Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Hybrid Deep Ensemble Intrusion Detection System with Adaptive Zero Trust Security for Industrial IoT Infrastructure

Manish Kumar Srivastava
National Forensic Sciences University
Kavalla Prasanna
National Forensic Sciences University
Ujjaval Patel
National Forensic Sciences University
Ravirajsinh Vaghela
ravirajsinh.vaghela@nfsucampus.ac.ug

National Forensic Sciences University Uganda Campus

Research Article

Keywords: Deep Neural Network, Dynamic Trust Management, Extreme Gradient Boosting, Hybrid Deep Learning-Enabled Intrusion Detection System, Industrial Internet of Things, Zero Trust Security

Posted Date: September 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10704863/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Hybrid Deep Ensemble Intrusion Detection System with Adaptive Zero Trust Security for Industrial IoT Infrastructure

[1]Manish Kumar Srivastava, [2]Kavalla Prasanna, [3]Ujjaval Patel, [4]Ravirajsinh Vaghela (corresponding author)

[1,2,3]*School of Cyber Security & Digital Forensics,*
National Forensic Sciences University,
Gandhinagar,Gujarat,India

[4]*Department of Cybersecurity & Digital Forensics,*
National Forensic Sciences University,
Jinja, Uganda.

Email: [1]manish.phdcs2426@nfsu.ac.in, [2]kavalla.btmtcs2116@nfsu.ac.in, [3]ujjaval.patel@nfsu.ac.in,
[4]ravirajsinh.vaghela@nfsucampus.ac.ug

Abstract:
Industrial IoT (IIoT) devices are used in Critical Information Infrastructure (CII) of the organization for seamless communication and improve productivity of the organization. The traditional cyber security approaches for such type of attacks are inadequate to counter sophisticated malware attacks. To tackle such scenario, this paper introduces a novel approach of Hybrid Deep Ensemble Intrusion Detection System (HDE- IDS) framework with dynamic Zero Trust Security enforcement. The proposed HDE- IDS framework is a hybrid combination of Deep Neural Networks (DNN) and Ensemble (XGBoost) to mitigate the cyber security threat of IIoT infrastructure. The proposed framework incorporates the transformer encoder which is being used to provide mechanism to distinguish benign traffic from malicious one using DNN. While the XGBoost ensemble classifier performs final malware attack type detection and classification and contributes to the adaptive risk scoring mechanism used by the zero-trust security.

The proposed approach uses CIC-IoT 2023 malware datasets for training and TON_IoT dataset for validation with the experimental evaluation. The proposed methodology outperforms classical approaches with much improved performance parameters. Statistical analysis reveals that HDE-IDS model is stable, consistent, and higher detection accuracy. Further, the model is validated by TON_IoT dataset and used to generalize the model for live unseen traffic. The proposed approach upscale the cyber resilience IIoT infrastructure by strong integration of intelligent malware detection with Zero Trust Security (ZTS). The proposed approach provides real time detection and isolation, adaptive trust management, provide scalability, reduces false positives and false negatives, and able to achieve best stability and consistency.

Keywords— Deep Neural Network, Dynamic Trust Management, Extreme Gradient Boosting, Hybrid Deep Learning-Enabled Intrusion Detection System, Industrial Internet of Things, Zero Trust Security.

I. INTRODUCTION

Many IIoT based cyber physical systems are used for seamless communication along with continuous data monitoring and control. Many utilities sector uses IIoT devices such as smart sensors, meters, and actuators as CII to provide real monitoring to perform advanced data analytics and predictive maintenance for operation and control. While IIoT based sensors not only enhances the monitoring, operational efficiency and real-time decision making, but also increases the attack surface [1]. The attack like advance malware, ransomware, fileless attacks and lateral movement campaigns posing critical cyber security risks to safety

of the IIoT based CII installation of the smart city [2]. Traditional signature-based malware detection and protection system are ineffective for these zero day and polymorphic threats against this IIoT based CII. These systems are failed to detect like Advanced Persistent Threats (APTs) which introduce customize malware [3]. Artificial Intelligence (AI) and Machine learning (ML) based malware intrusion detection systems are being used predominantly [4]. AI based method leverage behavioral analytics, anomaly detection and deep learning, and ensemble classification to identify malicious patterns with high dimensional IIoT traffic data. However, most of the AI base malware detection system focus on attack type detection and classification accuracy without integrating dynamic trust evaluation, contextual risk assessment, and adaptive policy enforcement. Moreover, industrial CII environments require reliability, explainability, and low latency decision making are constraints that many AI systems fails to satisfy [5].

Zero Trust Security (ZTS) is work on the concept of "*No Trust but verify and continuously authenticate*". Fig. 1 depict NIST Zero Trust Architecture [6]. ZTS enforcement on the CII is for the real time explicit verification, least privilege access, micro segmentation, and restricted lateral movements[7]. Unlike traditional perimeter-based security models, ZTS continuously validates the user identities, device posture, behavioral attributes, and contextual risk before granting access. In ZTS, dynamic enforcement can limit lateral movement of the malware, isolate compromised devices, and adapt security policies in response to real time threat intelligence.

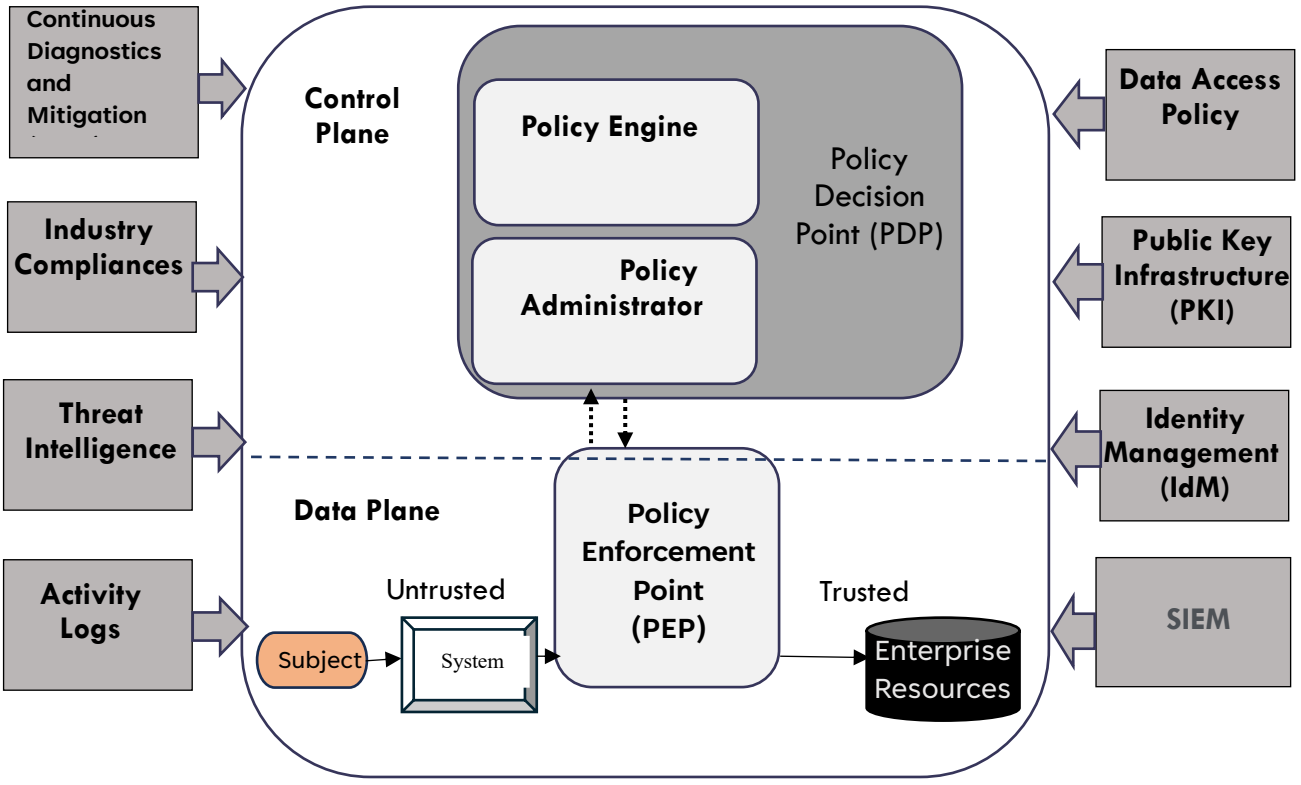


Fig.1 NIST Architecture for Zero Trust Architecture

A. Literature Survey

The increasing emphasis on securing IIoT infrastructures has led to the development of numerous intelligent security frameworks addressing diverse challenges and application scenarios.
These contributions exhibit considerable variations in terms of architectural design, detection capabilities, evaluation methodologies, and practical applicability. A systematic comparison of representative studies

is presented below in Table 1 to provide an overview of the existing research landscape and to identify the limitations that motivate the present work.

Table 1 Comparative Analysis of Existing AI-Based Security Frameworks

Reference	Framework / Technique	Application Domain	Key Contribution	Research Gap
[8] Farhan et al.	XAI-CPPS	CPPS	XAI-driven intrusion detection with improved interpretability.	No Federated Learning/ZTA integration; limited industrial validation.
[9] Gummadi et al.	XAI-IoT	IoT Anomaly Detection	Explainable anomaly detection with transparent security decisions and reduced false positives.	Limited cross-dataset validation; no Federated Learning integration.
[10] Jeon et al.	Multi-SPPNet	Smart IoT Malware Detection	Lightweight static malware detector with high detection accuracy.	High model complexity; lacks Zero Trust/privacy integration.
[11] Wazid et al.	Explainable DL Malware Detection	Intelligent Transport Systems	Hybrid CNN-LSTM model learns spatial and temporal attack patterns.	Domain-specific framework with limited cross-dataset scalability.
[12] Kalakoti et al.	XAI Botnet Detection	IoT Botnet Detection	Introduces quantitative explainability metrics for botnet detection.	Limited explainability metrics and no resource constrained IoT validation.
[13] Pirbhulal et al.	AD-5GIoT	5G-enabled IoT	Hybrid DNN/CNN-LSTM ensemble model for anomaly detection.	Centralized training and limited zero-day attack validation.
[14] Pallavi et al.	Polymorphic Malware Detection	IoT Malware Detection	Real-time polymorphic malware detection using RF, SVM and DNN.	Heavy dependence on synthetic data and limited hardware validation.
[15] Lefoane et al.	NMF + LSA Hybrid Feature Selection	Malware Detection	Combines dimensionality reduction and latent feature extraction for improved malware detection.	Traditional ML-based approach lacking Zero Trust alignment.
[16] Shtayat et al.	Explainable Ensemble Deep Learning (EEDL)	IIoT Intrusion Detection	CNN-LSTM-DNN ensemble with explainability for IIoT intrusion detection.	High computational overhead and limited lightweight IIoT evaluation.
[17] Alamro et al.	Bayesian Optimized Transfer Learning (BOTL)	IoT Cyber Attack Detection	Bayesian optimized transfer learning improves cyber-attack detection on resource constrained IoT devices.	Limited transfer scope and absence of immutable security architecture.

[18] Al-Sadi et al.	Federated Learning + Zero Trust	Smart City IoT	Combines Federated Learning with Edge AI for adaptive Zero Trust Security.	Early-stage prototype with limited state-of-the-art comparison.
[19] Alanazi et al.	Blockchain-enabled Zero Trust Framework	IoT Security	Privacy-preserving Zero Trust framework integrating Blockchain and domain-adaptive anomaly detection.	Scalability under large IoT deployments not fully evaluated.
[20] Hussain et al.	Privacy-preserving Zero Trust Framework	IoT Security	Integrates Zero Trust, Federated Learning, Differential Privacy, and Blockchain for secure IoT.	Noise introduced by Differential Privacy and scalability challenges.
[21] Ullah et al.	ZTID-IoV	Internet of Vehicles (IoV)	Combines Zero Trust, Neurosymbolic AI, and Federated Meta Learning for vehicle intrusion detection.	Scalability across heterogeneous IoT environments not validated.
[22] Nawshin et al.	Differential Privacy Malware Detection	Resource-constrained IoT	Integrates Zero Trust and Differential Privacy for privacy-aware malware detection.	No real-time implementation and no zero-day attack analysis.
[23] Zukaib et al.	Meta-IDS	Medical IoT	Meta-learning IDS using MAML for handling imbalanced and heterogeneous medical IoT traffic.	Limited explainability and absence of Blockchain-based trust management.
[24] Khan et al.	Context-aware Zero Trust Framework	Internet of Vehicles (IoV)	Context-aware trust evaluation integrated with ML-based intrusion detection.	Limited to vehicular environments; broader IoT deployment not evaluated.
[25] Sakraoui et al.	AI-enabled Blockchain Authentication	6G-enabled IoT	Lightweight AI-assisted authentication integrated with blockchain trust management.	Limited to authentication; full IDS evaluation not provided.
[26] Kaur et al.	Federated Learning with DRL-GRU	IIoT Intrusion Detection	Integrates Federated Learning, DRL, and GRU for energy-efficient and privacy-preserving intrusion detection.	High computational and communication overhead with limited explainability.
[27] Mao et al.	FeCoGraph	Few-shot Malware Detection	Federated contrastive learning using graph structural information improves few-shot detection.	Scalability challenges and dependence on graph quality.
[28] Jorgensen et al.	Extensible ML Framework	Encrypted Traffic Classification	Uses uncertainty quantification for robust encrypted traffic classification.	Experimental setup limited to VPN traffic environments.

[29] Liu et al.	ATVITSC	Network Traffic Classification	Combines Transformer-based packet representation with 1D CNN local byte features for improved classification.	Limited dataset diversity and scalability evaluation.
[30] Srivastava et al.	Zero Trust-enabled IDS	IoT Networks	Integrates Random Forest with Zero Trust conduit-zone segmentation to prevent lateral movement attacks.	Static zone configuration without adaptive segmentation mechanism.
[31] V. M. U. et al.	AI-enabled IoT Security Survey	IoT Security	AI-integrated layered architecture for secure IoT orchestration.	No scalability analysis or experimental validation.
[32] Mwangi et al.	Binary Diversification	IoT Malware Detection	Demonstrates vulnerability of ML malware detectors to binary transformations.	Limited datasets and structure-based analysis only.
[33] Moustafa et al.	XAI-based IDS Survey	IoT Intrusion Detection	Reviews Explainable AI techniques for IDS transparency.	No Zero Trust integration and limited validation.
[34] Hussain et al.	Hybrid Malware Detection Survey	Malware Detection	Survey's static, dynamic, and hybrid deep learning methods.	No unified benchmark or validation framework.
[35] Wang et al.	AI-enabled MEC Security	Mobile Edge Computing	Reviews AI-based security and privacy mechanisms for MEC.	Ignores adversarial attacks and real-time orchestration.
[36] Rahman et al.	Blockchain-AI CPS Protection	Industry 4.0 CPS	AI and Blockchain framework for APT detection.	High computational cost and no zero-day evaluation.
[37] Liu et al.	Zero Trust Survey	IoT Networks	Reviews Zero Trust architecture and enabling technologies.	Limited AI implementation and experimental validation.
[38] Zhu et al.	Blockchain-enabled Zero Trust	Healthcare IoMT	Secures healthcare communication using Blockchain and Zero Trust.	High overhead and limited validation.
[39] Mahmood et al.	Dynamic Trust Framework	IoT & 6G	Adaptive trust scoring for edge devices.	Limited zero-day and device variability analysis.
[40] Iftikhar et al.	Edge-centric Smart City Security	Smart City IoT	Integrates SDN, Edge Computing, and Zero Trust.	Controller dependency and interoperability issues.
[41] Díaz et al.	Zero Trust Token Authorization	Distributed Firewall	Token-based scalable Zero Trust authorization.	No AI integration or trust evaluation.

[42] Peepliwal et al.	Blockchain-ZTA Framework	Decentralized Clinical Trials	Combines Blockchain, Zero Trust, and Eigen Trust-PBFT.	High overhead and limited insider threat analysis.
[43] Han et al.	HGFF	Wireless Sensor Networks	Optimizes mobile sink routing using DRL and graphs.	Limited adaptability to topology changes.
[44] Chai et al.	Few-shot Incremental Learning	Malware Detection	Dynamic prototype-based incremental malware learning.	Computationally expensive and dataset imbalance issues.
[45] Djaidja et al.	Attention-based RNN NIDS	Network Intrusion Detection	Enables early intrusion detection using RNN with attention.	High complexity and limited zero-day generalization.
[46] Zha et al.	Adaptive NIDS	Network Intrusion Detection	Adaptive real-time intrusion detection for dynamic networks.	Depends on synthetic data and limited zero-day evaluation.
[47] Yang et al.	Open-set IIoT IDS	Industrial IoT	Lightweight open-set intrusion detection for unknown attacks.	Concept drift and lack of industrial validation.

The detailed organization of the paper is outlined here. Section I is the introduction and includes a review of the related literature. Section II analyses and summarizes the gaps in the present scenario of cyber security in Industrial Internet of Things in CII setup. Section III outlines the proposed framework. Section IV represents the research methodology with mathematical modelling. Section V demonstrates the experimental setup, evaluation procedure and results. Section VI represents Comparative performance analysis with some existing techniques followed by conclusion in Section VII.

II. GAP ANALYSIS

Based on the literature surveyed it is being observed that security is a predominant challenge in the CII infrastructure composed of IIoT. The spread of attack vector and increase of attack surface causes high vulnerability causing threat to CII infrastructure. The present countermeasure techniques of these attacks are limited and is challenging due to employment of conventional signature-based detection and protection techniques and not being adaptive to the varying attack patterns of the malware, polymorphic attack, and distributed denial of service attacks. The major gaps identified are detailed as follows:

1. Poor detection of polymorphic malware.
2. High false positive rate effecting operation continuity.
3. Lack of adaptive trust evaluation mechanisms.
4. Absence of dynamic integration between detection systems and zero trust security.
5. Insufficient real world experimental validation in IIoT CII scenarios.

The gaps identified is being investigated for the security of the IIoT devices from the sophisticated attacks in this research. The proposed framework is tested and validated with the regress evaluation with the methods detailed in upcoming sections.

III. PROPOSAL

The gap analysis revealed that CII based industrial security structure is altogether different from traditional cyber security. To understand the CII based infrastructure the paper inducted Purdue Enterprise

Reference Architecture (PERA) depicted in Fig. 2 for a city water utility. The PERA architecture helps in understanding the gaps observed and the suggesting the proposal. Based on the PERA architecture it being observed that IIoT installation is a perimeter less installation where attack surface is large and attack vector is predominant.

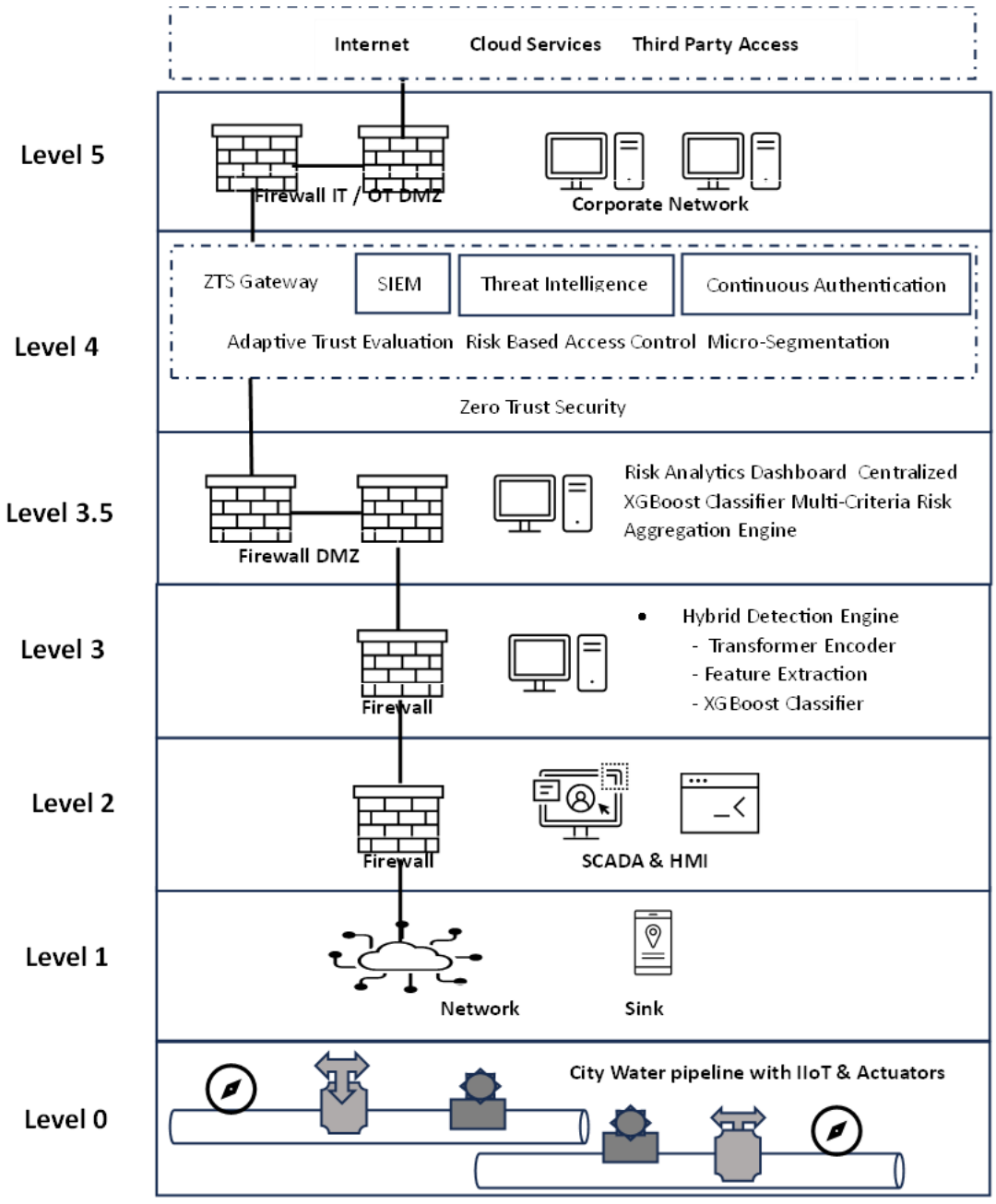


Fig.2 Proposed PERA Diagram

To cater such threats with large attack surface, need an artificial intelligence (AI) driven framework which should be capable of detecting and protecting the sophisticated threats mentioned in the gap analysis. This paper proposes a novel framework HDE-IDS which is a combination of Deep Neural Networks (DNN) and Ensemble (XGBoost) based Intrusion Detection System with Dynamic Zero Trust security with IIoT environment. The proposed framework experimentally:

1. Design multi-criteria feature extraction
2. AI-driven malware detection.

i. Deep Neural Networks
ii. Transformer based sequence modelling
iii. Ensemble learning methods (XGBoost)

3. Dynamic ZTS enforcement that
   i. Continuously recalculates device trust scores
   ii. Dynamically adjusts access privileges.
   iii. Applies micro-segmentation to restrict lateral movements
4. Experimental validation.
5. Comparison of performance with existing framework.

IV. RESEARCH METHODOLOGY

The proposed HDE-IDS framework is designed to protect Industrial IoT (IIoT) devices against sophisticated malware attacks. The IIoT system continuously monitors parameters and enabling the concern authority to perform real-time monitoring, analytics, predictive maintenance, and operational control. The proposed methodology follows an experimental research design.

The core objective is to design an HDE- IDS framework where:

1. A transformer encoder which is being used to extract temporal behavioural representations from IIoT network data.
2. An XGBoost classifier performs final malware detection using deep latent features.
3. A multi-criteria risk model dynamically updates Zero Trust enforcement decisions.

The proposed research methodology proposed to secure the IIoT installation using HDE- IDS framework. HDE- IDS framework is the combination of Deep Neural Networks (DNN), and Ensemble (XGBoost) and is depicted in Fig. 3. The proposed framework undergoes offline training, provide online inferences, and is based on adaptive risk score model. The HDE-IDS integrated with zero trust based adaptive security for real time IIoT protection.

Fig. 3 depicting the proposed research methodology, where malware attack detection being done by integrating advance deep learning, ensemble machine learning, feature engineering, risk aware trust evaluation, and adaptive feedback learning for continuous model improvement. The proposed framework incorporates both offline and online inference stages to support data generalization, real-time threat detection and adaptive security decision using zero-trust enforcement.

A. Dataset Used

Two Dataset being used in this paper, for offline IIoT traffic CICIoT2023 dataset being used and for online TON_IoT dataset is being used. Here the primary training dataset is CICIoT2023, which contains various attack categories and being used to train Transformer and XGBoost. TON_IoT dataset is being used for the online inferences and generalisation of the trained model.

The CICIoT2023 dataset contains 33 different attack variations that have been classified into 07 standard categories namely DDoS, DoS, Reconnaissance, Brute Force attacks, Web-Based attacks, Spoofing attacks, and Mirai Botnet Variants along with benign traffic. Table 2 shows 10 attack type that are representative and are chosen based on rank of count and percentage from the CICIoT2023 dataset.

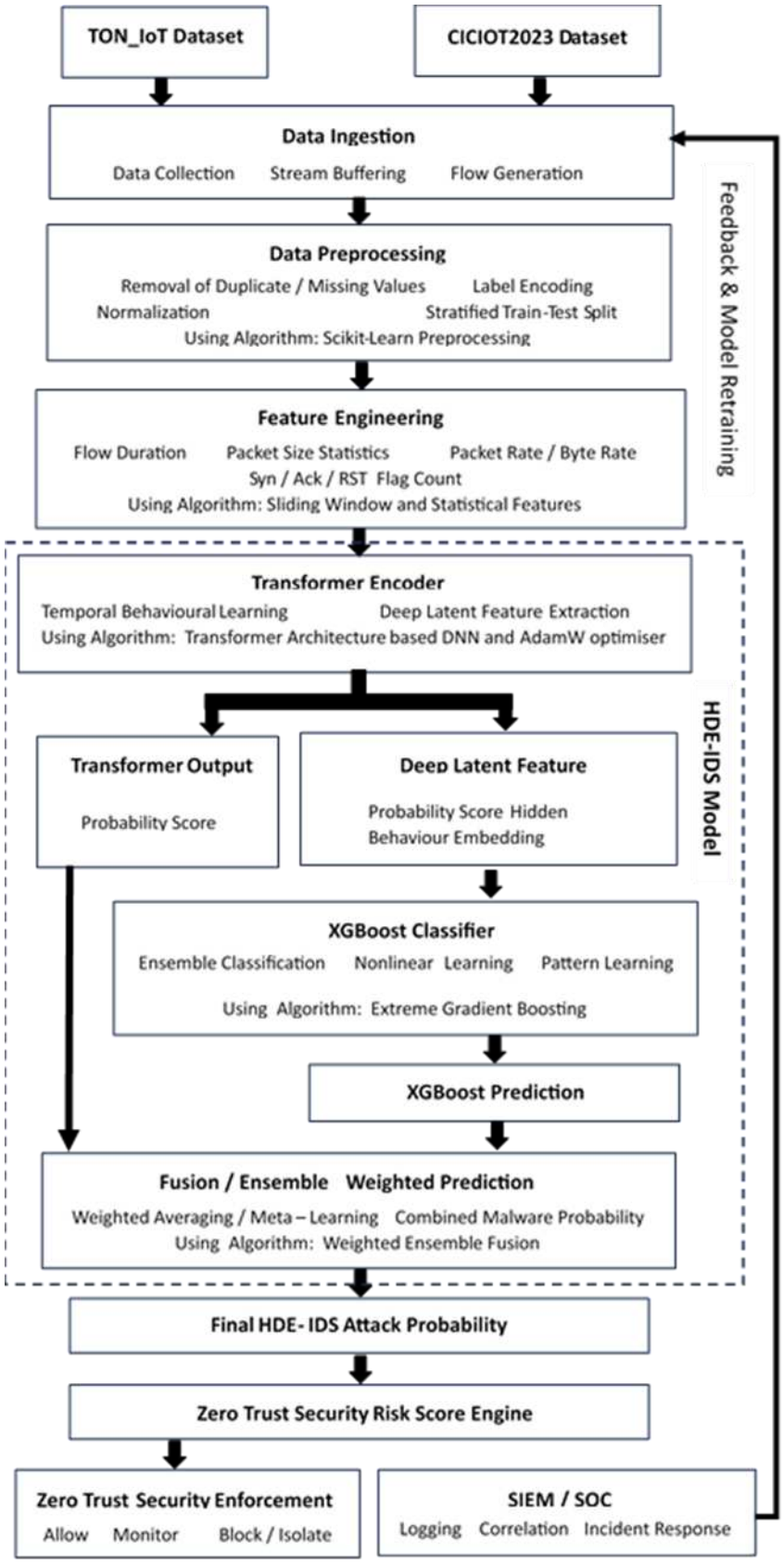


Fig.3 Proposed Research Methodology

Table 2 Attack Representatives from CICIoT2023 dataset

Rank	Label	Count	Percentage (%)
1	DDoS-UDP_Flood	1650000	9.5
2	DDoS-TCP_Flood	1520000	8.8
3	DDoS-ICMP_Flood	1430000	8.2
4	DDoS-SYN_Flood	1320000	7.6
5	DoS-UDP_Flood	1180000	6.8
6	DoS-TCP_Flood	1060000	6.1
7	DoS-SYN_Flood	970000	5.6
8	PortScan	890000	5.1
9	Merai-Griep Flood	760000	4.4
10	Merai-UDP Plain	690000	4.0

During training, each network flow record from training dataset of CICIOT2023 is transform into a structured feature vector. These structured feature vectors are arranged as temporal sequence of network flows, allowing the Transformer encoder to learn sequential traffic behaviour patterns. The encoder applies the self-attention mechanism, which evaluates relationships between different flows in the sequence and assigns importance weights to each element. This enables the model to identify abnormal traffic bursts, coordinate attack behaviour, or unusual communication interval which are common in IIoT malware attacks.

To evaluate the HDE-IDS framework for city water utility IIoT setup, the TON_IoT dataset is incorporated to simulate online IIoT traffic and device telemetry data. The TON_IoT dataset used to check the effectiveness of the framework in real world scenarios. The dataset contains sensor logs, operating system events and network traffic generated from IIoT devices. The TON_IoT dataset are also process through the same line – Data ingestion, Pre-processing, Feature engineering, Transformer embedding, and XGBoost classification before the hybrid prediction converted into the risk score within the Zero trust security.

B. Data Ingestion

In this stage of research methodology data is being collected from multiple sources like CICIoT2023 dataset for offline training and TON_IoT dataset for online inferences. SIEM (Security Information and Event Management) and SOC (Security Operation Centre) telemetry data is also collected also collected at this stage. The telemetry data contains firewall logs, network flow, IPS/IDS alerts, authentication logs, Endpoint Detection logs, IIoT devices telemetry and Switch/Router logs.

Data ingestion stage is act as data acquisition layer of the framework and implemented using CSV based batch loading for CICIoT2023 and TON_IoT. Packet capture mechanism for Network flow traffic and streaming mechanism for SIEM / SOC telemetry data. Finally, all captured data from CICIoT2023, TON_IoT and SIEM / SOC telemetry data for ETL (Extract, Transform and Load) for further storage and processing.

C. Data Pre-processing

This stage of research methodology, data pre-processing is important because of data preparation and cleansing. Pre-processing is required to improve the data quality of the data and model training. In the data pre-processing missing value handling, duplicate data handling, normalisation, and noise removal being

performed. In this process Scikit-learn being used for the feature distribution for make data suitable for deep learning and machine learning models. To make model unbiased for majority attack class and rare attack class get ignored data is also prepared for class imbalance weights calculated and applied before model training.

The Data Pre-processing stage is also performing data transformation and normalization which shall be utmost important for HDE- IDS framework. Data is normalized for further processing and needs to be statistically unbiased. The Data pre-processing ensure consistency of the data, temporal integrity, and processing readiness of IIoT traffic before model training and inference. Let the ingested dataset be represented as:

$$D = \{(x_i, y_i, t_i)\}_{i=1}^{N} \quad (1)$$

Eq. (1) defines the dataset consisting of N samples, where $x_i \in \mathbb{R}^d$ is the d-dimensional feature vector (like flow, telemetry) of i^{th} network flow, $y_i \in Y_i$ is the class label (attack/benign), and t_i is the timestamp.

a) Missing Value handling

To address incomplete observations, mean imputation is applied. Let x_{ij} be the j^{th} feature of sample i. If x_{ij} is missing, common imputation is

Where $\Omega_j = \{i = x_{ij} \text{ is observed}\}$. (2)

- ***Median Imputation***

$$x_{ij} \longleftarrow \text{median } \{ x_{ij} : i \in \Omega_j \} \quad (3)$$

Eq. (3) replaces missing feature values with the median of that feature's observed values.

b) Duplicate Removal

Let s_i, denote a unique flow signature. The cleaned dataset is defined as

$$D' = \{(x_i, y_i, t_i) \in D: s_i \notin | \{s_k\}_{k<1}\} \quad (4)$$

Eq. (4) remove duplicated flow records based on signature uniqueness.

c) Temporal Ordering (for sequence learning)

$$D_{sorted} = \text{sort } (D', \text{ by } t_i \quad (5)$$

Eq. (5) ensures temporal coherence required for Transformer based sequence learning.

d) Feature Normalization

To prevent magnitude dominance across heterogeneous features, Z-score standardization is applied.

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (6)$$

Eq. (6) standardizes feature j to zero mean and unit variance.

$$\sigma_j = \sqrt{\frac{1}{|\Omega j|-1} \sum_{i \in \Omega j} (x_{ij} - \mu_j)2} \quad (7)$$

Eq. (7) computes the standard deviation of feature j.

Alternatively, Min-Max scaling is defined as:

$$\overline{x_{ij}} = \frac{x_{ij} - \min(x.j)}{\max(x.j) - \min(x.j)} \quad (8)$$

Eq. (8) rescales features to the interval [0,1].

e) Categorical Encoding

For categorical variance c with K classes,

A categorical feature c ∈ {1, ……., K}, one-hot vector e(c) $\in \{0,1\}^K$

one hot encoding is:

$$e_k\,(c) = \begin{cases} 1 & if\ k = c \\ 0 & otherwise \end{cases} \qquad (9)$$

Eq. (9) transform categorical attributes into binary vector suitable for machine learning models

f) Sequence Construction (Transformer Input)

Multi-level label mapping (fine ⟶ coarse ⟶ binary)

If $\mathcal{Y}_{fine}$ is the set of 33 attack type plus benign, define mapping:

$$g: \quad \mathcal{Y}_{fine} \longrightarrow \mathcal{Y}_{coarse}$$

Example of coarse classes:

$\mathcal{Y}_{coarse}$ = {Benign, DDOS, DOS, Recon, Brute Force, Web, Spoofing, Mirai}

Binary Mapping:

$$h(y) = \begin{cases} 0 & y = Benign \\ 1 & y \in \mathcal{Y}_{fine} \setminus \{\text{Benign}\} \end{cases}$$

For time series modelling, create sliding windows of length L with stride s.

Let the order samples be $\{x_t\}_{t=1}^{T}$. Then the k-th sequence is:

$$S_k = [x_t, x_{tk+1}, \ldots\ldots\ldots x_{tk+L-1}] \in \mathbb{R}^{Lxd} \qquad (10)$$

Eq. (10) defines the k^{th} input sequence used for transformer training.

While the window index is computed as:

$$t_k = 1 + (k-1)s \qquad (11)$$

Eq. (11) determines the starting position of each window using stride s.

If label is per-window, common choice is:

Last-event label

$$y_k = y_{tk+L-1} \qquad (12)$$

Eq. (12) assign the sequence label using the last flow in the window.

If label is assigned the class that appears most frequently with in a sequence window the common choice is:

Majority Label

$$y_k = argmax_{c \in \mathcal{Y}} \sum_{r=0}^{L-1} 1(y_{t_{k+r}} = c) \qquad (13)$$

Eq. (13) shows y_k is the label of sequence k, and c is the candidate class in label set $\mathcal{Y}$, and the indicator function counts how many times class c appears in the window of length L.

g) Class Imbalance weighing

To mitigate imbalance, class weights are defined as

$$w_c = \frac{N}{K.\ Nc} \tag{14}$$

Eq. (14) assigns inverse-frequency weights, where N be the total samples, K is total number of classes, and Nc is the number of samples in class c.

Weighted cross-entropy (multi-class) loss is expressed as:

$$£ = -\sum_{i=1}^{N} w_{y_i} log p\theta(y_i \mid x_i) \tag{15}$$

For binary

$$£ = -\sum_{i=1}^{N} [w_1 y_i\, log p_i + w_0(1 - y_i)\log(1 - p_i)] \tag{16}$$

Eq. (16) penalizes misclassification of minority classes more strongly.

D. Feature Engineering

Sophisticated frameworks like HDE-IDS required to understand temporal dependencies like current packet dependence on previous packet over time scale or sequence for better analysis and attack detection. This stage of research methodology usually transformed pre-processed traffic data into meaningful security related features such as flow duration, packet size statistics, packet rate, byte rate, TCP flags such as SYN, ACK, RST, protocol encoding and communication frequency patterns.

This feature engineering stage deployed sliding window sequence generation to transform network traffic into sequential time series presentations as the HDE-IDS framework required to learn temporal dependencies. After the sequential time series presentations, statistical feature extraction and temporal sequence generation techniques are used to convert the discriminative representations that improve the model performance and reduce redundancy. Let the cleaned dataset obtained after pre-processing be defined as:

$$\mathrm{D} = \{(x_i, y_i)\}_{i=1}^{N} \tag{17}$$

Eq. (17) define the dataset consisting of N samples where $x_i \in \mathbb{R}^{\mathrm{d}}$ is the feature vector, $y_i \in \mathrm{Y}$ denotes associated class label.

The feature engineering stage develop an optimized subset of informative features:

$$\mathrm{F}^* = \{ f_1, f_2, \ldots\ldots, f_m \} \qquad \mathrm{m} < \mathrm{d} \tag{18}$$

Eq. (18) represents the reduced feature set F^* selected from the d features to improve predictive efficiency.

a) Statistical Feature Construction

For network traffic flows containing n packets with packet size values $p_1, p_2, \ldots\ldots p_n$, the mean packet size is computed as:

$$\mu_p = \frac{1}{n}\sum_{i=1}^{n} p_i \tag{19}$$

Eq. (19) calculate the average packet size of a network flow.

The variance of packet size is defined as

$$\mu_p{}^2 = \frac{1}{n}\sum_{i=1}^{n} (p_i - \mu_p)^2 \tag{20}$$

Eq. (20) measures the variability of packet sizes within the flow.

b) Feature Relevance Estimation

Feature importance can be evaluated using Information Gain based on entropy reduction

$$\text{IG}(f_j) = \text{H}(\text{Y}) - \text{H}(\text{Y} \mid f_j) \quad (21)$$

Eq. (21) measures how much knowledge of feature f_j reduces uncertainty in the class label Y.

The Entropy of class label is defined as

$$H(Y) = -\sum_{c \in Y} P(c) log_2 \text{P(c)} \quad (22)$$

Eq. (22) quantifies the uncertainty associated with class distribution.

The conditional Entropy after observing feature f_j is

$$H(Y \mid f_j) = -\sum_{v \in V_j} P(v) \sum_{c \in Y} P(c \mid v) log_2 P(c|v) \quad (23)$$

Eq. (23) measures the remaining uncertainty in the class labels after splitting on feature f_j.

Feature representation defined as:

$$\text{X} = [\, x_1, x_2, \ldots\ldots, x_N] \in \mathbb{R}^{N \times m} \quad (24)$$

Eq. (24) represents the engineered feature matrix.

c) Ensemble - Based feature Importance

For tree-based ensemble models like XGBoost, feature importance is calculated based on the reduction in loss when a feature is used for splitting.

$$Gain(f_j) = \sum_{k=1}^{K} \Delta L_k \quad (25)$$

Eq. (25) computes the cumulative loss reduction contributed by feature f_j across K decision tree splits.

So, the feature engineering stage derives statistical descriptors, evaluates feature relevance using Entropy based metrics, and selects an optimized feature subset based on importance scores. This process enhances detection performance while reducing feature redundancy and computational complexity in framework.

E. Transformer Encoder

The extracted sequential features are input to the Transformer Encoder, which is primarily deep learning component of the HDE-IDS framework. The Transformer Encoder architecture is a Transformer-based Deep Neural Network (DNN) consist of multi head self-attention layers and feed forward neural networks. Its primary objective is to learn temporal behavioural representations, sequential attack dependencies, communication patterns and deep latent features from the IIoT traffic. At the time of training, it uses AdamW optimiser with backpropagation and binary cross-entropy loss minimisation.

The Transformer learns these patterns through supervised learning, where each sequence is associated with a known class label (i.e., benign or attack). The model produces a predicted probability for each class, and the difference between the predicted output and true label is measured using a cross-entropy loss function. The network parameters are then updated through gradient optimization algorithm like AdamW, to minimize the classification error. Through repeated training iterations the Transformer encoder gradually learns meaningful temporal representations of IIoT traffic that help distinguish benign traffic from malicious one.

After training, the Transformer encoder generates contextual embedding vectors representing the temporal behaviour of each traffic sequence. These embeddings captures both short term and long-range dependencies among network flow, making them highly effective for identifying distributed or stealthy attacks. The learned embeddings are then forwarded to the XGBoost ensemble classifier, which performs final attack type detection and classification and contributes to the adaptive risk scoring mechanism used by the zero-trust security.

To capture temporal dependences, the dataset is converted to into sequences as:

$$S_k = [\, x_{t_k}, x_{t_{k+1}}, \ldots\ldots\ldots, x_{t_{k+L-1}}] \quad (26)$$

Eq. (26) defines the k^{th} input sequence with window length L.

The sequence is converted into embedding vectors.

$$E_k = \text{Embedding } (S_k) \quad (27)$$

Where E_k is Embedded sequence representation.

Then positional encoding is added as $\grave{E}_k$

$$\grave{E}_k = E_k + \text{PE} \quad (28)$$

Where PE is positional encoding and helping the encoder to understand order of the packets.

In Transformer process, embedded sequence is projected into three metrices i.e., Query (Q), Key (K), and Value (V) to learn relationship between different elements in sequence.

$$\text{Q} = \grave{E}_k W_Q \quad (29)$$

$$\text{K} = \grave{E}_k W_K \quad (30)$$

$$\text{V} = \grave{E}_k W_V \quad (31)$$

Q represents what information does the model is looking from the previous packet

K represents what information does previous packet carrying

V represents what actual information is

Now the self-attention mechanism is applied by the Transformer to determine which particular previous packets are important. The self-attention mechanism depicted as

$$\text{Attention (Q, K, V)} = \text{softmax } (\frac{QK^T}{\sqrt{d_k}})\text{V} \quad (32)$$

Eq. (32) computes the scaled dot-product attention used by the Transformer encoder.

Where QK^T called first similarity score and measures relation between current packet and previous packets and $\sqrt{d_k}$ called scaling factor used to normalize large values. While Softmax is a function used to convert similarity scores between packets (from Q and K) into normalized attention weights. It helps the model to know which previous packet is important enough to understand the current packet.

Transformer output depicted as:

$$h_k = \text{Transformer } (S_k) \quad (33)$$

Eq. (33) generates the temporal embedding vector h_k.

F. Deep Latent Feature Extractor

Transformer encoder generates the deep latent feature representations or hidden behavioural patterns after processing of sequential traffic data. These latent embeddings able to capture complex attack characteristics due to malware activity, abnormal communication patterns, and temporal dependencies.

These learned embeddings served as compressed high-level representations for secondary classification at XGBoost classifier. The temporal embedding vector h_k become the deep latent feature Z.

$$Z = (h_k) \tag{34}$$

and p_{deep} is the deep model probability score. The deep model probability score is defined as

$$p_{deep} = \sigma\,(\mathrm{W}h_k + \mathrm{b}) \tag{35}$$

Where $\mathrm{W}h_k$ is the trainable weight matrix while b is the bias, and σ is the weight parameter and sigmoid function (used to convert normal output to probability value between 0 and 1, usually it is 0.5).

G. XGBoost Classifier & Prediction

The deep latent feature vector Z extracted by the deep latent feature extractor is used as input to the XGBoost ensemble classifier. The XGBoost is a supervised machine learning ensemble algorithm based on gradient boosted multiple decision trees. It learns nonlinear decision boundaries from the Transformer latent features and provide reliable attack type detection and classification by optimising classification loss iteratively through boosting. This stage improves attack discrimination capability, reduces classification bias, and enhances generalization across heterogeneous datasets. The output from the XGBoost classifier is

$$\hat{y}_k = \sum_{m=1}^{M} f_m\,(Z) \tag{36}$$

Eq. (36) represents the aggregated score from M decision trees and f_m (Z) is the is the output of each decision tree.

The XGBoost probability score is defined as

$$p_{XGBOOST} = \sigma\,(\textstyle\sum_{m=1}^{M} f_m(Z)) \tag{37}$$

The transformer layer learns temporal representations of network traffic, while the XGBoost layer performs robust ensemble classification using the learned embeddings. The combination of deep sequential learning and gradient boost decision trees enhances detection accuracy and generalization capability for the IIoT malware detection.

H. Fusion / Ensemble

The Fusion and Ensemble layer integrates the prediction from the Transformer based deep learning model and the XGBoost ensemble classifier to generate a robust a reliable decision score for IIoT malware detection. The Fusion mechanism combines these predictions to improve detection accuracy and reduce false positives. The ensemble mechanism improves model robustness through multiple learners, while the fusion mechanism integrates heterogeneous model outputs to leverage complementary predictive capabilities. The weighted Fusion mechanism allows the framework to balance contributions from both the models. The hybrid prediction or fused prediction is computed using weighted fusion as

$$p_{hybrid} = \alpha\, p_{deep} + (1 - \alpha)\, p_{XGBOOST} \tag{38}$$

Eq. (38) combines the probability scores from the deep model and XGBoost classifier.

I. Final HDE-IDS Prediction

The hybrid prediction or attack prediction p_{hybrid} provide by fusion / ensemble stage is final prediction of the HDE-IDS framework. From p_{hybrid} final classification decision $\hat{y}$ is calculated to understand whether incoming traffic is malicious or benign as

$$\hat{y} = \begin{cases} 1 & p_{hybrid} \geq \tau \\ 0 & p_{hybrid} < \tau \end{cases} \tag{39}$$

Eq. (39) assigns the input traffic instance to the malicious class when the hybrid probability exceeds classification threshold τ which is usually 0.5.

J. Zero Trust Security Risk Score Engine

The hybrid prediction or attack prediction p_{hybrid} generated by the framework is then forwarded to Risk score engine of the Zero trust security, which dynamically calculate risk scoring R of the incoming traffic. Risk scoring is performed using adaptive threshold parameters θ_1 and θ_2. The risk scoring engine integrates multiple decision criteria, including deep learning predictions, ensemble model outputs, device trust levels, contextual behaviour indicators and anomaly severity indicators to calculate risk score.

Let the individual decision components be defined as:

1. p_{deep} – probability from the Transformer model
2. p_{ens} - probability from the XGBoost Ensemble model
3. T_d – device trust score
4. A_s – anomaly severity score
5. C_t – contextual behaviour score

The adaptive criteria fusion function defined as:

$$R = w_1 p_{deep} + w_2 p_{ens} + w_3 (1 - T_d) + w_4 A_s + w_5 C_t \quad (40)$$

Eq. (40) combines multiple security indicators into a unified risk score R.

The contribution weights are constrained by

$$\sum_{i=1}^{5} w_i = 1 \quad (41)$$

Eq. (41) ensures that weights are normalized and represent proportional influence.

The contextual behaviour score can be calculated from statistical traffic deviation.

$$C_t = \frac{|x - \mu|}{\sigma} \quad (42)$$

Eq. (42) measures deviation of observed traffic from expected normal behaviour.

Device trust score is computed using historical behaviour statistics

$$T_d = \frac{N_{benign}}{N_{total}} \quad (43)$$

Eq. (43) represents the reliability of the device based on the past benign interactions.

The threshold risk θ used to determine the security action decisions. The ROC based optimal threshold approach is being used to calculate θ. The threshold is selected to maximise the difference between the True Positive Rate (TPR) and False Positive Rate (FPR).

$$\theta^* = \arg\max_{\theta} (TPR(\theta) - FPR(\theta)) \quad (44)$$

Eq. (44) selects the threshold that maximizes the detection capability while minimizing the false alarms.

Where

$$\text{TPR} = \frac{TP}{TP+FN} \tag{45}$$

Eq. (45) represents the probability of correctly detecting the malicious traffic.

$$\text{FPR} = \frac{FP}{FP+TN} \tag{46}$$

The system determines the security action using risk thresholds

$$Decision = \begin{cases} Allow, & R < \theta_1 \\ Monitor, & \theta_1 \leq R < \theta_2 \\ Block, & R \geq \theta_2 \end{cases} \tag{47}$$

- Where θ_1 and θ_2 are derived using ROC
- Where θ_1 is a lower risk threshold (monitor boundary)
- And θ_2 is an upper risk threshold (block boundary)

Eq. (48) defines the adaptive Zero Trust decision rule.

K. Zero Trust Security Enforcement

Based on the input provided by risk score engine through θ_1 and θ_2 to zero trust security policy engine(PE) of the zero-trust security. The updation of PE enable policy administrator (PA) to enforce the decision of either block, monitor or allow through policy enforcement point (PEP). The zero trust security enforcement actions are based on the input provided by the HDS-IDS framework.

L. SIEM / SOC

Security telemetry for SIEM / SOC continuously feed the data ingestion stage along with all other data. SIEM / SOC feed provides data like firewall log, security logs, and endpoint detections provide additional contextual indicators for threat detection. The SIEM / SOC feed integration provides network threat visibility and network monitoring. SIEM / SOC passively collecting the information of the IIoT based CCI infrastructure and updating to the HDS-IDS framework with feedback with actual network events and incidents.

M. Feedback and Model Retraining

At this final stage of the research methodology, new observed zero-day attacks, new malware detected, and incidents reported from SOC are forwarded to HDE-IDS framework through a continuous feedback loop for retraining of the Transformer encoder and XGBoost classifier. This continuous learning from the continuous occurring events enable the framework to adapt dynamically to new threats and improved upon.

V. EXPERIMENTAL SETUP

The proposed framework HDE-IDS has been also evaluated through simulation experiment which emulate realistic IIoT benign network traffic and malicious attacks.

A. Technical Setup

The HDE-IDS framework setup implemented on a workstation having configuration of Intel Core i7 processor, 32 GB RAM with Ubuntu Linux operating system. The development environment is Jupyter Notebook and python 3.x. Transformer encoder model developed using PyTorch and the ensemble classification components implemented using XGBoost. Data pre-processing, feature engineering and performance evaluation is performed using NumPy, Scikit-learn and Pandas.

B. Dataset Preparation

The experimental evaluation utilizes the CICIoT2023 dataset, which contains labelled IoT network traffic containing both benign behaviour and multiple cyber-attack categories. The CICIoT2023 dataset is divided into three subsets – training set, validation set, and test sets for ensuring each subset carries a similar distribution of benign and malicious traffic. Table 3 shows the data split with percentage.

Table 3 CICIOT2023 Data Split

Dataset	Split in Percentage (%)
Training	70
Validation	15
Test	15

The CICIoT2023 dataset includes multiple attack categories like Distributed Denial of Services (DDoS), Mirai botnet attacks, reconnaissance activities, and other IIoT specific threats. Each record represents a network flow characterised by statistical features such as packet rate flow duration, byte count and protocol information. The TON-IoT dataset is being used as independent dataset to evaluate the generalization capability of the proposed framework HDE-IDS under realistic IIoT conditions. The TON-IoT is also used to simulate the online traffic behaviour and sensor telemetry data. This dataset contains telemetry logs, operating system traces and network traffic collected from IIoT.

C. Model Training Setup

The Transformer encoder is trained to learn temporal representations of network traffic sequences derived from CICIoT2023 dataset. Each sequence contains multiple network flows arranged chronologically. The model parameters are optimised using the AdamW gradient optimization algorithm, which provides stable convergence and improved generalization for deep neural networks. The XGBoost classifier receives the Transformer embeddings as input features and performs the attack type detection and classification. The ensemble model uses gradient boosting decision trees to learn nonlinear relationships among features. The final prediction of HDS-IDS framework is based on combined probability score of the Deep latent feature score and XGBoost prediction score.

D. Evaluation Procedure

The performance of the HDE-IDS framework is evaluated using Accuracy, Precision, Recall, F1-score, and area under the ROC curve (AUC). These metrics measure the ability of the system to correctly identify malicious traffic while minimizing false alarms. Let TP, TN, FP, and FN represent true positives, true negatives, false positive, and false negative, respectively.

Accuracy is calculated as

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision is defined as

$$Precision = \frac{TP}{TP+FP}$$

Recall is calculated as

$$Recall = \frac{TP}{TP+FN}$$

The F1-score which balances the Precision and Recall is given by

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

The ROC-AUC metric is also used to evaluate the model's ability to discriminate between benign and malicious traffic across different threshold values.

E. Results and Analysis

The proposed HDE- IDS detection system evaluated with TON_IoT Dataset. The approach adopted in this model is to integrate both Transformer and XGBoost models which improves the detection performance compared to standalone performances. The Transformer Encoder extracts high level temporal embeddings, while XGBoost refines attack type detection and classification using gradient boosting.

The performance of the experiment is excellent, and the model detection performance observed is high across all evaluation metrics. Refer to the Fig. 4 & Fig. 5 the overall accuracy reached 99.1% with a precision of 99.0 % and recall of 99.2%. The F1-score is high and is of 99.1%. This high F1-score achieved by the model shows that it effectively minimizes both false positive and false negative. The ROC-AUC value of 0.999 ensure excellent separability between benign and malicious traffic.

As the value of ROC-AUC very close to unity, it confirms that the model able to maintain the very strong attack type detection and classification thresholds across varying decision boundaries, making it suitable for adaptive risk-based decision in critical IIoT system.

For CII IIoT industrial implementation, low false positive rate is desire to minimise false alarms and generate unnecessary alerts. The confusion matrix analysis shows that the model able to achieve low false positive rate which is good for IIoT based CII infrastructure. Moreover, the proposed HDE-IDE detection system enhances robustness by reducing variance and bias simultaneously.

This result in improved generalization performance, and evident in the TON-IoT evaluation. The predicted probabilities from the model are transformed into a dynamic risk score, which is being used by the ZTS engine to enforce security policy. The dual threshold (θ_1, θ_2) used in the risk score create a buffer zone (uncertainty). θ_1 controls false positive while $\theta_2$controls false negative. The adaptive thresholding mechanism ensures that security decisions are context aware and dynamically adjusted and fine-tune based on traffic behaviour, model confidence.

```
Confusion Matrix Parameters:
TP = 496
TN = 495
FP = 5
FN = 4

Performance Metrics:
Accuracy  : 99.10%
Precision : 99.00%
Recall    : 99.20%
F1 Score  : 99.10%
ROC-AUC   : 0.9991
```

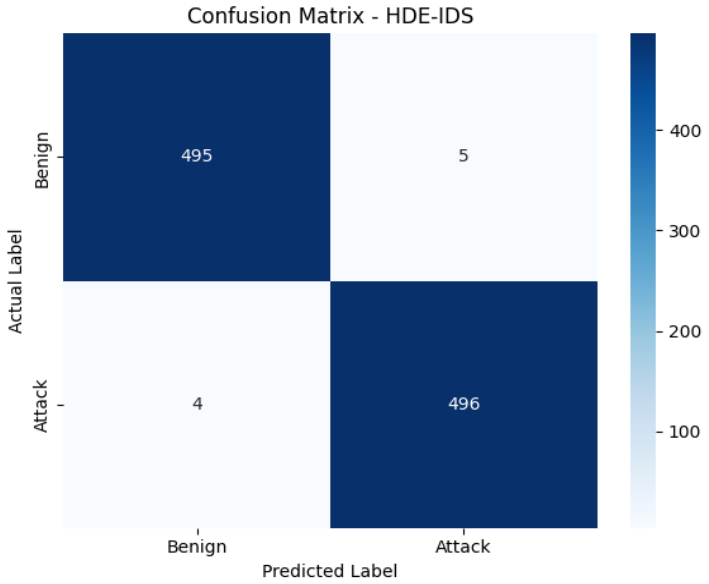
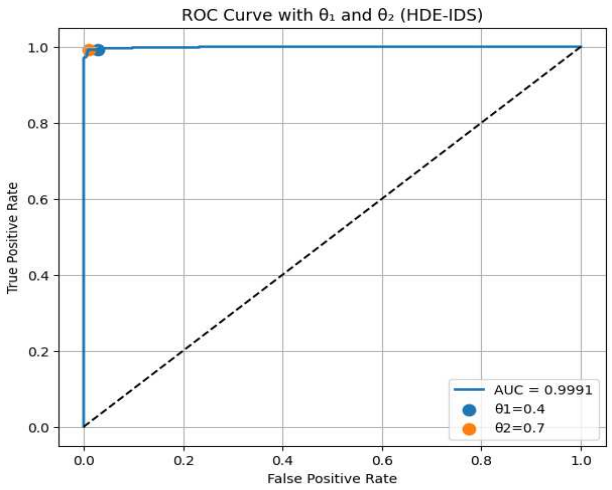


Fig.4 Confusion Matrix for proposed HDE-IDS




```
Confusion Matrix:
 [[495   5]
 [ 4 496]]

Metrics:
Accuracy  : 99.10%
Precision : 99.00%
Recall    : 99.20%
F1 Score  : 99.10%
AUC       : 0.9991

Thresholds:
θ1 = 0.4
θ2 = 0.7
```

Fig.5 ROC Curve for proposed HDE-IDS

The performance of the HDE-IDS for the generalization is evaluated using TON_IoT dataset. The purpose of this evaluation is to generalize live test for IIoT. Refer to the Fig. 6 & Fig. 7 the accuracy achieved is 89.0% with a precision of 89.8% and recall of 88.0 %. The F1-score achieved is of 88.89%. The ROC-AUC value of 0.9586 ensure excellent detection between benign and malicious traffic in the operation mode. The precision and recall value suggested that proposed model balanced false positive and false negative rate which is critical for CII water utility IIoT real-time network. The generalize result validate the proposed HDE-IDS model for city water utility project.

The effectiveness of the proposed HDE-IDS framework is further validated through its deployment in a representative Industrial IoT environment. The deployment integrates an ultrasonic water level sensor, an ESP8266-based edge communication module, the proposed Hybrid Deep Ensemble Intrusion Detection System (Transformer–XGBoost), an adaptive risk scoring and Zero Trust security engine with the real-time monitoring dashboard. This deployment serves as a proof-of-concept for illustrating how the proposed framework can be integrated into Industrial Internet of Things (IIoT) infrastructure to provide continuous monitoring and adaptive cyber defense.

```
==================================
HDE-IDS Generalization Results
==================================
Accuracy  : 89.00%
Precision : 89.80%
Recall    : 88.00%
F1 Score  : 88.89%
ROC-AUC   : 0.9586

Confusion Matrix:
[[450  50]
 [ 60 440]]

θ1 = 0.4
θ2 = 0.7
```

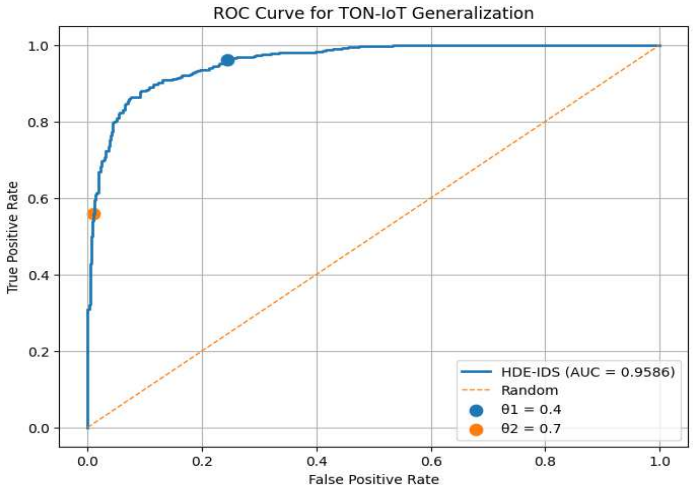
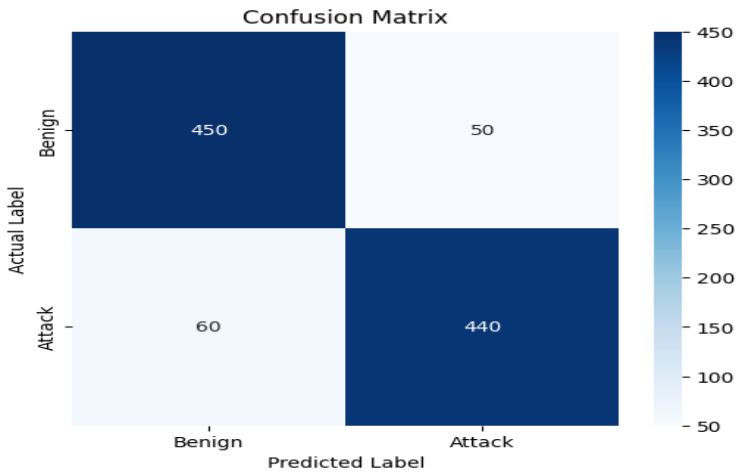


Fig.6 ROC Curve for Generalization of proposed HDE-IDS




```
Accuracy  : 89.00%
Precision : 89.80%
Recall    : 88.00%
F1 Score  : 88.89%
ROC-AUC   : 0.9586

Confusion Matrix:
[[450  50]
[ 60 440]]
```

Fig.7 Confusion Matrix for Generalization of proposed HDE-IDS

The below Fig.8 presents the industrial deployment architecture of the proposed framework. The deployment begins with an ultrasonic sensor integrated with an ESP8266 NodeMCU, which continuously monitors the water level and transmits the acquired measurements to the inference server through HTTP-based wireless communication. Upon arrival at the server, each communication request is processed by the data acquisition module, where the application-layer payload and communication metadata, including request size, arrival timestamp, flow duration, packet rate, inter-arrival time, cumulative traffic statistics, and sensor measurements, are collected to characterize the communication behavior.

The extracted information is subsequently processed by the feature extraction module to generate the input feature representation required by the proposed Hybrid Deep Ensemble Intrusion Detection System. The generated feature vector is analyzed by the Transformer–XGBoost-based HDE-IDS to distinguish between normal and malicious communication patterns before forwarding the prediction to the adaptive risk scoring and Zero Trust policy enforcement modules.

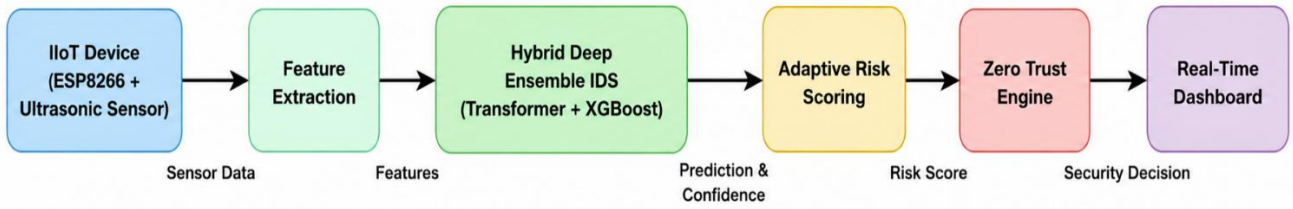


Fig.8 Industrial Deployment Architecture of the Proposed HDE-IDS Framework

The prediction generated by the inference engine is subsequently forwarded to the adaptive risk assessment module, which computes the corresponding risk score. The computed adaptive risk score is further utilized by the Zero Trust engine to assign an appropriate trust level and enforce the corresponding security policy. Depending on the estimated risk, the communication is either permitted to ALLOW, continuously observed as MONITOR, or automatically isolated as BLOCK, enabling dynamic security enforcement without manual intervention.

The proposed deployment architecture establishes a unified cyber–physical monitoring framework by integrating industrial process monitoring with intelligent cybersecurity analytics. Unlike conventional intrusion detection systems that operate independently of process monitoring, the proposed framework continuously correlates sensor observations with network behavior, allowing cyber-attacks to be detected while simultaneously preserving the visibility of the underlying industrial process. This integration provides

an additional layer of resilience for critical infrastructure by combining operational monitoring and adaptive cyber defence within a single architecture.



Fig.9 Real-Time Communication Between ESP8266 Sensor Node and HDE-IDS Inference Server

The real-time communication established between the ESP8266 sensing node and the inference server is presents below in Fig. 9. The serial monitor shows successful wireless communication by continuously transmitting water level measurements and receiving the corresponding inference response generated by the HDE-IDS framework. Along with the measured process parameter, the response includes the predicted traffic category, confidence score, adaptive risk value, device trust level, Zero Trust decision, and Top-3 predicted traffic classes. This bidirectional communication demonstrates the capability of the deployed framework to perform real-time security inference simultaneously with industrial process monitoring.

The operational behavior of the proposed framework under normal communication conditions is illustrated in Fig. 10. During legitimate sensor communication, the incoming traffic is correctly identified as Benign Traffic, resulting in a low adaptive risk score and a HIGH device trust level. Consequently, the Zero Trust engine authorizes communication by enforcing the ALLOW policy, thereby ensuring uninterrupted transmission of sensor information to the monitoring platform. The dashboard further provides confidence values and alternative traffic predictions, allowing operators to continuously assess the security posture of the deployed infrastructure while maintaining complete visibility of the industrial process.

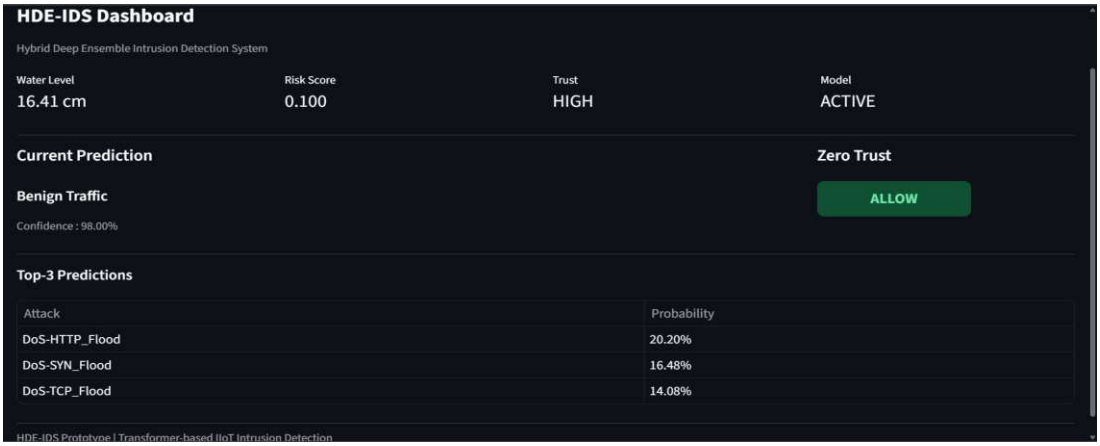


Fig.10 Dashboard Visualization Under Normal Industrial Operation

To evaluate the robustness of the proposed framework under adversarial conditions, an HTTP Flood attack was generated using the attack simulation utility shown in Fig. 11. The simulator continuously transmits malicious requests towards the inference server to emulate a denial-of-service attack against the industrial communication channel. The generated attack traffic serves as the input for validating the adaptive detection capability of the proposed HDE-IDS framework under realistic operating conditions.

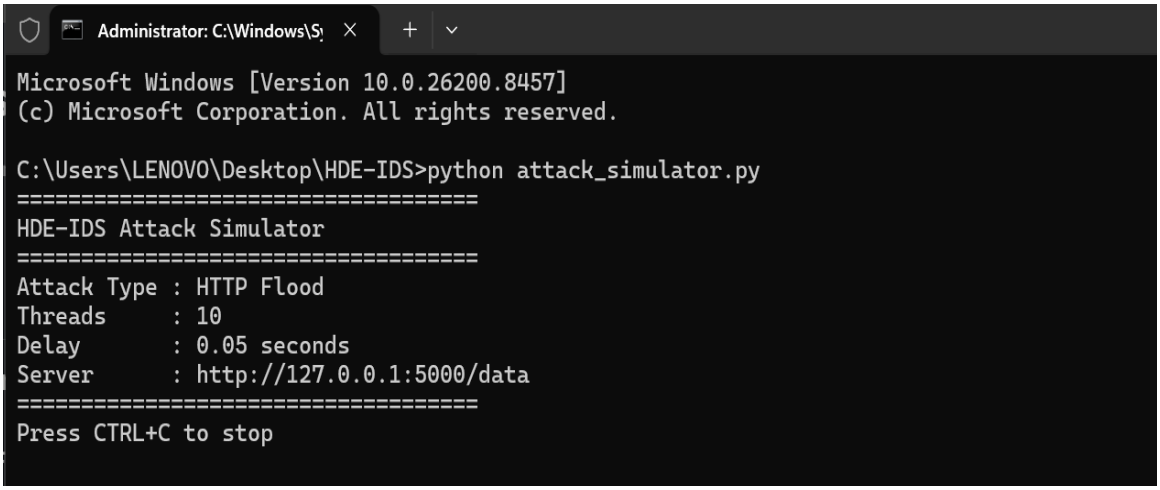


Fig.11 HTTP Flood Attack Simulation for Industrial Validation

The corresponding dashboard response during the attack scenario is presented in Fig. 12. The proposed Hybrid Deep Ensemble model successfully identifies the malicious communication as DoS-HTTP Flood, resulting in a significant increase in the adaptive risk score. Based on the estimated security risk, the device trust level is automatically reduced, and the Zero Trust engine dynamically escalates the security policy to BLOCK, thereby preventing malicious communication from propagating through the industrial network. This coordinated interaction between intelligent intrusion detection, adaptive risk assessment, and policy enforcement demonstrates the autonomous cyber defense capability of the proposed framework for protecting critical IIoT infrastructure.



Fig.12 Dashboard Response During DoS-HTTP Flood Attack Detection

The runtime performance metrics obtained from the deployed prototype are presented in Fig. 13. The observed runtime behavior indicates that the proposed HDE-IDS framework is capable of continuously monitoring industrial communication, performing real-time intrusion analysis, and enforcing adaptive Zero Trust policies while maintaining stable system operation. These observations demonstrate the practical feasibility of integrating intelligent intrusion detection with dynamic access control in resource constrained IIoT environments.

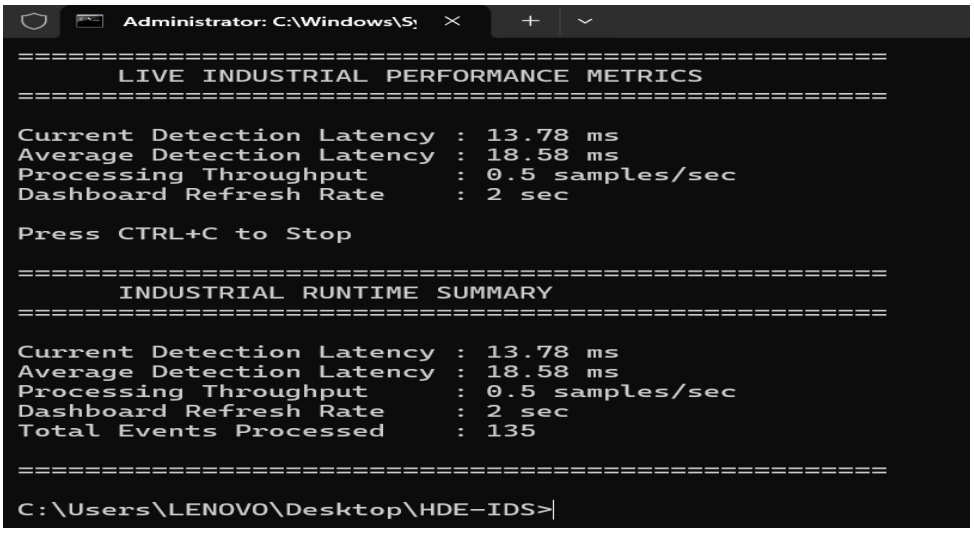


Fig.13 Runtime Performance Metrics of the Deployed HDE-IDS Prototype

The industrial implementation illustrates the integration of the proposed HDE-IDS framework within a representative smart water utility sector, where intelligent intrusion detection, adaptive risk scoring, and Zero Trust policy enforcement operate as a unified cyber–physical security mechanism. The observed operational behavior under both normal and attack scenarios indicates that the proposed architecture can continuously monitor industrial communication while dynamically adapting security decisions according to the estimated risk level. Such a deployment highlights the potential of the proposed framework for enhancing the resilience and operational security of future Industrial Internet of Things infrastructures.

F. Top 10 Attack Detection Performance Analysis

To evaluate the detection capability of the HDE-IDS system, its detection capability in terms of F1 score compared with performance of DNN and XGBoost. The top 10 attack representatives from CICIoT2023 dataset detailed in table 2 being used to evaluate the detection capability of each model. The table 4 shows the detail detection analysis.

Table 4 Top 10 Attack Detection Performance Analysis

Label	DNN	XGBoost	HDE-IDS (Proposed)
DDoS-UDP_Flood	0.982	0.990	0.995
DDoS-TCP_Flood	0.983	0.991	0.996
DDoS-ICMP_Flood	0.982	0.990	0.995
DDoS-SYN_Flood	0.978	0.988	0.993
DoS-UDP_Flood	0.980	0.989	0.994

DoS-TCP_Flood	0.976	0.987	0.992
DoS-SYN_Flood	0.972	0.984	0.990
PortScan	0.981	0.989	0.994
Merai-Griep Flood	0.980	0.988	0.993
Merai-UDP Plain	0.979	0.988	0.993

Table 5 shows the statistical analysis of DNN, XGBoost and HDE-IDS in terms of performance mean F1 score ($\bar{z}$) Standard deviation (σ) and Variance (σ^2).

Table 5 Statistical Analysis of DNN, XGBoost, and proposed HDE-IDS

Model	Mean F1-Score	Standard Deviation	Variance
DNN	0.9793	1.14×10^{5}	0.0034
XGBoost	0.9884	4.44×10^{-6}	0.0021
HDE-IDS (Proposed)	0.9935	3.25×10^{-6}	0.0018

From the table 5, it can be easily visualized that the proposed model HDE-IDS is performing better compared to DNN and XGBoost. The proposed HDE-IDS able to achieve higher F1-Score mean to handle top 10 attack of CICIoT2023 with minimum Standard deviation (σ) and minimum Variance (σ^2). Smaller the variance of HDE-IDS means highly consistent in detecting the attacks and robust performance under distribution changes compared to XGBoost and DNN.

VI. COMPARATIVE PERFORMANCE ANALYSIS

The proposed model HDE-IDS able to achieve F1 score of 99.1% on CICIoT2023 datasets. A comparative of the latest models of IoT security is performed and comparison is listed in the Table 6 with numerous existing algorithms.

Table 6 Comparative Performance Analysis of latest research with proposed HDE-IDS

Reference	Parameters			
	Accuracy	**Precision**	**Recall**	**F1 score**
Proposed Framework	**99.10%**	**99.00%**	**99.20%**	**99.10%**
[8] Farhan et al.	98.00%	97.00%	97.00%	97.00%
[11] Wazid et al.	99.40%	99.30%	99.20%	99.00%
[13] Pirbhulal et al.	99.00%	99.00%	99.00%	98.00%
[14] Pallavi et al.	99.00%	98.00%	98.00%	98.00%
[15] Lefoane et al.	99.00%	99.00%	98.00%	98.00%
[18] Al-Sadi et al.	98.00%	97.00%	97.00%	99.80%
[19] Alanazi et al.	99.00%	98.00%	98.00%	98.00%
[21] Ullah et al.	99.00%	99.00%	98.00%	98.00%

[24] Khan et al.	98.70%	98.20%	98.40%	98.30%
[25] Sakraoui et al.	98.00%	97.00%	97.00%	97.00%
[26] Kaur et al.	99.95%	99.99%	99.98%	99.99%
[27] Mao et al.	99.89%	96.92%	86.61%	91.12%
[30] Srivastava et al.	98.00%	97.50%	98.20%	97.80%

The comparative performance analysis shows that the proposed model HDE-IDS able to achieve Accuracy 99.10%, Precision 99.00%, Recall 99.20%, F1 score 99.10%. The proposed model able to demonstrate highly balanced detection performance. The proposed model outperforms existing algorithms with superior performance parameters and ability to discriminate between sophisticated malicious attack and benign traffic with minimum false alarm. The proposed model's precision and recall shows that model is able to minimize both false positive and false negative. These excellent figures with reduced false alarms, with minimum false positive and false negative reveals effectiveness of hybrid Transformer- XGBoost architecture.

VII.CONCLUSION

The paper proposes Hybrid Deep Ensemble Intrusion Detection System (HDS-IDS) framework, integrating the transformer encoder, XGBoost classifier and Zero trust security for the IIoT based CII. The proposed framework overcome to various limitations of the present traditional security frameworks in terms of handling varying, complex sophisticated cyber security threats in CII installations. The proposed framework able to achieve effective detection of malware.

The model is prepared by using CICIoT2023 dataset and able to achieve Accuracy of 99.10%, Precision 99.00 %, Recall 99.20%, and F1 score 99.10% which shows that model detection performance is excellent compared to individual DNN and XGBoost or other surveyed models. The model is validated by TON_IoT dataset and used to generalize the model for live unseen traffic. The dual threshold (θ_1, θ_2) used in the risk score enable the framework to add layer of adaptive zero trust parameters for dynamically evaluate the trust level for informed access control in live environment of CII.

Moreover, the statistical analysis of the model being performed and compared with individual DNN and XGBoost and was observed that mean F1 score of the HDE-IDS model is higher compared to DNN and XGBoost which indicate that proposed model has higher detection accuracy. The standard deviation of the HDE-IDS model is lower which indicate that proposed model has higher stability in detecting the attacks and least sensitive to attack variation compared to XGBoost and DNN. The variance of the HDE-IDS model is lower which indicate that proposed model is highly consistent in detecting the attacks and robust performance under distribution changes compared to XGBoost and DNN.

It has been observed that the proposed framework is highly scalable, better generalized, adaptive, dynamic trust based, and ability to handle sophisticated attacks in CII based installations. The proposed framework combined with zero trust security make overall cyber security system a highly reliable, with excellent detection accuracy.

Data Availability:

The datasets used in this study are publicly available. The CICIoT2023 dataset is available at https://doi.org/10.3390/s23135941.

The TON_IoT dataset is also publicly available. The TON_IoT dataset is available at https://research.unsw.edu.au/projects/toniot-datasets.

Figures:
All figures and illustrations presented in this manuscript were created by the authors.

Declarations:

The authors gratefully acknowledge the financial support received through the Seed Money Grant Scheme of National Forensic Sciences University awarded to Dr. Ravirajsinh Vaghela for carrying out this research work.

Funding Declaration- This research did not receive funding.
Clinical Trial Number- Not applicable.
Consent to publish declaration- Not applicable
Ethics and Consent to participate Declarations- Not applicable
Ethics approval- Not applicable.
Consent to participate- Not applicable.
Consent for publication- Not applicable.
Competing interests- The authors declare that they have no competing interests.

Author Contribution Declaration-

Manish Kumar Srivastava: conceived the research idea and prepared the original manuscript draft.
Kavalla Prasanna: implemented the proposed framework, conducted the experiments, and analyzed the results.
Ujjaval Patel: supervised the research, provided technical guidance, validated the methodology, and critically reviewed and edited the manuscript.
Ravirajsinh Vaghela: contributed to the research design, manuscript review, and overall guidance. All authors reviewed and approved the final manuscript.

REFRENCES

[1] S. Baho and J. Abawajy, “Analysis of consumer IoT device vulnerability quantification frameworks,” *Electronics*, vol. 12, no. 3, pp. 1–31, Feb. 2023.

[2] J. Fan *et al.*, “Understanding security in smart city domains from the ANT-centric perspective,” *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11199–11223, Jul. 2023.

[3] Manish Kumar Srivastava and U. Patel, “Cyber Security Risk Analysis of IoT Ecosystem,” in *Proceedings of the 2025 International Conference on Sustainable Energy Technologies and Computational Intelligence (SETCOM)*, Gandhinagar, India, 2025.

[4] N. T. Y. Huan and Z. A. Zukarnain, “A survey on addressing IoT security issues by embedding blockchain technology solutions: Review, attacks, current trends, and applications,” *IEEE Access*, vol. 12, pp. 69765–69782, 2024.

[5] M. Ozkan-Okay et al., "A Comprehensive Survey: Evaluating the Efficiency of Artificial Intelligence and Machine Learning Techniques on Cyber Security Solutions," in IEEE Access, vol. 12, pp. 12229-12256, 2024.

[6] Inam Ullah Khan, Fida Muhammad Khan, Zeeshan Ali Haider, Fahad Alturise, “Integrating AI, Blockchain, and Edge Computing for Zero-Trust IoT Security: A Comprehensive Review of Advanced Cybersecurity Framework”, Computers, Materials and Continua, Volume 85, Issue 3, 2025.

[7] National Institute of Standards and Technology, *Zero Trust Architecture (SP 800-207)*, Aug. 2020.

[8] M. Farhan, M. R. Naeem, A. Almadhor, A. K. Bashir, Z. Zhu and T. R. Gadekallu, "Explainable AI-Driven Security Framework for Cyber-Physical Production Systems in Industry 4.0: Leveraging Immersive Embedded CIoT," in IEEE Transactions on Consumer Electronics, vol. 71, no. 4, pp. 10446-10453, Nov. 2025.

[9] A. Namrita Gummadi, J. C. Napier and M. Abdallah, "XAI-IoT: An Explainable AI Framework for Enhancing Anomaly Detection in IoT Systems," in IEEE Access, vol. 12, pp. 71024-71054, 2024.

[10] J. Jeon, B. Jeong, S. Baek and Y. -S. Jeong, "Static Multi Feature-Based Malware Detection Using Multi SPP-net in Smart IoT Environments," in IEEE Transactions on Information Forensics and Security, vol. 19, pp. 2487-2500, 2024.

[11] M. Wazid et al., "Explainable Deep Learning-Enabled Malware Attack Detection for IoT-Enabled Intelligent Transportation Systems," in IEEE Transactions on Intelligent Transportation Systems, vol. 26, no. 5, pp. 7231-7244, May 2025.

[12] R. Kalakoti, H. Bahsi and S. Nõmm, "Improving IoT Security With Explainable AI: Quantitative Evaluation of Explainability for IoT Botnet Detection," in IEEE Internet of Things Journal, vol. 11, no. 10, pp. 18237-18254, 15 May15, 2024.

[13] S. Pirbhulal, H. Abie and M. Muzammal, "AD-5GIoT: AI-Based Anomaly Detection System for 5G-IoT Networks," ICC 2025 - IEEE International Conference on Communications, Montreal, QC, Canada, 2025.

[14] P. Pallavi, S. Joshi and A. K. Bairwa, "Real-Time Detection of Polymorphic Malware Using AI and Synthetic Data Augmentation," 2025 XXX International Conference on Information, Communication and Automation Technologies (ICAT), Sarajevo, Bosnia and Herzegovina, pp. 1-6, 2025.

[15] M. Lefoane, I. Ghafir, S. Kabir, I. -U. Awan, K. El Hindi and A. Mahendran, "Non-Negative Matrix Factorization and Latent Semantic Analysis for Hybrid Feature Selection: A Proposed Machine Learning System for the Detection of Malicious Executable Files," in IEEE Access, vol. 13, pp. 138867-138882, 2025.

[16] M. M. Shtayat, M. K. Hasan, R. Sulaiman, S. Islam and A. U. R. Khan, "An Explainable Ensemble Deep Learning Approach for Intrusion Detection in Industrial Internet of Things," in IEEE Access, vol. 11, pp. 115047-115061, 2023.

[17] H. Alamro et al., "Modeling of Bayesian-Based Optimized Transfer Learning Model for Cyber-Attack Detection in Internet of Things Assisted Resource Constrained Systems," in IEEE Access, vol. 12, pp. 177298-177311, 2024.

[18] D. Al-Sadi and A. Rjoub, "Federated Edge AI for Adaptive Zero-Trust Security in Resource-Constrained IoT Networks," 2025 IEEE International Smart Cities Conference (ISC2), Patras, Greece, pp. 1-6, 2025.

[19] F. Alanazi, M. Zareei and M. García Martínez, "A Blockchain-Enabled Privacy-Preserving IoT Framework With Domain-Adaptive Anomaly Detection in Zero-Trust Environments," in IEEE Access, vol. 13, pp. 206883-206895, 2025.

[20] A. Hussain et al., "Ensuring Zero Trust IoT Data Privacy: Differential Privacy in Blockchain Using Federated Learning," in IEEE Transactions on Consumer Electronics, vol. 71, no. 1, pp. 1167-1179, Feb. 2025.

[21] F. Ullah, G. Srivastava, L. Mostarda and U. Raza, "ZTID-IoV: Zero-Trust Intrusion Detection in IoV Using Neurosymbolic AI Approach With Federated Meta-Learning," in IEEE Transactions on Consumer Electronics, vol. 71, no. 4, pp. 12037-12046, Nov. 2025.

[22] Faria Nawshin, Devrim Unal, Mohammad Hammoudeh, Ponnuthurai N. Suganthan, " AI-powered malware detection with Differential Privacy for zero trust security in Internet of Things networks", Ad Hoc Networks, Volume 161, 2024.

[23] U. Zukaib, X. Cui, C. Zheng, M. Hassan and Z. Shen, "Meta-IDS: Meta-Learning-Based Smart Intrusion Detection System for Internet of Medical Things (IoMT) Network," in IEEE Internet of Things Journal, vol. 11, no. 13, pp. 23080-23095, 1 July1, 2024.

[24] Izhar Ahmed Khan, Marwa Keshk, Yasir Hussain, Dechang Pi, Bentian Li, Tanzeela Kousar, Bakht Sher Ali, "A context-aware zero trust-based hybrid approach to IoT-based self-driving vehicles security", Ad Hoc Networks, Volume 167, 2025.

[25] TL2AB: Sabrina Sakraoui, Makhlouf Derdour, Ahmed Ahmim, Reham Almukhlifi, Marwa Ahmim, Insaf Ullah, "Trusted lightweight authentication using AI and blockchain for 6G networks", Internet of Things, Volume 33, 2025.

[26] A. Kaur, "Intrusion detection approach for industrial Internet of Things traffic using deep recurrent reinforcement learning assisted federated learning," IEEE Transactions on Artificial Intelligence, vol. 6, no. 1, pp. 37–50, Jan. 2025.

[27] Q. Mao et al., "FeCoGraph: Label-aware federated graph contrastive learning for few-shot network intrusion detection," IEEE Transactions on Information Forensics and Security, vol. 20, pp. 2266–2280, 2025.

[28] S. Jorgensen et al., “Extensible machine learning for encrypted network traffic application labeling via uncertainty quantification,” IEEE Transactions on Artificial Intelligence, vol. 5, no. 1, pp. 420–433, Jan. 2024.

[29] Y. Liu *et al.*, “ATVITSC: A novel encrypted traffic classification method based on deep learning,” *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 9374–9389, 2024.

[30] M. K. Srivastava, K. Prasanna and U. Patel, "IoT Threat Detection Technique using AI, Zero Trust & Micro-segmentation," 2026 Third International Conference on Networking and Communications (ICNWC), Chennai, India, pp. 1-10, 2026.

[31] V. M. U et al., "AI-Powered IoT: A Survey on Integrating Artificial Intelligence With IoT for Enhanced Security, Efficiency, and Smart Applications," in IEEE Access, vol. 13, pp. 50296-50339, 2025.

[32] M. B. Mwangi and S. -M. Cheng, "An Adversarial Attack on ML-Based IoT Malware Detection Using Binary Diversification Techniques," in IEEE Access, vol. 12, pp. 185172-185186, 2024.

[33] N. Moustafa, N. Koroniotis, M. Keshk, A. Y. Zomaya and Z. Tari, "Explainable Intrusion Detection for Cyber Defences in the Internet of Things: Opportunities and Solutions," in IEEE Communications Surveys & Tutorials, vol. 25, no. 3, pp. 1775-1807, thirdquarter 2023.

[34] S. Hussain, M. F. A. Razak and A. Firdaus, "Deep Learning Based Hybrid Analysis of Malware Detection and Classification: A Recent Review," in Journal of Cyber Security and Mobility, vol. 13, no. 1, pp. 91-134, January 2024.

[35] C. Wang, Z. Yuan, P. Zhou, Z. Xu, R. Li and D. O. Wu, "The Security and Privacy of Mobile-Edge Computing: An Artificial Intelligence Perspective," in IEEE Internet of Things Journal, vol. 10, no. 24, pp. 22008-22032, 15 Dec.15, 2023.

[36] Z. Rahman, X. Yi and I. Khalil, "Blockchain-Based AI-Enabled Industry 4.0 CPS Protection Against Advanced Persistent Threat," in IEEE Internet of Things Journal, vol. 10, no. 8, pp. 6769-6778, 15 April15, 2023.

[37] C. Liu *et al.*, “Dissecting zero trust: Research landscape and its implementation in IoT,” *Springer Open*, pp. 1–28, 2024.

[38] H. Zhu, X. Xue, M. Xu, B. -G. Kim, X. Lyu and S. Rani, "Zero-Trust Blockchain-Enabled Secure Next-Generation Healthcare Communication Network," in IEEE Transactions on Network and Service Management, vol. 22, no. 4, pp. 3201-3212, Aug. 2025.

[39] S. Mahmood, M. Gohar, O. A. Khashan, N. Alzahrani, A. Ghani and F. Al-Turjman, "Securing Edge Devices in IoT and 6G: A Trust-Based Approach for Resource-Constrained Environments," in IEEE Open Journal of the Communications Society, vol. 6, pp. 6568-6610, 2025.

[40] Abeer Iftikhar, Faisal Bashir Hussain, Kashif Naseer Qureshi, Muhammad Shiraz, Mehdi Sookhak, “Securing edge based smart city networks with software defined Networking and zero trust architecture”, Journal of Network and Computer Applications, Volume 244, 2025.

[41] Dr. Daniel Díaz-Sánchez, Dr. Florina Almenarez-Mendoza, Dr. Celeste Campo-Vázquez, Dr. Carlos García-Rubio, “Zero-trust token authorization with trapdoor hashes for scalable distributed firewalls”, Future Generation Computer Systems, Volume 176, 2026.

[42] Ashok Kumar Peepliwal, Hari Mohan Pandey, Surya Prakash, Sudhinder Singh Chowhan, Vinesh Kumar, Rahul Sharma, Anand A. Mahajan, “ A prototype model of zero trust architecture blockchain with Eigen Trust-based practical Byzantine fault tolerance protocol to manage decentralized clinical trials, Blockchain: Research and Applications, Volume 5, Issue 4, 2024

[43] X. Han, X. Mu, and J. Zhong, “HGFF: A deep reinforcement learning framework for lifetime maximization in wireless sensor networks,” IEEE Transactions on Artificial Intelligence, vol. 6, no. 4, pp. 859–873, Apr. 2025.

[44] Y. Chai *et al.*, “MalFSCIL: A few-shot class-incremental learning approach for malware detection,” *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 2999–3014, 2025.

[45] T. E. T. Djaidja *et al.*, “Early network intrusion detection enabled by attention mechanisms and RNNs,” *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 7783–7793, 2024.

[46] C. Zha et al., “A-NIDS: Adaptive network intrusion detection system based on clustering and stacked CTGAN,” IEEE Transactions on Information Forensics and Security, vol. 20, pp. 3204–3219, 2025.

[47] X. Yang *et al.*, “A lightweight and dynamic open-set intrusion detection for industrial Internet of Things,” *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 2930–2943, 2025.