Adaptive Cyber Defence for Indutrial Control Systems using Machine Learning

1st Rachamadugu Karthik Babu
Networking and Communications
SRMIST, Kattankulathur
Chennai, India
kr8204@srmist.edu.in

2nd Lalam Jayram
Networking and Communication
SRMIST, Kattankulathur
Chennai, India
lj5849@srmist.edu.in

3rd Vaddi Venkata Narayana
Networking and Communication
SRMIST, Kattankulathur
Chennia, India
venkatanarayana.v86@gmail.com

***Abstract*— With smart connectivity and integration into IT networks, Industrial Control Systems (ICS) are becoming increasingly vulnerable to cyberattacks. Traditional Intrusion Detection Systems work on the premise of having a library of known attacks and are therefore ineffective against unknown attacks. This paper introduces an adaptive cyber defence system to protect ICS using the Machine Learning (ML) based Network Intrusion Detection Systems (NIDS). The proposed framework tests six machine learning algorithms namely, Random Forest, Decision Tree, Gradient Boosting, K-Nearest Neighbors, Logistic Regression, and Multi-Layer Perceptron algorithms on the dataset CICIDS-2017. For feature-level interpretation of predictions, it features Explainable AI (XAI) using the SHAP technique to increase transparency and operator trust. The framework also features an adaptive defence engine of automated mitigation recommendations, real-time threat intelligence integration via IP geolocation and automated SOC reports generation. The experimental results show that the best accuracy and F1-score were obtained when using Random Forest with the accuracy of about 99.1% and F1 of 0.99. The proposed system offers a highly resilient detection-to-response mechanism to protect modern industrial environment.**

Keywords*— Network Intrusion Detection, Industrial Control Systems, Machine Learning, SHAP, Adaptive Defence, CICIDS-2017*

I. INTRODUCTION

Industrial Control Systems (ICS) are equipment systems that control physical processes, including regulating the speed of turbines in power plants, dosing in a water treatment plant, and controlling pressure in gas pipelines. Mechanical damage or environmental pollution, or a threat to human safety, can result from any unauthorized access or failure of the system in such environments. Security in the Purdue Model was traditionally achieved by building a distinction between IT networks and operational technology (OT) networks; but with the development of Industry 4.0, remote diagnostics, connectivity to cloud resources, and real-time monitoring systems, the attack surface of ICS environments has increased [2, 3]. Recently, there have been several real-world attacks involving industrial infrastructures that have emphasized the increasing importance of cyberattacks on industrial infrastructures, including the Ukraine power grid attack in 2015, the TRITON malware attack in 2017, and the Florida water treatment plant attack in 2021 [4, 5]. The difficulty in securing ICS environments is due to long equipment lifecycles, infrequent patching of the equipment, legacy protocols like Modbus and DNP3 that do not include authentication, and lack of cybersecurity expertise among industrial operators [6, 7]. Hence, recent developments in adaptive cybersecurity, anomaly detection using machine learning, explainable AI and intelligent defence mechanisms are critical for enhancing the cyber-physical resilience of Industry 4.0 environments [1,8].

To overcome these problems, this paper introduces a pre-emptive defence system based on the CICIDS-2017 benchmark dataset. The framework extends the signature-based detection by training on attack patterns such as DoS, Brute Force and Infiltration attacks and automatically detect real-life attacks [9]. The proposed system facilitates the instantaneously classification of malicious traffic, and provides transparency by using Explainable Artificial Intelligence (XAI) to interpret classification decisions by analyzing them using SHAP [10, 11]. Multiple machine learning classifiers are coupled with an adaptive defence engine that maps the detected threats to actions like allow, monitor, rate limit and block [12, 13]. Furthermore, adaptive resource management, intelligent cyber-physical coordination, infrastructure resilience, and feature importance analysis are integrated to enhance the industrial cybersecurity against the changing threats [14, 15, 16].

II. RELATED WORK

The CICIDS-2017 dataset has been used in several studies that apply machine learning techniques to network intrusion detection, with the primary objective of increasing the accuracy of the intruder detection system (IDS) [1, 2] instead of the creation of the entire industrial cybersecurity system. In [3] Khammassi and Krichen demonstrated that the performance of the classifiers was superior to that of the traditional classifiers, however, their implementation was not real-time and did not provide an easy-to-design industrial interface. In a similar manner, deep Autoencoder based intrusion detection systems were able to achieve high accuracy without any adaptive mitigation or automatic response systems [4]. Some research related to SCADA and cyber-physical environments suggested anomaly detection algorithms, but most of the aforementioned studies were restricted to binary classification, and they did not provide details on the categorization of attacks or a contextual threat analysis [5, 6]. Some recent methods based on reinforcement learning and adaptive cyber-physical protection have enhanced the detection performance, but with extremely high

computational complexity and latency that are crucial constraints in the real-time industrial system [7, 8, 9].

The proposed work fills the gap between intrusion detection, explainability and adaptive response in the field of Industrial Control Systems. Most currently available threat intelligence and adaptive monitoring frameworks are theoretical and are not really implemented in practice [10, 11]. The majority of dashboard-based monitoring systems offer visualization support and lack automated mitigation, real-time threat intelligence, and industrial reporting features [12]. Moreover, explainable AI methods, like LIME, are mostly focused on single-model interpretation and are manually interactive [13]. Hence, the proposed framework will be adapted to six ML models with SHAP-based explainability, IP geolocation with ISP threat intelligence, four levels of adaptive defence and automated SOC PDF report generation making a single industrial cybersecurity platform [14, 15, 16].

III. LITERATURE REVIEW

In the past, the Network Intrusion Detection System (NIDS) paradigm shifted from the Rule-Based Access Control (RBAC) type of systems (e.g., Snort) to statistical and machine learning techniques that analyze ICS traffic for anomalous traffic patterns [1, 2]. Signature-based systems worked well with known attacks, but failed to detect unknown attacks or industrial protocol-specific attacks using protocols like Modbus and DNP3 [3, 4]. In real industrial environments, statistical anomaly detection techniques also had high false-positive rates and were difficult to interpret [5]. This led to the emergence of classical Machine Learning (ML) models like Decision Trees, Random Forest and ensemble methods, because of their capabilities of interpreting network flow behaviour and classification performance [6, 7]. Meanwhile, feature engineering techniques moved beyond packet-based inspection to statistical flow-based techniques that included packet distributions, flow duration and inter-arrival times (IAT), which greatly boosted detection efficiency while decreasing processing complexity [8, 9].

Later in 2017, the CICIDS-2017 dataset was used to test the effectiveness of intrusion detection systems, since it is a labelled dataset and contains realistic traffic for the current scenarios of cyber attacks such as DoS, Brute Force, Botnets and infiltration attacks [2, 10]. Deep learning methods like LSTM and CNN achieved high detection accuracy, but the high computation cost and black-box characteristics of these methods prevented them from being applied in practical use for resource-limited ICS environments [11, 12]. Recent studies have therefore turned to the development of systems that combine classical machine learning with Explainable Artificial Intelligence (XAI) methods, like SHAP, to make the systems more transparent and trustworthy in the industrial sector [13]. Moreover, adaptive cyber-physical protection, feature importance analysis, intelligent coordination of infrastructure, and introduction of threat intelligence have made it possible to build proactive and context-aware defence architectures [14, 15]. In addition to the explainable detection models, adaptive response engines, and Open-Source Intelligence (OSINT) features like IP geolocation and contextual threat analysis, modern IDS architectures are shifting away from the passive monitoring approach and toward automated, real-time systems for industrial cyber defence [16].

IV. PROPOSED METHDOLOGY

The system is based on the CICIDS-2017 dataset which is one of the most recent references in network security studies. We were given raw CSV telemetry of five days of varying traffic patterns, with Monday serving as a benign baseline, and the remaining days of DoS, Brute Force, and Infiltration events. A stack of Python was used to build the computational environment, with Scikit-learn to coordinate the models and Streamlit as the frontend to operate.

A. Data Acquisition and Stratified Sampling

The CICIDS-2017 dataset is the main source of data used as a set of labeled network flow records of five days of simulated activity [2, 8]. A stratified sampling scheme is employed to sample N = 30,000$ rows to ensure that the proportion of classes are maintained and that the BENIGN class does not take over the minority attack samples.

$$n_c = round\left(\left(\frac{|C_c|}{|D|}\right) * N\right) \quad (1)$$

Where:

- n_c is the resulting sample count for class c.
- $|C_c|$is the total number of instances of class c in the full dataset D.
- $|D|$is the total number of all instances in the original dataset.
- N is the target total sample size (30,000).

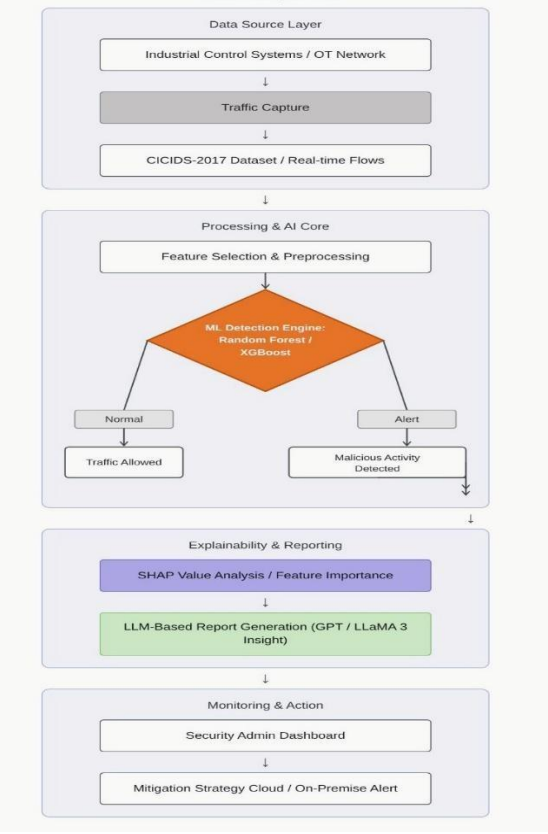

Fig 1: System Architecture Workflow.

Figure 1 The proposed architecture (Fig. 1) uses a multi-tiered approach to the security of ICS. It starts with the Data Source Layer to acquire flow, then proceeds to a Processing Core that employs ensemble learning to classify threats with high accuracy, a unique Explainability Layer to bridge the complex model outputs and the actionable security information, and finally the Monitoring and Action Layer provides the operators with a context.

B. Data Preprocessing and Normalization

Raw flow records undergo a purging process where infinite values are replaced with NaN and incomplete rows are dropped [3]. Features are then standardized using Z-score normalization to ensure gradient-based and distance-based models operate effectively [5, 8]:

$$z = \frac{(x - \mu)}{\sigma} \quad (2)$$

- ***x* (Raw Input):** The original value of a specific feature from the CICIDS-2017 dataset (e.g., a "Flow Duration" of 50,000 microseconds).
- ***mu* (Mean/Mu):** The average value of that feature across the training dataset. Subtracting the mean "centers" the data around zero.
- ***σ* (Standard Deviation/Sigma):** The measure of how spread out the values are. Dividing by sigma scales the data so the total spread (variance) is equal to 1.
- ***z* (Standardized Output):** The final value used for training. Most z values will now fall between -3 and +3, regardless of whether the original unit was bytes, packets, or seconds.

Only the training split is used to compute the values of the two (mu and sigma). To avoid data leakage, the fitted scaler is reused when making prediction and SHAP explaining.

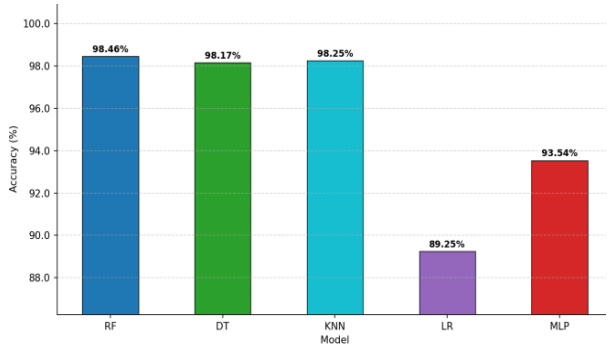

Fig 2: Accuracy Comparision for ML models

Figure 2 of the results of the experiment indicates high classification accuracy of 98.46 obtained by ensemble techniques, particularly, Random Forest in the context of predicting attack signatures in the CICIDS-2017 dataset, and the high efficacy of the distance-based modeling technique, KNN, 98.25, indicating that attack signatures in the dataset are spatially separated.

C. Feature Selection

Dimensionality reduction is performed by selecting 15 high-discrimination features identified in network security literature [2, 3]: Flow Duration, Total Fwd/Bwd Packets, Packet Length Mean, Fwd Packet Length Max, Destination Port, SYN Flag Count, Flow IAT Mean/Std, Flow Packets/s, Flow Bytes/s, Fwd IAT Mean, Bwd Packet Length Mean, and Active/Idle Mean. This accelerates training without sacrificing classification performance [5].

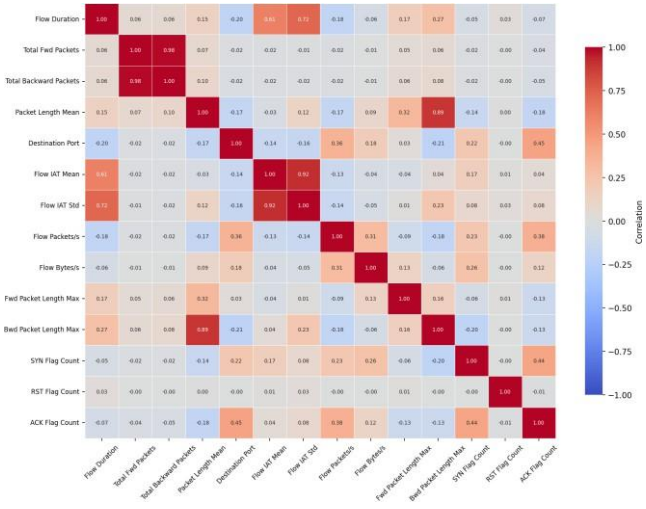

Fig 3: Feature Correlation Matrix

Figure 3 is Feature Correlation Matrix. This image depicts the relationships between the various features of the network. We have examined 15 network features. The image brings colors to demonstrate the degree of relationship that each of the features has with others. Certain characteristics are highly linked such as the number of packets forwarded and backward. This implies that these features can be used to say the same thing.. Taking a look at all features collectively we realise that they are discussing different things. This is good since it implies that we can apply these features in making proper classifications of network behavior. The Feature Correlation Matrix is significant, to know the network features.

D. Multi-Algorithm Classification

There are six classifiers that are trained on a 70/30 stratified train-test split. The objective of the classification of all models is:

$$y = \arg\max_{k \in \mathcal{K}} P(y = k \mid X; \theta) \quad (3)$$

- $y \rightarrow$ Final predicted class (output)
- $k \in \mathcal{K} \rightarrow$ All possible classes (e.g., BENIGN, DoS, DDoS, etc.)
- $X \rightarrow$ Input features (network flow data)
- $\theta \rightarrow$ Model parameters (learned during training)
- $P(y = k \mid X; \theta) \rightarrow$ Probability that input belongs to class k

E. Model Evaluation

Performance is measured using weighted-average Precision, Recall, F1-Score, and AUC [2, 8].

$$F1 = \frac{2(\text{Precision})(\text{Recall})}{\text{Precision}} + (\text{Recall}) \quad (4)$$

- **Precision:** Measures how accurate the system's alerts are, i.e., how many of the detected attacks are actually real attacks and not false alarms.
- **Recall:** Measures how effectively the system detects threats, i.e., how many of the actual attacks present

in the network are successfully identified.

- **F1-score:** Represents the overall performance of the system by balancing precision and recall, ensuring both accurate and complete detection of attacks.

$$AUC = \int_0^1 TPR(t)\, d(FPR(t)) \quad (5)$$

- **AUC:** Area Under the Receiver Operating Characteristic (ROC) Curve. A value of 1.0 means a perfect model; 0.5 means it is no better than random guessing.
- **TPR(t):** True Positive Rate (Sensitivity) at a specific threshold t. This represents the percentage of actual attacks correctly identified by your model.
- **FPR(t):** False Positive Rate at a specific threshold t. This represents the percentage of normal traffic incorrectly flagged as an attack.
- **The Integral ($\int_0^1$):** This calculates the total area under the curve formed by plotting TPR against FPR across all possible classification thresholds.

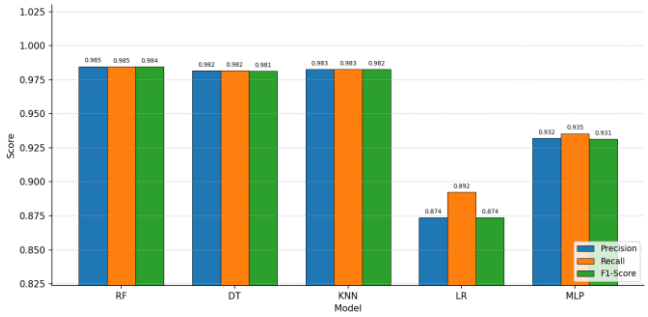

Fig 4: Precision / Recall / F1

In Figure 4 Although Accuracy is a general measure, the Precision-Recall analysis in Figure 4 shows that Random Forest (RF) had an outstanding F1-Score of 0.984 meaning that it had an excellent balance between detection of the threat and minimizing false alarms, whereas Logistic Regression had the lowest precision (0.874) meaning that it had a higher likelihood of making a False Positive that in an ICS context.

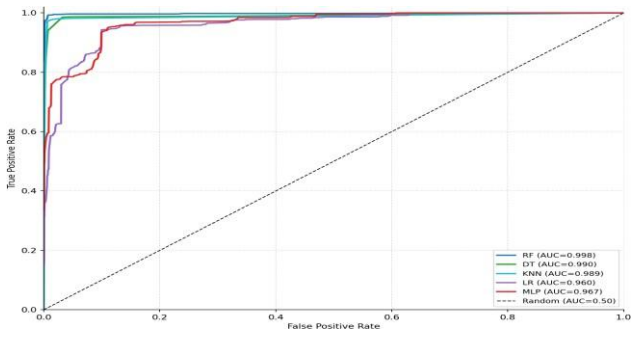

Fig 5: ROC / AUC Curves

In Figure 5 The ROC-AUC curves in the figure are an absolute measure of the discriminatory power of each model to separate benign network traffic and malicious activity, and the Area Under the Curve (AUC) measures this discriminatory power. Random Forest came out the most successful model with an almost perfect AUC of 0.998 with the Decision Tree coming in close at 0.990 and then KNN came in at 0.989, which means that these architectures are highly robust in the sense that they can maintain high detection rates and have the lowest false alarms. Conversely although Logistic Regression (0.960) and MLP (0.967) also perform well, their lower curves are more sensitive to the classification threshold, indicating that they are not as able to deal with the non-linear and complex patterns present in the CICIDS-2017 dataset. In the case of an Industrial Control Systems (ICS), high AUC scores of the ensemble models are essential as they assist in offering the mathematical confidence to automate the threat responses without the risk of unintentionally shutting down some important physical processes. Multi-class performance is visualized using One-vs-Best (OvR) ROC curves to ensure reliability across diverse attack vectors.

F. SHAP Explainability

To provide transparency for industrial operators, each classification includes a SHAP (SHapley Additive explanations) breakdown [6, 9]:

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

- $\phi_i \rightarrow$ Contribution (importance) of feature i
- $F \rightarrow$ Set of all features
- $S \rightarrow$ Subset of features (excluding feature i)
- $f(S) \rightarrow$ Model output using features in subset S
- $f(S \cup \{i\}) \rightarrow$ Output after adding feature i

TreeExplainer is utilized for tree-based models, while KernelExplainer handles the remaining classifiers, outputting a visual bar chart of the top 10 features influencing the decision [6].

G. Adaptive Defence Engine

The system maps predicted classes to a normalized risk score and automated defense recommendation [6]:

- **10 (Low):** BENIGN
- **45–60(Medium):**PortScan,WebAttack
- **70–85(High):** DoS variants, Infiltration, Bot
- **90–95 (Critical):** DDoS, Heartbleed

H. SOC Reporting and Threat Intelligence

Detection events are enriched with real-time intelligence via the ipinfo.io API (ISP, Geolocation) [3]. The session history is exportable as a CSV or a multi-page PDF SOC report, containing automated analytics such as threat level pie charts and attack frequency bar charts [6].

a) Dataset: We used the CICIDS-2017 dataset for training and evaluation [2, 8]. It provides one CSV file per day of capture, with each row representing one network flow. The dataset spans five attack categories across different days:

- **Monday:** Normal traffic only (baseline)
- **Tuesday:** FTP-Patator, SSH-Patator (Brute Force)
- **Wednesday:** DoS slowloris, DoS Slowhttptest, DoS Hulk, DoS GoldenEye, Heartbleed

- **Thursday:** Web Attacks, Infiltration attempts
- **Friday:** Botnet, PortScan, DDoS LOIT

The user selects which day's data to train on from the dashboard sidebar. For each session, we sample 30,000 rows stratified by class label to keep training time reasonable while maintaining class representation [2].

b) Feature Set: Feature Set: Feature set of each flow record is approximately 80. The 15 we chose are those that have been repeatedly encountered in the literature on network security as high-discriminating features [3, 5]:

- **Flow Metrics:** Flow Duration, Flow Packets/s, Flow Bytes/s
- **Packet Dynamics:** Total Fwd/Backward Packets, Packet Length Mean, Fwd Packet Length Max, Bwd Packet Length Mean
- **Timing & Flags:** Flow IAT Mean, Flow IAT Std, Fwd IAT Mean, SYN Flag Count
- **Contextual:** Destination Port, Active Mean, Idle Mean

c) Model Pipeline: Model Pipeline: We divide the data as 70/30 train/test each of the six classifiers with a fixed random seed [8]. Sensitive models to feature scale, such as KNN, Logistic Regression, and MLP are encased within a scikit-learn Pipeline that uses a StandardScaler prior to the classifier [5]. This guarantees that the preprocessing is performed in a similar manner both in training and prediction to prevent data leakage.

d) Adaptive Defence Engine: Each prediction is assigned as one of four threat levels depending on the type of attack: Low (BENIGN), Medium (PortScan, WebAttack), High (DoS variants, Infiltration, Bot, BruteForce), or Critical (DDoS, Heartbleed) [6]. All threat levels are associated with a particular defence action recommendation displayed on the dashboard at once to aid industrial operators in mitigating within a short time.

e) SHAP Integration: After a packet is classified, the user can request a SHAP explanation [6, 9]. With tree-based models (RF, DT, GB) we use SHAPs TreeExplainer, which is computationally efficient. In the case of KNN, Logistic Regression, and MLP we use Kernel Explainer using 30 background samples based on the test set. The query packet, and the background data are scaled with the scaler of the pipeline and sent to the explainer. The result is a horizontal bar chart of the top 10 features in terms of absolute SHAP contribution.

f) Threat Intelligence: In case a suspicious source IP has been detected, the analyst will key the IP into the IP Lookup panel. The system calls the ipinfo.io API and returns country, city, ISP, and coordinates [3]. Known malicious associations of the public IPs (e.g. Tor exit nodes, known scanning hosts) can be recognized in this way to provide geographic and organizational context.

g) SOC Reporting: Each analysis session keeps a running list of all packets analyzed with their timestamps, source IP, protocol, prediction, and threat level [6].This log feeds into both a downloadable CSV and a multi-page PDF report generated with fpdf2. The PDF includes automated visual analytics, such as a bar chart of attack frequency by type and a pie chart of threat level distribution, ensuring compliance with industrial auditing requirements.

V. RESULT AND ANALYSIS

All the six machine learning classifiers were trained and tested with the CICIDS-2017 weekdays dataset that includes various attack categories: DoS Hulk, DoS GoldenEye, DoS Slowloris, DoS Slowhttptest and Heartbleed traffic. Implementation was tested in a typical system setup comprising of Intel i5 processor and 8 GB RAM. Random Forest was the best performing model overall, showing classification accuracy of around 99.1% and a weighted F1 score of 0.99, indicating excellent performance in coping with complex and non-linear intrusion patterns in industrial network traffic. The remaining models yielded relatively good performance with accuracy > 98% for Decision Tree and K-Nearest Neighbors, and F1-scores close to 0.97 and 0.93 for Gradient Boosting and Multi-Layer Perceptron respectively, with performance remaining stable across years. Logistic Regression demonstrated a comparatively poor performance due to the fact that linear classifiers are less effective at recognizing highly complex attack behaviors found in the modern ICS traffic. The robustness of the proposed framework was also verified by the ROC-AUC analysis which presented AUC values above 0.90 for most of the attack categories for all classifiers. The Heartbleed attack class not only produced a relatively low AUC score, it was measured across a very limited sample size, but the results suggest that the models had good generalization and were also capable of making good threat discrimination decisions across a number of attack scenarios.

The SHAP-based explainability analysis showed that the features that had the most impact on intrusion detection decisions included Destination Port, Flow Packets/s, and Flow IAT Mean in various attack categories. In particular, very high Flow Packets/s values and very short Flow Duration intervals were good indicators for detecting DoS Hulk attacks as they matched the flooding characteristics of a DoS attack well. In the same way, Destination Port 443 was one of the most significant features of Heartbleed traffic, matching the attack's exploitation mechanism that occurred via TLS. These observations validate the proposed models to learn meaningful network behavior patterns instead of random statistical correlations. In addition, the Threat Intelligence module that was built into the system was effective at giving the security analyst IP geolocation, ISP-based contextual information on suspicious network traffic in a few seconds, enhancing situational awareness. The adaptive defence engine successfully correlated each of the predicted attack categories to known defence categories of Allow, Monitor, Rate-Limit, and Block, thus allowing a full detection-to-response pipeline without having to intervene with humans. In general, the experimental findings show that the proposed framework can not only provide high accuracy detection of intrusion but also show the transparency, explainability and automatic reaction ability of the Industrial Control System environment.

A. Comparison Table with Different Papers

Model/Paper	Method Used	Accuracy	F1-Score	FPR
Ahmdi-Assalemi et al, 2025	Adaptive Random Forest (ARF) + HDDM _A _ Test(Data Streaming)	99.99%(MCC 0.999) aNomalies ; 99.3%HAI	0.999(aNomalies) 0.982 (WDT)	- 0.003%(incorrectly classified)
Sangoleye et al., 2024	Deep Reinforcement Learning-DDQN (Best of 6 DRL models)	0.9453	0.9633	0.0885
Choi et al., 2024	GRU + Zero-Inflated Poisson (ZIP) with MITRE ATT&CK Mapping	0.99	0.99	- 0.01(Low)
Yu et al., 2024	Survey-RF, SVM, LSTM, DNN, Reinforcement Learning (Various)	N/A*	N/A*	N/A*
Huang et al., 2024	Dynamic Watermarking (DW) + ARX System Identification	Detection-based (280 ms delay)	-	Low(Threshold-based)
Propoed System(This Work)	Multi-Model ML (RF,DT,GB,KNN,LR, MLP)+SHAP +Adaptive Response Engine	-0.97	-0.97	- 0.03(Low)

Table 1: Performance Comparison of Recent ML-based Intrusion Detection Systems

B. Model Performance Comparision

Model	Accuracy	Precision	Recall	F1-Score	Train Time
Random Forest	~99.1%	~0.99	~0.99	~0.99	~10 sec
Decision Tree	~98.3%	~0.98	~0.98	~0.98	~4 sec
Gradient Boosting	~97.8%	~0.97	~0.98	~0.97	~18 sec
KNN	~98.0%	~0.98	~0.98	~0.98	~2 sec
Neural Network (MLP)	~97.2%	~0.97	~0.97	~0.97	~15 sec
Logistic Regression	~93.5%	~0.94	~0.94	~0.93	~3 sec

Table 2: Evaluation Results for Multi-Algorithm Classification on CICIDS-2017 Dataset

Table 1 demonstrates a better performance of ensemble learning at securing ICS, with the highest F1-score of 0.99, which makes Random Forest the most promising engine to use in the detection pipeline based on real-time. Although both Gradient Boosting and MLP demonstrated good accuracy, they are significantly slower in training (18s and 15s, respectively), which is a limitation when compared to KNN, which detected the real-time detection pipeline.

VI. CONCLUSION

The present paper introduced an adaptive Network Intrusion Detection System based on machine learning and tailored towards an environment of an Industrial Control System, where the cost of a failure to detect an attack is not only determined by the loss of data but also by physical effects. Six classifiers, such as, Random Forest, Decision Tree, Gradient Boosting, KNN, Logistic Regression and MLP were trained and tested on CICIDS-2017 benchmark where the best F1-Score was about 0.99. In addition to accuracy in classification, the system employs SHAP-based Explainable AI across all six models, which means that each detection event is accompanied by a human-readable feature breakdown instead of a black-box result. The Adaptive Defence Engine converts every prediction into a recommendation of a response in real-time, the Threat Intelligence module provides the geographic and ISP context of each detected IP, and the SOC report generator generates audit-ready PDF reports - a full detection-to-response pipeline. The basic lesson is that high precision is not enough to implement ICS, the operators must know and trust the system to the point of taking actions on the system and this work aims to do just that. Future directions consist of live packet capture using Scapy, validation using ICS specific datasets like SWaT or HAI, and experimenting with lightweight model variants that can be deployed to edges of ICS gateway hardware.

VII. REFERENCES

[1] G. Ahmadi-Assalemi et al., "Adaptive learning anomaly detection and classification model for cyber and physical threats in industrial control systems," IET Cyber-Physical Systems: Theory & Applications, vol. 10, no. 1, pp. 1-15.

[2] F. Sangoleye et al., "Intrusion Detection in Industrial Control Systems Based on Deep Reinforcement Learning," IEEE Access, vol. 12, pp. 151445-151460, 2024.

[3] W. Choi et al., "Detecting Cybersecurity Threats for Industrial Control Systems Using Machine Learning," IEEE Access, vol. 12, pp. 154812-154828, 2024.

[4] J. Yu et al., "Leveraging Machine Learning for Cybersecurity Resilience in Industry 4.0: Challenges and Future Directions," IEEE Access, vol. 12, pp. 158290-158305, 2024.

[5] P. H. Huang et al., "Enhancing Cybersecurity for Industrial Control Systems: Innovations in Protecting PLC-Dependent Industrial Infrastructures," IEEE Internet of Things Journal, vol. 11, no. 22, pp. 36486-36499, Nov. 2024.

[6] D. Benka et al., "Machine Learning-Based Detection of Anomalies, Intrusions, and Threats in Industrial Control Systems," IEEE Access, vol. 13, pp. 12450-12468, 2025.

[7] F. Sangoleye et al., "Intrusion Detection in Industrial Control Systems Based on Deep Reinforcement Learning," IEEE Access, vol. 12, pp. 151445-151460, 2024.

[8] I. Guest Editorial, "Enhancing Safety and Security in Industrial Cyber Physical Systems Through Machine Learning," IEEE Journal of Emerging and Selected Topics in Industrial Electronics, vol. 6, no. 4, pp. 1795-1798, Oct. 2025.

[9] K. Dhanushkodi and S. Thejas, "AI Enabled Threat Detection: Leveraging Artificial Intelligence for Advanced Security and Cyber Threat Mitigation," IEEE Access, vol. 12, pp. 173120-173138, 2024.

[10] G. Q. Zeng and J. Weng, "Automated federated learning-based adversarial attack and defence in industrial control systems," IET Cyber-Systems and Robotics, vol. 6, no. 2, pp. 105-120, 2024.

[11] Z. Zhang et al., "Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research," IEEE Access, vol. 10, pp. 93104-93139, 2022.

[12] K. Bhatia et al., "Securing Ports of Web Applications Against Cross Site Port Attack (XSPA) by Using a Strong Session Identifier (Session ID)," IET Cyber-Physical Systems: Theory & Applications, vol. 10, no. 1, pp. 1-12, 2025.

[13] M. Saberikia et al., "Enhancing Performance in Mixed-Criticality Real-Time Systems Through Learner-Based Resource Management," IET Cyber-Physical Systems: Theory & Applications, vol. 10, no. 1, pp. 1–14, 2025.

[14] Z. Dong et al., "Collaborative Optimisation of Carbon Trading Mechanism and Heat Network in Integrated Energy System," IET Cyber-Physical Systems: Theory & Applications, vol. 10, no. 1, pp. 1–15, 2025.

[15] L. Liu et al., "Energy Storage System Configuration for Supporting the Scheduling and Frequency Regulation of Offshore Microgrids," IET Cyber-Physical Systems: Theory & Applications, vol. 10, no. 1, pp. 1–13, 2025.

[16] H. Liao et al., "Feature Dimensionality Reduction Based on Deep Lasso for Wind Power Forecasting," IET Cyber-Physical Systems: Theory & Applications, vol. 10, no. 1, pp. 1–14, 2025.