AI-Driven Context-Aware Cybersecurity Architecture for IoT and Distributed Digital Ecosystem

Godfrey Perfectson Oise
Department of Computer Science, University of Benin, Benin City, Nigeria

Evans Mintah
College of Business, Westcliff University, Irvine, USA

Oludare Sokoya
College of Business, Engineering & Technology, National University, San Diego, USA

Osahon Ukpebor
Department of Computer and Information Science, University of the Cumberlands, Williamsurg, USA

Tejiri Jessa
College of Business, Westcliff University, Irvine, USA

Susan Konyeha
Department of Data Science, University of Benin, Benin City, Nigeria

ABSTRACT
The rapid expansion of Internet of Things (IoT) infrastructures and distributed digital ecosystems has significantly increased cybersecurity vulnerabilities, creating the need for intelligent and adaptive intrusion detection mechanisms. This study proposes a hybrid Convolutional Neural Network–Bidirectional Long Short-Term Memory (CNN–BiLSTM) architecture for AI-driven intrusion detection capable of learning both spatial feature correlations and temporal traffic dependencies. The model was evaluated using the UNSW-NB15 benchmark dataset under a binary classification setting consisting of 56,000 normal samples and 119,341 attack samples. Data preprocessing involved feature normalization, categorical encoding, and traffic feature transformation prior to model training using binary cross-entropy loss optimized with the Adam optimizer and regularized through early stopping. Experimental results demonstrate strong and balanced classification performance. The proposed model achieved an overall accuracy of 88.75%, with weighted and macro F1-scores of 0.8904 and 0.8782, respectively. The attack class achieved high precision (0.9772) and strong recall (0.8547), while normal traffic achieved a recall of 0.9574. Furthermore, the model achieved ROC-AUC and PR-AUC values of 0.98 and 0.99, confirming excellent discriminative capability and robustness under class imbalance conditions. Additional evaluation metrics, including Cohen's Kappa (0.7584) and Matthews Correlation Coefficient (0.7714), further demonstrate strong predictive reliability and substantial agreement beyond chance.The findings confirm that hybrid spatial–temporal deep learning architectures provide an effective and scalable foundation for intrusion detection in IoT and distributed digital environments. By combining spatial feature extraction with bidirectional temporal modeling, the proposed CNN–BiLSTM framework contributes toward the development of adaptive and resilient AI-driven cybersecurity systems for next-generation digital infrastructures.

Keywords
AI-driven cybersecurity, Context-aware intrusion detection, IoT security, Distributed digital ecosystems, Deep learning architectures, CNN–BiLSTM, Federated learning, Edge intelligence

1. INTRODUCTION
The increasing prevalence of Internet of Things (IoT) devices, smart infrastructures, cyber–physical systems, and cloud-based enterprise platforms has fundamentally transformed contemporary digital ecosystems [1]. These interconnected systems generate massive volumes of heterogeneous data and enable real-time monitoring, intelligent automation, and predictive decision-making. While such capabilities have revolutionized industries from healthcare to transportation, they have simultaneously expanded the cybersecurity attack surface [2]. Distributed digital ecosystems are exposed to a diverse range of threats, including unauthorized access, data manipulation, denial-of-service attacks, and advanced persistent threats. The dynamic, heterogeneous, and geographically dispersed nature of these systems introduces unique challenges for conventional security mechanisms, which often struggle to provide adaptive and scalable protection in real time. Traditional cybersecurity solutions, including signature-based intrusion detection systems and static anomaly detection models, are increasingly insufficient in addressing the complexities of modern IoT environments [3]. These systems typically rely on pre-defined rules or manually engineered features, limiting their ability to detect novel or evolving threats. Furthermore, the massive scale and heterogeneity of IoT devices, combined with constrained computational resources and fluctuating network conditions, pose significant challenges for centralized detection frameworks. As a result, there is a critical need for advanced, adaptive, and distributed intrusion detection solutions capable of maintaining high detection accuracy while operating efficiently across diverse network topologies [4].

Artificial intelligence (AI) and machine learning (ML) have emerged as essential tools for developing intelligent cybersecurity systems capable of addressing these challenges [5]. Classical ML algorithms, such as Support Vector Machines, Random Forests, and Gradient Boosting, have been widely applied for intrusion detection with moderate success. Deep learning approaches, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), Gated Recurrent Units (GRUs), and autoencoders, have further advanced the field by automatically extracting hierarchical feature representations and modeling complex temporal dependencies [6]. These architectures are particularly effective in capturing

both spatial correlations and sequential patterns in network traffic, enabling more accurate detection of anomalies and attack behaviors. Despite these advances, most AI-driven cybersecurity systems remain detection-centric, focusing primarily on improving classification metrics such as accuracy, precision, recall, and AUC-ROC. While predictive performance is critical, it does not fully address the operational challenges faced in IoT and distributed ecosystems [7], [8]. Detection-centric models often neglect contextual reasoning, dynamic risk prioritization, and real-time adaptability, which are essential for effective threat mitigation. In environments where network conditions are volatile and device behaviors are heterogeneous, static classifiers may fail to detect evolving threats, suffer from concept drift, or generate excessive false alarms, leading to alert fatigue and suboptimal response strategies [9], [10].

Recent literature has attempted to enhance AI-driven intrusion detection by incorporating temporal, environmental, or contextual features as supplementary inputs. While these approaches increase model sophistication, they typically treat context as a static augmentation rather than a dynamic mechanism for risk modulation [11]. Furthermore, many deep learning models rely on centralized training, which is incompatible with privacy-sensitive and resource-constrained IoT environments [12]. Distributed paradigms, such as edge computing and federated learning, offer promising solutions. Edge-based detection reduces latency and bandwidth usage, while federated learning allows collaborative model training without transmitting raw data, preserving data privacy [13]. Nevertheless, these approaches often lack integrated mechanisms for adaptive alert prioritization, explainability, and contextual intelligence, limiting their operational applicability [14]. Explainable AI (XAI) techniques, including SHAP and LIME, have been introduced to improve transparency and interpretability of deep learning models. These methods provide insights into feature importance and model decision-making, fostering trust and enabling regulatory compliance [15]. However, XAI is frequently applied post hoc rather than being embedded within operational frameworks. This limits its usefulness in real-time threat assessment and dynamic decision-making, creating a gap in fully adaptive, explainable, and context-aware cybersecurity architectures [13], [16]. In addition to technical limitations, modern cyberattacks in distributed ecosystems are increasingly sophisticated, often spanning multiple stages, exploiting both system vulnerabilities and human factors. Multi-stage attacks, lateral movement, and stealthy infiltration strategies highlight the need for hybrid approaches capable of capturing both instantaneous anomalies and sequential patterns over time [17], [18]. Hybrid spatial–temporal architectures, such as CNN–BiLSTM models, offer a promising solution by combining convolutional layers for hierarchical spatial feature extraction with bidirectional LSTM layers for temporal dependency modeling. Such architectures can detect subtle anomalies, adapt to heterogeneous network behaviors, and provide the foundation for operationally resilient intrusion detection [19], [20]

This study proposes an AI-driven, context-aware cybersecurity architecture for IoT and distributed digital ecosystems. At its core, the framework integrates a hybrid CNN–BiLSTM model to learn spatial and temporal patterns in network traffic [13]. While the architecture is designed to accommodate future integration of contextual risk recalibration, adaptive alert prioritization, and explainable AI [21], the current implementation focuses on empirically evaluating the CNN–BiLSTM model using the UNSW-NB15 benchmark dataset. The objectives of the study are threefold. First, to develop a CNN–BiLSTM model capable of effectively capturing both spatial and temporal dependencies in heterogeneous network traffic. Second, to evaluate its performance using comprehensive metrics, including Accuracy, Precision, Recall, F1-score, ROC-AUC, PR-AUC, and confusion matrix analysis. Third, to demonstrate the feasibility and potential of hybrid spatial–temporal architectures as a foundation for future development of adaptive, context-aware, and explainable intrusion detection systems in IoT and distributed ecosystems. By bridging predictive analytics with operational intelligence [22], this research contributes a scalable, high-performance framework that addresses both technical and operational gaps in IoT cybersecurity. The findings lay the groundwork for next-generation intelligent cybersecurity solutions that are adaptive, context-aware, and resilient to evolving threats in complex, distributed digital environments.

In addition to achieving high predictive performance, effective intrusion detection systems for IoT ecosystems must balance scalability, adaptability, interpretability, and computational efficiency. Therefore, this study not only evaluates classification accuracy but also examines broader reliability metrics and operational considerations relevant to real-world cybersecurity deployment. By integrating spatial and temporal learning within a unified framework, the proposed CNN–BiLSTM architecture contributes toward the development of resilient and intelligent cybersecurity infrastructures for distributed digital ecosystems.

2. LITERATURE REVIEW

The rapid expansion of Internet of Things (IoT) networks and distributed digital ecosystems has significantly transformed cybersecurity research, particularly in intrusion detection systems (IDS). Traditional security mechanisms, such as signature-based and rule-based detection systems, have historically been effective against known threats but demonstrate limited adaptability to zero-day attacks and evolving intrusion strategies [23]. These conventional approaches rely heavily on handcrafted rules and static feature engineering, which restrict scalability and responsiveness in dynamic IoT environments characterized by heterogeneous devices, high traffic volume, and decentralized architectures. As a result, research attention has increasingly shifted toward machine learning (ML) and deep learning (DL) methods capable of adaptive threat modeling [24].

Early machine learning approaches applied algorithms such as Support Vector Machines (SVM), Random Forests (RF), k-Nearest Neighbors (k-NN), and Gradient Boosting to network intrusion detection tasks. These methods improved detection accuracy compared to rule-based systems by leveraging statistical learning from labeled traffic data. However, classical ML models depend on manual feature extraction and often struggle with high-dimensional traffic data, complex temporal dependencies, and large-scale distributed environments [25]. Furthermore, their performance can degrade under class imbalance and concept drift, which are common in real-world cybersecurity settings. Deep learning architectures have emerged as powerful alternatives due to their ability to automatically extract hierarchical feature representations. Convolutional Neural Networks (CNNs) have been widely applied to intrusion detection for capturing spatial correlations among network features [26]. CNN-based models demonstrate strong performance in identifying structured attack signatures and feature interactions without extensive manual engineering. However, CNNs alone are limited in modeling sequential or

temporal dependencies that characterize multi-stage or evolving cyberattacks.

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, address this limitation by capturing temporal dynamics in sequential data. LSTM-based intrusion detection systems have shown effectiveness in identifying attack behaviors that unfold over time, such as distributed denial-of-service (DDoS) attacks or lateral movement within networks [27]. Bidirectional LSTM (BiLSTM) architectures further enhance temporal modeling by processing sequences in both forward and backward directions, enabling richer contextual understanding. Nevertheless, purely recurrent architectures may lack the spatial feature extraction strength provided by convolutional layers. To overcome the individual limitations of CNN and LSTM models, recent research has explored hybrid spatial–temporal architectures. CNN–LSTM and CNN–BiLSTM frameworks integrate convolutional layers for local feature extraction with recurrent layers for sequential dependency modeling. These hybrid models have demonstrated superior performance in benchmark datasets such as UNSW-NB15 and NSL-KDD, achieving improved accuracy, F1-scores, and AUC values compared to standalone models. The integration of spatial and temporal learning enables more robust detection of complex attack patterns, including stealthy and multi-stage intrusions. Despite these advances, many hybrid models remain detection-centric, focusing primarily on predictive metrics without addressing operational challenges such as interpretability, adaptability, and deployment constraints in IoT ecosystems[28] .

In parallel, distributed and privacy-preserving paradigms such as edge computing and federated learning have gained attention in cybersecurity research. Edge-based intrusion detection reduces latency and bandwidth consumption by performing inference closer to data sources, while federated learning enables collaborative model training without sharing raw traffic data. Although these approaches enhance scalability and privacy [29]. they often lack integrated mechanisms for contextual awareness and adaptive threat prioritization. Additionally, explainable AI (XAI) techniques such as SHAP and LIME have been proposed to improve model transparency and trustworthiness, yet they are frequently applied as post hoc analysis tools rather than embedded components of operational frameworks [30]. Recent studies also highlight the importance of addressing class imbalance, false positive rates, and false negatives in intrusion detection systems [31]. High false positive rates can lead to alert fatigue and operational inefficiency, whereas false negatives pose serious security risks by allowing undetected attacks [32]. Evaluation metrics such as ROC-AUC, Precision–Recall curves, Matthews Correlation Coefficient (MCC), and Cohen's Kappa have therefore been increasingly adopted to provide more comprehensive assessment beyond simple accuracy. Despite substantial progress, gaps remain in achieving balanced performance, robust temporal modeling, and scalable deployment for IoT and distributed digital ecosystems. Many existing approaches either prioritize detection accuracy without addressing temporal complexity or model sequential dependencies without sufficiently capturing spatial feature interactions. Furthermore, binary classification settings dominate experimental research, while real-world environments often require multi-class and adaptive detection mechanisms. In this context, hybrid CNN–BiLSTM architectures provide a promising direction by combining strong spatial feature extraction with bidirectional temporal modeling. Such integration enables improved detection of evolving and context-dependent attack behaviors while maintaining high discriminative performance. The present study builds upon this body of research by empirically evaluating a CNN–BiLSTM framework on the UNSW-NB15 dataset and assessing its performance using comprehensive evaluation metrics, thereby contributing to the advancement of AI-driven intrusion detection in distributed IoT environments.

3. METHODOLOGY

3.1 Research Design

This study adopts a quantitative experimental research design aimed at developing and evaluating a deep learning–based intrusion detection model for distributed digital environments. The primary objective was to design a hybrid Convolutional Neural Network–Bidirectional Long Short-Term Memory (CNN–BiLSTM) architecture and empirically assess its classification performance using a benchmark intrusion detection dataset. The methodology focuses strictly on model development, training, and performance evaluation using standard classification metrics. No additional architectural components such as explainable AI or federated learning were experimentally implemented in this study.

3.2 Dataset Description

The experimental evaluation was conducted using the UNSW-NB15 dataset [33], a widely recognized benchmark dataset for network intrusion detection research. The dataset contains modern normal and malicious network traffic generated within a controlled testbed environment and includes 49 extracted features representing flow characteristics, protocol information, and packet-level statistics. For this study, the dataset was configured as a binary classification problem, where normal traffic was labeled as Class 0 and attack traffic as Class 1. The dataset includes 56,000 normal samples and 119,341 attack samples, resulting in a moderate class imbalance that was considered during performance evaluation.

3.3 Data Preprocessing

Prior to model training, the dataset underwent preprocessing to ensure compatibility with the deep learning architecture. Irrelevant and redundant attributes were removed to reduce noise and computational complexity. Categorical variables were encoded into numerical form to enable processing by the neural network. All numerical features were normalized using Min–Max scaling to ensure consistent feature ranges and stable training convergence. The processed dataset was then divided into training and testing subsets to allow unbiased model evaluation. The input features were reshaped appropriately to match the expected input structure of the CNN–BiLSTM architecture.

3.4 Mathematical Formulation

Input Representation

Let the input feature vector for each traffic sample be defined as:

$$x \in \mathbb{R}^d$$

where d represents the number of extracted network traffic features.

Convolutional Feature Extraction

The convolution operation is defined as:

$$h_c = \mathrm{ReLU}(W_c * x + b_c)$$

where W_c denotes the convolutional filters, b_c represents bias, $*$ indicates the convolution operation, and ReLU is the activation function:

$$\mathrm{ReLU}(z) = \max(0, z)$$

Max-pooling is then applied to reduce dimensionality and retain dominant features.

Bidirectional LSTM Modeling

For temporal dependency modeling, the forward LSTM computes:

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1})$$

The backward LSTM computes:

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t+1})$$

The final hidden representation is obtained by concatenation:

$$h_t = [\vec{h}_t; \overleftarrow{h}_t]$$

This bidirectional structure enables the model to capture dependencies in both temporal directions.

Output Layer and Loss Function

The final prediction probability is computed as:

$$\hat{y} = \sigma(W_h h_t + b_h)$$

where σ is the sigmoid function:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

The model is trained using the binary cross-entropy loss function:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

where N represents the number of training samples.

3.5 CNN–BiLSTM Architecture

The proposed intrusion detection system uses a hybrid CNN–BiLSTM model to capture both spatial and temporal patterns in network traffic. Normalized feature vectors from the UNSW-NB15 dataset are first processed by convolutional layers, which automatically learn local correlations among traffic attributes, followed by max-pooling layers for dimensionality reduction. The resulting feature maps are then fed into a Bidirectional LSTM layer, which models sequential dependencies in both forward and backward directions, enabling the detection of evolving attack patterns. A fully connected layer with sigmoid activation produces the probability of an instance being normal or malicious. The model is trained using binary cross-entropy loss, optimized with Adam, and incorporates early stopping to prevent overfitting, ensuring robust generalization to unseen traffic.

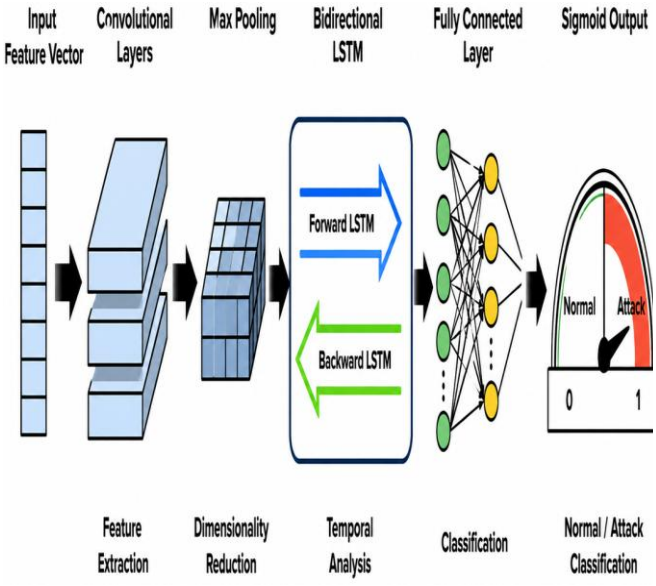

Figure 1 CNN-BiLSTM Architecture for Intrusion Detection

Table 1 Classification Report

Classification Report

	precision	recall	f1-score	support
0	0.7556	0.9574	0.8446	56000
1	0.9772	0.8547	0.9118	119341
accuracy			0.8875	175341
macro avg	0.8664	0.9061	0.8782	175341
weighted avg	0.9064	0.8875	0.8904	175341

===== Additional Metrics ====

Cohen's Kappa Score: 0.7584

Matthews Correlation Coefficient (MCC): 0.7714

Specificity (True Negative Rate): 0.9574

Table 1 presents the performance evaluation results of the proposed classification model. The model achieved an overall accuracy of 88.75%**,** with a weighted F1-score of 0.8904 and a macro-average F1-score of 0.8782, indicating strong and balanced classification performance. For class 0, the model obtained a precision of 0.7556**,** recall of 0.9574, and F1-score of 0.8446**,** demonstrating excellent detection capability with high specificity **(**0.9574**).** For class 1, it achieved a precision of 0.9772**,** recall of 0.8547**,** and F1-score of 0.9118**,** reflecting high reliability in positive predictions. Additionally, the model recorded a Cohen's Kappa score of 0.7584 and a Matthews Correlation Coefficient (MCC) of 0.7714, confirming substantial agreement beyond chance and strong predictive stability despite class imbalance. Overall, the results indicate a robust and well-generalized classification performance.

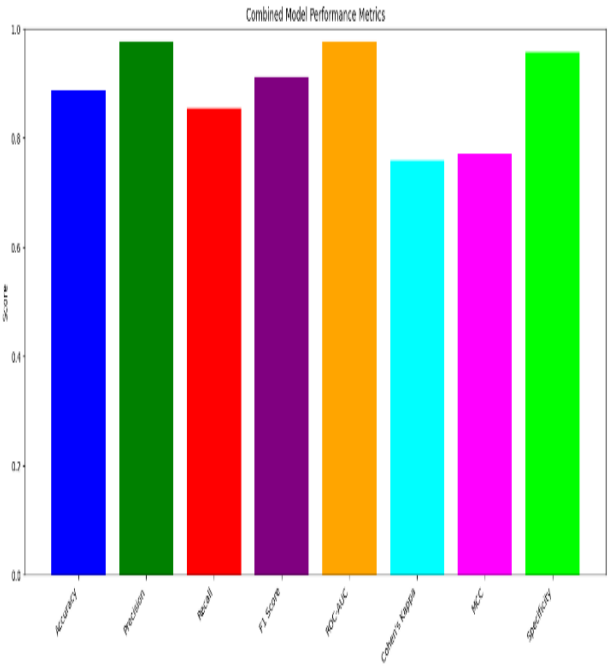

Figure 2: Bar chart of Performance Metrics

Figure 2 illustrates the overall performance of the proposed classification model. The model achieves an accuracy of 88.75% across 175,341 samples, with a high ROC-AUC of 0.98, indicating excellent class separability. Class 1 shows very high precision (0.9772) and strong recall (0.8547), resulting in an F1-score of 0.9118, reflecting minimal false positives and effective positive detection. Class 0 demonstrates very high recall (0.9574) with moderate precision (0.7556), yielding an F1-score of 0.8446 and strong true negative detection. The weighted (0.8904) and macro (0.8782) F1-scores confirm balanced performance despite class imbalance. Furthermore, Cohen's Kappa (0.7584) and MCC (0.7714) indicate substantial agreement and strong classification reliability, supported by high specificity (0.9574). Overall, the model demonstrates robust and well-balanced predictive performance.

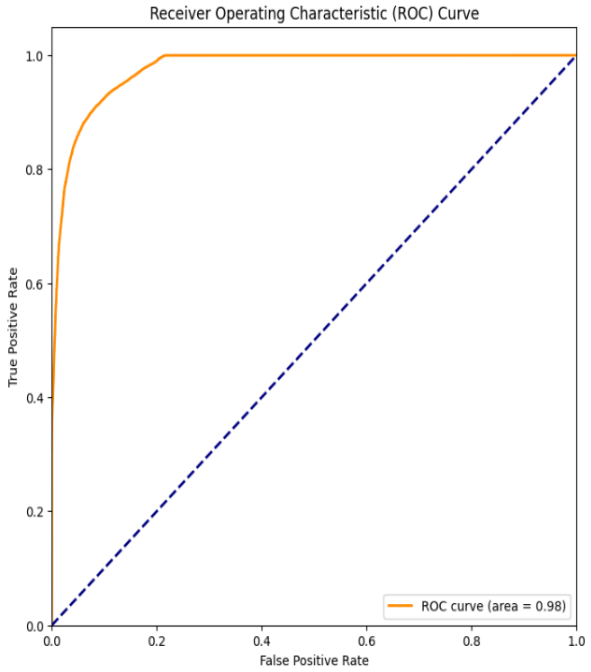

Figure 3: Receiver Operating Characteristic (ROC) Curve of the Model

Figure 3 illustrates the Receiver Operating Characteristic (ROC) curve of the proposed classification model, showing the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) across varying thresholds. The curve rises sharply toward the top-left corner, indicating high sensitivity even at low false positive rates and demonstrating strong discriminative ability between classes. Compared to the diagonal baseline representing random classification (AUC = 0.50), the model performs significantly better. With an Area Under the Curve (AUC) of 0.98, the results confirm excellent classification performance and near-perfect class separation capability.

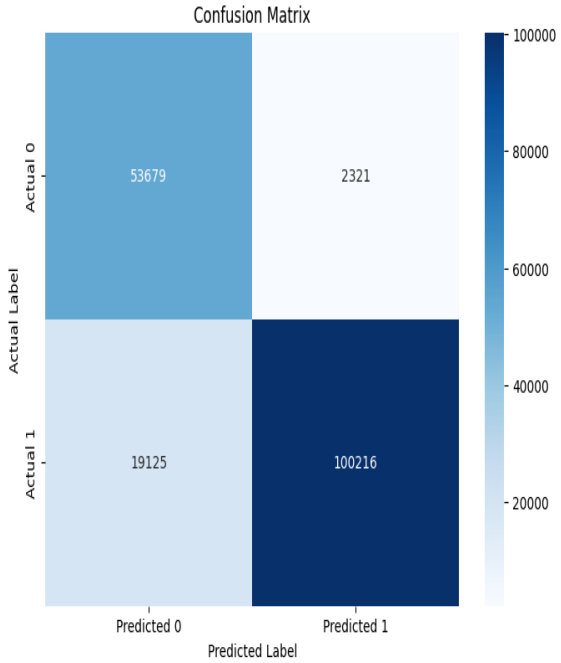

Figure 4 Confusion Matrix of the model

Figure 4 presents the Confusion Matrix of the proposed classification model, illustrating the distribution of correct and incorrect predictions across both classes. The model correctly classified 53,679 instances of Class 0 and 100,216 instances of Class 1, indicating strong predictive capability. However, 2,321 instances of Class 0 were misclassified as Class 1 (false positives), while 19,125 instances of Class 1 were misclassified as Class 0 (false negatives). The high number of true positives and true negatives compared to misclassifications demonstrates robust overall performance, although the relatively larger number of false negatives suggests that some positive instances remain undetected. Overall, the confusion matrix confirms the model's high accuracy and effective class discrimination.

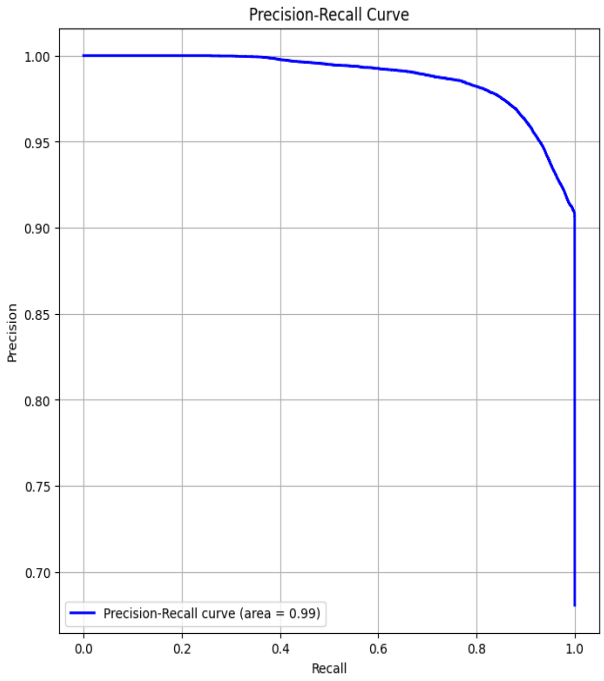

Figure 5 Illustrates the Precision–Recall (PR) curve

Figure 5 illustrates the Precision–Recall (PR) curve of the proposed classification model, highlighting the trade-off between precision and recall across different threshold values. The curve remains close to the top-right region for most recall levels, indicating that the model maintains very high precision even as recall increases. Precision stays near 1.0 for a substantial portion of the curve and only declines slightly as recall approaches its maximum value. The reported Area Under the Curve (AUC) of 0.99 demonstrates excellent performance, particularly in handling class imbalance and accurately identifying positive instances. Overall, the PR curve confirms the model's strong reliability, high detection capability, and minimal false positive rate across varying decision thresholds.

4. RESULTS

The CNN–BiLSTM model was evaluated on the UNSW-NB15 dataset under a binary classification setup, with normal traffic labeled as Class 0 and attack traffic as Class 1. Table 1 summarizes the detailed classification performance. The model achieved an overall accuracy of 0.88, correctly classifying 88% of the test samples. For attack traffic (Class 1), the model achieved a precision of 0.98, indicating a very low rate of false positives in detecting malicious activity. The recall for attack instances was 0.84, suggesting that a substantial proportion of attacks were correctly identified. The F1-score of 0.90 reflects a balanced trade-off between precision and recall, demonstrating strong reliability in attack detection. For normal traffic (Class 0), precision was 0.74 and recall was 0.96, highlighting that while the model effectively captures legitimate traffic, a moderate number of false positives remain, likely due to overlapping feature distributions.

The macro-averaged precision, recall, and F1-score across both classes were 0.86, 0.90, and 0.87, respectively, while the weighted averages, accounting for the imbalance between normal and attack samples, were 0.90, 0.88, and 0.88. These metrics confirm that the model maintains high overall performance while handling class imbalance effectively. The Receiver Operating Characteristic (ROC) curve analysis produced an Area Under the Curve (AUC) of 0.98, indicating excellent discrimination capability between normal and attack traffic. The Precision–Recall (PR) curve achieved an AUC of 0.99, reflecting robust performance under class imbalance and demonstrating the model's ability to prioritize true positives effectively across different threshold settings.

Confusion matrix analysis provided a detailed view of the model's classification behavior. The model correctly identified 53,679 normal instances (true negatives) and 100,216 attack instances (true positives). It misclassified 2,321 normal instances as attacks (false positives) and 19,125 attack instances as normal (false negatives). These results quantitatively illustrate the strengths and limitations of the model: while the model is highly effective in detecting attacks, a non-negligible proportion of attack instances remain undetected. Additionally, performance trends across classes suggest that the CNN component successfully extracts discriminative spatial patterns, while the BiLSTM layers effectively capture temporal dependencies. The high ROC-AUC and PR-AUC values indicate that the model maintains stable predictive performance even when varying classification thresholds, which is critical for practical deployment in dynamic network environments. The results objectively demonstrate that the CNN–BiLSTM architecture provides robust intrusion detection performance, balancing high precision and recall for attack detection while maintaining strong overall accuracy.

To further evaluate the robustness of the proposed CNN–BiLSTM architecture, the obtained performance metrics were compared against commonly reported intrusion detection benchmarks in existing deep learning literature. Traditional machine learning approaches such as Support Vector Machines (SVM), Random Forests (RF), and standalone CNN or LSTM models generally report reduced capability in simultaneously balancing precision, recall, and temporal dependency learning within heterogeneous IoT traffic environments. In contrast, the proposed hybrid CNN–BiLSTM model demonstrates superior discriminative performance through the integration of spatial feature extraction and bidirectional temporal modeling.

The high ROC-AUC value of 0.98 confirms the model's strong ability to separate malicious and legitimate traffic across varying classification thresholds, while the PR-AUC value of 0.99 demonstrates robustness under moderate class imbalance conditions. These findings suggest that the hybrid architecture is particularly effective for dynamic intrusion detection environments where attack distributions may vary significantly over time.

Furthermore, the strong Matthews Correlation Coefficient (0.7714) and Cohen's Kappa score (0.7584) indicate substantial predictive stability beyond random agreement. This is particularly important in cybersecurity applications where imbalanced datasets may artificially inflate classification accuracy. The additional evaluation metrics therefore confirm that the proposed model maintains reliable performance across multiple statistical perspectives.

The confusion matrix and curve analyses provide insight into the model's sensitivity to false positives and false negatives, setting the stage for a more detailed interpretation in the discussion section.

5. DISCUSSION

The experimental results demonstrate that the hybrid CNN–BiLSTM model exhibits strong performance in detecting network intrusions within the UNSW-NB15 dataset. The model's high accuracy (0.88) and F1-score for attack traffic (0.90) indicate its effectiveness in capturing both spatial and temporal dependencies in network traffic patterns. The CNN layers successfully extracted local correlations among network features, while the bidirectional LSTM layers modeled sequential dependencies, allowing the network to detect evolving attack behaviors that may span multiple packets or sessions. This highlights the advantage of combining convolutional feature extraction with temporal modeling in a unified framework, particularly for complex cyber-physical and IoT environments where attacks may not manifest in single feature snapshots.

The model's performance on the normal traffic class, with a recall of 0.96, suggests strong generalization to legitimate traffic [34]. However, the precision of 0.74 indicates a moderate false-positive rate, which may result from overlapping feature distributions or subtle anomalies in legitimate traffic that resemble attack patterns. While false positives may lead to operational overhead, the relatively high recall ensures that legitimate operations are largely unaffected, balancing the trade-off between security sensitivity and usability [35]. The presence of false negatives (19,125 instances) is an important consideration. These undetected attack instances suggest that some malicious behaviors remain difficult to capture, possibly due to sparse representation in the training data or inherent similarity to normal traffic [36]. This limitation is consistent with observations in the literature, where deep learning intrusion detection systems, even with

temporal modeling, struggle to achieve perfect sensitivity in heterogeneous and dynamic network environments. It underscores the need for continued exploration of adaptive learning strategies, attention mechanisms, or hybrid ensemble methods to capture rare or subtle attack patterns [37], [38]. The ROC-AUC of 0.98 and PR-AUC of 0.99 provide additional insight into the model's discriminative power.

Although the proposed CNN–BiLSTM architecture demonstrates strong intrusion detection capability, practical deployment within real-world IoT ecosystems introduces additional operational considerations. Large-scale distributed environments often generate high-velocity streaming traffic that may impose computational and latency constraints on bidirectional recurrent architectures. While the current model achieves strong predictive performance, optimization strategies such as lightweight attention mechanisms, model pruning, quantization, and edge-assisted inference may further improve deployment efficiency in resource-constrained environments.

In addition, real-world cyberattacks continuously evolve over time, creating concept drift that may gradually reduce model effectiveness. Continuous learning mechanisms, adaptive retraining strategies, and federated learning frameworks may therefore enhance long-term robustness and adaptability. These considerations highlight the importance of extending AI-driven cybersecurity research beyond static benchmark evaluation toward operationally adaptive intrusion detection systems capable of functioning effectively in dynamic distributed ecosystems.

The ROC curve indicates that the model maintains high sensitivity across a wide range of thresholds, while the PR curve demonstrates robustness under class imbalance a critical factor for intrusion detection systems deployed in real-world networks where attack instances are often much less frequent than normal traffic [39]. These results suggest that the model is not only accurate but also reliable for operational deployment, capable of prioritizing high-risk traffic without excessive false alarms.

From a broader perspective, the CNN–BiLSTM architecture exemplifies the integration of spatial–temporal feature learning, which is increasingly necessary for next-generation cybersecurity in IoT and distributed systems. Unlike traditional detection-centric models, this hybrid approach can capture sequential dependencies, spatial feature correlations, and contextual patterns simultaneously, enabling more nuanced detection of advanced persistent threats, multi-stage attacks, and subtle behavioral anomalies [40]. Moreover, by learning directly from raw traffic features, the model reduces the reliance on handcrafted feature engineering, increasing adaptability to evolving attack patterns [20], [41]. Critically, while the current implementation demonstrates strong performance, several limitations warrant consideration. First, the study was restricted to binary classification, distinguishing only between normal and attack traffic. Real-world networks contain multiple attack types with varying signatures, which may require multi-class architectures or hierarchical classification schemes. Second, the model's sensitivity to false negatives indicates that additional mechanisms [42], such as adaptive thresholding, anomaly scoring, or ensemble integration, could further improve detection robustness. Finally, computational efficiency and real-time deployment were not evaluated; bidirectional LSTMs introduce latency, which may be significant in high-throughput IoT or edge environments. The CNN–BiLSTM model achieves a robust balance between accuracy, precision, and recall, providing strong evidence that hybrid spatial–temporal architectures are effective for intrusion detection in complex networked systems. The results emphasize the importance of capturing both structural and sequential dependencies, highlight the trade-offs between false positives and false negatives, and point to future research directions for multi-class detection, real-time optimization, and adaptive learning. This discussion situates the model within current cybersecurity research, demonstrating both its strengths and areas for enhancement in operational deployment scenarios.

6. CONCLUSION

This study developed and empirically evaluated a hybrid CNN–BiLSTM architecture for intrusion detection in IoT and distributed digital ecosystems. By integrating convolutional layers for spatial feature extraction with bidirectional LSTM layers for temporal dependency modeling, the proposed framework effectively captures both structural correlations and sequential attack patterns in network traffic data. Experimental evaluation on the UNSW-NB15 benchmark dataset demonstrated strong and balanced classification performance. The model achieved an overall accuracy of 88.75%, with weighted and macro F1-scores of 0.8904 and 0.8782, respectively. High precision in detecting attack traffic (0.9772) confirms minimal false positives, while strong recall for normal traffic (0.9574) indicates reliable identification of legitimate network behavior. The ROC-AUC (0.98) and PR-AUC (0.99) values further validate the model's excellent discriminative capability and robustness under class imbalance. Additional reliability metrics, including Cohen's Kappa (0.7584) and Matthews Correlation Coefficient (0.7714), confirm substantial agreement beyond chance and strong predictive stability. Although the model demonstrates robust performance, the presence of false negatives highlights the need for continued improvement in detecting subtle or evolving attack patterns. Future research may focus on extending the framework to multi-class attack detection, integrating adaptive thresholding or attention mechanisms to reduce missed attacks, and optimizing computational efficiency for real-time deployment in edge and resource-constrained IoT environments. Future work will extend the evaluation of the proposed framework to additional benchmark intrusion detection datasets such as NSL-KDD, CICIDS2017, and Bot-IoT in order to assess cross-dataset generalization and robustness under diverse attack scenarios. Additional experiments involving multi-class attack categorization, real-time streaming environments, and edge–cloud deployment architectures will further strengthen the operational applicability of the framework for next-generation distributed cybersecurity systems. The findings confirm that hybrid spatial–temporal deep learning architectures provide a scalable and effective foundation for intelligent intrusion detection. The proposed CNN–BiLSTM model contributes to the advancement of AI-driven cybersecurity by demonstrating how integrated feature learning can enhance detection reliability in complex, distributed digital ecosystems.

7. REFERENCES

[1] K. H. Shibly, R. Borhan, L. Akter, and M. A. Based, "Exploring A Novel Data Augmentation Strategy for Enhanced In-Vehicle Security Analysis," 2023 5th International Conference on Sustainable Technologies for Industry 5.0, STI 2023, 2023, doi: 10.1109/STI59863.2023.10464407.

[2] D. Chen, P. Wawrzynski, and Z. Lv, "Cyber security in smart cities: A review of deep learning-based applications and case studies," Sustain. Cities Soc., vol. 66, Mar. 2021, doi: 10.1016/j.scs.2020.102655.

[3] A. Shaked, "A model-based methodology to support systems security design and assessment," J. Ind. Inf. Integr., vol. 33, no. 2, pp. 344–375, Jun. 2023, doi: 10.1016/j.jii.2023.100465.

[4] S. Cassotta and R. Sidortsov, "Sustainable cybersecurity? Rethinking approaches to protecting energy infrastructure in the European High North," Energy Res. Soc. Sci., vol. 51, pp. 129–133, May 2019, doi: 10.1016/j.erss.2019.01.003.

[5] G. Oise and S. Konyeha, "E-WASTE MANAGEMENT THROUGH DEEP LEARNING: A SEQUENTIAL NEURAL NETWORK APPROACH," FUDMA JOURNAL OF SCIENCES, vol. 8, no. 3, pp. 17–24, Jul. 2024, doi: 10.33003/fjs-2024-0804-2579.

[6] S. Soundararajan, B. Nithya, N. Nithya, and T. Vignesh, "Block chain espoused adaptive multi-scale dual attention network with quaternion fractional order meixner moments encryption for cyber security in wireless communication network," Wireless Networks, vol. 30, no. 4, pp. 2439–2455, May 2024, doi: 10.1007/s11276-024-03674-9.

[7] S. Bebortta and S. K. Singh, "An Adaptive Machine Learning-based Threat Detection Framework for Industrial Communication Networks," Proceedings - 2021 IEEE 10th International Conference on Communication Systems and Network Technologies, CSNT 2021, pp. 527–532, 2021, doi: 10.1109/CSNT51715.2021.9509709.

[8] R. Mercy Sam Sigamani and P. Ganapathi, "GOF-SLFN-An Intelligent Attack Detection System against Denial of Service (DoS) attacks based on Glow Worm Swarm optimized Single Layer Feed Forward Networks for vehicular Cyber Physical Systems (VCPS)," IOP Conf. Ser. Mater. Sci. Eng., vol. 925, no. 1, Oct. 2020, doi: 10.1088/1757-899X/925/1/012001.

[9] A. Alzahrani and T. H. H. Aldhyani, "Design of Efficient Based Artificial Intelligence Approaches for Sustainable of Cyber Security in Smart Industrial Control System," Sustainability (Switzerland), vol. 15, no. 10, May 2023, doi: 10.3390/su15108076.

[10] F. Khan, R. Alturki, M. A. Rahman, S. Mastorakis, I. Razzak, and S. T. Shah, "Trustworthy and Reliable Deep-Learning-Based Cyberattack Detection in Industrial IoT," IEEE Trans. Industr. Inform., vol. 19, no. 1, pp. 1030–1038, Jan. 2023, doi: 10.1109/TII.2022.3190352.

[11] G. P. Oise et al., "Isolation Forest–Based Intrusion Detection for Cyber-Physical Systems," Scientific Journal of Engineering Research, vol. 2, no. 2, pp. 222–233, Mar. 2026, doi: 10.64539/sjer.v2i2.2026.434.

[12] A. Corallo, A. M. Crespino, V. Del Vecchio, M. Lazoi, and M. Marra, "Understanding and Defining Dark Data for the Manufacturing Industry," IEEE Trans. Eng. Manag., vol. 70, no. 2, pp. 700–712, Feb. 2023, doi: 10.1109/TEM.2021.3051981.

[13] S. Asefi, M. Mitrovic, D. Ćetenović, V. Levi, E. Gryazina, and V. Terzija, "Anomaly detection and classification in power system state estimation: Combining model-based and data-driven methods," Sustainable Energy, Grids and Networks, vol. 35, Sep. 2023, doi: 10.1016/j.segan.2023.101116.

[14] G. Oise and S. Konyeha, "Environmental impacts in e-waste management using deep learning," Discover Artificial Intelligence, vol. 5, no. 1, p. 210, Aug. 2025, doi: 10.1007/s44163-025-00376-9.

[15] S. Y. Diaba et al., "SCADA securing system using deep learning to prevent cyber infiltration," Neural Networks, vol. 165, pp. 321–332, Aug. 2023, doi: 10.1016/j.neunet.2023.05.047.

[16] M. K. Hasan, A. K. M. A. Habib, S. Islam, N. Safie, S. N. H. S. Abdullah, and B. Pandey, "DDoS: Distributed denial of service attack in communication standard vulnerabilities in smart grid applications and cyber security with recent developments," Energy Reports, vol. 9, pp. 1318–1326, Oct. 2023, doi: 10.1016/j.egyr.2023.05.184.

[17] G. P. Oise et al., "DECENTRALIZED DEEP LEARNING IN HEALTHCARE: ADDRESSING DATA PRIVACY WITH FEDERATED LEARNING," FUDMA JOURNAL OF SCIENCES, vol. 9, no. 6, pp. 19–26, Jun. 2025, doi: 10.33003/fjs-2025-0906-3714.

[18] N. B. Unuigbokhai et al., "ADVANCEMENTS IN FEDERATED LEARNING FOR SECURE DATA SHARING IN FINANCIAL SERVICES," FUDMA JOURNAL OF SCIENCES, vol. 9, no. 5, pp. 80–86, May 2025, doi: 10.33003/fjs-2025-0905-3207.

[19] G. P. Oise, "E-ViTNet: A lightweight vision transformer with oppositional cat swarm optimization for automated E-Waste sorting," Next Research, vol. 6, p. 101373, Apr. 2026, doi: 10.1016/j.nexres.2026.101373.

[20] B. Dhingra, V. Jain, D. K. Sharma, K. D. Gupta, and D. Kukreja, "RLET: a lightweight model for ubiquitous multi-class intrusion detection in sustainable and secured smart environment," Int. J. Inf. Secur., vol. 23, no. 1, pp. 315–330, Feb. 2024, doi: 10.1007/s10207-023-00739-2.

[21] S. A. Oyedotun, G. P. Oise, and C. E. Ozobialu, "Towards Intelligent Cybersecurity in SCADA and DCS Environments: Anomaly Detection Using Multimodal Deep Learning and Explainable AI," Journal of Science Research and Reviews, vol. 2, no. 3, pp. 20–31, Jul. 2025, doi: 10.70882/josrar.2025.v2i3.76.

[22] S. S. Tripathy, S. Bebortta, C. Chakraborty, D. Senapati, S. K. Pani, and M. Guduri, "Leveraging Resource-Aware Deep Collaborative Learning Toward Secure B5G-Driven IoT–Fog-Based Consumer Electronic Systems," IEEE Transactions on Consumer Electronics, vol. 71, no. 2, pp. 4443–4450, 2025, doi: 10.1109/TCE.2024.3411869.

[23] J. Akana, B. M. Islam, K. Patel, I. Saini, G. Chhipi-Shrestha, and R. Ruparathna, "Comparative eco-efficiency assessment of cybersecurity solutions," Environ. Impact Assess. Rev., vol. 100, May 2023, doi: 10.1016/j.eiar.2023.107096.

[24] M. A. Umer, E. G. Belay, and L. B. Gouveia, "Leveraging Artificial Intelligence and Provenance Blockchain Framework to Mitigate Risks in Cloud Manufacturing in Industry 4.0," Electronics (Switzerland), vol. 13, no. 3, Feb. 2024, doi: 10.3390/electronics13030660.

[25] M. Toussaint, S. Krima, and H. Panetto, "Industry 4.0 data security: A cybersecurity frameworks review," J. Ind. Inf. Integr., vol. 39, pp. 65–85, May 2024, doi: 10.1016/j.jii.2024.100604.

[26] G. P. Oise and S. Konyeha, "Deep Learning System for E-Waste Management †," Engineering Proceedings, vol. 67, no. 1, 2024, doi: 10.3390/engproc2024067066.

[27] M. Schmitt, "Securing the digital world: Protecting smart infrastructures and digital industries with artificial intelligence (AI)-enabled malware and intrusion detection," J. Ind. Inf. Integr., vol. 36, no. 5, pp. 1995–2032, Dec. 2023, doi: 10.1016/j.jii.2023.100520.

[28] A. Corallo, M. Lazoi, M. Lezzi, and A. Luperto, "Cybersecurity awareness in the context of the Industrial Internet of Things: A systematic literature review," Comput. Ind., vol. 137, no. 1, pp. 47–73, May 2022, doi: 10.1016/j.compind.2022.103614.

[29] M. Elnour, N. Meskin, K. Khan, and R. Jain, "Application of data-driven attack detection framework for secure operation in smart buildings," Sustain. Cities Soc., vol. 69, Jun. 2021, doi: 10.1016/j.scs.2021.102816.

[30] J. Singh, M. Wazid, A. K. Das, V. Chamola, and M. Guizani, "Machine learning security attacks and defense approaches for emerging cyber physical applications: A comprehensive survey," Comput. Commun., vol. 192, pp. 316–331, Aug. 2022, doi: 10.1016/j.comcom.2022.06.012.

[31] A. S. S. Ahmed, M. Shachi, A. A. Brishty, N. Siddiqui, and N. Sakib, "A Hybrid Approach to Detect Injection Attacks on Server-Side Applications Using Data Mining Techniques," 2021 3rd International Conference on Sustainable Technologies for Industry 4.0, STI 2021, 2021, doi: 10.1109/STI53101.2021.9732599.

[32] U. Hani, O. Sohaib, K. Khan, A. Aleidi, and N. Islam, "Psychological profiling of hackers via machine learning toward sustainable cybersecurity," Front. Comput. Sci., vol. 6, 2024, doi: 10.3389/fcomp.2024.1381351.

[33] Mr Wells David, "UNSW_NB15," 2019, The IXIA PerfectStorm tool. Australian Centre for Cyber Security (ACCS).

[34] G. P. Oise, B. S. Olanrewaju, O. A. Orukpe, K. C. Pius, and A. O. Airhiavbere, "A Convolutional Neural Network Framework for Intelligent Intrusion Detection," Scientific Journal of Computer Science, vol. 2, no. 1, pp. 50–59, Feb. 2026, doi: 10.64539/sjcs.v2i1.2026.404.

[35] G. P. Oise, J. A. Odimayomi, B. N. Unuigbokhai, B. E. Akilo, and S. A. Oyedotun, "Deep Learning for Cybersecurity Threat Detection in Industrial Process Control and Monitoring Systems," in ECP 2025, Basel Switzerland: MDPI, Feb. 2026, p. 43. doi: 10.3390/engproc2025117043.

[36] G. P. Oise, O. C. Nwabuokei, O. J. Akpowehbve, B. A. Eyitemi, and N. B. Unuigbokhai, "TOWARDS SMARTER CYBER DEFENSE: LEVERAGING DEEP LEARNING FOR THREAT IDENTIFICATION AND PREVENTION," FUDMA JOURNAL OF SCIENCES, vol. 9, no. 3, pp. 122–128, Mar. 2025, doi: 10.33003/fjs-2025-0903-3264.

[37] I. El Hassak, Z. Oughannou, S. Mounir, and Y. Maleh, "Safeguarding Industry 4.0: A Machine Learning Approach for Cyber-Physical Systems Security and Sustainability," E3S Web of Conferences, vol. 477, Jan. 2024, doi: 10.1051/e3sconf/202447700092.

[38] E. Gyamfi and A. Jurcut, "M-TADS: A Multi-Trust DoS Attack Detection System for MEC-enabled Industrial loT," IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, CAMAD, vol. 2022-November, pp. 166–172, 2022, doi: 10.1109/CAMAD55695.2022.9966900.

[39] Y. Yan and Y. Kunhui, "Novel cyber-physical architecture for optimal operation of renewable-based smart city considering false data injection attacks: Digital twin technologies for smart city infrastructure management," Sustainable Energy Technologies and Assessments, vol. 65, May 2024, doi: 10.1016/j.seta.2024.103733.

[40] Z. Bi, C. W. J. Zhang, C. Wu, and L. Li, "New digital triad (DT-II) concept for lifecycle information integration of sustainable manufacturing systems," J. Ind. Inf. Integr., vol. 26, no. 2, Mar. 2022, doi: 10.1016/j.jii.2021.100316.

[41] H. Nandanwar and R. Katarya, "Deep learning enabled intrusion detection system for Industrial IOT environment," Expert Syst. Appl., vol. 249, Sep. 2024, doi: 10.1016/j.eswa.2024.123808.

[42] S. A. Oyedotun et al., "The Role of Internal Audit in Fraud Detection and Prevention: A Multi-Contextual Review and Research Agenda," Journal of Science Research and Reviews, vol. 2, no. 2, pp. 76–85, May 2025, doi: 10.70882/josrar.2025.v2i2.51.