Stacking architecture-endpoint detection: a hybrid multi-layered architecture for endpoint threat detection

Abd Rahman Wahid[1], Desi Anggreani[1], Muhyiddin A. M. Hayat[1], Aedah Abd Rahman[2], Muhammad Faisal[1]
[1]Department of Informatics, Faculty of Engineering, Muhammadiyah University of Makassar, Makassar, Indonesia
[2]School of Science and Technology, Asia e University, Kuala Lumpur, Malaysia

Article Info

Article history:

Received Jul 15, 2025
Revised Sep 25, 2025
Accepted Nov 4, 2025

Keywords:

Adversarial robustness
Artificial intelligence
Endpoint detection
Graph network
Hybrid learning

ABSTRACT

Modern endpoint threat detection systems face persistent challenges in balancing detection accuracy, resilience against zero-day attacks, and the interpretability of artificial intelligence (AI) models. Although deep learning (DL) approaches often achieve high accuracy on benchmark datasets, they remain vulnerable to adversarial perturbations and operate as opaque "black boxes," thereby reducing trust and limiting practical adoption in critical infrastructures. This research introduces stacking architecture-endpoint detection (STACK-ED), a hybrid multi-layered architecture for endpoint threat detection. STACK-ED integrates three complementary paradigms: supervised learning for known attack patterns, self-supervised Fgraph-based learning for structural relationships, and unsupervised anomaly detection for emerging or unknown threats. The outputs are consolidated by a meta-learner, followed by a post-hoc correction (PHC) mechanism to minimize false negatives. The framework was evaluated on a combined benchmark dataset (CSE-CIC-IDS2018 and UNSW-NB15, hereafter referred to as HIDS-Set). Experimental results demonstrate state-of-the-art performance, achieving an F2-score of 98.89% after hybrid integration and active learning, with the primary optimization objective being the reduction of undetected attacks. Furthermore, the Shapley additive explanations (SHAP) method enhances interpretability by revealing feature contributions, while the PHC successfully recovered 62.64% of missed zero-day candidates. The findings position STACK-ED not only as a highly accurate detection model but also as an adaptive, resilient, and transparent framework, offering practical implications for enterprise-grade endpoint defense and future zero-trust cybersecurity systems.

Corresponding Author:

Abd Rahman Wahid
Department of Informatics, Faculty of Engineering, Muhammadiyah University of Makassar
Jl. Sultan Alauddin, Makassar, Sulawesi Selatan, Indonesia
Email: 105841116522@student.unismuh.ac.id

1. INTRODUCTION

The digital security landscape is increasingly shaped by the dynamics of a global cyber arms race, where zero-day or unknown attacks have become routine threats to critical infrastructure and enterprise environments [1], [2]. In Indonesia alone, more than 403 million traffic anomalies were recorded in 2023 [3], threatening essential sectors such as banking and public services with both economic and social repercussions [4]. Conventional signature-based endpoint detection (ED) has proven ineffective, relying on databases of known threats and failing to respond to emerging adaptive attacks [5]. To overcome these limitations, recent studies have explored artificial intelligence (AI)-based endpoint detection, which promises adaptive anomaly recognition and intelligent decision-making. Yet, first-generation AI models still face paradoxes deep learning (DL) achieves high accuracy on benchmarks but remains opaque, vulnerable to adversarial perturbations, and inconsistent in detecting zero-day threats [6], [7]. These constraints hinder adoption in mission-critical cybersecurity and emphasize the need for systems that are both robust and explainable. Despite notable advances such as ensemble learning, graph neural networks (GNNs), adversarial robustness (AdvR), and explainable AI (XAI), existing works remain fragmented, focusing on isolated techniques rather than cohesive integration. A critical research gap, therefore, persists: the absence of a framework that unites hybrid learning, structural representation, anomaly detection, interpretability, and adversarial resilience.

This study addresses that gap by introducing stacking architecture (STACK)-ED, a hybrid multi-layered architecture [8] that orchestrates multiple detection paradigms in a unified framework. At the first layer, STACK-ED employs supervised models, namely CatBoost [9] and XGBoost [10] optimized with Optuna [11], alongside self-supervised graph-based components (GINEConv [12]) and unsupervised anomaly detectors (Autoencoder [13], Isolation Forest [14], and one-class support vector machine (SVM) [15]). The outputs are fused at the second layer through a meta-learner optimized via GridSearchCV [16]. At the third layer, a post-hoc correction (PHC) mechanism integrates sensitive outlier detectors, including Mahalanobis distance [17], hierarchical density-based spatial clustering of applications (HDBSCAN) [18], local outlier factor (LOF) [19], and cosine similarity [20] to recover missed zero-day attacks. This multi-faceted design surpasses single-model and conventional ensemble approaches by offering an adaptive, explainable, and resilient endpoint defense mechanism. The main contribution of this study is the design of a holistic, hybrid, and layered framework that integrates supervised, self-supervised, and unsupervised paradigms for endpoint threat detection, the integration of post-processing correction mechanisms that significantly improve zero-day detection and reduce false negatives, the use of Shapley additive explanations (SHAP) to improve the transparency and reliability of the model's decision-making process, and the validation of the proposed framework on the HIDS-Set hybrid benchmark dataset (HIDS-Set), achieving best performance with an F2 score of 98.89% while recovering 62.64% of missed zero-day candidates.

2. LITERATUR RIVIEW

ED is an evolving research domain that must adapt to the growing sophistication of cyber threats [21]. Various studies have explored ensemble learning, AdvR, XAI, yet most remain isolated, addressing only one capability in depth. To contextualize this research, Table 1 compares representative ED approaches across five core capabilities: hybrid architecture, GNN, AdvR, PHC, and XAI.

Table 1. Summary and comparison of related research works in endpoint detection systems

Study (Author, year)	Hybrid arch	GNN	AdvR	PHC	XAI
Tama *et al.* (2019) [22]	✓	✗	✗	✗	✗
Vinayakumar *et al.* (2019) [6]	✗	✗	✗	✗	✗
Mohamed *et al.* (2023) [23]	✓	✗	✗	✗	✗
Ghosh (2025) [24]	✓	✓	✗	✗	✗
Magoo and Garg (2021) [25]	✗	✗	✓	✗	⌀
Kharoubi *et al.* (2025) [26]	✗	✗	✗	✗	✗
Alghazali and Hanoosh (2022) [27]	✓	✗	✗	✗	✗
Vishwakarma and Kesswani (2025) [28]	✓	✗	✗	✗	✓
Zhong *et al.* (2024) [29]	✗	✓	✗	✗	⌀
He *et al.* (2024) [30]	✗	✗	✓	✗	✗
Arreche *et al.* (2024) [31]	✗	✗	✗	✗	✓
Roshan and Zafar (2024) [32]	✗	✗	✓	✗	✗
Sun *et al.* (2024) [33]	✗	✗	✗	✓	⌀
This study (2025)	✓	✓	✓	✓	✓

Legend: ✓: Fully discussed/implemented; ⌀: Partially discussed/implied; ✗: Not addressed

2.1. Ensemble learning based approach

Ensemble learning has long been applied to enhance ED accuracy and robustness by leveraging multiple classifiers. Tama *et al.* [22] introduced a two-stage ensemble that improved anomaly detection, while Mohamed *et al.* [23] adopted ensemble voting for IoT contexts. Alghazali and Hanoosh [27] further integrated random forest as a meta-learner in hybrid DL. Despite these advances, such models often depend on tabular features and overlook adversarial resilience or interpretability. Recent extended detection and response (XDR) surveys underscore the need for multi-layered ensembles coupled with endpoint telemetry [29], but few frameworks achieve holistic fusion across these capabilities.

2.2. Deep learning and graph neural networks

DL has revolutionized ED through automatic feature extraction from network traffic. Vinayakumar *et al.* [6] demonstrated DL's feasibility for intrusion detection, and Kharoubi *et al.* [26] employed convolutional neural networks (CNNs) for IoT traffic classification. Yet, DL remains opaque and sensitive to adversarial perturbations. GNNs have emerged as promising alternatives. Ghosh [24] utilized graph convolutional networks (GCNs) to capture relational dependencies, while Zhong *et al.* [29] highlighted their role in system-on-Chip (SoC) telemetry pipelines. Moreover, linking GNN-based detection with frameworks such as MITRE ATT&CK provides a more threat-informed context [31]. Still, most GNN-based studies focus on structural modeling without integrating adversarial defense or PHC.

2.3. Adversarial robustness and explainable AI

AdvR remains a key challenge, as attackers exploit imperceptible perturbations to evade detection. Magoo and Garg [25] analyzed DL vulnerability to fast gradient sign method (FGSM) and projected gradient descent (PGD) attacks, while He *et al.* [30] and Roshan and Zafar [32] proposed generalized adversarial defenses. Parallelly, the demand for transparency has fostered the adoption of XAI. Vishwakarma and Kesswani [28] incorporated SHAP into stacking ensembles, and Arreche *et al.* [31] designed explainable intrusion detection system (IDS) frameworks. However, most of these approaches still treat interpretability and robustness separately, lacking synergy in unified architectures.

2.4. Research gap

From the above review, it is evident that although ensemble learning, GNN-based detection, adversarial resilience, and XAI have advanced considerably, existing works continue to optimize these components in isolation. Post-execution correction strategies, such as those proposed by Sun *et al.* [33], have yet to be applied comprehensively to endpoint detection. This study introduces STACK-ED, a unified hybrid framework integrating supervised, self-supervised, and unsupervised paradigms with adversarial adaptation, XAI, and PHC. The framework bridges accuracy, resilience, and transparency characteristics crucial for modern SoC, enterprise cybersecurity, and zero-trust environments.

3. PROPOSED METHOD

The research methodology was designed to build the STACK-ED architecture. It involved several stages, beginning with data preparation. The process continued through to the final model evaluation.

3.1. Data pipeline and pre-processing

The proposed data pipeline, illustrated in Figure 1, was designed to guarantee the integrity and representativeness of traffic flows prior to modeling. By combining multiple stages of cleansing, transformation, and dimensionality reduction, the pipeline aims to prepare data that faithfully reflects both normal operations and malicious behaviors while minimizing noise and redundancy.

In the datasets, two benchmark intrusion detection corpora were used to ensure diversity of attack types and network conditions. The UNSW-NB15 dataset [34] provides statistical flow descriptors, widely recognized for capturing background traffic patterns. Complementarily, the CSE-CIC-IDS2018 dataset contains session-level features (e.g., flow inter-arrival time, packet length maxima, and forward packet rate) that emphasize behavioral signatures of modern cyberattacks such as distributed denial of service (DDoS), brute force, and infiltration attempts. Sampling strategy. To maintain balanced representation across classes, a stratified sampling of 10% was applied to each dataset. This produced a combined corpus of 278,770 records, equally distributed between benign and attack instances. The stratification ensured that rare but critical attack categories were retained proportionally, thereby avoiding class skew. Label standardization and feature refinement. After merging, all labels were standardized into a binary format (0=normal, 1=attack). Redundant identifiers were removed, categorical variables were encoded using one-hot representation [35], and numeric variables were scaled to the interval [0,1] with MinMax transformation [36]. These steps prevent dominance of high-magnitude features and promote equal contribution across the feature space. Feature engineering and selection. To enrich discriminatory power, derived indicators were created, such as the source-to-destination byte ratio, packet frequency, inter-arrival variance, and traffic intensity [37]. From the augmented set, 50 predictive attributes were selected via the mutual information (MI) criterion [38], which quantifies the reduction in uncertainty of the target label when a given feature is observed. Features such as destination port, flow duration, inter-arrival statistics, and forward packet rates exhibited strong non-linear association with attack presence and were therefore prioritized. Balancing and dimensionality reduction. To address residual imbalance, the BorderlineSMOTE technique [39] was employed, synthesizing samples along decision boundaries to enhance the detection of minority classes.

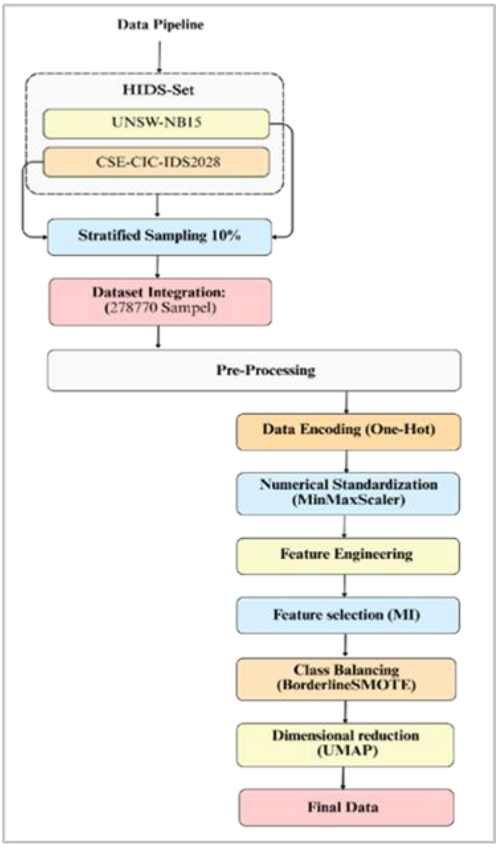

Figure 1. Flow data pipeline and pre-processing

Subsequently, latent representation was reduced to 10 dimensions using uniform manifold approximation and projection (UMAP) [40]. UMAP preserves both local and global structure by optimizing the cross-entropy objective [41]:

$$\frac{min}{y}\sum_{i,j} w_{ij} \log\left(\frac{w_{ij}}{v_{ij}}\right) + \left(1 - w_{ij}\right) \log\left(\frac{1-w_{ij}}{1-v_{ij}}\right) \tag{1}$$

where w_{ij} and v_{ij} denote pairwise similarities in the original and projected space, respectively. This transformation captures hidden topological patterns, enabling subsequent models to operate on compact yet expressive manifolds. Table 2 presents examples of the final pre-processed features, ranging from statistical flow measures to engineered ratios and encoded categorical variables. These structured representations form the standardized input for the STACK-ED architecture.

Table 2. Sample structure of final feature set after pre-processing)

Features	Sample 1	Sample 2	Sample 3	Description
Flow duration	0.0015	0.8920	0.0008	Flow duration (seconds)
Sbytes	0.0021	0.0005	0.0033	Source byte
PktLen Max	0.125	0.045	0.150	Maximum packet length
Fwd Pkts/s	0.0035	0.8540	0.0042	Maximum packet length
Flow IAT Mean	0.0014	0.0120	0.0007	Forward packet frequency
Iat variance	0.0001	0.0023	0.0000	Average inter-arrival
Sbytes, dbytes, ratio	0.4520	1.0000	0.3891	Source/destination byte ratio
Protokol UDP	0	1	0	UDP (one-hot)
Protokol TCP	1	0	1	TCP (one-hot)
Label	0	1	0	0=normal, 1=attack

3.2. Architectural design

The STACK-ED architecture was conceived as a layered intelligence system inspired by the workflow of human cybersecurity analysts. Rather than relying on a single detection mechanism, it integrates multiple analytical perspectives and consolidates them through a hierarchical decision process. As depicted in Figure 2, the architecture consists of three functional levels: i) specialist model components (base-learners), ii) intelligence fusion by a meta-learner, and iii) a PHC mechanism that provides resilience against false negatives and zero-day attacks.

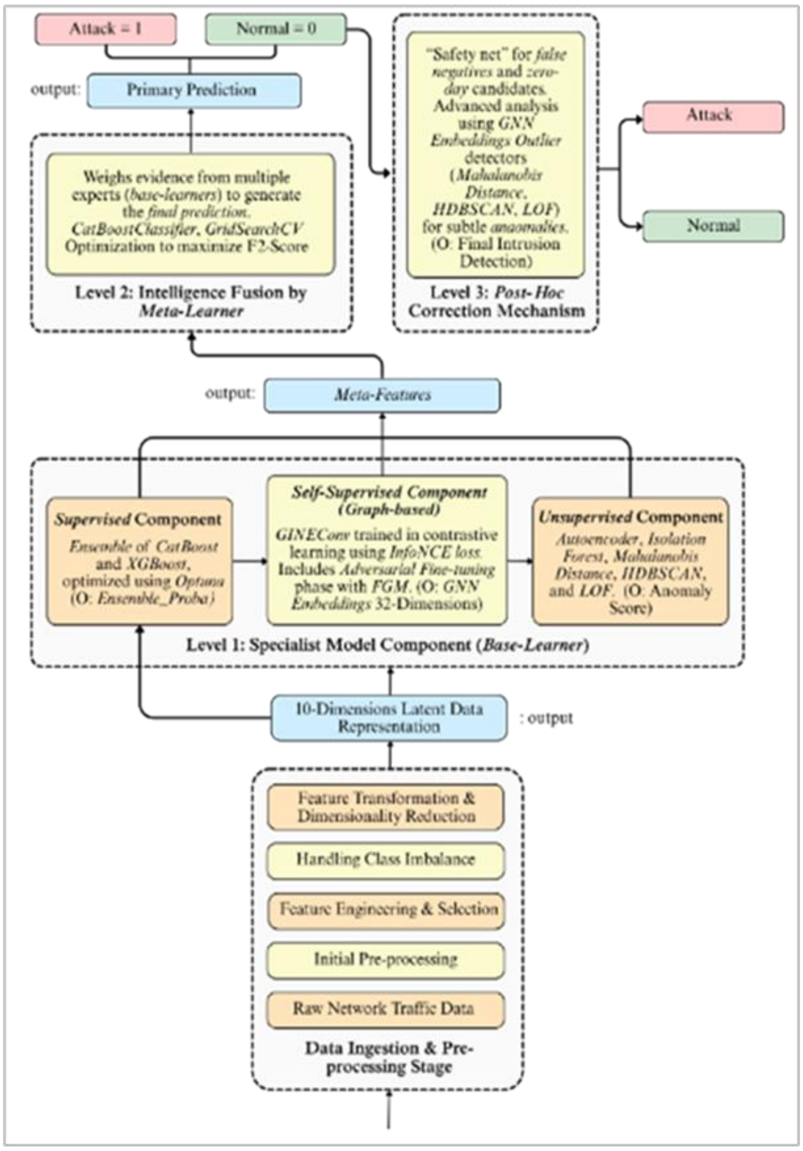
Attack = 1
Normal = 0
output:
Primary Prediction
Weighs evidence from multiple experts (*base-learners*) to generate the *final prediction*. *CatBoostClassifier*, *GridSearchCV* Optimization to maximize F2-Score
Level 2: Intelligence Fusion by *Meta-Learner*
"Safety net" for *false negatives* and *zero-day* candidates. Advanced analysis using *GNN Embeddings Outlier* detectors (*Mahalanobis Distance*, *HDBSCAN*, *LOF*) for subtle *anaomalies*. (O: Final Intrusion Detection)
Level 3: *Post-Hoc* Correction Mechanism
Attack
Normal
output:
Meta-Features
Supervised Component
Ensemble of *CatBoost* and *XGBoost*, optimized using *Optuna* (O: *Ensemble_Proba*)
Self-Supervised Component (Graph-based)
GINEConv trained in contrastive learning using *InfoNCE loss*. Includes *Adversarial Fine-tuning* phase with *FGM*. (O: *GNN Embeddings* 32-Dimensions)
Unsupervised Component
Autoencoder, *Isolation Forest*, *Mahalanobis Distance*, *HDBSCAN*, and *LOF*. (O: Anomaly Score)
Level 1: Specialist Model Component (*Base-Learner*)
10-Dimensions Latent Data Representation
: output
Feature Transformation & Dimensionality Reduction
Handling Class Imbalance
Feature Engineering & Selection
Initial Pre-processing
Raw Network Traffic Data
Data Ingestion & Pre-processing Stage

Figure 2. Architecture stacking ensemble STACK-ED

3.2.1. Level 1 specialist model components (base-learner)

The first level receives the 10-dimensional latent data representation and processes it through three complementary perspectives:

- Supervised component (pattern recognition expert). This component employs an ensemble of CatBoost and XGBoost classifiers, optimized via Bayesian hyperparameter search (Optuna), to achieve a balanced bias-variance trade-off [42]. It specializes in identifying patterns consistent with previously known intrusions.
- Graph-based self-supervised component (network structure expert). Leveraging GINEConv within a contrastive learning framework, this component captures structural dependencies among network flows.

Through adversarial fine-tuning, it produces robust 32-dimensional embeddings capable of discriminating subtle variations in traffic relations.

- Unsupervised anomaly detection component (behavioral deviation expert). Operating on the graph embeddings, this component detects outliers without requiring labeled data. It integrates autoencoders, Isolation Forest, Mahalanobis distance, and LOF to capture behavioral deviations indicative of novel or rare attacks.

3.2.2. Level-2 intelligence fusion by meta-learner

At the second level, the system functions as an intelligence fusion hub. Here, outputs from all base-learners are transformed into meta-features and aggregated by a meta-learner. This mechanism parallels the role of a security operations center, where evidence from different analysts is weighted and synthesized to improve overall situational awareness. Table 3 summarizes the meta-features used in this layer, including supervised probabilities, initial graph predictions, and anomaly scores. These features are then processed by a CatBoostClassifier optimized with GridSearchCV to maximize the F2-score, which emphasizes recall in security-critical contexts.

$$Y_{pred} = F_{meta}(m) \text{ where} = [h_1(X), h_2(X), \dots, h_k(X)] \quad (2)$$

Here, Y_{pred} is the primary prediction from STACK-ED, F_{meta} is a function learned by the meta-learner, X is the original input data vector, and h_i(X) is output from the base-learner.

Table 3. Meta-features as input for meta-learners

Meta-feature	Model description and source
Ensemble proba	Attack probability of supervised component (CatBoost+XGBoost)
GINE Prediction	Initial binary prediction of the graph representation component
Autoencoder score	Anomaly score (reconstruction error) of the autoencoder
Mahalanobis score	Mahalanobis distance from the sample to the center of the “normal” distribution in GNN space
Isolation forest score	Anomaly score from the isolation forest algorithm

3.2.3. Level 3 post-hoc correction mechanism

The third level functions as a safeguard against undetected intrusions. It is only activated when the meta-learner classifies a sample as normal ($Y_{pred} = 0$). In such cases, the sample undergoes a secondary investigation using graph embeddings and sensitive outlier detectors. Techniques such as Mahalanobis distance (with adaptive thresholds), HDBSCAN clustering (noise detection), LOF (outlier detection), and cosine similarity are applied to reassess subtle anomalies. If any detector flags the sample as suspicious, the final prediction is corrected to attack.

3.3. Specialist components

This section elaborates on the three specialist components that constitute Level 1 of the STACK-ED architecture. Each component represents a distinct analytical perspective, contributing complementary evidence for the higher-level fusion stage. Beyond their individual analytical roles, these specialists are designed to reflect how real-world security teams distribute expertise across different detection modalities. Each component captures a unique facet of endpoint behavior pattern recognition from supervised models, structural context from graph-based learning, and behavioral deviation from unsupervised techniques, ensuring that no single viewpoint dominates the decision-making pipeline. By integrating these heterogeneous perspectives at the foundational layer, STACK-ED mitigates model bias, reduces single-point failure risks, and establishes a richer evidence base for the meta-learner to synthesize into a robust final judgment.

3.3.1. Supervised component

The supervised component serves as a pattern recognition expert that captures regularities in previously observed attack behaviors. It employs an ensemble of gradient boosting classifiers (CatBoost and XGBoost) combined through a soft voting strategy, where prediction probabilities are averaged to form a consensus. Hyperparameter tuning was performed using Optuna, a Bayesian optimization framework that systematically explores the search space to minimize false negatives. This design ensures a robust bias-variance trade-off, which is critical for cybersecurity contexts where undetected attacks may lead to severe consequences. Table 4 summarizes the hyperparameter ranges explored for both CatBoost and XGBoost during the optimization process. The primary output of this component is the Ensemble Proba, representing the probability of a traffic flow being malicious. This output forms a key input (meta-feature) for the meta-learner.

Table 4. Hyperparameter search space for supervised components using optuna

Model	Hyperparameters	Type	Range/value
CatBoost	Iteration	Integer	[200.500]
	Depth	Integer	[4, 6]
	Learning_rate	Float	[0.01, 0.05]
	12_leaf_reg	Float	[1, 10]
XGBoost	n_estimators	Integer	[200. 500]
	Max_depth	Integer	[4, 6]
	Learning_rate	Float	[0.01, 0.05]
	reg_lambda	Float	[1, 10]

3.3.2. Graph-based self-supervised components

The self-supervised component acts as a network structure expert. Network traffic is abstracted into graph representations, where each flow is modeled as a small graph with two nodes (endpoints) connected by an edge that carries statistical attributes (e.g., bytes, duration, and inter-arrival time). Node features are derived from UMAP embeddings, ensuring a compact yet discriminative latent representation. The graph model employed is GINEConv, trained through a two-phase strategy:

- Contrastive learning. Training begins with the InfoNCE loss [41], designed to maximize similarity between embeddings of normal-normal pairs while separating normal-attack pairs. This encourages the formation of embeddings that are both robust and discriminative [43], where representations are learned by distinguishing informative positive samples from noise-based negatives through the InfoNCE objective.

$$L_{NCE} = E_x\left[log\frac{\exp\,(\frac{sim(a,p)}{\tau}}{\exp\left(\frac{sim(a,p)}{\tau}\right)+\sum_{i=1}^{N}\exp\left(\frac{sim(a,n_i)}{\tau}\right)}\right] \quad (3)$$

Where *a/p* denotes anchor-positive (normal) pairs, n_i represents attack embeddings, and τ=0.07 is temperature parameter.

- Adversarial fine-tuning. To further enhance robustness, adversarial samples are generated using the fast gradient method (FGM) [42]. These perturbations are introduced as additional negative examples, training the GNN to resist evasion attempts that subtly manipulate network features.

The conceptual workflow of this adversarial training process is illustrated in Figure 3, where robust embeddings are optimized iteratively through contrastive and adversarial pairings. The final output of this component is a 32-dimensional embedding that captures both relational structure and resilience to adversarial perturbations.

3.3.3. Unsupervised anomaly detection components

The unsupervised component functions as a behavioral deviation expert, focusing on the detection of rare or previously unseen attacks. Unlike supervised learning, it operates entirely on the 32-dimensional GNN embeddings and does not rely on labels. Instead of relying on a single method, a weighted ensemble of anomaly detectors is constructed, including Autoencoder reconstruction error, Mahalanobis distance, Isolation Forest, and one-class SVM. The contribution of each detector is weighted according to its F2-score performance on the training data, ensuring that methods with higher discriminative capacity exert greater influence. Formally, the ensemble score is defined as (4).

$$w_i = \frac{F2_i}{\sum_{j=1}^{k} F2_j} \; ; \; S_{comb} = \sum_{i=1}^{k} w_i \, S_i \quad (4)$$

Where S_i is the normalized anomaly score from detector i, and k is the total number of detectors. Table 5 presents the relative weights assigned to each detector in the final ensemble, reflecting their proportional contribution.

The outcome of this component is a set of anomaly scores that form the unsupervised meta-features. These serve as a critical complement to the supervised and self-supervised perspectives, enriching the evidence available to the meta-learner.

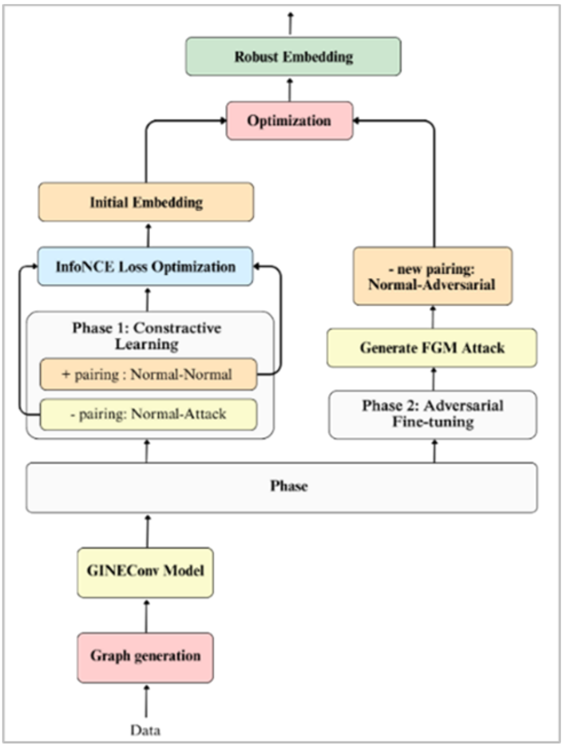
Robust Embedding
Optimization
Initial Embedding
InfoNCE Loss Optimization
Phase 1: Constractive Learning
+ pairing : Normal-Normal
- pairing: Normal-Attack
- new pairing: Normal-Adversarial
Generate FGM Attack
Phase 2: Adversarial Fine-tuning
Phase
GINEConv Model
Graph generation
Data

Figure 3. Workflow of adversarial fine-tuning for the GINEConv graph model

Table 5. Relative weights of anomaly detection components

Detector	Actual weight value
GNN (initial prediction)	0.409
Mahalanobis	0.200
Autoencoder	0.198
Isolation forest	0.109
One-class SVM	0.084

3.4. Intelligence fusion by meta-learner

At the second level of the STACK-ED framework, intelligence fusion is performed by a meta-learner that integrates the diverse evidence produced by the specialist components. This mechanism emulates the role of an intelligence analysis center, where inputs from multiple experts are consolidated into a single, more reliable decision. The meta-learner is implemented using the CatBoostClassifier, selected for its ability to handle heterogeneous feature distributions and for its robustness to overfitting. Inputs to this learner are the meta-features derived from supervised, self-supervised, and unsupervised components (see Table 3), including probabilities, anomaly scores, and structural predictions. To ensure that the model generalizes well and emphasizes recall in security-sensitive scenarios, the optimization process employed a 3-fold cross-validation with the F2-score as the primary objective. The F2-score was chosen because it penalizes false negatives more heavily, reflecting the high cost of undetected attacks in real-world applications. Table 6 summarizes the hyperparameter search space explored using GridSearchCV.

The optimization procedure systematically evaluated all parameter combinations, and the configuration that maximized the F2-score during cross-validation was selected. The final meta-learner thus represents an optimized decision fusion engine, designed not only to maximize predictive accuracy but also to enhance robustness against undetected intrusions. Subsequent validation of this model, including PHC, is presented in section 4.

Table 6. Hyperparameter search space for meta-learner (GridSearchCV)

Hyperparameters	Type	Value explored
Iterations	Integer	[300, 500]
Depth	Integer	[4, 6, 8]
Learning_rate	Float	[0.01, 0.05, 0.1]

3.5. Interpretability and post-hoc correction

A key design principle of STACK-ED is its ability to evolve over time and remain transparent in its decision-making. To achieve this, two complementary mechanisms are introduced: interpretability through XAI techniques and a PHC process that safeguards against false negatives. These mechanisms ensure that STACK-ED does not operate as a static or opaque model, but rather as a system capable of continuous self-assessment and refinement. Interpretability enables security analysts to trace how evidence is weighted, reducing the cognitive barrier often associated with complex machine learning pipelines and strengthening operational trust. Meanwhile, the PHC layer acts as a secondary analytical sweep, allowing the system to revisit borderline cases with deeper scrutiny by leveraging structural embeddings and sensitive anomaly detectors. Together, these components not only enhance transparency and resilience but also position STACK-ED as a practically deployable framework aligned with modern expectations for accountable and adaptive AI-driven security systems.

3.5.1. Interpretability and efficiency improvement (active learning)

Interpretability is addressed using SHAP [44], which provides a theoretically grounded means to quantify the contribution of each meta-feature to the final decision. By decomposing predictions into additive feature attributions, SHAP enables human analysts to validate and understand the reasoning of the model. These interpretability insights are further utilized to support an Active Learning cycle [45]. In this cycle, instances that the model is most uncertain about are prioritized for further inspection and retraining. Uncertainty is measured using Shannon entropy, where a probability distribution close to uniform indicates maximum confusion as (5).

$$H(p) = -p\, log_2(p) - (1-p)log_2(1-p) \tag{5}$$

Here, p represents the predicted probability of the positive class. When p ≈0.5, the entropy reaches its maximum, signaling that the sample is highly ambiguous. By iteratively incorporating such high-uncertainty samples into the training process, STACK-ED maintains continuous improvement and better generalization over time.

3.5.2. Post-hoc correction mechanism

To address the problem of false negatives cases where attacks are incorrectly classified as normal, STACK-ED introduces a PHC mechanism as a last line of defense. This mechanism leverages embeddings produced by the graph-based component and subjects them to a series of highly sensitive anomaly detectors. The process begins by isolating samples predicted as normal by the meta-learner. These are re-examined in the GNN latent space (32-dimensional embeddings), where subtle deviations from normal patterns can be more easily detected. Several complementary detectors are applied Mahalanobis distance, which measures statistical distance from the distributional center of normal data. HDBSCAN, a density-based clustering algorithm capable of labeling low-density points as noise. LOF, which evaluates the local density deviation of a sample relative to its neighbors, and Cosine similarity, which captures deviations in directional similarity among embeddings. If any detector identifies an instance as anomalous, the prediction is corrected to attack. This layered correction process serves as a safeguard against evasive and zero-day attacks, reinforcing STACK-ED's resilience beyond the initial classification stage.

3.6. Adversarial robustness evaluation

Beyond predictive accuracy, an essential dimension of evaluating intrusion detection models lies in their robustness against adversarial manipulation. In real-world scenarios, attackers may deliberately introduce subtle perturbations to network traffic in order to evade detection. To assess the resilience of STACK-ED, an AdvR evaluation was conducted. The procedure follows the principle of white-box adversarial testing, where the adversary is assumed to have knowledge of the target model. Perturbations are generated using the FGM, a widely adopted evasion technique that applies small but strategically crafted modifications to input features. For each perturbation strength, parameterized by ϵ, adversarial samples are produced and subsequently used to test the model's stability. Two models were subjected to this evaluation.

3.6.1. Supervised ensemble (model baseline)

The baseline model corresponds to the supervised ensemble component (CatBoost+XGBoost) operating without higher-level fusion or correction mechanisms. This configuration provides a benchmark to quantify the inherent vulnerability of conventional supervised approaches under adversarial perturbations, particularly in detecting subtle or evasive attack patterns. By isolating the supervised ensemble from the additional resilience layers implemented in STACK-ED, this baseline enables a clear assessment of the inherent limitations of standalone boosting-based classifiers and highlights the performance gains introduced by the multi-layered architecture.

3.6.2. STACK-ED

The complete STACK-ED architecture, integrating supervised, self-supervised, and unsupervised components along with the meta-learner, was tested against the same adversarial samples. For each perturbed instance, the meta-features were reconstructed from the embeddings and anomaly scores of the base-learners, ensuring that the fusion process remained consistent under attack conditions. Performance degradation was measured using the F2-score, chosen because it emphasizes recall and directly penalizes false negatives, an especially critical property in cybersecurity, where undetected attacks carry severe implications. By comparing the rate of F2-score reduction across increasing values of ϵ, the evaluation provides insight into the relative resilience of STACK-ED versus the baseline model. In summary, the proposed methodology integrates multiple perspectives into a unified STACK-ED framework, beginning with a structured data pipeline, followed by layered architecture design, and culminating in intelligence fusion, interpretability, and resilience mechanisms. Each methodological component was deliberately selected to address the limitations of conventional detection systems: supervised learning for known patterns, graph-based self-supervised learning for structural insights, unsupervised anomaly detection for novel threats, and PHC as a safeguard against false negatives. The inclusion of explainability (via SHAP), active learning, and AdvR evaluation further ensures that STACK-ED not only achieves high predictive performance but also maintains transparency and adaptability in adversarial environments. Having established this methodological foundation, the next section presents the experimental results and discusses how the proposed framework performs in practice across multiple evaluation dimensions.

4. RESULT

This section presents the results of a series of empirical evaluations designed to validate the performance of STACK-ED across multiple dimensions of endpoint threat detection. The reported findings demonstrate how the proposed hybrid architecture enhances predictive accuracy, strengthens resilience against both conventional and adversarial threats, and improves decision reliability through multi-perspective intelligence fusion. Beyond numerical performance, the results also highlight the model's interpretability and adaptability, emphasizing how STACK-ED maintains stable performance in dynamic environments while offering transparent insights that support practical cybersecurity decision-making.

4.1. Overall performance of STACK-ED model

The performance of STACK-ED was evaluated progressively across its constituent stages, reflecting the layered complexity of the architecture. Table 7 reports the results in terms of Accuracy, F2-score, Precision, Recall, and ROC-AUC for each stage, ranging from individual components to the final hybrid model enhanced by active learning.

Table 7. Summary of evaluation metric performance for each stage of the STACK-ED model

Model stage	Accuracy	F2-score	Precision	Recall	ROC-AUC
Supervised (T3)	0.9626	0.9662	0.9572	0.9685	0.9626
Self-supervised (T4)	0.8383	0.8298	0.8472	0.8256	0.8383
Unsupervised- training (T5)	0.8239	0.8226	0.8252	0.8219	0.8239
Unsupervised-validation (T5)	0.8224	0.8224	0.8224	0.8224	0.8224
Hybrid training (T6)	0.9866	0.9911	0.9796	0.9940	0.9866
Hybrid validation (T6)	0.9857	0.9898	0.9793	0.9924	0.9857
Hybrid+AL (T7)	0.9829	0.9889	0.9733	0.9929	0.9829

A clear trend of progressive performance improvement can be observed across the layered stages. The supervised ensemble at stage T3 serves as a strong baseline, achieving an F2-score of 0.9662 and accuracy of 0.9626, which demonstrates its reliability in identifying known attack patterns. The graph-based self-supervised component (T4), while not a final classifier, contributes significantly by producing embeddings that capture relational structures, with an F2-score of 0.8298. Similarly, the unsupervised anomaly detection component (T5) achieves an F2-score of 0.8224, reflecting its capacity to identify anomalies without labeled data. The most substantial performance gain occurs at stage T6, where the meta-learner integrates outputs from all base-learners. The hybrid fusion achieves an F2-score of 0.9898 and accuracy of 0.9857, evidencing the effectiveness of stacking in synthesizing diverse forms of evidence.

Finally, with the incorporation of an active learning cycle (T7), STACK-ED sustains state-of-the-art performance (F2-score of 0.9889) while enhancing its adaptability and generalization to evolving attack scenarios. These findings emphasize the novelty of STACK-ED; its staged progression demonstrates that robustness and adaptability are not the result of a single model, but rather of the orchestrated integration of multiple complementary paradigms.

4.2. Evolution of GNN component training

The graph-based self-supervised component was trained through two sequential phases. To capture the dynamics of this learning process, the evolution of the training loss for the GINEConv model is illustrated in Figure 4, providing insight into how the model adapts as training progresses. During constructive learning, the GNN rapidly learns discriminative relational patterns from network traffic, reflected by a consistent decline in loss values. When adversarial perturbations are introduced in the second phase, the model undergoes a controlled adjustment period, signaling its effort to reconcile structural consistency with robustness requirements. Collectively, these learning dynamics highlight the importance of combining contrastive objectives with adversarial exposure to produce embeddings that remain stable and effective under both normal and hostile conditions. As shown in Figure 4, the loss decreases rapidly during the initial epochs, reflecting the efficiency of constructive learning in capturing discriminative structural features from network flows. Around epoch 300, the curve exhibits a transient increase before stabilizing, indicating the model's adaptation process as adversarial perturbations are introduced. This stabilization suggests that the embeddings produced by the GNN progressively become more robust, balancing sensitivity to normal patterns with resistance to adversarial manipulations. Although the GNN is not used as a final classifier, its role is pivotal in generating resilient embeddings that serve as the foundation for both anomaly detection and higher-level fusion in STACK-ED. This demonstrates how the integration of self-supervised graph learning contributes to the overall robustness of the architecture.

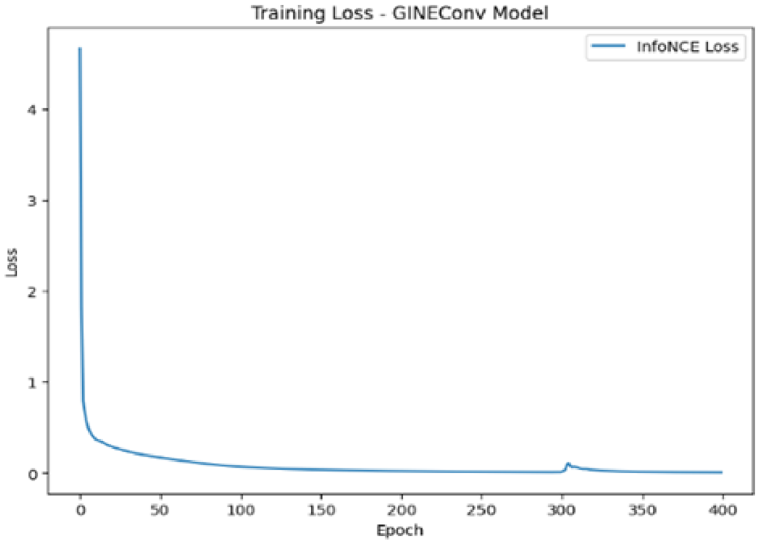

Figure 4. Training loss curve of the GINEConv model during constructive and adversarial phases

4.3. Transparency and adaptive enhancement

A central challenge in adopting machine learning for cybersecurity lies in the opacity of model decisions. To address this, STACK-ED incorporates interpretability and adaptive learning mechanisms that enhance both transparency and long-term efficiency. Figures 5 and 6 present the SHAP analysis applied to the meta-learner. The bar plot in Figure 5 indicates that Ensemble Proba exerts the greatest average influence on the model's final decisions, followed by the Mahalanobis Score and Autoencoder Score. This finding highlights the dominant role of supervised evidence, complemented by anomaly detection features, in shaping the meta-learner's outputs. The summary plot in Figure 6 further illustrates the directional contributions of meta-features: positive SHAP values correspond to "attack" predictions, while negative values contribute to "normal" classifications. Together, these visualizations demonstrate how STACK-ED integrates diverse evidence in a transparent and traceable manner, thereby strengthening trust in the system's decision-making process.

Beyond interpretability, STACK-ED introduces adaptability through an active learning cycle. Figure 7 illustrates the probability distribution of samples identified for additional labeling using uncertainty sampling. The concentration of samples near probability 0.5 suggests that the procedure effectively targets instances where the model exhibits maximum uncertainty. By selectively incorporating such cases into training, STACK-ED improves generalization while significantly reducing labeling overhead, eliminating the need for exhaustive annotation of all new data. These findings underscore that STACK-ED is not only accurate but also transparent and adaptable, providing both explainable evidence for its decisions and mechanisms for continuous improvement in evolving threat landscapes.

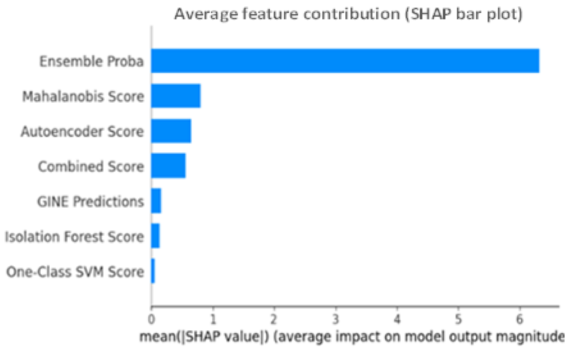

Figure 5. General performance of the STACK-ED model

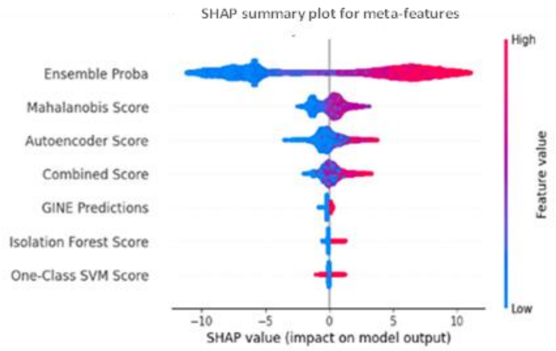

Figure 6. SHAP summary plot for meta-features

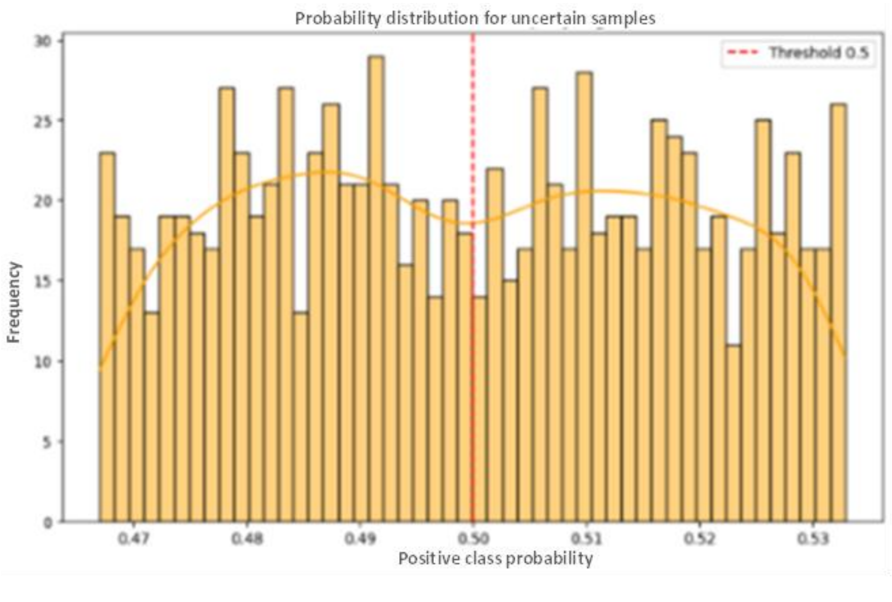

Figure 7. Probability distribution for uncertain samples

4.4. Zero-day candidate detection through post-hoc correlation

A distinctive contribution of STACK-ED lies in its ability to address false negatives, attacks that evade initial classification. To mitigate this challenge, a PHC mechanism is incorporated as a secondary safeguard, leveraging the structural embeddings generated by the GNN to re-examine instances initially predicted as normal. Figure 8 depicts the distribution of Mahalanobis distance scores for samples that were misclassified in the primary detection stage. Although these instances passed the initial screening, their embeddings exhibit distinguishable deviations from the distributional center of normal traffic. By applying sensitive anomaly detectors, such as Mahalanobis distance thresholds, HDBSCAN clustering, LOF, and cosine similarity, the post-hoc mechanism reclassified a significant proportion of these samples as attacks.

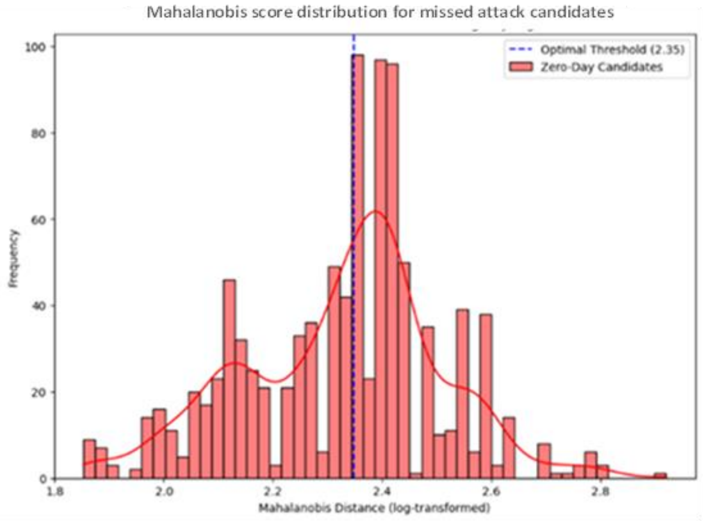

Figure 8. Mahalanobis score distribution for missed attack candidates

Quantitatively, the mechanism corrected approximately 617 out of 985 missed samples, representing a recovery rate of about 62.6%. This finding provides strong evidence that PHC enhances STACK-ED's robustness, effectively functioning as a "safety net" against zero-day or evasive threats that are often overlooked by conventional models.

4.5. Resistance to adversarial attacks

Evaluating robustness under adversarial conditions is essential to ensure the reliability of intrusion detection systems in realistic threat environments. Figure 9 and Table 8 present the comparative performance of STACK-ED and the supervised ensemble baseline when exposed to adversarial perturbations of varying strengths (ϵ). Evaluating robustness under adversarial conditions is essential to ensure that intrusion detection systems remain reliable when confronted with realistic and intentionally manipulated attack scenarios. In this study, STACK-ED and the supervised ensemble baseline were subjected to adversarial perturbations of varying strengths (ϵ) to assess their stability under evasion-oriented attacks. Figure 9 and Table 8 present a comparative overview of how both models respond as the magnitude of perturbations increases, highlighting differences in vulnerability across methods. This evaluation is crucial because adversarial manipulations often mimic subtle alterations in traffic patterns that can mislead conventional supervised models. By quantifying performance degradation under controlled adversarial conditions, this analysis provides a clear understanding of the resilience benefits introduced by STACK-ED's layered design and adversarially trained graph-based embeddings.

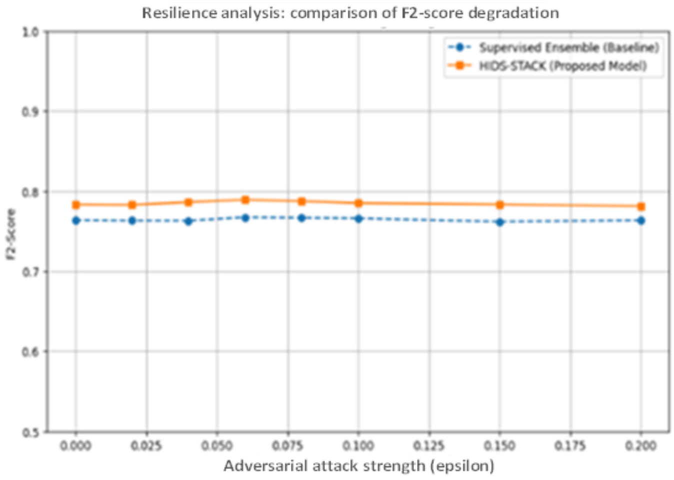

Figure 9. Comparative degradation of F2-scores

Across the entire range of disturbance strengths, STACK-ED consistently maintains a higher F2-score and exhibits less degradation compared to supervised ensembles. This indicates that the integration of graph-based adversarial adaptation and intelligence fusion enables STACK-ED to withstand evasion strategies more effectively than conventional supervised models. Even at higher disturbance magnitudes (ϵ=0.20), the performance gap remains visible, highlighting the resilience of the proposed architecture. These results emphasize adversarial resilience as one of STACK-ED's distinguishing features. The system maintains high detection rates under clean conditions and demonstrates measurable resistance when subjected to targeted manipulation, confirming the contribution of adversarial training to the GNN component and the resilience of the layered fusion design. In summary, the experimental results across all dimensions, overall performance, interpretability, zero-day detection, and AdvR demonstrate that STACK-ED advances beyond traditional approaches by unifying accuracy, resilience, and transparency within a single framework. The following section provides a deeper discussion of these findings, situating them in the broader context of related research and practical cybersecurity applications.

Table 8. Comparison of F2-scores under adversarial perturbations

Adversarial attack power (ϵ)	Supervised ensemble (baseline)	STACK-ED (proposed model)
0.00	0.7639	0.7835
0.02	0.7633	0.7831
0.04	0.7632	0.7862
0.06	0.7674	0.7892
0.08	0.7667	0.7874
0.10	0.7661	0.7851
0.15	0.7620	0.7836
0.20	0.7638	0.7814

5. DISCUSSION

The results presented in the preceding section comprehensively demonstrate STACK-ED’s capabilities and situate its contributions within the broader research landscape of intrusion detection. The high F2-score achieved at stage T7 (0.9889) confirms the effectiveness of the hybrid stacking design in achieving accuracy, robustness, and adaptability simultaneously. The evolution of the GNN components illustrates how adversarial adaptation enhances the quality of latent representations, allowing the system to withstand perturbations that often degrade conventional DL models. This finding extends prior graph-based intrusion detection studies by integrating adversarial training and multi-perspective fusion. The PHC mechanism further strengthens resilience against zero-day attacks, successfully recovering 62.6% of missed candidates, a capability rarely implemented in earlier ensemble-based approaches.

In addition, the AdvR analysis confirms STACK-ED’s superiority over standard supervised ensembles across multiple perturbation levels, validating the benefit of adversarially trained GNN embeddings. These outcomes refine previous evaluations of adversarial vulnerability in intrusion detection by offering a defense-oriented and multi-layered approach. Interpretability, achieved through SHAP analysis, provides insight into the meta-learner’s decision process. The identification of Ensemble Proba, Mahalanobis Score, and Autoencoder Score as dominant features confirms the combined role of supervised and anomaly-based evidence. This aligns with ongoing developments in XAI and strengthens analyst confidence through transparent inference. Moreover, the active learning cycle enables adaptability by focusing on uncertain samples, improving generalization while reducing labeling costs, in line with established perspectives on query informativeness in active learning [46] and information-theoretic views of uncertainty based on entropy measures [47]. Beyond its academic significance, STACK-ED also offers practical implications for enterprise cybersecurity. While commercial endpoint detection and response (EDR) systems such as CrowdStrike Falcon emphasize real-time telemetry and cloud-native orchestration, empirical assessments of production EDR platforms show persistent limitations in detecting stealthy or highly adaptive APT-style threats [48].

Complementary research in graph-based defense has demonstrated that physics-informed GNNs and structural reasoning can improve attack path prediction and enhance cyber defense posture [49]. Positioned within this context, STACK-ED contributes as a conceptual blueprint for next-generation EDR solutions combining supervised classifiers, graph-based embeddings, unsupervised detectors, and PHC into a unified, explainable, and resilient framework. Nevertheless, practical deployment introduces considerations of computational overhead, particularly due to GNN embeddings and multi-stage ensemble inference. Although this study’s experiments were conducted under controlled conditions, real-time EDR operations demand low-latency responses where even sub-second delays can reduce efficacy.

Optimization strategies such as model compression, distributed inference, or GPU-accelerated execution will therefore be crucial to achieve operational scalability. In summary, STACK-ED represents an adaptive and holistic defense paradigm that bridges fragmented research directions by integrating hybrid learning, graph representation, adversarial resilience, PHC, and interpretability into a single architecture. The framework advances the state of the art in endpoint detection while providing a foundation for future enterprise-grade, explainable, and resilient cybersecurity systems.

6. CONCLUSION

This study proposed STACK-ED, a hybrid stacking architecture for endpoint detection that integrates supervised, self-supervised, and unsupervised components, fused by a meta-learner and reinforced by a PHC mechanism. The framework addresses key challenges in modern intrusion detection by combining accuracy, resilience against adversarial manipulation, interpretability, and adaptability. Experimental evaluations confirm that STACK-ED achieves high detection performance while enhancing robustness to zero-day attacks and adversarial perturbations, thereby advancing the state of the art in multi-layered defense strategies. Despite these contributions, several limitations remain. The computational complexity of graph-based components and the reliance on high-quality preprocessing may hinder real-time deployment in resource-constrained environments. These challenges highlight opportunities for further research, including the optimization of distributed and parallel processing, the incorporation of continuous learning mechanisms, and the evaluation of STACK-ED under diverse black-box adversarial scenarios. Future work may also explore the integration of curriculum learning, retrieval-augmented generation (RAG), and experience replay techniques to enable more efficient adaptation to streaming data and emerging attack patterns without requiring full retraining. Additionally, federated learning could be investigated to allow collaborative model training across multiple organizations without compromising sensitive data, thereby extending STACK-ED's applicability in privacy-preserving contexts. Deception technologies, such as honeypots and moving target defenses, may complement STACK-ED by generating adversarial uncertainty and enriching anomaly signals for detection. Finally, integration with SIEM platforms and cloud-native architectures would provide operational pathways to embed STACK-ED within enterprise-scale monitoring ecosystems, enabling seamless orchestration and real-time incident response in production environments.

ACKNOWLEDGMENTS

Sincere appreciation is extended to COCONUT Computer Club and Universitas Muhammadiyah Makassar for computing facilities, fellow researchers, and the inspiring environment. The ongoing moral support from family and colleagues is also greatly appreciated for continuously motivating the completion of this work.

FUNDING INFORMATION

This research is supported by Universitas Muhammadiyah Makassar through the research scheme funding number 107/05/A.5II/XI/47/2025.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Abd Rahman Wahid	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Desi Anggreani		✓		✓	✓					✓		✓		✓
Muhyiddin A. M. Hayat			✓			✓		✓	✓		✓		✓	
Aedah Abd Rahman				✓	✓		✓			✓	✓			✓
Muhammad Faisal				✓		✓			✓			✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis

I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review and **E**diting

Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in [UNSW-NB15 and CSE-CIC-IDS2018] at http://doi.org/[doi], reference number [34], and Mendeley at https://data.mendeley.com/datasets/29hdbdzx2r/1.

REFERENCES

[1] H. Song, D. B. Rawat, S. Jeschke, and C. Brecher, *The internet of things—current trends, applications and future challenges*. MDPI, 2024. doi: 10.3390/books978-3-7258-1621-7.

[2] N. Moustafa, J. Hu, and J. Slay, "A holistic review of network anomaly detection systems: a comprehensive survey," *Journal of Network and Computer Applications*, vol. 128, pp. 33–55, 2019, doi: 10.1016/j.jnca.2018.12.006.

[3] Badan Siber Sandi Negara (BSSN), "Annual report on Indonesia's cybersecurity situation in 2023 (in Indonesian: *Lanskap keamanan siber Indonesia 2023*)." 2024. [Online] Available: https://www.bssn.go.id/wp-content/uploads/2024/03/Lanskap-Keamanan-Siber-Indonesia-2023.pdf

[4] World Economic Forum, "The global risks report 2025, 20th edition." *Insight Report-World Economic Forum*, 2025. [Online] Available: https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf

[5] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16–24, 2013, doi: 10.1016/j.jnca.2012.09.004.

[6] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/access.2019.2895334.

[7] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *International Conference on Learning Representations (ICLR)*, 2014.

[8] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.

[9] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[10] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[11] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: a next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701.

[12] W. Hu *et al.*, "Open graph benchmark: datasets for machine learning on graphs," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[13] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504–507, 2006, doi: 10.1126/science.1127647.

[14] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*, 2008, pp. 413–422. doi: 10.1109/ICDM.2008.17.

[15] B. Schölkopf, J. C. Platt, J. S.-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001, doi: 10.1162/089976601750264965.

[16] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.

[17] P. C. Mahalanobis, "On the generalized distance in statistics," *National Institute of Science of India*, vol. 2, pp. 49–55, 1936.

[18] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Advances in Knowledge Discovery and Data Mining*, vol. 7819, Springer, 2013. doi: 10.1007/978-3-642-37456-2_14.

[19] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," *SIGMOD Record*, vol. 29, no. 2, pp. 93–104, 2000, doi: 10.1145/335191.335388.

[20] S. Agarwal, "Data mining: data mining concepts and techniques," in *2013 International Conference on Machine Intelligence and Research Advancement (ICMIRA)*, 2013, pp. 203–207. doi: 10.1109/ICMIRA.2013.45.

[21] A. Aldweesh, A. Derhab, and A. Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: a survey, taxonomy, and open issues," *Knowledge-Based Systems*, vol. 189, p. 105124, 2020, doi: 10.1016/j.knosys.2019.105124.

[22] B. A. Tama, M. Comuzzi, and K.-H. Rhee, "TSE-IDS: a two-stage classifier ensemble for intelligent anomaly-based intrusion detection system," *IEEE Access*, vol. 7, pp. 94497–94507, 2019, doi: 10.1109/access.2019.2928048.

[23] H. Mohamed, A. Hamza, and H. Hefny, "An efficient intrusion detection approach using ensemble deep learning models for IoT," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 1, 2023. doi: 10.22266/ijies2023.0228.31.

[24] S. Ghosh, "Network traffic analysis based on cybersecurity intrusion detection through an effective automated separate guided attention federated graph neural network," *Applied Soft Computing*, vol. 169, p. 112603, 2025, doi: 10.1016/j.asoc.2024.112603.

[25] C. Magoo and P. Garg, "Machine learning adversarial attacks: a survey beyond," in *Machine Learning and the Internet of Medical Things in Healthcare*, Wiley, 2021. doi: 10.1002/9781119764113.ch13.

[26] K. Kharoubi *et al.*, "Network intrusion detection system using convolutional neural networks: NIDS-DL-CNN for IoT security," *Cluster Computing*, vol. 28, p. 219, 2025, doi: 10.1007/s10586-024-04904-7.

[27] A. Alghazali and Z. Hanoosh, "Using a hybrid algorithm with intrusion detection system based on hierarchical deep learning for smart meter communication network," *Webology*, vol. 19, no. 1, p. 253, 2022, doi: 10.14704/web/v19i1/web19253.

[28] M. Vishwakarma and N. Kesswani, "StaEn-IDS: an explainable stacking ensemble deep neural network-based intrusion detection system for IoT," *IEEE Access*, vol. 13, pp. 109713–109728, 2025, doi: 10.1109/ACCESS.2025.3582391.

[29] M. Zhong, M. Lin, C. Zhang, and Z. Xu, "A survey on graph neural networks for intrusion detection systems: methods, trends and challenges," *Computers & Security*, vol. 141, p. 103821, 2024, doi: 10.1016/j.cose.2024.103821.

[30] K. He, D. D. Kim, and M. R. Asghar, "NIDS-Vis: improving the generalized adversarial robustness of network intrusion detection system," *Computers \& Security*, vol. 145, p. 104028, 2024, doi: 10.1016/j.cose.2024.104028.

[31] O. Arreche, T. Guntur, and M. Abdallah, “XAI-IDS: toward proposing an explainable artificial intelligence framework for enhancing network intrusion detection systems,” *Applied Sciences*, vol. 14, no. 10, p. 4170, 2024, doi: 10.3390/app14104170.

[32] K. Roshan and A. Zafar, “Boosting robustness of network intrusion detection systems: a novel two phase defense strategy against untargeted white-box optimization adversarial attack,” *Expert Systems with Applications*, vol. 249, p. 123567, 2024, doi: 10.1016/j.eswa.2024.123567.

[33] H. Sun, H. Shu, F. Kang, Y. Zhao, and Y. Huang, “Malware2ATT&CK: a sophisticated model for mapping malware to ATT&CK techniques,” *Computers & Security*, vol. 140, p. 103772, 2024, doi: 10.1016/j.cose.2024.103772.

[34] N. Moustafa and J. Slay, “UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set),” in *2015 Military Communications and Information Systems Conference (MilCIS)*, 2015, pp. 1–6. doi: 10.1109/MilCIS.2015.7348942.

[35] S. Khalid, T. Khalil, and S. Nasreen, “A survey of feature selection and feature extraction techniques in machine learning,” in *2014 Science and Information Conference*, 2014, pp. 372–378. doi: 10.1109/SAI.2014.6918213.

[36] H. Henderi, T. Wahyuningsih, and E. Rahwanto, “Comparison of Min-Max normalization and Z-Score normalization in the k-nearest neighbor (kNN) algorithm to test the accuracy of types of breast cancer,” *International Journal of Informatics and Information Systems*, vol. 4, no. 1, pp. 13–20, 2021.

[37] R. Saia, S. Carta, D. R. Recupero, G. Fenu, and M. M. Stanciu, “A discretized extended feature space (DEFS) model to improve the anomaly detection performance in network intrusion detection systems,” in *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019)*, pp. 322–329. doi: 10.5220/0008113603220329.

[38] T. M. Cover and J. A. Thomas, *Elements of information theory, 2nd edition*. John Wiley & Sons, 2006.

[39] H. Han, W. Y. Wang, and B. H. Mao, “Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning,” in *Advances in Intelligent Computing. ICIC 2005*, vol. 3644, Springer, 2005. doi: 10.1007/11538059_91.

[40] L. McInnes, J. Healy, and J. Melville, “UMAP: uniform manifold approximation and projection for dimension reduction.” Journal of Open Source Software, vol. 3, no. 29, 2018. doi: 10.21105/joss.00861.

[41] M. Jumarlis *et al.*, “A hybrid hue saturation lightness, gray level co-occurrence matrix, and k-nearest neighbour for palm-sugar classification,” *IAES International Journal of Artificial Intelligence*, vol. 13, no. 3, pp. 2934–2945, 2024, doi: 10.11591/ijai.v13.i3.pp2934-2945.

[42] I. Mulyadi *et al.*, “A hybrid model for palm sugar type classification: advancing image-based analysis for industry applications,” *Ingénierie des systèmes d’information*, vol. 29, no. 5, pp. 1937–1948, 2024, doi: 10.18280/isi.290525.

[43] A. van den Oord, Y. Li, and O. Babbar, “Representation learning with contrastive predictive coding.” *arXiv*, 2018. doi: 10.48550/arXiv.1807.03748

[44] A. Kurakin, I. Goodfellow, and S. Bengio, “Adversarial examples in the physical world,” *arXiv*, 2017. doi: 10.48550/arXiv.1607.02533.

[45] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” in *Proceedings of the 31st International Conference on Neural Information Processing System*, Curran Associates Inc. pp. 4768-4777, 2017.

[46] B. Settles, “Active learning literature survey,” *University of Wisconsin-Madison, Computer Sciences Technical Report 1648*, 2009. [Online] Available: https://research.cs.wisc.edu/techreports/2009/TR1648.pdf

[47] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

[48] G. Karantzas and C. Patsakis, “An empirical assessment of endpoint detection and response systems against advanced persistent threats attack vectors,” *Journal of Cybersecurity and Privacy*, vol. 1, no. 3, pp. 387–421, 2021, doi: 10.3390/jcp1030021.

[49] M. François, P.-E. Arduin, and M. Merad, “Physics-informed graph neural networks for attack path prediction,” *Journal of Cybersecurity and Privacy*, vol. 5, no. 2, p. 15, 2025, doi: 10.3390/jcp5020015.

BIOGRAPHIES OF AUTHORS

Abd Rahman Wahid is a Computer Science student at Muhammadiyah University Makassar, Indonesia, a researcher, and an active member of COCONUT Computer Club, Indonesia. His research interests include cybersecurity, including networks, applications, infrastructure, and cryptography, as well as artificial intelligence across various sectors. He is working to develop a future cybersecurity system model that can operate synergistically, adaptively, and dynamically. He can be contacted at email: 105841116522@student.unismuh.ac.id.

Desi Anggreani is completed her bachelor's degree at the University of Muslim Indonesia (UMI) Makassar in the Computer Science Program, Faculty of Computer Science. Subsequently, the author continued her master's degree at the University of Malang in the Electrical Engineering Program, Faculty of Engineering. The author's field of expertise includes information technology, artificial intelligence (AI), and data science. The author is actively involved in research and development activities, particularly in the application of digital technology across various sectors. The primary focus of the author's research lies in the development of AI-based applications, decision support systems, and AI-based forecasting. She can be contacted at email: desianggreani@unismuh.ac.id.

Muhyiddin A. M. Hayat Holds a Master's of Engineering from the University of Hassanuddin in 2016. He also completed his bachelor of Computer Science from Universitas Veteran Republik Indonesia in 2007. Since 2016, He has lectured in the Department of Informatics at Universitas Muhammadiyah Makassar. His research interests include algorithm modeling, machine learning, and data mining. He can be contacted at muhyiddin@unismuh.ac.id.

Aedah Abd Rahman is a Professor in the School of Science and Technology, Asia e University, Kuala Lumpur, Malaysia. Have expertise in data mining and software engineering. She can be contacted at email: aedah.abdrahman@aeu.edu.my.

Muhammad Faisal obtained his bachelor of Information Systems (S.SI) in the Information Systems Department from STMIK Profesional, Makassar, Indonesia, in 2011, and his Master of Engineering (M.T) in Informatics Engineering Department from Universitas Hasanuddin, Gowa, Indonesia, in 2014. He holds a Ph.D. degree from the School of Science and Technology, Doctoral Programme, Asia E University Malaysia. He is currently a lecturer and researcher at the Department of Informatics, Universitas Muhammadiyah Makassar, Indonesia. His research interests include artificial intelligence, data mining, decision support systems, image processing, deep learning, and machine learning. He can be contacted at email: muhfaisal@unismuh.ac.id.