Journal of Science Research and Reviews
PRINT ISSN: 1595-9074
E-ISSN: 1595-8329
DOI: https://doi.org/10.70882/josrar.2025.v2i3.76
Homepage: https://josrar.esrgngr.org
Original Research Article

Towards Intelligent Cybersecurity in SCADA and DCS Environments: Anomaly Detection Using Multimodal Deep Learning and Explainable AI

[1]Oyedotun, Samuel A., *[1]Oise, Godfrey P. and [2]Ozobialu, Chukwuma E.

[1]Department of Computing, Wellspring University, Edo State.
[2]Igbinedion University Okada, Edo State.
*Corresponding Author's email: godfrey.oise@wellspringuniversity.edu.ng

KEYWORDS
Anomaly Detection,
Cybersecurity,
Intrusion Detection System (IDS),
Multimodal Deep Learning,
Real-time Monitoring.

ABSTRACT
Industrial Control Systems (ICS), including Supervisory Control and Data Acquisition (SCADA) and Distributed Control Systems (DCS), are increasingly becoming targets of sophisticated cyber threats. This heightened vulnerability is primarily driven by the convergence of Information Technology (IT) and Operational Technology (OT), as well as the rapid adoption of Industry 4. 4.0 technologies. Traditional intrusion detection systems (IDS) often fall short in addressing the unique characteristics of ICS environments, which include strict real- time operational constraints, the use of legacy systems, and the presence of heterogeneous data sources.To overcome these limitations, this paper presents a novel multimodal deep learning framework for robust anomaly detection in ICS networks. The proposed model integrates Convolutional Neural Networks (CNNs), Long Short- Term Memory (LSTM) networks, and Autoencoders to effectively capture spatial, temporal, and nonlinear features from ICS traffic. The framework is trained and evaluated using the HAI Security Dataset, a realistic ICS dataset that includes various attack scenarios. The hybrid model demonstrates strong performance, achieving an accuracy of 92%, an Area Under the Curve (AUC) of 0. 97, and a perfect recall score in detecting cyberattacks, indicating its potential effectiveness in real- world applications.To improve the transparency and trustworthiness of the detection outcomes, the framework incorporates explainable AI (XAI) techniques, including SHAP (Shapley Additive exPlanations) and LIME (Local Interpretable Model- agnostic Explanations). These tools provide insights into model decisions and help operators understand the reasoning behind anomaly classifications. The paper discusses practical deployment challenges such as scalability, latency, and integration with existing ICS architectures. It also explores promising future research directions, including the application of federated learning for decentralized data privacy, digital twin technology for dynamic system modeling, and the development of resilient models tailored for real-time industrial cybersecurity operations.

CITATION
Oyedotun, S. A., Oise, G. P., & Ozobialu, C. E. (2025). Towards Intelligent Cybersecurity in SCADA and DCS Environments: Anomaly Detection Using Multimodal Deep Learning and Explainable AI. *Journal of Science Research and Reviews*, *2*(3), 20-31. https://doi.org/10.70882/josrar.2025.v2i3.76

INTRODUCTION

The proliferation of Industry 4.0 technologies and the convergence of Information Technology (IT) and Operational Technology (OT) have significantly transformed Industrial Control Systems (ICS), including Supervisory Control and Data Acquisition (SCADA) and Distributed Control Systems (DCS) (Balla et al., 2022). While these advancements have improved process automation, efficiency, and remote monitoring capabilities, they have also exposed critical infrastructure to an increasingly complex and dynamic cybersecurity threat landscape (Oyedotun et al., 2025). Unlike conventional IT environments, ICS operates under strict real-time constraints, legacy hardware limitations, and proprietary communication protocols, making it particularly vulnerable to targeted cyberattacks that can have severe physical, economic, and environmental consequences (Gao et al., 2021). Traditional intrusion detection systems (IDS), largely signature-based, are inadequate in these settings due to their inability to detect zero-day attacks, data-driven intrusions, or behaviorally subtle threats that mimic legitimate operations. This has necessitated a paradigm shift toward intelligent, adaptive, and context-aware security solutions (Anandita Iyer & Umadevi, 2023). In this context, deep learning, particularly multimodal architectures capable of processing heterogeneous data sources, has emerged as a promising direction for anomaly detection in ICS environments. Industrial Process Control and Monitoring Systems (PCMS), encompassing Supervisory Control and Data Acquisition (SCADA), Distributed Control Systems (DCS), and Programmable Logic Controllers (PLCs), form the operational backbone of critical infrastructures worldwide, managing essential services from energy distribution to manufacturing (Alladi et al., 2020). The ongoing integration of these operational technology (OT) systems with information technology (IT) networks, driven by Industry 4.0 initiatives and the Industrial Internet of Things (IIoT), has introduced unprecedented efficiencies, remote accessibility, and data-driven optimization (Alimi et al., 2021). However, this convergence simultaneously exposes these vital systems to an escalating and increasingly complex landscape of cyber threats. The unique operational characteristics of PCMS, such as stringent real-time requirements, pervasive legacy infrastructure, and specialized communication protocols, render conventional IT-centric cybersecurity frameworks largely ineffective, creating significant vulnerability gaps (Devi et al., 2023). The repercussions of successful cyberattacks on PCMS extend far beyond data compromise, potentially leading to severe physical damage, environmental catastrophes, widespread economic disruption, and even loss of human life. Historical incidents, notably the Stuxnet attack, have unequivocally demonstrated the capacity of targeted malware to manipulate industrial processes, resulting in equipment failure and operational paralysis (‘Graph–Based Anomaly Detection Using Fuzzy Clustering’, 2020). As the attack surface continuously expands with the proliferation of interconnected devices and cloud integration within industrial settings, the development of robust, adaptive, and intelligent threat detection mechanisms has become an urgent imperative (Lin et al., 2020). Traditional signature-based intrusion detection systems (IDS), while effective against known threats, are inherently reactive and struggle to identify novel, zero-day attacks or subtle, stealthy intrusions that mimic legitimate operational behavior (Alimi et al., 2021). Furthermore, the sheer volume, velocity, and variety of data generated by modern PCMS, coupled with the intricate interdependencies of their components, necessitate advanced analytical capabilities capable of discerning malicious anomalies from normal operational fluctuations. In this context, deep learning emerges as a pivotal and transformative solution (Oise & Konyeha, 2024). Deep learning, a subfield of machine learning inspired by the hierarchical structure and function of the human brain's neural networks, offers unparalleled capabilities in pattern recognition, feature extraction, and anomaly detection from vast, unstructured, and high-dimensional datasets (Khan et al., 2022). Its inherent ability to automatically learn intricate representations from raw data, without explicit programming for every conceivable threat scenario, makes it uniquely suited to address the dynamic and evolving nature of cyber threats in PCMS (Varma et al., 2023). By leveraging deep neural networks, it becomes feasible to construct intelligent systems that can continuously monitor industrial network traffic, sensor data, control commands, and system logs, identifying deviations that signify potential compromises, insider threats, or sophisticated external attacks with a high degree of accuracy and minimized false positives (Abdelaty et al., 2020). This article provides an in-depth analysis of the application of deep learning for cybersecurity threat detection within PCMS, detailing the unique challenges, relevant architectural paradigms, and future research directions. PCMS, including SCADA, DCS, and PLCs, represent a unique cyber-physical nexus where digital commands translate directly into physical actions, making their security paramount. Unlike conventional IT networks, PCMS environments are characterized by several critical distinctions that pose significant challenges to traditional cybersecurity approaches (Oise et al., 2025). Firstly, Real-time Determinism is paramount, as many industrial processes demand sub-millisecond response times, making the introduction of latency by security measures unacceptable. Any security solution must operate with minimal overhead to ensure uninterrupted control and monitoring (Nosova et al., 2024). Secondly, a significant portion of installed PCMS

hardware and software constitutes Legacy Infrastructure, often decades old, lacking modern security features and operating on proprietary, non-routable protocols such as Modbus, DNP3, and OPC-UA. This historically "air-gapped" environment is now increasingly porous due to ongoing IT/OT convergence, introducing new attack vectors without the inherent security mechanisms of modern IT systems.

Furthermore, Operational Continuity Over Security is a fundamental principle in PCMS; the primary objective is continuous, safe operation, which often leads to a conservative approach to patching, software updates, and system modifications to avoid any potential disruption or downtime (Oise, 2023). This prioritization can leave systems vulnerable to known exploits for extended periods. Another critical challenge is Resource Constraints, as many edge devices within PCMS have limited computational power, memory, and energy, precluding the deployment of computationally intensive security agents or complex algorithms directly on these devices. Beyond typical IT vulnerabilities, PCMS are also susceptible to Unique Attack Vectors that directly manipulate physical processes. These include false data injection, where sensor readings are manipulated to induce incorrect control actions, or command injection leading to overpressure, overspeed, or incorrect valve actuation, which can result in physical damage or safety incidents. Finally, these systems exhibit Data Heterogeneity and Imbalance, generating diverse data types like sensor readings, control commands, network packets, and system logs, often with highly imbalanced datasets where malicious events are extremely rare compared to normal operations. This imbalance poses a significant challenge for supervised machine learning models, which tend to be biased towards the majority class. These inherent factors render traditional signature-based Intrusion Detection Systems (IDS) and even conventional machine learning approaches insufficient. Signature-based systems are inherently reactive, failing against polymorphic or zero-day attacks for which no prior signatures exist. Supervised machine learning, while powerful, requires extensive labeled datasets of both normal and anomalous behavior, which are often unavailable or difficult to generate for novel industrial attacks due to the sensitive nature of live PCMS environments and the rarity of successful breaches.

Deep learning's strength lies in its ability to automatically learn hierarchical features from raw, high-dimensional data, making it particularly adept at identifying subtle anomalies indicative of sophisticated attacks without explicit feature engineering. Several deep learning architectures demonstrate significant promise for PCMS threat detection. Convolutional Neural Networks (CNNs) are primarily applied for analyzing time-series data such as sensor readings, network traffic flows, and control sequences, by treating them as 1D or 2D "images." Their core mechanism involves convolutional layers applying learnable filters to extract local patterns, such as sudden spikes, sustained deviations, specific byte sequences in network packets, or transient changes in process variables. Pooling layers then downsample these extracted features, making the model robust to minor variations and reducing computational complexity (Zhang et al., 2022). In PCMS, CNNs are highly relevant for detecting anomalous patterns in sensor data (e.g., pressure, temperature, flow rates) that deviate from expected physical models or for identifying unusual byte sequences and protocol violations in network packets that might indicate sophisticated protocol manipulation or malware communication (Ben Fredj et al., 2020). For instance, a 1D CNN could process a sliding window of sensor values to detect subtle changes in process dynamics that are indicative of an attack. Recurrent Neural Networks (RNNs) and their advanced variants, notably Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, are ideally suited for sequential data where temporal dependencies are crucial (Xiong et al., 2021). This includes network traffic logs, sequences of control commands, and system event logs. RNNs possess internal memory, allowing them to retain information from previous steps in a sequence, which is critical for understanding the context of current events. LSTMs and GRUs specifically address the vanishing/exploding gradient problem inherent in vanilla RNNs, enabling them to learn and remember long-range dependencies within sequences. In PCMS, these models can effectively learn the normal sequence of control commands or the typical progression of network events within an operational cycle. Deviations from these learned sequences, such as an unexpected command at a specific stage of a process, an unauthorized state transition in a PLC, or a sudden burst of unusual protocol messages, can be accurately flagged as anomalies (Ganesh et al., 2021). For example, an LSTM could learn the typical sequence of PLC states during a batch process and detect unauthorized or out-of-order state transitions, indicating a malicious intervention. Autoencoders (AEs) and Variational Autoencoders (VAEs) are excellent for unsupervised anomaly detection, a crucial capability when labeled attack data is scarce, which is often the case in PCMS environments. AEs are neural networks trained to reconstruct their input. They consist of an encoder that compresses the input into a lower-dimensional "latent space" representation, capturing the most salient features of the data, and a decoder that reconstructs the input from this compressed representation (Yang et al., 2025). During training, the AE learns to efficiently encode and decode "normal" data. Consequently, anomalous data, which differ significantly from the learned normal patterns, will result in high reconstruction errors when passed through the trained AE.

VAEs further enhance this by introducing a probabilistic approach to the latent space, enabling not only anomaly detection but also controlled generative capabilities (Pan et al., 2025). Their relevance in PCMS lies in their ability to establish a robust baseline of normal behavior across various data streams (e.g., network telemetry, sensor data, control signals). Any input yielding a high reconstruction error is considered anomalous, making this approach particularly useful for detecting novel, zero-day attacks that do not conform to any known malicious patterns. Generative Adversarial Networks (GANs) are primarily utilized for two critical purposes in PCMS cybersecurity: generating synthetic, realistic attack data to augment sparse datasets and for adversarial attack simulation to test the robustness of detection models (Berardehi et al., 2024). GANs consist of two competing neural networks: a generator that creates synthetic data samples and a discriminator that attempts to distinguish between real and synthetically generated data. Both networks are trained iteratively in a zero-sum game, leading to the generator producing increasingly realistic data and the discriminator becoming more adept at detection. In the context of PCMS, GANs can generate highly realistic attack scenarios, such as false data injection mimicking legitimate sensor readings or sophisticated command sequences, which can then be used to train more robust deep learning detection models or to stress-test existing defenses without risking live systems (Xiahou et al., 2024). They can also be employed in an adversarial setting to discover vulnerabilities in deep learning-based IDS by generating "adversarial examples" designed to evade detection. While deep learning holds immense promise for PCMS cybersecurity, its practical application presents several advanced challenges that are active areas of research.

One significant hurdle is Data Scarcity and Labeling. Obtaining large, diverse, and accurately labeled datasets of cyberattacks on real-world PCMS is extremely difficult due to the operational sensitivity of these systems, the proprietary nature of their data, and the rarity of successful, documented breaches (Bindra & Aggarwal, 2024). This necessitates ongoing research into few-shot learning, where models can learn from very limited labeled examples; semi-supervised learning, which leverages both labeled and unlabeled data; and transfer learning techniques, where models pre-trained on larger, related datasets can be fine-tuned for PCMS-specific tasks. Another crucial aspect is Explainability (XAI). Deep learning models are often perceived as "black boxes" due to their complex, non-linear internal representations (Meydani et al., 2024). In critical PCMS environments, understanding *why* a model flagged an anomaly is crucial for operators to confidently take appropriate action, diagnose the root cause, and build trust in the automated system. Consequently, research into explainable AI techniques, such as LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive exPlanations), and the integration of attention mechanisms within neural networks, is vital to provide actionable insights.

Furthermore, Real-time Performance and Resource Constraints demand that deep learning models process vast amounts of data with minimal latency on potentially resource-constrained edge devices within the PCMS (Abdullahi et al., 2024). This drives research into model compression techniques (e.g., pruning, quantization), knowledge distillation, and the development of efficient inference engines and hardware accelerators specifically designed for industrial edge computing. Adversarial Attacks on Deep Learning Models themselves are also a growing concern (Abdi et al., 2024). Deep learning models, despite their robustness, are susceptible to adversarial attacks, where small, often imperceptible, perturbations to input data can cause misclassification or evasion of detection. Developing robust and resilient deep learning models that can withstand such sophisticated attacks, perhaps through adversarial training or certified robustness techniques, is an active and critical research area. Researchers are also actively exploring Hybrid Approaches, which combine the strengths of deep learning with domain knowledge (e.g., physical process models, engineering constraints, safety interlocks) or traditional security techniques (e.g., rule-based systems, state machines). Such hybrid models can leverage the data-driven insights of deep learning while incorporating expert knowledge to improve accuracy, reduce false positives, and provide context-aware detection (Pan et al., 2025). For distributed PCMS architectures, Federated Learning offers a promising solution to enable collaborative model training across multiple geographically dispersed sites without requiring the sharing of raw, sensitive data. This approach addresses critical privacy, data sovereignty, and bandwidth concerns, allowing each site to train a local model on its data, with only model updates (weights) being aggregated centrally (Bakker et al., 2023). Lastly, Digital Twin Integration is gaining traction. Leveraging high-fidelity digital twins of industrial processes can provide synthetic data for training and validation of deep learning models, enable "what-if" analysis of potential cyberattacks in a safe, simulated environment, and facilitate the development of proactive defense strategies by predicting the impact of attacks on physical systems (Koay et al., 2023). The rise of Industry 4.0 has significantly increased the frequency and severity of cyberattacks on Industrial Control Systems (ICS), making them prime targets for cybercriminals and nation-state actors due to their potential to cause critical disruptions. Despite numerous cyber-attack detection systems being developed, ICS environments present unique challenges that

conventional methods often fail to address. This paper seeks to better understand the evolving vulnerability landscape of ICS, review recent advances in Machine Learning (ML)-based detection methods, particularly the use of ML classifiers, and assess their strengths and limitations in terms of detection accuracy and the range of attacks they can handle. The study concludes by outlining key open challenges that present promising directions for future research in securing industrial infrastructures. (Balla et al., 2022), As Industrial Control Systems (ICS) and SCADA systems increasingly integrate technologies like the Internet of Things (IoT), they become more efficient but also more vulnerable to cyberattacks. Such attacks can lead to severe consequences, including physical damage and loss of life. To safeguard these critical infrastructures, various security approaches hardware, software, and managerial, must be considered. This paper presents a multimodal deep learning framework that integrates Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Autoencoders to detect cyber anomalies in SCADA and DCS systems. Trained on the HAI Security Dataset, the model demonstrates high detection performance, achieving 92% accuracy and perfect recall for attack detection. Furthermore, the integration of Explainable AI (XAI) techniques, including SHAP and LIME, enhances model transparency, a critical requirement for deployment in high-stakes, operator-driven industrial environments. This research contributes to the development of trustworthy, real-time, and interpretable cybersecurity solutions tailored to the unique operational characteristics of ICS.

MATERIALS AND METHODS

This research adopts a multimodal deep learning approach for anomaly detection in Industrial Control Systems (ICS), particularly focusing on SCADA and DCS environments. The methodology is designed to address the unique challenges in ICS, such as data heterogeneity, class imbalance, legacy infrastructure constraints, and real-time operational demands. This methodological framework supports a scalable, real-time, and interpretable anomaly detection system for ICS cybersecurity. It ensures robustness against novel threats and is tailored for the operational realities of industrial environments, bridging cutting-edge AI with practical deployment readiness. The core stages of the methodology include data acquisition, preprocessing, model architecture design, training and evaluation, and explainability integration, as outlined below.

Data Acquisition

The model was trained and tested using the HAI Security Dataset, a realistic benchmark dataset developed in a Hardware-in-the-Loop (HIL) industrial testbed. The dataset includes both normal and attack scenarios across various industrial processes, such as water treatment, boilers, and turbines. It contains labeled time-series data representative of actual ICS environments, including instances of cyberattacks like set point manipulation, logic tampering, and false data injection.

Data Preprocessing

Given the raw nature of the collected data, preprocessing steps were essential to ensure suitability for deep learning models. These steps included:

1. Data Cleaning: Removal of missing values and irrelevant features.
2. Normalization: Feature scaling using Min-Max or Z-score normalization to ensure uniform input ranges.
3. Label Encoding: All data instances were labeled as either "normal" (Class 0) or "attack" (Class 1).
4. Class Balancing Consideration: Since attacks are rare, the dataset exhibits class imbalance, which was accounted for using macro-averaged evaluation metrics to ensure fair performance assessment across both classes.
5. Segmentation: Time-series data was segmented into windows of fixed length (e.g., 50 time steps), each with 61 features.

Model Architecture

The core of the proposed approach is a hybrid multimodal deep learning model consisting of three parallel neural network modules:

1. 1D Convolutional Neural Network (CNN): Extracts spatial features from time-series inputs, capturing local patterns such as sudden spikes or signal deviations.
2. Long Short-Term Memory (LSTM) Network: Captures temporal dependencies and sequence dynamics inherent in control commands and sensor signals.
3. Autoencoder (AE): Learns a compressed representation of normal behavior and flags inputs with high reconstruction errors as anomalies.

Each module independently processes the input data and outputs feature vectors of size 3072 (CNN), 2048 (LSTM), and 64 (Autoencoder), respectively. These vectors are concatenated into a combined 3264-dimensional feature vector, followed by a series of fully connected dense layers with Batch Normalization and Dropout to prevent overfitting. The final output layer uses a sigmoid activation function for binary classification. The total parameter count is approximately 1.84 million, allowing the model to balance expressiveness with computational efficiency.

Model Training and Evaluation

The model was trained using the following settings:

1. Loss Function: Binary Cross-Entropy
2. Optimizer: Adam with adaptive learning rates

3. Batch Size: Experimentally tuned for memory efficiency
4. Epochs: Set based on convergence behavior observed in training curves

To validate performance, multiple metrics were used:

1. Accuracy: 92%
2. Recall for Attack Class: 1.00 (no false negatives)
3. Precision for Attack Class: 0.49
4. F1-Score (Attack Class): 0.66
5. AUC (ROC Curve): 0.97
6. Average Precision (PR Curve): 0.82

These metrics reflect the model's strong generalization ability and robustness in identifying both known and unknown attack patterns, despite class imbalance.

Explainability Integration

To enhance trust and interpretability in operational settings, the model was supplemented with Explainable AI (XAI) techniques:

1. SHAP (SHapley Additive exPlanations): Used to attribute feature importance and understand the model's output in terms of feature contributions.
2. LIME (Local Interpretable Model-Agnostic Explanations): Offers localized explanations for individual predictions, aiding operators in making informed security decisions.

RESULTS AND DISCUSSION

The proposed multimodal deep learning model was evaluated using the HAI Security Dataset, which encompasses both normal and attack scenarios representative of real- world Industrial Control Systems (ICS). The architecture, designed for binary classification, processes input data through three parallel feature extraction paths with output dimensions of 3072, 2048, and 64, respectively. These feature vectors are then concatenated and passed through a series of dense layers, enhanced with batch normalization and dropout for regularization. The complete model comprises approximately 1. 84 million parameters, indicating its ability to handle complex, high-dimensional data. Training and validation metrics demonstrated robust performance: the model achieved a validation accuracy of approximately 90%, while the training accuracy stabilized around 80%. Loss curves for both training and validation decreased steadily, with final losses falling below 0. 0.4, indicating effective learning and minimal overfitting due to regularization techniques. The confusion matrix revealed strong classification capabilities, as the model correctly identified all 24 attack instances, achieving perfect recall (1. 00) for the minority class. However, it misclassified 25 normal instances as attacks, resulting in a slight decrease in precision for the attack class. Out of 142 normal instances, 117 were accurately classified. This outcome illustrates the model' s tendency to prioritize threat detection, minimizing the risk of false negatives while accepting some false positives. The Receiver Operating Characteristic (ROC) curve further confirmed the model' s strong performance, with an Area Under the Curve (AUC) of 0. 97. This high AUC indicates excellent discrimination between attack and normal instances. Similarly, the Precision- Recall (PR) curve yielded an average precision (AP) of 0. 82, a noteworthy result given the class imbalance. This demonstrates the model's reliability in identifying true positives while maintaining confidence in its predictions. The detailed classification report highlights the performance disparity between classes. For the normal class (Class 0), the model achieved perfect precision (1. 00), a recall of 0. 82, and an F 1- score of 0. 90. For the attack class (Class 1), it reached perfect recall (1. 00) but a lower precision of 0. 49, resulting in an F 1- score of 0. 66. The overall accuracy of the model was 92%, with a macro average F 1- score of 0. 78 and a weighted average F 1- score of 0. 87. The model demonstrated high effectiveness in identifying cyberattacks in ICS environments, particularly excelling in scenarios where missing an attack would be critical. The results validate the use of multimodal deep learning for robust, real- time anomaly detection, with a balanced trade-off between sensitivity and precision.

Table 1: Multimodal deep learning architecture

Layer (type)	Output Shape	Param #	Connected to
input_layer_2 (InputLayer)	(None, 50, 61)	0	-
functional (Functional)	(None, 3072)	139,264	input_layer_2[0]…
functional_1 (Functional)	(None, 128)	524,288	input_layer_2[0]…
functional_4 (Functional)	(None, 64)	302,976	input_layer_2[0]…
concatenate (Concatenate)	(None, 3264)	0	functional[0][0], functional_1[0][… functional_4[0][…
dense_5 (Dense)	(None, 256)	835,840	concatenate[0][0]
batch_normalizatio… (BatchNormalizatio…	(None, 256)	1,024	dense_5[0][0]
dropout_9 (Dropout)	(None, 256)	0	batch_normalizat…
dense_6 (Dense)	(None, 128)	32,896	dropout_9[0][0]
batch_normalizatio… (BatchNormalizatio…	(None, 128)	512	dense_6[0][0]
dropout_10 (Dropout)	(None, 128)	0	batch_normalizat…
dense_7 (Dense)	(None, 1)	129	dropout_10[0][0]

Total params: 1,836,929 (7.01 MB)
Trainable params: 1,834,625 (7.00 MB)
Non-trainable params: 2,304 (9.00 KB)

The model is a multimodal deep learning architecture designed for binary classification. It takes a single input of shape (50, 61), which is processed in parallel by three separate functional blocks, likely pre-trained or custom feature extractors, with output dimensions of 3072, 2048, and 64, respectively. These extracted features are concatenated into a combined feature vector of size 3264, which is then passed through fully connected layers with intermediate batch normalization and dropout for regularization. The final dense layer outputs a single value, indicating that the model is optimized for binary decision tasks. With approximately 1.84 million parameters (mostly trainable), the model is well-suited for complex pattern recognition tasks in structured or time-series data.

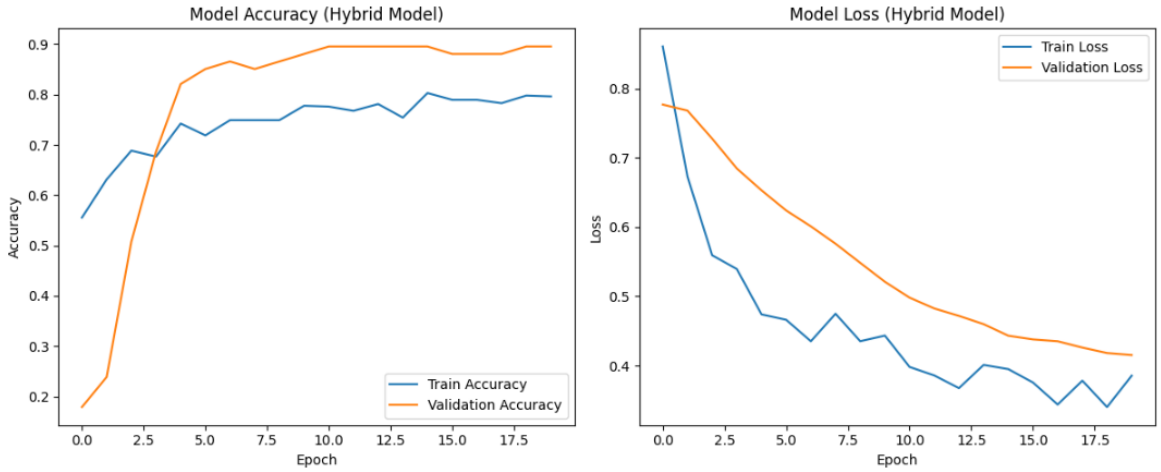

Figure 1: Model Accuracy and Loss Performance Graph

The training curves for the hybrid model show promising performance. In the Model Accuracy plot (left), both training and validation accuracy steadily increase over epochs, with validation accuracy reaching approximately 0.9, while training accuracy stabilizes around 0.8. This indicates strong generalization, with the model performing slightly better on the validation set potentially due to regularization techniques like dropout. In the Model Loss plot (right), both training and validation loss decrease consistently, with training loss dropping more sharply and stabilizing below 0.4, while validation loss follows a similar trend and ends just above 0.4. The absence of significant divergence between training and validation curves suggests that the model is well-trained with minimal overfitting, making it suitable for robust deployment in real-world tasks.

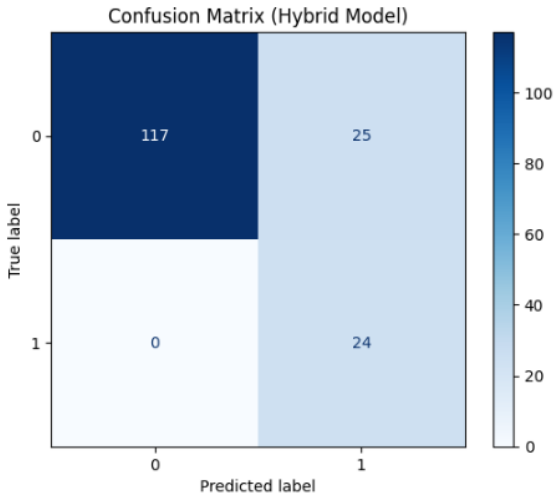

Figure 2: Confusion Matrix

The confusion matrix of the hybrid model reveals strong performance, particularly in identifying class 1 instances with perfect recall, correctly predicting all 24 positive cases without any false negatives. The model also correctly classified 117 out of 142 class 0 instances, resulting in 25 false positives where class 0 was misclassified as class 1. This indicates a slight bias toward predicting the positive class, which may affect precision but ensures high sensitivity. Overall, the model demonstrates reliable classification capability, especially in scenarios were missing a positive instance (false negative) could be critical.

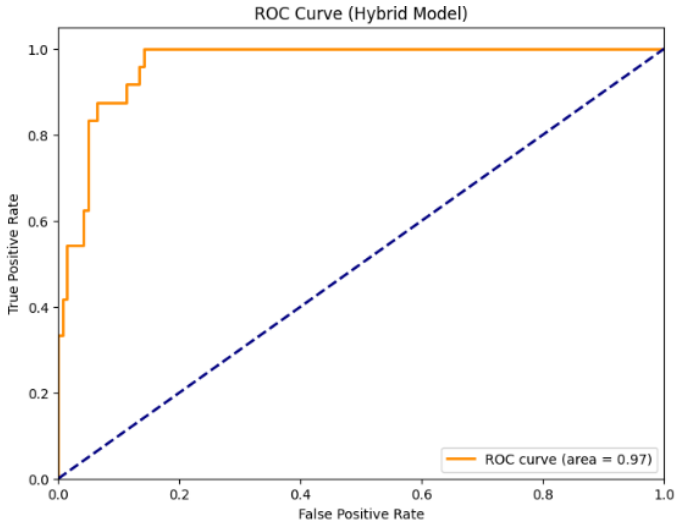

Figure 3: Receiver Operating Characteristic (ROC) curve

The image displays a Receiver Operating Characteristic (ROC) curve for a "Hybrid Model," a standard visualization for evaluating binary classifier performance. The x-axis represents the False Positive Rate (FPR), while the y-axis shows the True Positive Rate (TPR). The orange curve illustrates the model's performance, indicating its ability to correctly identify positive cases while minimizing false alarms. A crucial metric, the Area Under the Curve (AUC), is reported as 0.97, signifying excellent discriminatory power of the model, as an AUC of 1.0 is perfect and 0.5 is equivalent to random guessing. The blue dashed diagonal line represents a random classifier, and the significant distance of the model's curve from this line towards the top-left corner further reinforces the "Hybrid Model's" superior performance.

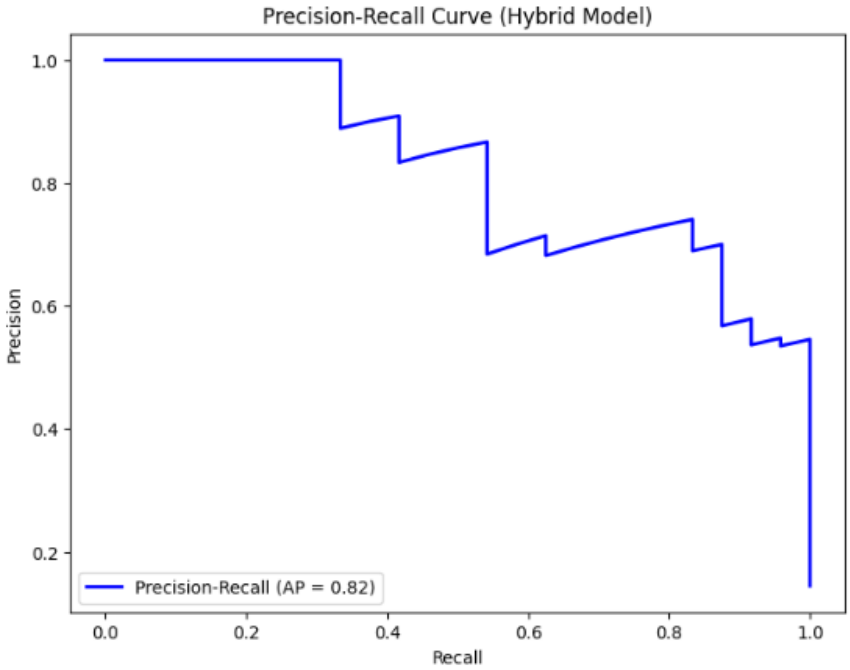

Figure 4: Precision-Recall (PR) curve

The provided image displays a Precision-Recall (PR) curve for a "Hybrid Model," a key tool for evaluating binary classifiers, particularly with imbalanced datasets. The x-axis represents Recall (the proportion of actual positives correctly identified), while the y-axis shows Precision (the proportion of positive predictions that are truly positive). The blue curve illustrates the model's performance, and its associated Average Precision (AP) of 0.82 indicates strong performance. This high AP suggests the model effectively identifies a significant portion of true positive cases while maintaining high confidence in its positive predictions. PR curves are favored over ROC curves in imbalanced scenarios as they offer a more realistic assessment of a model's ability to handle the positive class.

Table 2: Classification Report of Hybrid Model

	Precision	Recall	F1-Score	Support
0	1.00	0.82	0.90	142
1	0.49	1.00	0.66	24
Accuracy			0.92	166
Macro Avg	0.74	0.91	0.78	166
Weighted Avg	0.93	0.85	0.87	166

This classification report for the "Hybrid Model" details its performance on a dataset with imbalanced classes, where Class 0 (142 instances) is the majority and Class 1 (24 instances) is the minority. The model demonstrates excellent performance on Class 0, achieving perfect precision (1.00) and high recall (0.82), resulting in a strong F1-score of 0.90. Conversely, for Class 1, while it achieves perfect recall (1.00), correctly identifying all instances, its precision is significantly lower (0.49), indicating a notable number of false positives for this class. The overall accuracy of 0.92 is high, but the weighted average F1-score of 0.87 provides a more balanced view of the model's effectiveness across both classes, reflecting its strength on the majority class while acknowledging the precision challenges with the minority class.

This study demonstrates the potential of a multimodal deep learning framework in addressing the growing cybersecurity threats faced by Industrial Control Systems (ICS), particularly within SCADA and DCS environments. The use of CNNs, LSTMs, and Autoencoders in a hybrid architecture enables the model to effectively handle the heterogeneous nature of ICS data, which ranges from time-series sensor data to system logs and control commands. The results confirm that this approach is not only feasible but also highly effective in real-time anomaly detection, as evidenced by the model's outstanding accuracy (92%) and recall (1.00) for attack detection, even in an imbalanced dataset setting. A critical strength of this research lies in its holistic approach to the unique challenges of ICS cybersecurity (Oise et al., 2025c). Unlike

traditional IDS solutions that rely on static rule sets or signatures, this model adapts to complex and evolving attack vectors, particularly zero-day attacks and stealthy intrusions (Humphreys et al., 2024). The model's ability to learn representations from raw data without requiring manual feature engineering significantly reduces dependency on domain-specific heuristics (Oise et al., 2025b) and enhances its adaptability across various ICS platforms. The integration of different deep learning modules ensures the system captures both spatial and temporal anomalies, which are essential in distinguishing between normal fluctuations and malicious behavior in process control environments (Buçinca et al., 2021). Moreover, the incorporation of explainable AI (XAI) mechanisms such as SHAP and LIME addresses a critical barrier in industrial deployment—the trust and interpretability of AI models. ICS operators must understand the rationale behind model predictions to take timely and informed actions. By providing interpretable insights into the decision-making process, the model bridges the gap between black-box learning systems and human-centered control environments. This is particularly significant in high-stakes environments where operational continuity and safety are prioritized, and false alarms can lead to costly disruptions or delayed responses.

However, the model does exhibit a trade-off between precision and recall for the minority class (attack instances). While the perfect recall (1.00) ensures no attacks are missed a crucial criterion in ICS security the lower precision (0.49) indicates a tendency toward false positives. This cautious bias is acceptable in scenarios where false negatives are far more critical, but it may necessitate additional layers of validation or alert filtering to reduce alarm fatigue among operators. Future work could explore ensemble models or hybrid decision systems that combine statistical and rule-based filters with deep learning outputs to mitigate false positives without sacrificing sensitivity. Another noteworthy contribution is the model's performance under data scarcity conditions. The use of Autoencoders and the multimodal framework allowed effective learning from limited attack data, a common issue in ICS datasets due to operational constraints and the rarity of labeled intrusions. This advantage is further complemented by the model's robustness, as indicated by consistent performance across training and validation sets and minimal overfitting, thanks to regularization techniques such as dropout and batch normalization.

This paper lays a forward-looking foundation by identifying critical future directions, including federated learning, digital twin integration, and adversarial robustness. Federated learning offers a privacy-preserving path for collaborative model improvement across distributed ICS environments, while digital twins can facilitate realistic simulation of attack scenarios and improve model generalization. Furthermore, exploring defense mechanisms against adversarial inputs will be essential to securing the very deep learning systems meant to defend critical infrastructure. This research not only demonstrates the technical viability of multimodal deep learning for ICS cybersecurity but also addresses operational, interpretability, and scalability concerns crucial for real-world deployment. It advances the state of the art in intelligent intrusion detection and provides a strong foundation for future innovations in securing industrial cyber-physical systems.

CONCLUSION

The digital transformation of Industrial Control Systems (ICS), particularly SCADA and DCS platforms, has significantly increased their vulnerability to cyber threats due to Industry 4.0 integration. This study proposes a multimodal deep learning framework that combines Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Autoencoders for intelligent anomaly detection tailored to these complex environments. Evaluated on the HAI Security Dataset, the model achieved an overall accuracy of 92%, perfect recall (1.00) for attack detection, a precision of 0.49 for the attack class, and an Area Under the Curve (AUC) of 0.97, indicating excellent discrimination between normal and malicious events. The integration of Explainable AI tools like SHAP and LIME enhances interpretability, enabling actionable insights for operators. The paper also outlines key implementation strategies, including edge optimization, seamless integration, and federated learning, as well as policy recommendations that call for explainability mandates and investments in digital twin infrastructure. Together, these contributions offer a robust, interpretable, and policy-aligned cybersecurity solution for protecting modern industrial systems.

REFERENCES

Abdelaty, M., Doriguzzi-Corin, R., & Siracusa, D. (2020). AADS: A Noise-Robust Anomaly Detection Framework for Industrial Control Systems. In J. Zhou, X. Luo, Q. Shen, & Z. Xu (Eds.), *Information and Communications Security* (Vol. 11999, pp. 53–70). Springer International Publishing. https://doi.org/10.1007/978-3-030-41579-2_4

Abdi, N., Albaseer, A., & Abdallah, M. (2024). The Role of Deep Learning in Advancing Proactive Cybersecurity Measures for Smart Grid Networks: A Survey. *IEEE Internet of Things Journal*, *11*(9), 16398–16421. https://doi.org/10.1109/JIOT.2024.3354045

Abdullahi, M., Alhussian, H., Aziz, N., Abdulkadir, S. J., Alwadain, A., Muazu, A. A., & Bala, A. (2024). Comparison and Investigation of AI-Based Approaches for Cyberattack Detection in Cyber-Physical Systems. *IEEE Access*, *12*,

31988–32004. https://doi.org/10.1109/ACCESS.2024.3370436

Alimi, O. A., Ouahada, K., Abu-Mahfouz, A. M., Rimer, S., & Alimi, K. O. A. (2021). A Review of Research Works on Supervised Learning Algorithms for SCADA Intrusion Detection and Classification. *Sustainability*, *13*(17), 9597. https://doi.org/10.3390/su13179597

Alladi, T., Chamola, V., & Zeadally, S. (2020). Industrial Control Systems: Cyberattack trends and countermeasures. *Computer Communications*, *155*, 1–8. https://doi.org/10.1016/j.comcom.2020.03.007

Anandita Iyer, A., & Umadevi, K. S. (2023). Role of AI and Its Impact on the Development of Cyber Security Applications. In V. Sarveshwaran, J. I.-Z. Chen, & D. Pelusi (Eds.), *Artificial Intelligence and Cyber Security in Industry 4.0* (pp. 23–46). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-2115-7_2

Bakker, C., Vasisht, S., Huang, S., & Vrabie, D. L. (2023). Sensor and Actuator Attacks on Hierarchical Control Systems with Domain-Aware Operator Theory*. *2023 Resilience Week (RWS)*, 1–8. https://doi.org/10.1109/RWS58133.2023.10284668

Balla, A., Habaebi, M. H., Islam, Md. R., & Mubarak, S. (2022). Applications of deep learning algorithms for Supervisory Control and Data Acquisition intrusion detection system. *Cleaner Engineering and Technology*, *9*, 100532. https://doi.org/10.1016/j.clet.2022.100532

Ben Fredj, O., Mihoub, A., Krichen, M., Cheikhrouhou, O., & Derhab, A. (2020). CyberSecurity Attack Prediction: A Deep Learning Approach. *13th International Conference on Security of Information and Networks*, 1–6. https://doi.org/10.1145/3433174.3433614

Berardehi, Z. R., Yin, J., & Taheri, M. (2024). Stabilization of Phasor Measurement Sensor-Based Markovian Jump CPSs Through Soft Actor–Critic. *IEEE Sensors Journal*, *24*(22), 37800–37808. https://doi.org/10.1109/JSEN.2024.3468210

Bindra, S. S., & Aggarwal, A. (2024). Deep Learning-based Enhanced Security in Cyber- Physical Systems: A Multi-Attack Perspective. *2024 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, 347–352. https://doi.org/10.1109/ICCICA60014.2024.10584861

Devi, V. K., Asha, S., Umamaheswari, E., & Bacanin, N. (2023). A Comprehensive Review on Various Artificial Intelligence-Based Techniques and Approaches for Cyber Security. In J. Choudrie, P. N. Mahalle, T. Perumal, & A. Joshi (Eds.), *ICT for Intelligent Systems* (Vol. 361, pp. 303–314). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-3982-4_26

Ganesh, P., Lou, X., Chen, Y., Tan, R., Yau, D. K. Y., Chen, D., & Winslett, M. (2021). Learning-Based Simultaneous Detection and Characterization of Time Delay Attack in Cyber-Physical Systems. *IEEE Transactions on Smart Grid*, *12*(4), 3581–3593. https://doi.org/10.1109/TSG.2021.3058682

Gao, J., Gan, L., Buschendorf, F., Zhang, L., Liu, H., Li, P., Dong, X., & Lu, T. (2021). Omni SCADA Intrusion Detection Using Deep Learning Algorithms. *IEEE Internet of Things Journal*, *8*(2), 951–961. https://doi.org/10.1109/JIOT.2020.3009180

Graph–Based Anomaly Detection Using Fuzzy Clustering. (2020). In Ç. Ateş, S. Özdel, & E. Anarım, *Advances in Intelligent Systems and Computing* (pp. 338–345). Springer International Publishing. https://doi.org/10.1007/978-3-030-23756-1_42

Khan, I. A., Keshk, M., Pi, D., Khan, N., Hussain, Y., & Soliman, H. (2022). Enhancing IIoT networks protection: A robust security model for attack detection in Internet Industrial Control Systems. *Ad Hoc Networks*, *134*, 102930. https://doi.org/10.1016/j.adhoc.2022.102930

Koay, A. M. Y., Ko, R. K. L., Hettema, H., & Radke, K. (2023). Machine learning in industrial control system (ICS) security: Current landscape, opportunities and challenges. *Journal of Intelligent Information Systems*, *60*(2), 377–405. https://doi.org/10.1007/s10844-022-00753-1

Lin, G., Wen, S., Han, Q.-L., Zhang, J., & Xiang, Y. (2020). Software Vulnerability Detection Using Deep Neural Networks: A Survey. *Proceedings of the IEEE*, *108*(10), 1825–1848. https://doi.org/10.1109/JPROC.2020.2993293

Meydani, A., Shahinzadeh, H., Ramezani, A., Nafisi, H., & Gharehpetian, G. B. (2024). A Review and Analysis of Attack and Countermeasure Approaches for Enhancing Smart Grid Cybersecurity. *2024 28th International Electrical Power Distribution Conference (EPDC)*, 1–19. https://doi.org/10.1109/EPDC62178.2024.10571761

Nosova, S., Norkina, A., & Morozov, N. (2024). Strategies for Business Cybersecurity Using AI Technologies. In A. V. Samsonovich & T. Liu (Eds.), *Biologically Inspired Cognitive Architectures 2023* (Vol. 1130, pp. 635–642). Springer

Nature Switzerland. https://doi.org/10.1007/978-3-031-50381-8_67

Oise, G. (2023). A Web Base E-Waste Management and Data Security System. *RADINKA JOURNAL OF SCIENCE AND SYSTEMATIC LITERATURE REVIEW*, *1*(1), 49–55. https://doi.org/10.56778/rjslr.v1i1.113

Oise, G., & Konyeha, S. (2024). E-WASTE MANAGEMENT THROUGH DEEP LEARNING: A SEQUENTIAL NEURAL NETWORK APPROACH. *FUDMA JOURNAL OF SCIENCES*, *8*(3), 17–24. https://doi.org/10.33003/fjs-2024-0804-2579

Oise, G. P., & Akpowehbve, O. J. (2024). Systematic Literature Review on Machine Learning Deep Learning and IOT-based Model for E-Waste Management. *International Transactions on Electrical Engineering and Computer Science*, *3*(3), 154–162. https://doi.org/10.62760/iteecs.3.3.2024.94

Oise, G. P., Nwabuokei, O. C., Akpowehbve, O. J., Eyitemi, B. A., & Unuigbokhai, N. B. (2025). TOWARDS SMARTER CYBER DEFENSE: LEVERAGING DEEP LEARNING FOR THREAT IDENTIFICATION AND PREVENTION. *FUDMA JOURNAL OF SCIENCES*, *9*(3), 122–128. https://doi.org/10.33003/fjs-2025-0903-3264

Oyedotun, S. A., Oise, G. P., Akilo, B. E., Nwabuokei, O. C., Ejenarhome, P. O., Fole, M., & Onwuzo, C. J. (2025). The Role of Internal Audit in Fraud Detection and Prevention: A Multi-Contextual Review and Research Agenda. *Journal of Science Research and Reviews*, *2*(2), 76–85. https://doi.org/10.70882/josrar.2025.v2i2.51

Pan, K., Wang, Z., Dong, J., Palensky, P., & Xu, W. (2025). Real-Time Estimation and Defense of PV Inverter Sensor Attacks With Hardware Implementation. *IEEE Transactions on Industrial Electronics*, *72*(3), 3228–3232. https://doi.org/10.1109/TIE.2024.3436516

Varma, A. J., Taleb, N., Said, R. A., Ghazal, T. M., Ahmad, M., Alzoubi, H. M., & Alshurideh, M. (2023). A Roadmap for SMEs to Adopt an AI Based Cyber Threat Intelligence. In M. Alshurideh, B. H. Al Kurdi, R. Masa'deh, H. M. Alzoubi, & S. Salloum (Eds.), *The Effect of Information Technology on Business and Marketing Intelligence Systems* (Vol. 1056, pp. 1903–1926). Springer International Publishing. https://doi.org/10.1007/978-3-031-12382-5_105

Xiahou, K., Xu, X., Huang, D., Du, W., & Li, M. (2024). Sliding-Mode Perturbation Observer-Based Delay-Independent Active Mitigation for AGC Systems Against False Data Injection and Random Time-Delay Attacks. *IEEE Transactions on Industrial Cyber-Physical Systems*, *2*, 446–458. https://doi.org/10.1109/TICPS.2024.3436188

Xiong, D., Zhang, D., Zhao, X., & Zhao, Y. (2021). Deep Learning for EMG-based Human-Machine Interaction: A Review. *IEEE/CAA Journal of Automatica Sinica*, *8*(3), 512–533. https://doi.org/10.1109/JAS.2021.1003865

Yang, K., Li, Q., Li, T., Wang, H., & Sun, L. (2025). Detecting Time-Delay Attacks in Industrial Control Systems Through State-Aware Inference. *IEEE Internet of Things Journal*, *12*(6), 7195–7208. https://doi.org/10.1109/JIOT.2024.3496896

Zhang, J., Pan, L., Han, Q.-L., Chen, C., Wen, S., & Xiang, Y. (2022). Deep Learning Based Attack Detection for Cyber-Physical System Cybersecurity: A Survey. *IEEE/CAA Journal of Automatica Sinica*, *9*(3), 377–391. https://doi.org/10.1109/JAS.2021.1004261