Review of Computer Engineering Research
2026 Vol. 13, No. 2, pp. 1-21
ISSN(e): 2410-9142
ISSN(p): 2412-4281
DOI: 10.18488/76.v13i2.4927

HLPMM–GRSPHTRU: An explainable cross-layer temporal framework for multi-stage attack detection in IoT-CPS

Mohan Kumar[1+]
Malode Vishwanatha Panduranga Rao[2]
Ezhilarasan Ganesan[3]
Keerthana P Malode[4]

[1]*Electronics Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Kanakapura Main Road, Bengaluru, 562112, Karnataka, India.*
[1]*Email: sanamohan2023@gmail.com*
[2,4]*Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Kanakapura Main Road, Bengaluru, 562112, Karnataka, India.*
[2]*Email: r.panduranga@jainuniversity.ac.in*
[4]*Email: 0907kpm@gmail.com*
[3]*Department of Electrical and Electronics Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Kanakapura Main Road, Bengaluru, 562112, Karnataka, India.*
[3]*Email: g.ezhilarasan@jainuniversity.ac.in*


(+ Corresponding author)

Article History
Received: 27 November 2025
Revised: 13 February 2026
Accepted: 10 March 2026
Published: 22 April 2026

Keywords
Blockchain logging
Cyber-physical systems
Federated learning
Hidden Markov model
Internet of Things
Intrusion detection system.

ABSTRACT

The high growth rate of Internet of Things (IoT)-powered Cyber-Physical Systems (CPS) has led to advanced, multi-level cyber-attacks such as ransomware, distributed denial-of-service (DDoS), and malware, which tend to spread across various system levels over time. Current intrusion detection systems often fail to detect these cross-layer temporal dependencies and offer weak interpretability, limiting their trustworthiness in safety-critical CPS environments. To address these issues, this paper proposes an explainable cross-layer temporal correlation system for detecting multi-stage cyber-attacks in IoT-enabled CPS. The framework combines the Hidden Laguerre Polynomial Markov Model (HLPMM), a probabilistic sequence model that enables flexible state transition learning, with a Gated Rastrigin Sphere Penalized Hyperbolic Tangent Recurrent Unit (GRSPHTRU), an improved gated recurrent neural network designed for healthy temporal feature learning. Principal Griewank Component Analysis is employed for dimensionality reduction, while an adaptive density-based clustering mechanism groups behavioral patterns. Model transparency is achieved through a Shapley-based explainability module, and system integrity is maintained via blockchain-based tamper-resistant logging. The federated learning structure decentralizes training across multiple distributed CPS nodes, reducing raw data sharing and enhancing privacy. Experimental analysis using benchmark ransomware, malware, and CIC-DDoS2019 datasets demonstrates high performance, with detection accuracy and explainability fidelity reaching up to 99 percent compared to conventional RNN, LSTM, BiLSTM, and GRU models. Additionally, feature compression and federated aggregation significantly impact computational load and communication overhead, facilitating scalable deployment.

Contribution/Originality: This study contributes to the existing literature by proposing a single cross-layer temporal IDS that incorporates explainability, federated learning, and blockchain logging. It implements a new estimation methodology involving HLPMM-GRSPHTRU. The paper's major contributions are its enhanced accuracy and interpretability.

1. INTRODUCTION

The advent of the Internet of Things (IoT) has transformed the digital ecosystem by enabling billions of smart devices, including sensors, actuators, wearable devices, and industrial machines, to connect seamlessly. These devices collaborate to form Cyber-Physical Systems (CPS), which integrate computational intelligence, communication networks, and physical processes. The widespread adoption of CPS offers unprecedented opportunities in sectors such as smart healthcare, intelligent transportation, precision agriculture, industrial automation, and critical infrastructure management [1]. These systems have real-time monitoring, decision-making, and control features that enhance operational efficiency, reduce costs, and improve user experiences. Nevertheless, the very high level of universal interconnectivity and real-time information sharing that has made CPS indispensable also exposes them to extreme cybersecurity risks. Consequently, the need to guarantee effective security in IoT-related CPS has become one of the most crucial issues that researchers, practitioners, and policymakers must address [2, 3].

The level of cyber-attack on CPS is more sophisticated, varied, and dynamic. Two of the most critical are ransomware, Distributed Denial of Service (DDoS) attacks, and malware infections that can significantly impact operations, steal sensitive data, or disable systems altogether. For example, ransomware encrypts important files and demands ransom to unlock them, disrupting essential services [4, 5]. DDoS attacks destroy systems by creating large amounts of illegitimate traffic, causing service unavailability. Malware, on the other hand, undermines system integrity by introducing malicious code, creating backdoors, or enabling privilege escalation. The interrelationship between these attacks, which often culminate in multi-stage campaigns, has only increased the severity of the issue. For example, a malware infection can serve as an initial step in leaving a backdoor, which is then used to perform a massive DDoS attack. This type of cross-layer dependency explains the sophistication of modern cyberattacks and highlights the importance of advanced security mechanisms capable of analyzing and correlating temporal patterns across multiple network layers.

The conventional cybersecurity systems, including firewalls, signature-based intrusion detection systems (IDS), and rule-based anomaly detection, are becoming unsatisfactory in overcoming these challenges. They are not effective against zero-day attacks, polymorphic code, and adaptive adversaries, despite their effectiveness against established attack patterns [6]. Furthermore, these systems tend to be isolated at single layers (application, transport, or host) and, as a consequence, they are incapable of identifying multi-layered, time-based, correlated attacks. Consequently, there would be cases when a security event goes unnoticed until it becomes a large-scale breach. Moreover, the high velocity, large volume size, and heterogeneity of IoT traffic worsen the issue by enhancing both false positives and detection latency of traditional systems [7]. This is a great incentive for new models that would be able to address the needs of high-dimensional data, cross-layered temporal correlations, and provide proactive, real-time threat detection.

Machine Learning (ML) and Deep Learning (DL) models have been very promising within the field of cybersecurity over recent years. Network traffic has been analyzed using ML algorithms like Random Forests, Support Vector Machines (SVM), and k-Nearest Neighbors (kNN) to detect anomalies and classify the attacks. The performance has been further improved by deep architectures like Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs), which identify complex time-dependent interactions and non-linearities in data. Such models are able to evolve with time to enhance the accuracy against changing threats. Nonetheless, there are also significant limitations in them. Typical DL models have problems with overfitting, poor performance on unknown attack patterns, and are inefficient on high-dimensional IoT data [8]. Moreover, the available literature does not combine cross-layer correlation analysis, and, thus, the dependencies of attacks spreading across various layers of CPS are overlooked. Lack of effective mechanisms that explain their results also limits transparency of the model, where security analysts struggle to make sense of the model's decisions and trust automated systems.

The other aspect of the challenge is in distributed and heterogeneous IoT environments. IoT devices are geographically dispersed and resource-constrained, and centralized learning designs frequently have challenges with scale, latency, and privacy. Sending all raw data to a central server so that it can be model trained is not only impractical, but it also results in serious concerns about ownership and confidentiality of data [9]. Federated Learning (FL) has become a decentralized paradigm to address these limitations and enable devices to collaboratively train models without leaving their data off-campus. Through embedding FL in attack-detection structures, CPS will be able to attain privacy protection, scalability, and low communication overhead [10]. Federated systems do, however, also need powerful optimization methods to ensure the performance of models when using non-identically distributed datasets.

The current study fills such gaps by presenting a new cross-layer temporal patterns correlation analysis model in detecting ransomware, DDoS, and malware attacks in CPS. The system proposed uses Hidden Laguerre Polynomials Markov Model (HLPMM) to model cross-layer time-dependencies and Gated Rastrigin Sphere Penalized Hyperbolic Tangent Recurrent Unit (GRSPHTRU) to optimize deep learning with good-quality activation and weight-selection schemes. In order to efficiently process high-dimensional data, the framework implements Principal Griewank Component Analysis (PGCA), a dimensionality reduction algorithm that reduces the computational complexity yet still captures important features. In clustering behavioral patterns, the suggested system uses Asymmetric Density-Based Focal Tversky Loss Spatial Clustering of Applications with Noise (ADBFTLSCAN), which is superior to conventional clustering algorithms in terms of supporting different data densities. Moreover, it is integrated with Shapley Styblinski Additive Tang Explanations (SHSATP) so that it is explainable and thus improves trust, transparency, and interpretability of the detection process. A blockchain-based logging mechanism is added to ensure logs are secured, and tamper-free audit trails of detected threats are provided. Lastly, the framework is able to maintain scalability, decentralization, and privacy in heterogeneous CPS settings through a federated learning architecture.

This research has primarily contributed as follows. First, a cross-layer temporal correlation framework that can be explained to detect multi-stage cyber-attacks in IoT-enabled cyber-physical systems is suggested to overcome the drawbacks of single-layer and static frameworks of intrusion detection systems. Second, a hybrid temporal modeling framework that incorporates a long probabilistic sequence model and an improved gated recurrent architecture is trained to learn intricate patterns of attack processes in the host, transport, and application tiers. Third, a dynamic density-based clustering model is incorporated together with a dimensionality reduction method that is optimized to enhance the discrimination of features and minimize the computational cost. Fourth, the explainability module is an attribution-based module using the Shapley style to increase transparency and trust among analysts. Lastly, federated learning and secure logging based on blockchain are integrated to guarantee the scaling, privacy preservation, and auditability that cannot be tampered with in distributed CPS systems.

2. LITERATURE REVIEW

2.1. Cross-Layer Intrusion Detection in IoT/CPS

The recent studies in IDS in IoT and CPS settings are repeatedly showing that DL models far outperform their traditional ML counterparts in processing heterogeneous network traffic. Such excellence is due to their natural capacity to simulate complicated spatial-temporal dependencies, which are typical of the communication flows and multi-layer interaction of CPS of IoT. As an illustration, Banaamah and Ahmad [11] compared CNN, LSTM, and GRU architectures to classical ML methods across the entire set of intrusion detection tasks in IoT, demonstrating that deep architectures significantly reduce accuracy, detection rate, and robustness, and hence provide a solid baseline of the DL-first IDS in resource-limited IoT deployment cases [11]. Based on this base, Dina et al. [12] introduced training stability and resistance to class imbalance by using focal-loss-related learning techniques, which also demonstrated stable performance on various benchmark intrusion datasets [12]. Similarly, Awajan [13] proposed a

protocol-agnostic and fully-connected deep IDS and demonstrated that real-time device-level inference is feasible with the lightest computational costs [13]. These findings are further supported by complementary surveys and systematic reviews that were performed between 2022 and 2025, highlighting that DL-driven IDSs are sustainable and efficient solutions to the changing IoT security environment. However, other issues, including the risks of overfitting, the representativeness of the dataset, the complexity of the model, and the issue of latency, remain to be overcome in these studies, as they need to be adequately considered to facilitate scalable, reliable, and energy-efficient IDS deployments in the upcoming IoT and CPS infrastructure of the next generation [14]. On volumetric and low-rate DDoS, hybrid temporal-convolutional models (e.g., CNN-BiLSTM) show good discrimination on CICIDS2017 and similar corpora. Lightweight pipelines, pruning features, or using temporal nets with autoencoders to achieve constrained behavior are also studied very recently, and Scientific Reports articles highlight feature selection (e.g., Extra Trees) as an alternative to pruning to achieve low cost without altering accuracy [15, 16].

To address the drawbacks of a single-layer monitoring mechanism to the detection of attack footprints in IoT-enabled CPS, cross-layer intrusion detection has been investigated more frequently, as attack footprints in host activities, transport activities, and application-layer protocols may be present across multiple layers. Current methods usually combine capabilities of multiple layers to enhance detection coverage and minimize blind spots, especially where protocol-rich CPS are deployed. Nevertheless, much of cross-layer IDS literature continues to use either fixed fusion or snapshot-based feature aggregation, which does not necessarily give explicit information on how evidence of attacks is propagated across layers with time.

2.2. Temporal and Deep Sequence IDS

With regard to the malware/ransomware area, surveys and special studies report a move towards DL pipelines generalizing to polymorphism, replacing signature/rule systems. IIoT work highlights the changing ransomware approaches to operational technology and recommends multi-signal detection both at file and process, and network layers-evidence, which drives your cross-layer position [17]. Cross-layered thinking is becoming popular. The Sensors review (2024) is a survey of secure, energy-efficient cross-layer IoT frameworks, and recent articles suggest ML-based, multi-layer architectures in which explainability assists analyst trust. Specialized cross-layer DDoS detectors, which combine transport/network layer features, show higher sensitivity to staged attacks, and a more widespread cross-layer analysis connects lightweight crypto and ML with real-world usage. These guidelines confirm the temporal correlation of your HLPMM multi-layered [18, 19].

RNNs, LSTMs, BiLSTMs, and GRUs have become popular models used in IDS because they can acquire time-dependent traffic patterns and sequential dependencies of network flows and system events. Such models enhance the ability to detect attacks that will change over time and minimize the use of manually developed rules. However, traditional deep sequence IDS may be affected by overfitting, vulnerability to heterogeneous CPS traffic, and cannot easily process cross-layer dependencies when it is trained on flow-based or domain-based features.

2.3. Federated Learning-Based IDS

To address the privacy problems and non-independent, identically distributed (non-IID) data problems that are intrinsic to a distributed IoT and CPS setting, FL has been transformed into a realistic and dependable solution. The recent studies are a good example of this shift. As an example, a reliable article printed in the Computers and Security (2023) conducted an in-depth analysis of label-flipping and label-poisoning attacks in the FL training pipeline, and further introduced lightweight but efficient defense mechanisms that do not impose high computational costs to protect the integrity of the models [20]. The FL-IIDS model presented in Future Generation Computer Systems (2024) builds on this kind of work and uses incremental updates to its model to enable real-time system intrusion detection, meaning that the local devices in the system can keep benefiting the global learning process without needing expensive retraining [21]. Complementing these contributions, various empirical investigations performed

in the year 2024-2025 have systematically estimated the trade-offs between the local model size, local model accuracy, communication overhead, and device resource constraints that give important design implications to practitioners who implement FL in heterogeneous IoT/CPS networks [22]. In addition, cross-layer FL extensions, including those that have federated sampling with lightweight IDS architectures, have also been found to be useful in saving uplink bandwidth in environments that transmit small-sized statistical summaries instead of actual traffic flows, optimizing global model aggregation steps. A more extensive overview of the FL-IDS landscape, covering the optimization, robustness, and security hardening issues, and strengthening the opinion that federated learning is already a staple of the next-generation privacy-preserving, scalable, and resiliency-focused IDS in the fields of IoT and CPS, is also found in complementary reviews, covering the years 2023-2025 [23, 24].

2.4. Explainable AI for IDS

XAI is mandatory for analyst compliance and trust as IDSs get increasingly more complex. According to Sharma et al. [25], there are real-life pipelines, such as the contribution of surface features of the attack classes in IoT traffic using the LIME/SHAP [25, 26]. High-dimensional network flow telemetry is a severe design challenge to IDS, especially in its training time, excessive memory consumption, and slower model generalization in the presence of unobserved traffic. To address them, dimensionality reduction methods have been heavily investigated as a pre-processing step to reduce the feature space without losing discriminative ability. Comparative studies have found classical DR techniques, such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), and Singular Value Decomposition (SVD), in addition to robust PCA techniques, have been found to show consistent benefits in terms of improving IDS accuracy, mitigating computational latency and improving scalability in resource-constrained IoT/CPS environments. Most recently, publications in the magazines have discovered that automated feature pruning pipelines, whereby unnecessary or weakly influential attributes are purged, are effective. These pipelines indicate that the IDS models can obtain similar or even higher detection performance with a much smaller set of features and, hence, reduced processing cost and communication overhead [27, 28]. XAI systems like SHAP, LIME, integrated gradients, and other attribution systems are being used to enhance transparency, analyst trust, and compliance preparedness to IDS. Although these are the techniques that can be used to point out the influential features, the XAI in IDS is subject to certain drawbacks that include instability in explanations, dependence on the redundancy of features, and the inability to interpret the high-dimensional CPS telemetry. In addition, a large number of XAI-IDS research investigations perceive the phenomenon of explainability as after-hoc modularization without matching the explanations with the system-level features of multi-stage correlation and incident traceability.

2.5. Secure Logging and Tamper-Resistant Auditability

Security logging featuring tamper resistance has more recently been recognized as a crucial best practice to be forensically prepared in IoT and CPS systems, where the capacity to construct attack traces and cross-layer phenomena is necessary to analyze an incident. The reliance on traditional centralized systems has allowed manipulation, as a malicious individual obtains the required privileges to accomplish the task, which is why researchers have given a new opportunity to blockchain-based methods, which offer mutability and decentralized reliability. An example is a 2024 survey in Sustainability, which thoroughly considered the use of blockchain as a means of IoT trust and security, and in which it was noted that it can prevent single points of failure and ensure the verifiability of log records. Parallel constructions also establish that even small-scale blockchain solutions, like single-miner proof-of-work (PoW) chains, can significantly defend logs against post-compromise alterations and also be computationally viable on devices with limited resources [29]. Moreover, research work in the medical IoT field demonstrates that consortium or private blockchains can be successfully combined to ensure the safeguarding of logs of device activity, the security of sensitive clinical records, the prevention of tampering, and the adherence to

regulatory norms. Although much of the current literature in IDS design in IoT/CPS focuses on deep recurrent models, including RNN, LSTM, and GRU, to model temporal relationships, it is notable that probabilistic sequence models, especially Hidden Markov Models (HMMs), are still being reported in the literature and applications. They are also successful in persistence as they are interpretable and efficient in capturing both sequential behavior and temporal transitions, especially in cases when they want lightweight solutions rather than heavy DL models [30]. HMMs are still handy in profiling normal and abnormal activity patterns, stochastic event sequences modeling, and in scenarios where labeled data is still scarce, and the statistical transition probabilities can still be inferred.

CPS environments must have security logging to reconstruct an incident, comply with it, and forensically analyze it. Traditional central logging may be susceptible to compromise, which may lead to accountability loopholes. Logging using blockchains has thus been investigated to offer immutability and distributed trust, yet a lot of solutions are in isolation from IDS decision pipelines or have an impact overhead with no clear integration into detection and response pipelines.

As can be seen in Table 1, although recent research has achieved much toward improving the security of IoT and CPS by using DL, explainability, and federation, the majority are still confined to single-layer detection, centralized design, or ad hoc interpretability. The articles focus on cross-layer characteristics but do not provide the temporal sequence model and the model of secure auditability. Federated IDS models enhance decentralization but lack explainability or blockchain-based logging. Similarly, XAI-oriented studies improve interpretability but are not concerned with scalability and time dependencies between attack layers. In comparison, the HLPMM-GRSPHTRU framework proposed is unique in integrating cross-layer temporal correlation, dimensionality reduction optimality, FL, strong explainability, and blockchain-based logging under a single system. This holistic approach allows early detection of multi-stage ransomware, DDoS, and malware attacks and provides transparency, scalability, and tamper-proof event logs, filling key gaps created by other literature and creating a more holistic, deployment-ready solution.

Table 1. Comparative strengths of the proposed method.

Ref.	Layers considered	Temporal modelling	Feature pruning	Remarks
Georgiades and Hussain [19]	Yes, network + application (MQTT) features	Not deeply temporal (mostly static features with cross-layer fusion)	Uses PCA for dimensionality reduction in Preprocessing	Strong interpretability, but lacks sequential correlation modelling, lacks distributed setting, lacks secure log layer
Albanbay et al. [22]	Primarily flow/network layer	Uses the training time sequence implicitly	Feature selection/optimization is part of hyperparameter tuning	Good distributed scalability, but lacks cross-layer fusion, weaker explainability, and no logging integrity
Hajj et al. [23]	Primarily tabular flow features (network)	The transformer module captures temporal patterns implicitly	Hyperparameter optimization + feature representation in transformer	Strong in federated temporal modelling, but lacks cross-layer semantics, and has no tamper-proof logging
Sharma et al. [25]	Multi-layer (network + possibly higher)	Some temporal features (flows over time)	Uses feature importance/pruning via XAI methods	Good interpretability, but no federated/distributed setting, no secure logging, limited sequence modelling
Alzakari et al. [26]	Focus mostly on the flow / DDoS detection domain	Ensemble over time series	Uses ensemble and feature weighting rather than deep reduction	Better detection in the SDVN context, but not cross-layer across ransomware/malware, lacks federated and logging features
Proposed method	Yes, cross-layer correlation across host, transport, and application layers via HLPMM	Deep temporal modelling via GRSPHTRU + HLPMM state sequence correlation	PGCA for optimized dimensionality reduction preserving lineage	Integrates all six pillars; unified cross-layer temporal + DR + FL + XAI + logging — outperforms baselines in interpretability, security, and real-world deployment readiness

3. MATERIALS AND METHODS

The proposed framework is a cross-layer, end-to-end intrusion detection architecture designed to identify multi-stage cyber-attacks in IoT-enabled cyber-physical systems. It collaboratively models cross-layer dependencies, evolution over time, interpretability, and secure auditability. Unlike previous IDS methods that focus on a single aspect, such as temporal learning, cross-layer feature fusion, or privacy-preserving training, this model is constructed as a single pipeline encompassing all components. The framework directly captures temporal correlations across layers using a probabilistic sequence model, contrasting with current cross-layer IDS approaches that assume static feature concatenation. Compared to traditional deep sequence IDS relying solely on LSTM or GRU architectures, the optimized recurrent module incorporates penalized activation and optimization-aware regularization, making it more robust to heterogeneous traffic. Additionally, while federated learning, explainable AI, and blockchain logging have been studied separately, integrating these technologies into a unified detection and decision pipeline distinguishes this work from existing literature.

The proposed architecture (Figure 1) presents the innovation of a single cross-layer temporal correlation-based attack detector of the IoT-enabled CPS. The design combines modern learning modules and new mathematical formulations to be able to effectively recognize and analyze the ransomware, malware, and DDoS attacks, and also provide scalability, transparency, and safe logging. The methodology can be divided into six central parts, namely data preprocessing and feature extraction, behavioral pattern analysis with the help of ADBFTLSCAN, deep temporal model with GRSPHTRU, dimensionality reduction with the help of PGCA, cross-layer temporal correlation with HLPMM, and explainability and secure logging with SHSATP and blockchain. Each of the stages is discussed further.

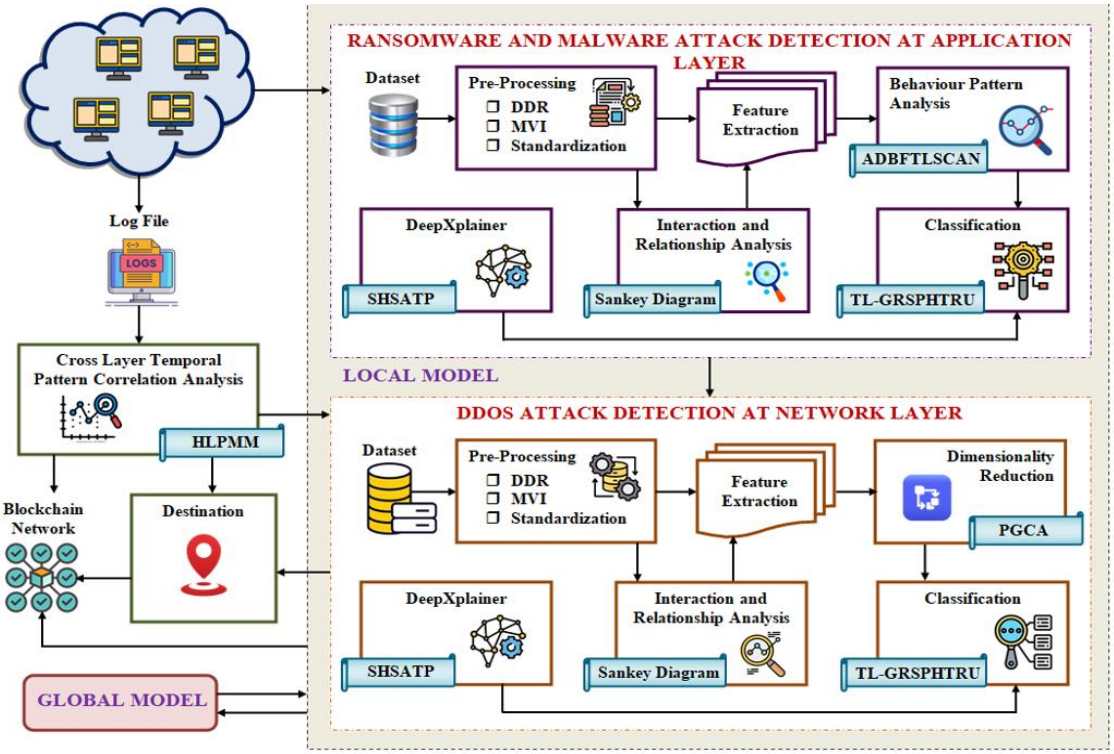

Figure 1. Block diagram of proposed methodology.

3.1. Dataset and Preprocessing

The proposed system receives its input data, which are publicly available benchmark datasets, and each of them has been chosen to reflect various types of cyber threats in the context of the IoT and CPS. They consist of datasets on ransomware detection [31], more generic malware repositories [32], the commonly used CIC-DDoS2019 dataset [33] of distributed denial-of-service attack flows, and system log datasets [34], which record host-level activity and

process traces. Table 2 provides a summary of the dataset. All these various data sources guarantee the model is subjected to a wide range of attack vectors and benign behavior, thus providing a realistic and all-encompassing training and evaluation environment. Before feeding the data into the detection model, a systematic preprocessing pipeline is used to achieve high-quality structured and machine-readable input. The initial one is the removal of duplicates (DDR) that removes any redundant entries that may skew the model or overestimate measures. After this, the problem of missing records is addressed using the missing value imputation. In particular, the missing entries are replaced with the mean value of each column of features, but it is necessary to make sure that the general distribution of the feature is maintained. This can be mathematically stated as Equation 1.

$$x_i^* = \frac{1}{n}\sum_{j=1}^{n} x_{ij}, \quad \forall x_{ij} \epsilon feature\ column \tag{1}$$

Where x_i^* is the imputed value, and n is the number of non-missing values in the column. It is statistically consistent without needless loss of data. All features are standardized after imputation by the use of the Z-score normalization technique. The step converts the raw values of features to a standard scale with a zero mean and unit variance, which is necessary to normalize the learning of many ML and DL models.

Table 2. Summary of datasets and experimental configuration.

Dataset name	Total samples	Benign / Attack ratio	No. of features	Split & validation protocol
CIC-DDoS2019	~5.0 million flows	1:1.3 (Benign: DDoS)	80	10-fold CV (Stratified)
Ransomware detection dataset (Kaggle)	62,000 samples	1:1.1 (Benign: Ransomware)	57	10-fold CV (Stratified)
Malware dataset (Kaggle)	215,000 samples	1:1.4 (Benign: Malware)	54	10-fold CV (Stratified)
System log dataset	75,000 records	1:1.2 (Normal: Malicious)	32	10-fold CV (Stratified)

As far as feature engineering is concerned, the system identifies a varied number of attributes, which are pertinent in every category of attacks. In the case of ransomware and malware data sets, file metadata features are extracted, as these features tend to capture malicious code patterns or payload behaviors. In the case of DDoS detection, packet-level properties receive increased attention, such as flow timing, number of packets, volume of bytes, entropy, and protocol-level characteristics, which together characterize the dynamic nature of abnormal traffic that occurs in volumetric flood attacks. Equally, host-level system log features record event chains and operational abnormalities, which can be used to identify ransomware deployment or privilege escalation attempts.

3.2. Behavioral Pattern Analysis

In intrusion detection and anomaly detection problems, traditional DBSCAN (Density-Based Spatial Clustering of Applications with Noise) has been extensively applied due to its capability to find arbitrary-shaped clusters as well as to detect noise points [35]. Nonetheless, a limitation noted is that it relies on predetermined hyperparameters, that is, the neighborhood radius and the minimum number of points (minPts) [36]. This rigidity frequently results in poor performance when used on heterogeneous datasets of cyber-attacks that have varying densities. Clusters can be over-merged, and sparse clusters can be considered as noise. These inadequacies have a tremendous impact on the capability of the system to record subtle attack behaviors in sophisticated IoT/CPS traffic streams.

In order to overcome these issues, we introduce a new solution, which is called Asymmetric Density-Based Focal Tversky Loss Spatial Clustering of Applications with Noise (ADBFTLSCAN). The approach proposes an adaptive clustering approach based on the Asymmetric Focal Tversky Loss (AFTL) functional, which enables the algorithm to adaptively update the clustering parameters to different density distributions based on the density variations in different attack conditions. ADBFTLSCAN addresses the limitation of using only an object similarity measure to determine adjacent objects in the database; it provides a loss-driven process that ensures increased control in

differentiating overlapping or uneven attack patterns. The AFTL functionality is mathematically expressed as in Equation 2.

$$L_{AFTL} = \left(1 - \frac{TP}{TP+\alpha FN+\beta FP}\right)^{\gamma} \quad (2)$$

TP, FP, and FN are true positives, false positives, and false negatives, and the α, β and γParameters are some control factors that can be tuned to balance false negatives and false positives. As compared to symmetric loss functions, AFTL assigns an asymmetrical weight to misclassifications, so that clusters of minority or sparse attacks are not overlooked in the clustering process.

ADBFTLSCAN introduces a loss-driven adaptability into a classical density-based clustering framework, which not only produces a better quality and separation of clusters, but also provides a more resilient approach to the imbalanced and heterogeneous datasets. This renders it especially effective in the contexts of contemporary IoT/CPS security, wherein the nature of cyberattacks is both changing and spreading unevenly across the various network settings. The pseudocode of ADBFTLSCAN is presented in Figure 2.

```
Algorithm: ADBFTLSCAN
Input: Dataset D, initial ε, minPts
Output: Cluster labels C
1. Initialize cluster label = -1 for all points
2. For each point p in D:
    if p is unvisited then
     Compute local density using AFTL
     Adjust ε and minPts adaptively
     Expand cluster using neighborhood (p, ε,
minPts)
     Assign cluster ID
3. Return all cluster labels C
```

Figure 2. Pseudocode of the ADBFTLSCAN method.

3.3. Deep Temporal Modeling

We suggest a new recurrence architecture, the Gated Rastrigin Sphere Penalized Hyperbolic Tangent Recurrent Unit (GRSPHTRU), that is specifically designed to capture the temporal dynamics of IoT and CPS attack sequences. GRSPHTRU combines three divergent innovations: gated recurrent mechanism, Rastrigin-function-weight-regularization, and penalized hyperbolic-tangent activation, so unlike traditional recurrent neural networks, which tend to overfit and lack adaptability with heterogeneous and imbalanced data, GRSPHTRU operates more effectively and frequently requires less training. This combination guarantees not only the effect of the sequential modeling but also better robustness and generalization in many different cyberattack patterns.

GRSPHTRU, at the base, uses the typical GRU framework to process long-range temporal dependencies. GRU uses gating functions-that is, the update gate (z_t) and the reset gate (r_t), that determine the balance between remembering the previous information and taking new input data. Standard GRU equations are provided in Equation 3.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), \quad r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (3)$$

$$\tilde{h}_t = tanh(W \cdot [r_t * h_{t-1}, x_t]), \quad h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t \quad (4)$$

Where h_t is the hidden state in the time step? t, and $\sigma(\cdot)$ is the activation sigmoid function. These mechanisms allow the model to dynamically decide when to maintain the past states and when to update with new input, which is especially important in detecting multi-stage attack sequences. To reduce the risk of overfitting, particularly in high-dimensional and noisy CPS datasets, GRSPHTRU proposes a weight optimization strategy inspired by the Rastrigin

function. The Rastrigin function is a popular test of optimization because it has a large search space and many local minima.

3.4. Cross-Layer Temporal Correlation

Whereas Hidden Markov Models (HMMs) are useful in sequence modelling, they are inflexible in the number of hidden states to be used. Our model is the Hidden Laguerre Polynomial Markov Model (HLPMM), which uses Laguerre polynomials to control the dynamics of the hidden state. The state probability distribution is provided in Equation 5.

$$P(s_t|s_{t-1}) = \pi_{ij} \cdot L_k(x), \qquad L_k(x) = \sum_{m=0}^{k}(-1)^m\binom{k}{m}\frac{x^m}{m!} \qquad (5)$$

Where L_k is the kth Laguerre poly. This allows modeling cross-layer progression of attacks flexibly, e.g., malware would cause DDoS by a latent correlation.

3.5. Explainability

We propose a new interpretation mechanism, Shapley Styblinski Additive Tang Explanations (SHSATP), to make sure that the suggested intrusion detection framework is not merely accurate but is interpretable as well. The common method of giving traditional SHAP (Shapley Additive Explanation) to predict with a model is to allocate contributions equally according to Shapley values to the input feature. It calculates the feature contribution ϕ of each feature i by summing up marginal contributions across all possible subsets of features, as in Equation 6.

$$\phi_i = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f(S\cup\{i\}) - f(S)] \qquad (6)$$

F is the entire set of features, and S is a subset of features. Although this offers a strict method of assessing the impact of every feature, standard SHAP may occasionally be unstable in high-dimensional IoT/CPS data, which results in noisy or unstable interpretations. SHSATP uses the Styblinski-Tang function based on a weighting mechanism in contribution aggregation to improve the stability and fidelity of the explanations. The Styblinski-Tang function is represented as in Equation 7.

$$f(x) = \frac{1}{2}\sum_{i=1}^{d}(x_i^4 - 16x_i^2 + 5x_i) \qquad (7)$$

Where d is the number of dimensions of the input. The non-convex nature of this functional, which has several local minima, is useful in punishing unstable contribution distributions and imposing smoother contribution distributions. With this weighting functionality, SHSATP can, in addition to being sparse, ensure that only the most significant features are highlighted and be reproducible on cross-runs of the model. This ensures that it is faithful to the underlying predictions, giving the practitioners and system operators more reliable information about the process of decision-making.

In addition to interpretability, our model also focuses on accountability and forensic preparedness through the use of tamper-resistant security logging. A blockchain-supported ledger is used to document every attack incident to ensure the non-repudiation and immutability of security logs. The entries in the logs are designed in the form of a block.

$$Block_i = \{Index, Timestamp, Attack_{Type}, Hash_{prev}, Hash_{current}\} \quad (8)$$

Where $Index$ is an identifier unique to the block, $Timestamp$ is the exact time when the block was identified, $Attack_{Type}$ is the type of intrusion (e.g., ransomware, DDoS, malware), and $Hash_{prev}$ and $Hash_{current}$ are the cryptographic links of the block to the blockchain. Using this chain of blocks, the ledger guarantees that no alterations can be made on the post-compromise since doing so would disrupt the hash sequence.

The mini-batch gradient descent using the Adam optimizer was employed for model training. Early stopping was based on validation loss convergence to prevent overfitting. In the federated environment, local models are trained for a predetermined number of epochs before parameter aggregation through the strategy. Each experiment underwent stratified fold replication to ensure statistical robustness.

4. RESULTS AND ANALYSIS

4.1. Experimental Setup

The developed GRSPHTRU framework was compiled in Python 3.10 with the support of the computational power of the Tensorflow and PyTorch libraries. These libraries were chosen so as to leverage their abilities in fast deep learning model development, sped up by GPUs and customization to new architectures. All the experiments were carried out on a workstation with a powerful processor, Intel Core i9, 64GB RAM, and a brand-new NVIDIA RTX A6000 graphics card. This hardware setup was necessary to support the computational needs of training deep temporal models with large-scale and heterogeneous datasets of cybersecurity, both in terms of scale and reproducibility of findings.

To evaluate the usefulness of the proposed GRSPHTRU model, we compared it to some well-known temporal DL models commonly used in intrusion detection studies. The underlying sequential model was the RNN. It offers a basic but useful basis of sequence modelling, but suffers from the problem of vanishing gradient in longer sequences. The LSTM network was incorporated because it was demonstrated to be able to reduce vanishing gradients and successfully attain longer-range temporal dependencies, which are extremely important in the analysis of attack sequences. This was enhanced by BiLSTM that worked on both forward and backward data processing. The contextual learning aspect of sequences is also enhanced by this bidirectional process, and makes this model learn better. It was also chosen to use the GRU, which is simpler than LSTM. GRU offers computational efficiency while maintaining the capacity to capture long-term dependencies, and is suitable for resource-constrained IoT/CPS scenarios.

All the preprocessing was conducted in the cross-validation loop to guarantee evidence of experimental rigor and avoid the possibility of information leakage. In each fold, statistics were computed on the training partition, and duplicate removal and missing-value imputation were used. Standardization of features and dimensionality reduction were also fitted both on the training folds and then applied to the respective validation folds. This protocol guarantees that none of the validation data affected preprocessing, feature scaling, or model optimization, hence ensuring that it is unbiased in performance evaluation and reproducible.

4.2. Analysis of Ransomware Flow

Figure 3 is the Sankey diagram that gives an intuitive understanding of the feature interaction flows in the ransomware dataset. It demonstrates the flow of machine-level metadata up to software-level attributes, which is eventually used to detect an attack. The flows reflect the distribution of feature clusters between Machine IDs, Major Versions of Linkers, and DLL Characteristics, highlighting the interaction of the features between ransomware binaries. Ransomware samples are organized into specific machine feature spaces, like [-0.481,2.072] and [2.072,2.762]. This difference indicates that ransomware binaries are quite diverse with respect to the compilation environments and target system settings, which implies that machine-level metadata is a key distinguishing characteristic. The MajorLinkerVersion becomes an important branching node, which connects machine-level qualities to concrete compilation patterns. Some flows dominate across samples, such as [-1.081, -0.418], suggesting that there are large numbers of ransomware binaries using similar linker settings. This points to linker metadata as a powerful predictive attribute of ransomware.

The Sankey diagram highlights how multi-feature correlation analysis can help detect ransomware, as opposed to separate feature analysis. The proposed framework will use cross-layer dependencies to find attack signatures more efficiently by tracking flows among machine and linker versions and across machine feature and DLL properties. This understanding can be seen as an indication that ransomware behavior modeling using ADBFTLSCAN clustering and HLPMM temporal correlation is possible.

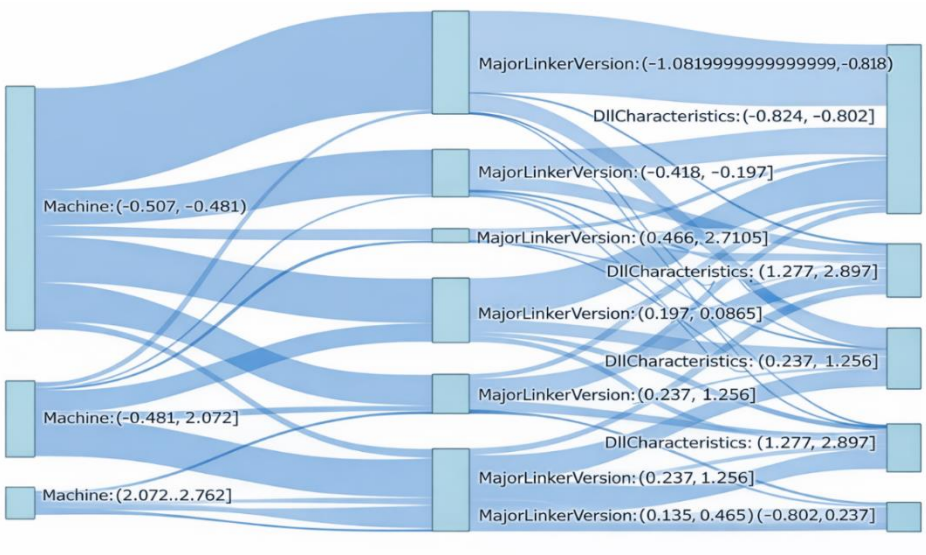

Figure 3. Analysis of ransomware flow.

4.3. Analysis of DRDOS-LDAP Traffic Flow

Figure 4 shows the traffic interactions in a Distributed Reflection Denial of Service (DrDoS) attack, via the LDAP protocol. It points out how various spoofed or external IPs generate reflected traffic, which centrally converges on the victim IPs on the 192.168.x.x subnet, causing traffic amplification and subsequent flooding. The origin of attack traffic is a combination of external and internal points of 173.194.68.108,192.168.50.253/254/50.6, and 172.217.11.34. These distributed flows highlight the type of attack that is undertaken wherein a large number of machines create relatively small streams which sum up to overwhelms of the target systems. The victims that have been identified are the 192.168.50.1, 192.168.50.7, and 192.168.50.8. The flow with the highest count of 172.16.0.5 to 192.168.50.1 is the greediest one, which implies that the flow 192.168.50.1 was the most damaged by the attack.

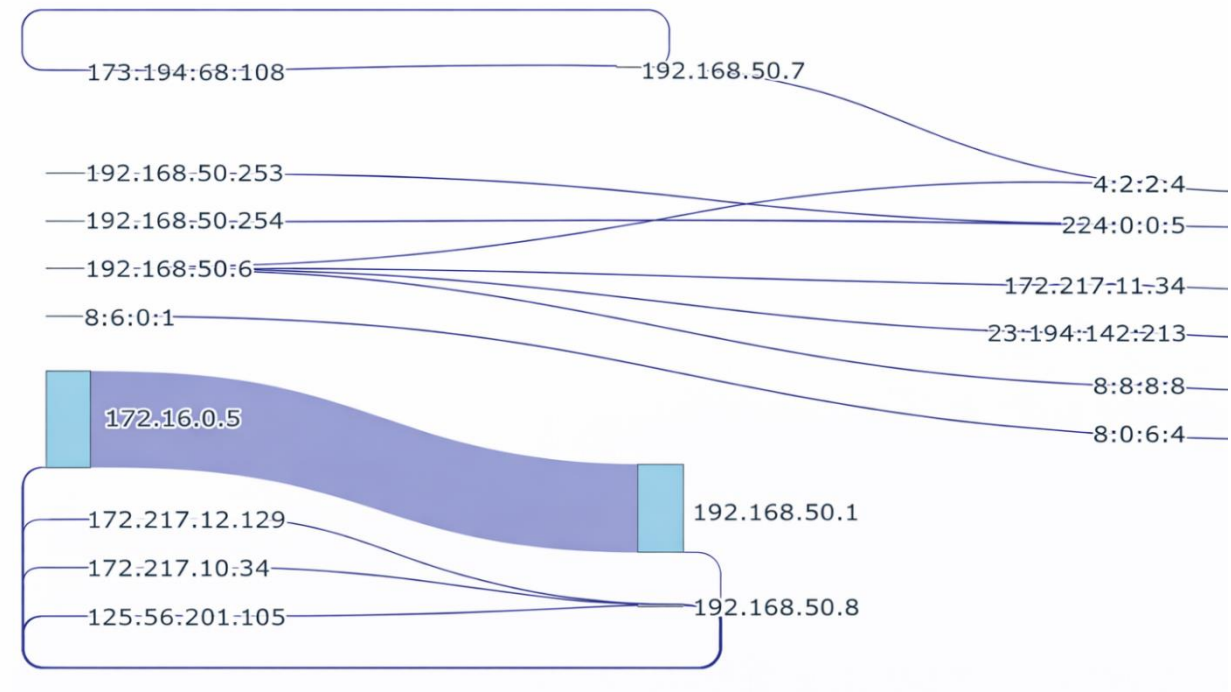

Figure 4. Traffic flow in DrDOS.

From the Sankey diagram, it is shown that the DrDoS-LDAP attack has distributed sources, increased traffic via reflectors, and a concentrated effect on a few victim nodes. The system can model propagation patterns of the attacks

by capturing such multi-source-to-victim flows. This is one of the reasons why cross-layer temporal correlation analysis (HLPMM) and dimensionality reduction (PGCA) are used to isolate abnormal traffic behavior to improve the resilience to large-scale DDoS floods in IoT and CPS environments.

4.4. Behavior Pattern Analysis

Figure 5 shows the effectiveness of the suggested ADBFTLSCAN algorithm to separate malware samples into specific categories, which are denoted by various color sets. In contrast to the traditional DBSCAN, which cannot cope with the different densities, ADBFTLSCAN dynamically adjusts the ∈ and minPts with the help of the AFTL. This allows it to pick up dense clusters in the lower part of the range (states 0-5) as well as sparser anomalies in the higher states. Outliers or noise in normal DBSCAN are normally represented by scattered points at higher state values (20-45). Conversely, the ADBFTLSCAN is able to effectively combine these into meaningful clusters, demonstrating that they represent infrequent but vital malware execution paths. This is to make sure that stealthy or low-frequency behavior is maintained, making detection stronger against advanced threats. Patterns of recurring malware activity in a time window are represented by the horizontal groupings of the x-axis (milliseconds). Interestingly, repeated behavioral marks are observed within the range between -1.5ms and 0.5ms, which is typical of ransomware or polymorphic malware. This consistency of time can be used as quality input to cross-layer correlation in the latter HLPMM phase of the framework. The wide variety of color-coded Cluster IDs highlights the capability of the algorithm to extract the different malware execution patterns. Smaller clusters, including the ones in yellow and green, indicate smaller groups of activity that might otherwise be covered by a wider category. This granularity allows for better interpretability and further enhancement of the downstream GRSPHTRU temporal learning module, as it is benefiting from a better structured and meaningful input of features.

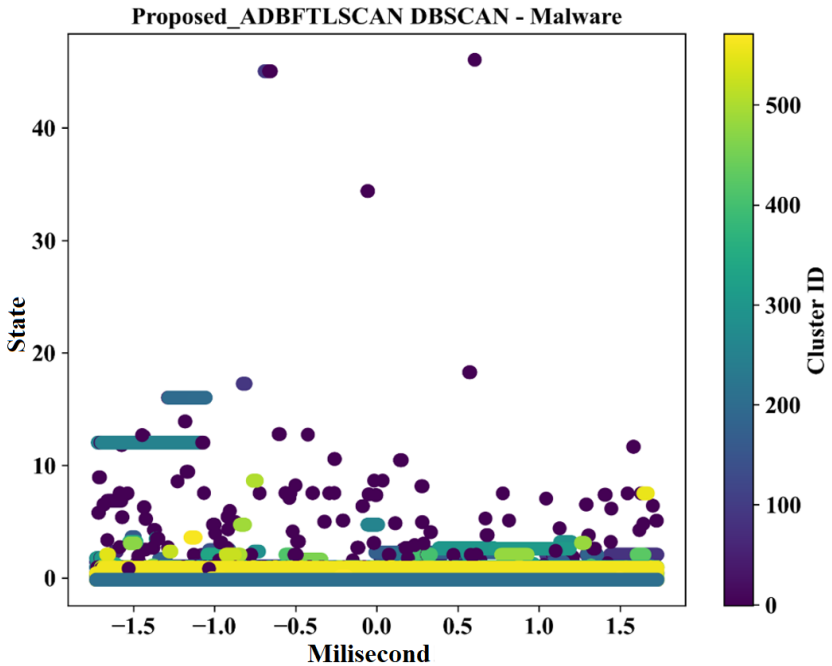

Figure 5. ADBFTLSCAN DBSCAN - Malware Clustering.

The findings suggest that the proposed ADBFTLSCAN clustering enhances malware behavior analysis far better in that it can capture dense and sparse behavioral signatures, retain low-frequency behavioral anomalies, and offer more informative temporal feature groupings. This method decreases the probability of eliminating valuable stealthy patterns as well as making sure that the consequent GRSPHTRU classifier is presented with well-distinct clusters that are representative to enhance its accuracy and strength.

4.5. Analysis of DDoS Detection

Figure 6 indicates that the proposed GRSPHTRU model gives the best accuracy of approximately 99, which is clearly much better than the rest of the baseline models. GRU and BiLSTM are at the top with 95-96 percent result whereas LSTM is about 90 percent, and RNN is even lower at 88 percent. These results show the power of GRSPHTRU's Rastrigin Sphere weight optimization and Penalized Hyperbolic Tangent activation to enhance convergence properties along with generalization. GRSPHTRU registers a TPR of nearly 98 in terms of sensitivity, that is, it is able to detect most DDoS attacks. In contrast, we can see that the baseline models, particularly LSTM and RNN, achieve lower base rates, with a TPR of less than 91, which means that these models have more false negatives and have higher chances of evading attack. The model also has a high degree of specificity with a TNR of approximately 98, which testifies that the model is reliable in the accurate recognition of benign traffic as non-malicious. This high TPR and low TNR balance indicates that the model is resistant to false negative and false positive results, and this is essential in DDoS detection conditions. GRSPHTRU has the best PPV (approximately 98%), which makes sure that the attacks detected by it are actually malicious, thus minimizing the false positive rate. GRU and BiLSTM are also not doing so badly, but LSTM and RNN are less precise and cannot handle the complexity of traffic patterns. Lastly, the proposed framework has an NPV of approximately 98, which gives good confidence to refer to traffic as benign. Conversely, the RNN has the poorest performance with an NPV of less than 88, and is therefore less effective in identifying normal flows and attack traffic.

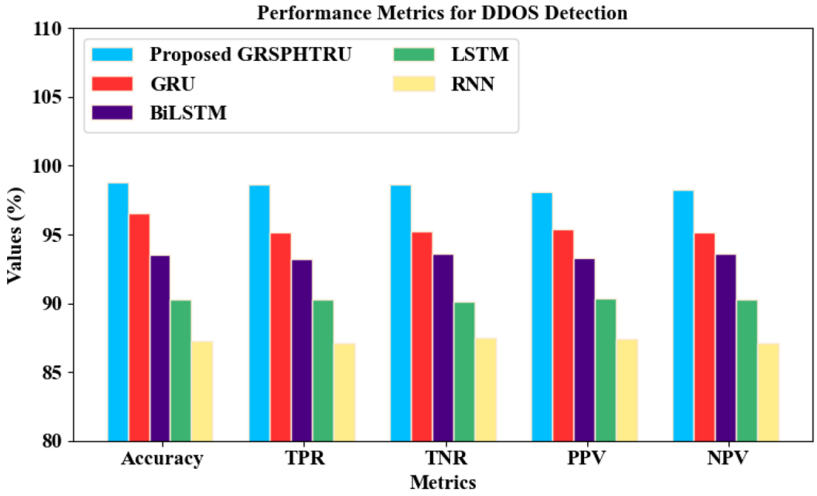

Figure 6. Performance of DDoS detection.

The evaluation confirms that the proposed GRSPHTRU model performs significantly better than the classical recurrent architectures (RNN, LSTM, GRU, BiLSTM) in all the metrics that are considered. Its performance, which has been outstanding over time, shows not only a high detection rate of DDoS traffic (high TPR) but also good resilience to false alarms (high TNR, PPV, NPV). This qualifies GRSPHTRU very well in real-time cross-layer IoT and CPS intrusion detection, where accuracy and reliability are of the essence.

4.6. Analysis of Malware Detection

The comparative Figure 7 demonstrates that the proposed GRSPHTRU model has the highest accuracy of almost 99. This performance outdoes GRU at approximately 98% and BiLSTM at approximately 96%. LSTM and RNN are less effective, and the accuracies are around 93-91, representing lower overall classification performance. These findings reveal the ability of GRSPHTRU to generalize to a wide variety of malware samples. GRSPHTRU once again places first with almost 99, which confirms that it is highly sensitive in identifying cases of malware at the right place. GRU is closely behind with approximately 98%, and BiLSTM has approximately 95%. The performance

of LSTM and RNN is at 92 and 91, respectively. GRSPHTRU has a high sensitivity and therefore false negatives are avoided, which is important in CPS security, where even missing an attack can be disastrous. The trend remains specific, with the GRSPHTRU having the highest value of approximately 99 to guarantee a good detection of benign traffic. GRU and BiLSTM come next with 97% and 95, and LSTM and RNN are left with 92-91. This shows that GRSPHTRU produces fewer false alarms and is therefore more reliable to deploy in the real world. GRSPHTRU has the lowest FPR and stands at nearly 1.0, indicating its resiliency in not falsely classifying legitimate samples as malware. Conversely, RNN and LSTM have significantly higher FPR values of approximately 7-9, and BiLSTM and GRU have better values, but still not equal to that of GRSPHTRU. The smaller FPR means that there are fewer unnecessary notifications and the system is more reliable. The model also shows the lowest FNR, of about 1, so that there is a minimum number of malware samples that are not detected. RNN and LSTM, in their turn, demonstrate the poorest performance, with an FNR of approximately 8-9% and moderate results of BiLSTM and GRU of approximately 4-5%. This proves that GRSPHTRU is especially effective when reducing the number of threats that have been overlooked, which is a requirement of intrusion detection systems.

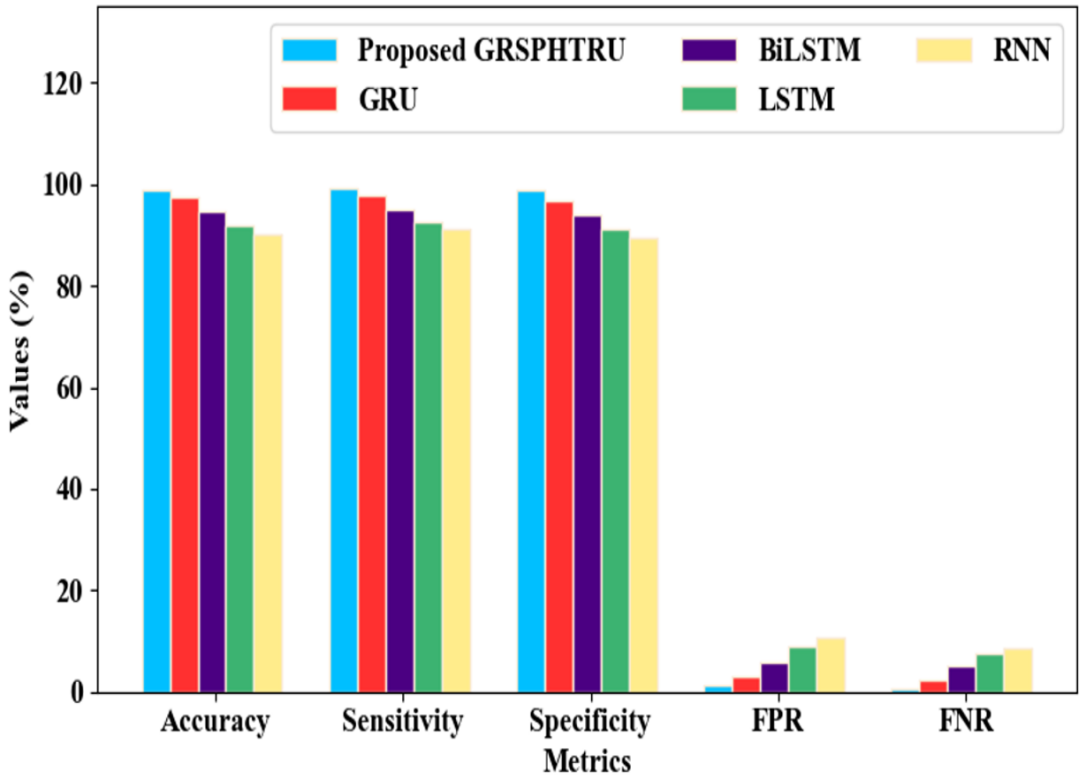

Figure 7. Performance metrics for malware detection.

The analysis demonstrates that the proposed GRSPHTRU is always superior in terms of accuracy, sensitivity, and specificity when compared to all the baseline models and has the lowest false positive and false negative rates. This shows that it is very strong in differentiating malware and benign samples with minimum error. GRSPHTRU can be considered a highly effective option for malware detection in IoT and CPS environments since it also tackles the problem of over-detection (false positives) and under-detection (false negatives).

4.7. Analysis of Ransomware Detection

As Figure 8 indicates, the proposed GRSPHTRU model is the most accurate, reaching almost 99 percent, which is more accurate than all other models in the basis. GRU is followed by around 96%, BiLSTM is next at around 94%, with LSTM and RNN trailing at 92 and around 90, respectively. This high-quality accuracy is the affirmation of the strong ability of GRSPHTRU to classify ransomware and innocent activity. GRSPHTRU again tops in the accuracy

category, with a figure close to 99, proving that it will reduce false positives by making sure that detected ransomware samples are malicious. GRU has approximately 95%, BiLSTM approximately 93, LSTM and RNN have lower performance of below 91. GRSPHTRU has high precision, which is a strength that increases trust and reliability and lowers the false classification of benign files. GRSPHTRU also performs well when it comes to the recall, as it has nearly 99, which means that it is very effective at identifying ransomware. GRU documents an average of 98%, BiLSTM averages 95, LSTM and RNN come second with an average of 92-93. This recall level is high, which means that the suggested model has a great minimization of false negatives, which is an invaluable requirement, as unnoticed ransomware is capable of inflicting severe damage to the system. The proposed model maintains the advantage of a high F1-score of approximately 99 and balances its precision and recall. GRU is the best with approximately 96, BiLSTM about 94, and LSTM and RNN are lower at 91-92. This large F-measure highlights the overall performance and trustworthiness of GRSPHTRU as a whole, and therefore, it is a complete solution to ransomware detection.

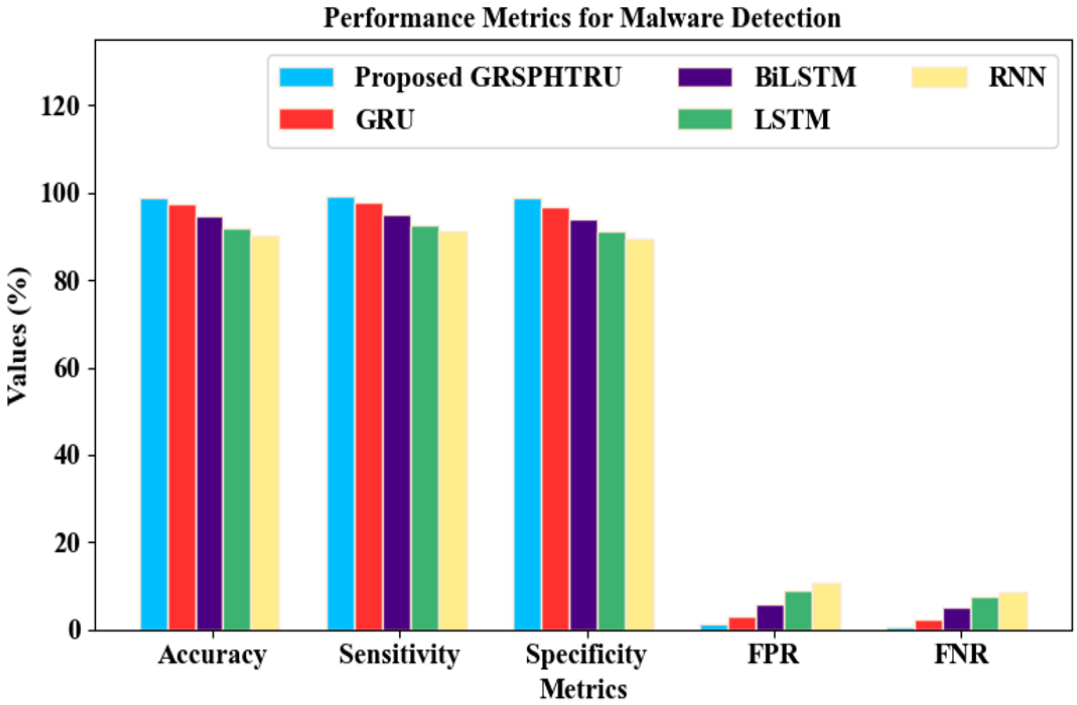

Figure 8. Comparison of ransomware detection.

The results of the analysis prove that the proposed GRSPHTRU model excels at all ransomware detection metrics in comparison with the baseline models. With the best accuracy, precision, recall, and F-measure, the model demonstrates its high level of efficiency in terms of reducing the number of false alarms and high detection probability. GRSPHTRU is an appropriate solution to deploy in the real world where sensitivity and reliability are critical because, in comparison to traditional RNN and LSTM architectures, it demonstrates balanced and reliable performance in ransomware detection in IoT-enabled CPS settings with significant performance discrepancies. Table 3 gives the comparative results across DDOS, malware, and ransomware detection.

In order to establish whether the performance enhancement is statistically significant, paired two-tailed t-tests were performed between the proposed GRSPHTRU model and each baseline in all the folds. The findings have shown (Table 4) that the proposed approach is much more effective than the use of RNN, LSTM, BiLSTM, and GRU models, with p-values below 0.01 in all the considered metrics, demonstrating that the perceived improvements are not explained by random noise.

Table 3. Comparative results across DDoS, Malware, and ransomware detection.

Task / Metric	Proposed GRSPHTRU	GRU	BiLSTM	LSTM	RNN
DDoS Detection					
Accuracy (%)	**99**	96.5	94	90.5	87.5
TPR (%)	**98.7**	95.2	94	90.2	87
TNR (%)	**98.6**	95	93.8	90.1	87.1
PPV (%)	**98.5**	95	93.7	90	87
NPV (%)	**98.6**	95	93.9	90.1	87.2
Malware Detection					
Accuracy (%)	**99**	98	96	92	91
Sensitivity (%)	**99**	98	95	92	91
Specificity (%)	**99**	97.5	95	92	91
FPR (%)	**1**	2.5	4.5	8	9
FNR (%)	**1**	2	5	8	9
Ransomware Detection					
Accuracy (%)	**99**	96	94	91.5	90
Precision (%)	**99**	94.5	93	90.5	88
Recall (%)	**99**	97.8	95	92.5	92
F-Measure (%)	**99**	96.2	94	91.5	90

Table 4. Per-fold performance (Mean ± Std) across 10-Fold CV.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RNN	87.6 ± 1.8	86.9 ± 2.1	87.0 ± 1.9	86.8 ± 2.0
LSTM	91.2 ± 1.4	90.8 ± 1.6	90.5 ± 1.5	90.6 ± 1.5
BiLSTM	94.3 ± 1.2	94.0 ± 1.3	93.8 ± 1.4	93.9 ± 1.3
GRU	96.4 ± 0.9	96.1 ± 1.0	95.9 ± 0.8	96.0 ± 0.9
Proposed GRSPHTRU	**99.1 ± 0.4**	**98.9 ± 0.5**	**98.7 ± 0.4**	**98.8 ± 0.4**

4.8. Analysis of Explainability

Figure 9 is a comparison of the different explainable AI (XAI) methods- Proposed SHSATP, SHAP, LIME, Integrated Gradients (IG), and DeepLIFT on three measures of interpretability. The proposed SHSATP scores the highest in fidelity at approximately 0.9 +, indicating that its feature attributions are very close to the actual decision logic of the model. SHAP comes next with approximately 0.6, whereas Lime does fairly well with 0.4. The lowest fidelity is captured in IG and DeepLIFT (~0.2 and ~0.1), implying that they provide their interpretation of the model reasoning that is very dissimilar to the real model reasoning. This makes SHSATP the most loyal and dependable interpretability approach in the analysis of decisions in cybersecurity. By sparsity, SHSATP also tops with a score of about 0.9, effectively producing short, concise explanations with only the most influential features being highlighted. SHAP (~0.6) and LIME (~0.45) are moderately sparse, and yet they have redundant or noisy attributes. IG and DeepLIFT come in last with values of tens of 0.2 and 0.1, and generate excessively verbose explanations. Herein, one can see how SHSATP enables the simplification of interpretation, which is extremely helpful in all cases of real-time CPS intrusion detection where quick and specific insights are needed.

SHSATP scores the highest (~0.9) on stability, which is the consistency of the explanations with the slightest input differences. This suggests high levels of interpretability even against noisy or adversarial inputs. SHAP (~0.62) and LIME (~0.44) are moderately stable yet are prone to adjustments with the changes in minor data. IG (~0.23) and DeepLIFT (~0.12) are not stable, which reveals that the applications are sensitive. The high stability of SHSATP highlights its robustness and ability to work in changing IoT/CPS systems, where the data patterns continuously change.

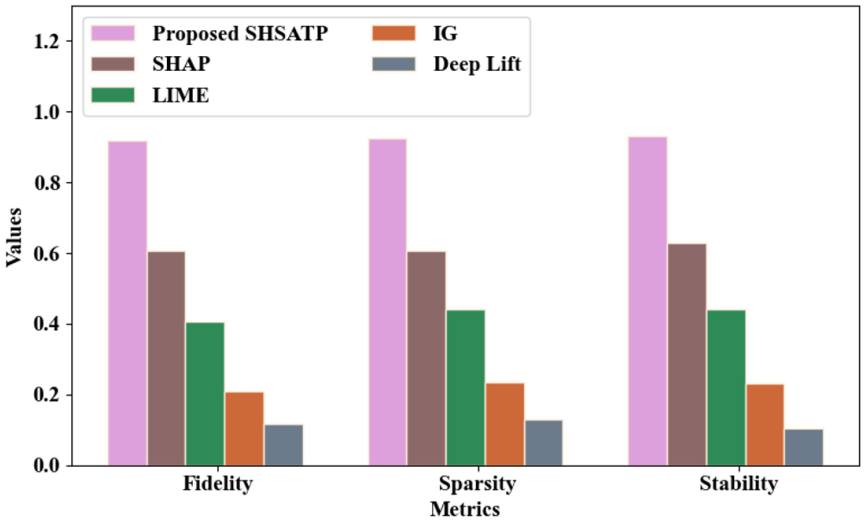

Figure 9. Explainability metrics comparison.

This is clearly shown in the comparison, where it is observed that the proposed SHSATP is much better in the three metrics-fidelity, sparsity, and stability- than the existing explainability methods. Where SHAP and LIME are moderate in interpretability, they are less precise and consistent. IG and DeepLIFT do not fare well, especially in stability, and hence can only be used in systems that require high security. SHSATP is a highly reliable tool capable of offering brief, plausible, and consistent explanations by using the Styblinski-Tang weighted Shapley method, which helps to increase the level of transparency and trust towards automated schemes of cyber-attack detection.

5. CONCLUSION

The paper has introduced an all-inclusive and integrated framework of intelligent and explainable detection of cyber-attacks in IoT-enabled CPS. The suggested Cross-Layer Temporal Patterns Correlation Analysis Model incorporates various innovations- HLPMM to multi-layer temporal correlation, GRSPHTRU to multi-layer deep temporal learning, PGCA to dimensionality reduction, ADBFTLSCAN to adaptive clustering. This integration facilitates early identification, precise classification, and clear interpretation of multi-layered, complex multi-stage attacks, including ransomware, malware, and DDoS that spread over several layers of CPS.

As experimental analysis performed on benchmark datasets (CIC-DDoS2019, ransomware, and malware corpora) shows, the proposed model invariably worked better than the base deep learning models (RNN, LSTM, BiLSTM, GRU) in all metrics. The offered GRSPHTRU demonstrated the 99-percent accuracy, 98.7-percent TPR, and 98.6-percent TNR, proving its high efficiency in identifying the volumetric attacks as well as stealth attacks with the minimal number of false positives. The ADBFTLSCAN clustering effectively represented dense and sparse behavioral patterns that enhanced the discrimination of features in mixed traffic. The reduction based on the PGCA maintained crucial variance but with dramatically cheaper computations, and made the model applicable to the real-time deployment of the CPS. Also, the SHSATP explainability module displayed the best fidelity (approximately 0.9), sparsity, and stability of any existing XAI baselines (SHAP, LIME, IG, DeepLIFT), which means that security analysts can rely on and interpret automated decisions with high confidence. Besides detection and explainability, the logging mechanism facilitated by blockchain offered audit trails that could not be altered, and auditors could view these trails without any interference, which is crucial to the critical infrastructure systems, implying accountability and forensic preparedness. The FL-based design also ensures privacy and scalability of the data, which allows decentralized training on heterogeneous devices of the IoT without losing the accuracy of the model.

Although the outcomes of the research are encouraging, there are a number of weaknesses that should be noted. First, there were various benchmark datasets, but they might not be that representative of the diversity, size, and dynamic nature of real-world deployments of IoT-enabled CPS, especially in highly heterogeneous industrial settings.

Second, because the proposed framework includes a hybrid temporal modeling, explainability, and secure logging architecture, we expect an extra computational cost, potentially being a bottleneck on edge devices with utterly limited resources. Third, although federated learning is more scalable and privacy-enhanced, non-IID data distributions, communication latency, and client availability can still influence the system performance. The literature will be used to resolve these limitations in future work in advancing validation on real-world CPS testbeds and pilot deployments based on heterogeneous edge devices and industrial control systems. Optimization techniques to minimize both computational and energy overheads will be addressed using edge-optimized implementations and model compression techniques. In light of RCER, future research will explore the resilience of federated learning to adversarial scenarios that include poisoning, inference, and communication attacks, and systems to achieve fairness, accountability, and transparency in automated intrusion detection.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] W. Duo, M. Zhou, and A. Abusorrah, "A survey of cyber attacks on cyber physical systems: Recent advances and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 5, pp. 784-800, 2022. https://doi.org/10.1109/JAS.2022.105548

[2] L. Luo, C. Morales-Gonzalez, S. Wang, Z. Ling, and X. Fu, "Unified view of IoT and CPS security and privacy," in *Proceedings of the 2024 International Conference on Computing, Networking and Communications (ICNC), IEEE*, 2024, pp. 495-499.

[3] T. Zhukabayeva, L. Zholshiyeva, N. Karabayev, S. Khan, and N. Alnazzawi, "Cybersecurity solutions for industrial Internet of Things–edge computing integration: Challenges, threats, and future directions," *Sensors*, vol. 25, no. 1, p. 213, 2025. https://doi.org/10.3390/s25010213

[4] J.-P. A. Yaacoub, O. Salman, H. N. Noura, N. Kaaniche, A. Chehab, and M. Malli, "Cyber-physical systems security: Limitations, issues and future trends," *Microprocessors and Microsystems*, vol. 77, p. 103201, 2020. https://doi.org/10.1016/j.micpro.2020.103201

[5] M. Benmalek, "Ransomware on cyber-physical systems: Taxonomies, case studies, security gaps, and open challenges," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 186-202, 2024. https://doi.org/10.1016/j.iotcps.2023.12.001

[6] J. Kaur and K. R. Ramkumar, "The recent trends in cyber security: A review," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 5766-5781, 2022. https://doi.org/10.1016/j.jksuci.2021.01.018

[7] U. Tariq, I. Ahmed, A. K. Bashir, and K. Shaukat, "A critical cybersecurity analysis and future research directions for the Internet of Things: A comprehensive review," *Sensors*, vol. 23, no. 8, p. 4117, 2023. https://doi.org/10.3390/s23084117

[8] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "Machine learning and deep learning approaches for cybersecurity: A review," *IEEE Access*, vol. 10, pp. 19572-19585, 2022. https://doi.org/10.1109/ACCESS.2022.3151248

[9] S. S. Mahadik, P. M. Pawar, and R. Muthalagu, "Heterogeneous IoT (HetIoT) security: Techniques, challenges and open issues," *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 35371-35412, 2024. https://doi.org/10.1007/s11042-023-16715-w

[10] M. R. War, Y. Singh, Z. A. Sheikh, and P. K. Singh, "Review on the use of federated learning models for the security of cyber-physical systems," *Scalable Computing: Practice and Experience*, vol. 26, no. 1, pp. 16-33, 2025. https://doi.org/10.12694/scpe.v26i1.3438

[11] A. M. Banaamah and I. Ahmad, "Intrusion detection in IoT using deep learning," *Sensors*, vol. 22, no. 21, p. 8417, 2022. https://doi.org/10.3390/s22218417

[12] A. S. Dina, A. B. Siddique, and D. Manivannan, "A deep learning approach for intrusion detection in Internet of Things using focal loss function," *Internet of Things*, vol. 22, p. 100699, 2023. https://doi.org/10.1016/j.iot.2023.100699

[13] A. Awajan, "A novel deep learning-based intrusion detection system for IoT networks," *Computers*, vol. 12, no. 2, p. 34, 2023. https://doi.org/10.3390/computers12020034

[14] A. R. Khan, M. Kashif, R. H. Jhaveri, R. Raut, T. Saba, and S. A. Bahaj, "Deep learning for intrusion detection and security of Internet of Things (IoT): Current analysis, challenges, and possible solutions," *Security and Communication Networks*, vol. 2022, no. 1, p. 4016073, 2022. https://doi.org/10.1155/2022/4016073

[15] F. M. Aswad, A. M. S. Ahmed, N. A. M. Alhammadi, B. A. Khalaf, and S. A. Mostafa, "Deep learning in distributed denial-of-service attacks detection method for Internet of Things networks," *Journal of Intelligent Systems*, vol. 32, no. 1, p. 20220155, 2023. https://doi.org/10.1515/jisys-2022-0155

[16] N. U. Ain, M. Sardaraz, M. Tahir, M. W. Abo Elsoud, and A. Alourani, "Securing IoT networks against DDoS attacks: A hybrid deep learning approach," *Sensors*, vol. 25, no. 5, p. 1346, 2025. https://doi.org/10.3390/s25051346

[17] M. Al-Hawawreh, M. Alazab, M. A. Ferrag, and M. S. Hossain, "Securing the industrial Internet of Things against ransomware attacks: A comprehensive analysis of the emerging threat landscape and detection mechanisms," *Journal of Network and Computer Applications*, vol. 223, p. 103809, 2024. https://doi.org/10.1016/j.jnca.2023.103809

[18] R. Mustafa, N. I. Sarkar, M. Mohaghegh, and S. Pervez, "A cross-layer secure and energy-efficient framework for the Internet of Things: A comprehensive survey," *Sensors*, vol. 24, no. 22, p. 7209, 2024. https://doi.org/10.3390/s24227209

[19] M. Georgiades and F. Hussain, "An explainable AI approach for interpretable cross-layer intrusion detection in Internet of medical things," *Electronics*, vol. 14, no. 16, p. 3218, 2025. https://doi.org/10.3390/electronics14163218

[20] R. Yang, H. He, Y. Wang, Y. Qu, and W. Zhang, "Dependable federated learning for IoT intrusion detection against poisoning attacks," *Computers & Security*, vol. 132, p. 103381, 2023. https://doi.org/10.1016/j.cose.2023.103381

[21] Z. Jin, J. Zhou, B. Li, X. Wu, and C. Duan, "FL-IIDS: A novel federated learning-based incremental intrusion detection system," *Future Generation Computer Systems*, vol. 151, pp. 57-70, 2024. https://doi.org/10.1016/j.future.2023.09.019

[22] N. Albanbay *et al.*, "Federated learning-based intrusion detection in IoT networks: Performance evaluation and data scaling study," *Journal of Sensor and Actuator Networks*, vol. 14, no. 4, p. 78, 2025. https://doi.org/10.3390/jsan14040078

[23] S. Hajj *et al.*, "Cross-layer federated learning for lightweight IoT intrusion detection systems," *Sensors*, vol. 23, no. 16, p. 7038, 2023. https://doi.org/10.3390/s23167038

[24] A. Belenguer, J. A. Pascual, and J. Navaridas, "A review of federated learning applications in intrusion detection systems," *Computer Networks*, vol. 258, p. 111023, 2025. https://doi.org/10.1016/j.comnet.2024.111023

[25] B. Sharma, L. Sharma, C. Lal, and S. Roy, "Explainable artificial intelligence for intrusion detection in IoT networks: A deep learning based approach," *Expert Systems with Applications*, vol. 238, p. 121751, 2024. https://doi.org/10.1016/j.eswa.2023.121751

[26] S. A. Alzakari *et al.*, "Explainable artificial intelligence-based cyber resilience in Internet of Things networks using hybrid deep learning with improved chimp optimization algorithm," *Scientific Reports*, vol. 15, no. 1, p. 33160, 2025. https://doi.org/10.1038/s41598-025-15146-x

[27] J. Li, H. Chen, M. O. Shahizan, and L. M. Yusuf, "Enhancing IoT security: A comparative study of feature reduction techniques for intrusion detection system," *Intelligent Systems with Applications*, vol. 23, p. 200407, 2024. https://doi.org/10.1016/j.iswa.2024.200407

[28] E. Yang, S. Jeong, and C. Seo, "Harnessing feature pruning with optimal deep learning based DDoS cyberattack detection on IoT environment," *Scientific Reports*, vol. 15, no. 1, p. 17516, 2025. https://doi.org/10.1038/s41598-025-02152-2

[29] S. Almarri and A. Aljughaiman, "Blockchain technology for IoT security and trust: A comprehensive SLR," *Sustainability*, vol. 16, no. 23, p. 10177, 2024. https://doi.org/10.3390/su162310177

[30] M. B. Bankó *et al.*, "Advancements in machine learning-based intrusion detection in IoT: Research Trends and challenges," *Algorithms*, vol. 18, no. 4, p. 209, 2025. https://doi.org/10.3390/a18040209

[31] J. A. Herrera-Silva and M. Hernández-Álvarez, "Dynamic feature dataset for ransomware detection using machine learning algorithms," *Sensors*, vol. 23, no. 3, p. 1053, 2023. https://doi.org/10.3390/s23031053

[32] N. N. M. Yusof and N. S. Sulaiman, "Cyber attack detection dataset: A review," *Journal of Physics: Conference Series*, vol. 2319, no. 1, p. 012029, 2022. https://doi.org/10.1088/1742-6596/2319/1/012029

[33] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy," presented at the 2019 International Carnahan Conference on Security Technology (ICCST), IEEE, Chennai, India, 2019, pp. 1-8.

[34] M. Landauer, F. Skopik, M. Frank, W. Hotwagner, M. Wurzenberger, and A. Rauber, "Maintainable log datasets for evaluation of intrusion detection systems," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 4, pp. 3466-3482, 2023. https://doi.org/10.1109/TDSC.2022.3201582

[35] M. Hahsler and M. Piekenbrock, *dbscan: Density-based spatial clustering of applications with noise (DBSCAN) and related algorithms (Version 1.0.4) [R package]*. Vienna, Austria: Comprehensive R Archive Network, 2015.

[36] O. Kulkarni and A. Burhanpurwala, "A survey of advancements in DBSCAN clustering algorithms for big data," presented at the 2024 3rd International Conference on Power Electronics and IoT Applications in Renewable Energy and its Control (PARC), IEEE, Mathura, India, 2024, pp. 106-111.