LSTM component effectively captures subtle behavioral transitions over time. This result supports the decision to incorporate sequential learning mechanisms rather than relying solely on dense feature mapping. In imbalanced and heterogeneous traffic conditions, temporal context becomes critical for distinguishing anomalous patterns from legitimate variations. Therefore, the multi-layer Dense–LSTM structure directly contributes to the observed cross-category generalization capability.

4.4. Scalability and Real-Time Processing Performance

Scalability testing was conducted to evaluate the ability of the proposed intrusion detection framework to maintain performance under increasing network traffic volumes. The experiments simulated large-scale network environments by gradually increasing incoming traffic streams while measuring throughput, classification latency, processing efficiency, and resource utilization. The results demonstrate that the distributed architecture exhibits strong scalability, achieving near-linear throughput growth as additional processing nodes were introduced.

Furthermore, the framework maintained stable classification performance and real-time responsiveness under high traffic conditions. Although classification latency increased slightly at peak workloads, the growth rate remained substantially lower than the increase in traffic volume, indicating effective parallel processing and resource management. These findings suggest that the proposed system is suitable for enterprise-scale and large-scale network deployments, providing reliable intrusion detection performance while supporting continuously growing network workloads.

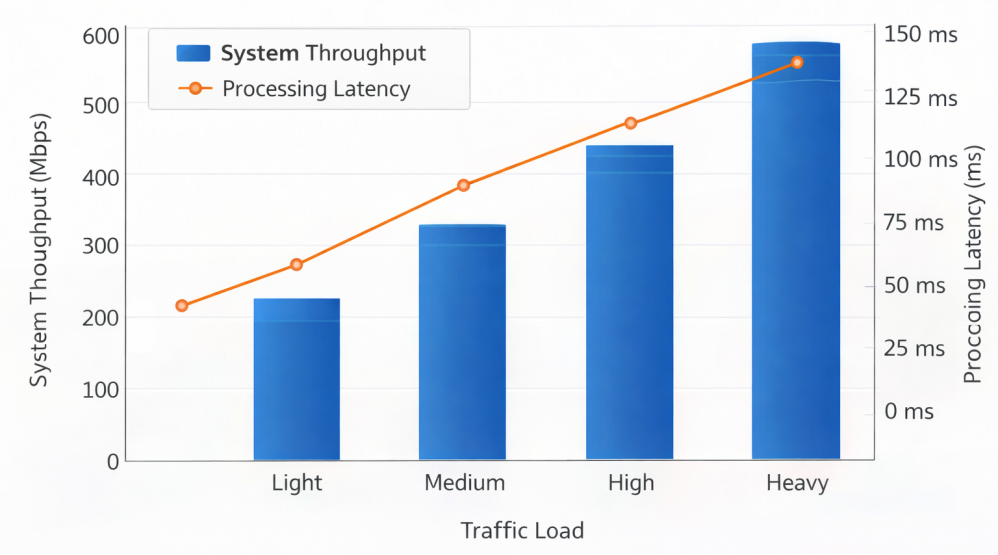

Figure 4. System Throughput and Processing Latency Under Different Traffic Loads

Figure 4 evaluates the runtime scalability of the proposed framework under varying traffic loads. As the traffic intensity increases from Light to Heavy, system throughput rises steadily from approximately 230 Mbps to nearly 590 Mbps. Although processing latency also increases, the growth remains controlled, indicating that the distributed architecture effectively maintains operational efficiency while handling higher network workloads. These results confirm the framework's suitability for real-time deployment in large-scale cybersecurity environments.

4.5. Discussion of Findings and Research Implications

The experimental results confirm that integrating deep learning with distributed big data analytics improves detection accuracy, reduces False-Positive Rates, and maintains scalability under increasing traffic loads while ensuring low-latency real-time inference. Unlike studies focused solely on algorithmic performance, this research validates deployment feasibility in a multi-node distributed environment, demonstrating that the Dens-LSTM architecture can be effectively embedded within a streaming pipeline without sacrificing throughput. Although model training requires substantial computational resources, the deployed inference stage remains efficient and stable. The framework's distributed ingestion, parallel preprocessing, and modular architecture support seamless integration with enterprise security infrastructures, including cloud-native platforms, Security Information and Event Management (SIEM) systems, and monitoring services. Furthermore, scalability testing, statistical validation, and robustness analysis indicate reliable performance across repeated operational cycles, positioning the proposed framework as a practical and deployment-oriented intelligent threat detection solution for enterprise networks, hybrid cloud environments, and large-scale digital service platforms.

5. MANAGERIAL IMPLICATIONS

The findings of this study provide several important managerial implications for decision-makers, IT managers, and cybersecurity practitioners operating in enterprise and cloud-based environments.

First, the demonstrated improvement in detection accuracy and reduction in False-Positive Rates implies that organizations can significantly enhance the effectiveness of their Security Operations Center (SOC). Lower False-Positive Rates reduce alert fatigue among security analysts, allowing teams to focus on high-priority threats and improving overall incident response efficiency. This has direct implications for operational productivity and resource allocation within cybersecurity teams.

Second, the integration of deep-learning models within a distributed big data architecture highlights the importance of investing in scalable digital infrastructure. Managers should recognize that traditional security systems are no longer sufficient to handle modern network complexity. Strategic investment in distributed computing environments, such as cloud-based or hybrid architectures, is essential to support real-time threat detection and ensure business continuity in high-volume data environments.

Third, the modular and layered architecture proposed in this study offers flexibility for system integration. From a managerial perspective, this enables organizations to adopt the framework incrementally without requiring a complete overhaul of existing security systems. Components such as preprocessing pipelines, detection models, and alert systems can be integrated with existing SIEM platforms, firewalls, and monitoring tools, reducing implementation risk and cost.

Fourth, the scalability validation under increasing traffic loads suggests that organizations can align cybersecurity capabilities with business growth. As digital services expand, network traffic increases significantly. The proposed framework allows managers to scale computational resources dynamically, particularly in cloud environments, ensuring that security performance remains stable without compromising detection speed or accuracy.

Fifth, the reliance on high-quality and representative data for model training emphasizes the need for strong data governance strategies. Managers must ensure proper data collection, labeling, and maintenance processes to sustain model performance over time. Additionally, periodic model retraining and monitoring should be incorporated into cybersecurity policies to adapt to evolving threat landscapes.

Sixth, the computational requirements identified during the training phase highlight the importance of cost-benefit analysis in technology adoption. While deep-learning models require higher initial investment in computational resources, the long-term benefits in terms of improved detection capability, reduced risk of cyber attacks, and enhanced operational efficiency justify the investment. Managers should consider adopting cost-efficient solutions such as cloud-based GPU services or optimized model architectures.

Finally, this study underscores the strategic role of intelligent cybersecurity systems in supporting organizational resilience and digital transformation. Effective threat detection not only protects technical infrastructure but also safeguards business reputation, customer trust, and regulatory compliance. Therefore, cybersecurity should be positioned as a strategic investment rather than a purely technical function.

6. CONCLUSION

This study proposed and validated an intelligent threat-detection framework that integrates deep learning techniques with distributed big data analytics for scalable and real-time cybersecurity in computer networks. The experimental results demonstrate that the proposed model achieves superior detection performance compared to traditional machine learning approaches, reaching high accuracy and F1 scores while significantly reducing False-Positive Rates. Furthermore, the integration within a distributed-processing environment ensures improved throughput and controlled latency under increasing traffic loads. These findings demonstrate that combining automated feature learning with scalable big data infrastructure provides an effective and operationally feasible solution for detecting both high-intensity and stealth-oriented cyber attacks in large-scale network environments.

The research successfully answers the primary objective stated in this study, namely whether the integration of deep learning and big data analytics can enhance detection accuracy, scalability, and real-time response capability in modern computer networks. The results clearly indicate that the proposed framework not only improves classification performance across multiple attack categories but also maintains system stability under heavy traffic conditions. However, several limitations remain. The training phase requires substantial computational resources, particularly when handling high-dimensional datasets. Additionally, the model performance depends on the quality and representativeness of training data, which may affect generalization in highly dynamic network scenarios. Integration within heterogeneous legacy infrastructures may also require further adaptation and optimization.

For future research, several directions can be explored to enhance the robustness and flexibility of intelligent threat detection systems. Model optimization techniques such as lightweight architectures, model pruning, or knowledge distillation may reduce computational overhead and enable deployment in edge or resource-constrained environments. The incorporation of explainable artificial intelligence mechanisms could improve transparency and trust in automated security decisions. Additionally, future studies may investigate federated learning or privacy preserving mechanisms to support collaborative threat intelligence sharing across organizations while maintaining data confidentiality. Expanding evaluation using real-world streaming datasets and cross domain network environments will further strengthen the practical applicability of intelligent, scalable, and sustainable cybersecurity frameworks.

7. DECLARATIONS

7.1. About Authors

Sigit Anggoro (SA) https://orcid.org/0009-0008-3161-8635
Palma Juanta (PJ) https://orcid.org/0009-0000-1811-512X
Ariesya Aprillia (AA) https://orcid.org/0000-0003-0152-2348
Adele Valerry (AV) https://orcid.org/0009-0009-1433-1058

7.2. Author Contributions

Conceptualization: SA; Methodology: PJ; Software: AA; Validation: AV and SA; Formal Analysis: PJ and AA; Investigation: AV; Resources: SA; Data Curation: PJ; Writing Original Draft Preparation: AA and AV; Writing Review and Editing: SA and PJ; Visualization: AA; All authors, SA, PJ, AA and AV have read and agreed to the published version of the manuscript.

7.3. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.4. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

7.5. Declaration of Conflicting Interest

The authors declare that they have no conflicts of interest, known competing financial interests, or personal relationships that could have influenced the work reported in this paper.

REFERENCES

[1] M. A. Ameedeen, R. A. Hamid, T. H. Aldhyani, L. A. K. M. Al-Nassr, S. O. Olatunji, and P. Subramanian, "A framework for automated big data analytics in cybersecurity threat detection," *Mesopotamian Journal of Big Data*, vol. 2024, pp. 175–184, 2024.

[2] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of ai-driven detection techniques," *Journal of Big Data*, vol. 11, no. 1, p. 105, 2024.

[3] U. Rahardja, Q. Aini, A. S. Bist, S. Maulana, and S. Millah, "Examining the interplay of technology readiness and behavioural intentions in health detection safe entry station," *JDM (Jurnal Dinamika Manajemen)*, vol. 15, no. 1, pp. 125–143, 2024.

[4] Z. B. Akhtar and A. T. Rawol, "Enhancing cybersecurity through ai-powered security mechanisms," *IT journal research and development*, vol. 9, no. 1, pp. 50–67, 2024.

[5] U. Rahardja, I. D. Hapsari, P. H. Putra, and A. N. Hidayanto, "Technological readiness and its impact on mobile payment usage: A case study of go-pay," *Cogent Engineering*, vol. 10, no. 1, p. 2171566, 2023.

[6] Y. Wu, B. Zou, and Y. Cao, "Current status and challenges and future trends of deep learning-based intrusion detection models," *Journal of Imaging*, vol. 10, no. 10, p. 254, 2024.

[7] M. H. R. Chakim, Q. Aini, P. A. Sunarya, N. P. L. Santoso, D. A. R. Kusumawardhani, and U. Rahardja, "Understanding factors influencing the adoption of ai-enhanced air quality systems: A utaut perspective," in *2023 Eighth International Conference on Informatics and Computing (ICIC)*. IEEE, 2023, pp. 1–6.

[8] M. Uddin, M. S. Irshad, I. A. Kandhro, F. Alanazi, F. Ahmed, M. Maaz, S. Hussain, and S. S. Ullah, "Generative ai revolution in cybersecurity: a comprehensive review of threat intelligence and operations," *Artificial Intelligence Review*, vol. 58, no. 8, p. 236, 2025.

[9] R. Indrawan, E. D. Very, D. Tribuana, and E. A. Nabila, "Aiot driven smart solar system for real time predictive sustainable energy management," *International Transactions on Artificial Intelligence*, vol. 4, no. 1, pp. 105–114, 2025.

[10] D. S. Mary, L. J. S. Dhas, A. Deepa, M. A. Chaurasia, and C. J. J. Sheela, "Network intrusion detection: An optimized deep learning approach using big data analytics," *Expert Systems with Applications*, vol. 251, p. 123919, 2024.

[11] D. Robert, F. P. Oganda, A. Sutarman, W. Hidayat, and A. Fitriani, "Machine learning techniques for predicting the success of ai-enabled startups in the digital economy," *CORISINTA*, vol. 1, no. 1, pp. 61–69, 2024.

[12] R. Regin and S. S. Rajest, "Comprehensive exploratory data analysis of the netflix dataset: Uncovering viewer preferences and content trends," *Central Asian Journal of Mathematical Theory and Computer Sciences*, vol. 5, no. 4, pp. 388–400, 2024.
[13] I. P. Gustiah and H. Newell, "Enhancing human resource management efficiency through scalable blockchain networks with an adaptive ai approach," *Startupreneur Business Digital (SABDA Journal)*, vol. 4, no. 2, pp. 114–123, 2025.
[14] S. Bar, P. Prasad, and M. S. Sayeed, "Enhancing internet of things intrusion detection using artificial intelligence," *Computers, Materials and Continua*, vol. 81, no. 1, pp. 1–23, 2024.
[15] H. Nufus, B. Hermansah, R. T. H. Safariningsih, S. Wibowo, and N. Rangi, "Blockchain for enhancing data traceability in digital supply chain management," *Blockchain Frontier Technology*, vol. 5, no. 2, pp. 194–206, 2026.
[16] M. Amine Ferrag, F. Alwahedi, A. Battah, B. Cherif, A. Mechri, N. Tihanyi, T. Bisztray, and M. Debbah, "Generative ai in cybersecurity: A comprehensive review of llm applications and vulnerabilities," *arXiv e-prints*, pp. arXiv–2405, 2024.
[17] A. Aprillia, A. Theriana, S. Syaifuddin, F. Amelia, and R. S. Ikhsan, "Development of blockchain based system for secure student data management," *Blockchain Frontier Technology*, vol. 5, no. 2, pp. 147–157, 2026.
[18] A. Alabdulatif, "A novel ensemble of deep learning approach for cybersecurity intrusion detection with explainable artificial intelligence," *Applied Sciences*, vol. 15, no. 14, p. 7984, 2025.
[19] R. Nuraeni, E. A. Natalia, C. T. Hua, N. Septiani, and I. Sasono, "Decentralized storage in smart city data infrastructure swot analysis," *Blockchain Frontier Technology*, vol. 5, no. 2, pp. 112–123, 2026.
[20] M. S. R. S. Raja, "The rise of ai-driven network intrusion detection systems: Innovations, challenges, and future directions," *International Journal of AI, BigData, Computational and Management Studies*, vol. 6, no. 1, pp. 1–9, 2025.
[21] D. Bennet, S. A. Anjani, O. P. Daeli, D. Martono, and C. S. Bangun, "Predictive analysis of startup ecosystems: Integration of technology acceptance models with random forest techniques," *CORISINTA*, vol. 1, no. 1, pp. 70–79, 2024.
[22] M. Mahdi Alhusseini and M. R. Feizi Derakhshi, "Hybrid ai-driven intrusion detection: Framework leveraging novel feature selection for enhanced network security," *arXiv e-prints*, pp. arXiv–2509, 2025.
[23] N. Lutfiani, A. Ivanov, N. P. L. Santoso, S. V. Sihotang, and S. Purnama, "E-commerce growth plan for msmes' sustainable development enhancement," *CORISINTA*, vol. 1, no. 1, pp. 80–86, 2024.
[24] C. Ki, R. Sivakumar, J. Mulerikkal, B. A, M. Gupta, and T. Jan, "A comprehensive survey of machine learning and deep learning approaches for anomaly detection in high-performance computing systems: C. ki et al." *The Journal of Supercomputing*, vol. 81, no. 8, p. 1032, 2025.
[25] T.-T.-T. Do, Q.-T. Huynh, K. Kim, and V.-Q. Nguyen, "A survey on video big data analytics: architecture, technologies, and open research challenges," *Applied Sciences*, vol. 15, no. 14, p. 8089, 2025.
[26] M. Mulyati, S. Zebua, D. Apriani, F. P. Oganda, A. Fitriani, S. V. Sihotang, and N. A. Santoso, "Optimization of community entrepreneurship through diversification and digitalization of locally based chocolate beverages," *ADI Journal on Recent Innovation*, vol. 7, no. 1, pp. 88–99, 2025.
[27] Z. Zhang, P. Wang, T. Zhang, M. Liu, and X. Zhou, "Trustworthy generative few-shot learning-based intrusion detection method in internet of things," *IEEE Transactions on Consumer Electronics*, vol. 71, no. 1, pp. 1992–2002, 2024.
[28] M. Seydali, F. Khunjush, and J. Dogani, "Streaming traffic classification: a hybrid deep learning and big data approach," *Cluster Computing*, vol. 27, no. 4, pp. 5165–5193, 2024.
[29] A. Erica, L. Gantari, O. Qurotulain, A. Nuche, and O. Sy, "Optimizing decision-making: Data analytics applications in management information systems," *APTISI Transactions on Management*, vol. 8, no. 2, pp. 115–122, 2024.
[30] A. Jain, "Ai-powered cloud security: A review of intrusion detection and prevention strategies." *International Journal of Advanced Research in Computer Science*, vol. 16, no. 6, 2025.
[31] P. Liu, L. Chen, H. Zhang, Y. Zhang, C. Liu, C. Li, and Z. Wang, "Pear: Positional-encoded asynchronous autoregression for satellite anomaly detection," *Pattern Recognition Letters*, vol. 176, pp. 96–101, 2023.
[32] R. Fahrudin, Y. F. DWI, F. A. YADI, A. Wilson, and T. Kuusk, "Addressing regulatory risks in fintech through decentralized technologies," *APTISI TRANSACTIONS ON MANAGEMENT: iLearning Journal Center*, vol. 8, no. 3, pp. 204–212, 2024.

[33] H. J. D. S. De Queiroz and H. MacLennan, "Threat detection and anomaly identification using deep learning," in *Revolutionizing Cybersecurity With Deep Learning and Large Language Models*. IGI Global Scientific Publishing, 2025, pp. 65–96.

[34] P. Li, G. Wu, Y. Zhou, and J. Leng, "Enhancing random surface anomaly detection in real-world using a four-stage one-class approach," *Pattern Recognition Letters*, vol. 194, pp. 32–40, 2025.

[35] R. PUPUT and N. Sintesa, "The influence of financial literacy on the investment behavior of indonesian migrant workers in japan," *APTISI TRANSACTIONS ON MANAGEMENT : iLearning Journal Center*, vol. 8, no. 3, pp. 175–185, 2024.

[36] O. Aouedi, V. A. Le, K. Piamrat, and Y. Ji, "Deep learning on network traffic prediction: Recent advances, analysis, and future directions," *ACM computing surveys*, vol. 57, no. 6, pp. 1–37, 2025.

[37] M. M. Rathore, A. Ahmad, and A. Paul, "Real time intrusion detection system for ultra-high-speed big data environments," *The Journal of Supercomputing*, vol. 72, no. 9, pp. 3489–3510, 2016.

[38] E. Susetyono, D. S. Priyarsono, A. Sukmawati, and P. Nurhayati, "Improving risk management maturity in ultra micro soe holding companies," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 1, pp. 310–324, 2026.

[39] H. B. Ahmad, H. Gao, and N. Latif, "Adaptive anomaly detection and classification in critical infrastructure systems: A real-time privacy-preserving multi-model framework," *High-Confidence Computing*, p. 100360, 2025.

[40] E. T. Rusmiati, L. Febrina, Y. Sari, and E. M. S. Sakti, "Adoption of ai driven ecological preaching systems using sem pls analysis," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 1, pp. 284–295, 2026.

[41] E. C. P. Neto, S. Iqbal, S. Buffett, M. Sultana, and A. Taylor, "Deep learning for intrusion detection in emerging technologies: a comprehensive survey and new perspectives," *Artificial Intelligence Review*, vol. 58, no. 11, p. 340, 2025.

[42] H. J. Hadi, Y. Cao, S. Li, Y. Hu, J. Wang, and S. Wang, "Real-time collaborative intrusion detection system in uav networks using deep learning," *IEEE Internet of Things Journal*, vol. 11, no. 20, pp. 33 371–33 391, 2024.

[43] I. H. Ji, J. H. Lee, M. J. Kang, W. J. Park, S. H. Jeon, and J. T. Seo, "Artificial intelligence-based anomaly detection technology over encrypted traffic: A systematic literature review," *Sensors*, vol. 24, no. 3, p. 898, 2024.

[44] E. Arif, S. Suherman, and A. P. Widodo, "Analyzing public sentiment on digital banks in indonesia via social media x," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 1, pp. 253–267, 2026.

[45] J. R. Trillo, F. González-López, J. A. Morente-Molinera, R. Magán-Carrión, and P. García-Sánchez, "Evaluation of explainable, interpretable and non-interpretable algorithms for cyber threat detection," *Electronics*, vol. 14, no. 15, p. 3073, 2025.

[46] J. Vansiya, A. Chandi, and R. A. Khan, "Ai-based intrusion detection & prevention models for smart home iot systems: A literature review," *Journal of Computer Science and Technology Studies*, vol. 7, no. 3, pp. 982–996, 2025.

[47] J. B. Hendrawidjaja, B. W. Soetjipto, R. D. Kusumastuti, and O. Jayanagara, "Ecosystem exchange, strategic capabilities, and firm performance with agility and innovation mediators," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 1, pp. 226–238, 2026.

[48] N. Tatipatri and S. Arun, "A comprehensive review on cyber-attacks in power systems: Impact analysis, detection, and cyber security," *IEEe Access*, vol. 12, pp. 18 147–18 167, 2024.

[49] Y. Zhang, R. C. Muniyandi, and F. Qamar, "A review of deep learning applications in intrusion detection systems: overcoming challenges in spatiotemporal feature extraction and data imbalance," *Applied Sciences*, vol. 15, no. 3, p. 1552, 2025.

[50] Y. H. Dulanlebit, H. Hernani, L. Liliasari, M. B. Amran, and G. A. Pangilinan, "Technopreneurship and market feasibility of modified carrageenan hydrogel for industrial heavy metal remediation," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 1, pp. 199–210, 2026.