STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, December 2026, pp 5538–5560.
Published online in International Academic Press (www.IAPress.org)

Enhancing IoT Intrusion Detection System using ASAE-XGBoost for Binary and Multi-class Attack Classification

Aya NEJARI [1,*], Abdelali ELMOUNADI [1,2], Abdelmounaim BADAOUI [1], Naoual BERBICHE [1]

[1] *LASTIMI Laboratory, ESTS, Mohammed V University, Salé, Morocco*
[2] *École Normale Supérieure of Rabat, Mohammed V University, Rabat, Morocco*

Abstract As the number of Internet of Things (IoT) devices continues to grow, the IoT network has become more vulnerable to cyber threats, which requires reliable intrusion detection systems (IDSs) to protect IoT environments. Despite their extensive use, traditional IDSs still face challenges in detecting unknown attacks and accurately classifying different attack categories. To bridge these challenges, we propose a hierarchical two-stage intrusion detection framework based on the integration of the Asymmetric Stacked AutoEncoder (ASAE) with XGBoost for binary anomaly detection and multi-class attack classification. The ASAE is trained only on normal traffic to identify anomalous traffic patterns, while the XGBoost classifier uses the learned latent representations to classify the detected attacks into their corresponding categories. We evaluate the proposed framework on two real-world datasets, TON_IoT Network and IOTID20, using a standardized experimental protocol. Results show that the proposed framework demonstrates reliable binary anomaly detection performance, accurate attack classification along with a low computational cost across both datasets. Reliable performance on both datasets confirm that our framework is a lightweight, robust and effective intrusion detection system in real-time diverse IoT network environments.

Keywords Internet of Things (IoT), Intrusion Detection System (IDS), Asymmetric Stacked AutoEncoder (ASAE), Anomaly detection, XGBoost, Multi-Class Attack Classification

DOI: 10.19139/soic-2310-5070-4365

1. Introduction

The Internet of Things (IoT) represents a transformative technology, which links billions of physical devices to the Internet, facilitating real-time communication between objects, systems and users, including industrial sensors and smart home devices [1]. This technology can be characterized as a large-scale interconnected system of sensors and actuators deployed across diverse application domains, including agriculture, healthcare, automotive, and smart grid infrastructures [2]. This interconnected system allows real-time automation, continuous monitoring and intelligent decision-making in IoT environments [3]. Nevertheless, its operational benefits are challenged by the open architecture, heterogeneity and wide distribution of IoT networks, which introduce significant cybersecurity challenges [4].
IoT environments face growing cyber threats that include Denial of Service (DoS), Distributed Denial of Service (DDoS), and Man-in-the-Middle (MiTM) attacks [5, 6, 7]. These threats represent a significant risk to production operations while simultaneously endangering sensitive data and causing harm to physical infrastructure. The traditional security measures including firewalls and signature based detection systems fail to effectively address such dynamic and distributed threats [8].

*Correspondence to: Aya NEJARI (Email: aya.nejari@research.emi.ac.ma). LASTIMI Laboratory, ESTS, Mohammed V University, Salé, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 2026 International Academic Press

The protection of IoT networks requires intrusion detection systems as an essential security measure [4]. The monitoring function of IDS systems or networks enables the detection of abnormal and dangerous activity. There are different types of IDS: some rely on known attack patterns (signature-based), others look for unusual activities in system behavior (anomaly-based), and some combine both to improve detection results [9, 10]. In IoT systems, intrusion detection plays a crucial role in ensuring safety by monitoring the traffic flowing through the network. However, these systems face significant challenges: they must manage encrypted data that is not always accessible, process vast volumes of information at high speeds, and operate on small devices with limited computational power. Addressing these issues is essential for maintaining effective security in dynamic and resource-constrained environments [11].

Machine Learning methods have been extensively used to improve the performance of Intrusion Detection Systems (IDS) by providing the ability to detect both known and unknown attacks [12, 13]. They can learn attack patterns from the data and then identify unusual network activities and behaviors [12]. Deep Learning models can also detect attacks in IoT networks more accurately by learning complex and non-linear patterns from IoT network traffic data, hence they are highly suitable for large scale and dynamic network environments [14]. Importantly, lightweight and hybrid IDS architecture are more and more seen as practical solutions for real-world deployment [15].

In intrusion detection, autoencoders are widely employed because they can learn compact latent representations that represent the normal behavior of network traffic without using labeled attack data [16]. During training, autoencoders minimize the reconstruction error between the input data and the reconstructed data and capture normal behavior of network traffic. Therefore, traffic deviating from the learned representation is classified as abnormal. Autoencoders rely on unsupervised learning and are trained on normal traffic, making them well suited for identifying abnormal traffic that deviates from the normal behavior learned during the training. This detection capability is especially important in IoT environments, where large-scale of labeled data for unknown (zero day) attacks are rarely available which increases the difficulty of identifying new threats [17]. This advantage allows autoencoders to be successfully deployed in IoT environments [16]. However, autoencoders still face several limitations. Their detection performance depends on reconstruction errors, reducing their ability to accurately separate normal and anomalous traffic share similar characteristics. Moreover, thresholds derived from reconstruction errors are affected by class imbalance, evolving attack patterns and noisy data reducing detection reliability in complex IoT environments [16, 18].

To address these limitations, recent studies focused on hybrid intrusion detection framework that merge anomaly detection models with supervised machine learning classifiers [19]. These frameworks combine the strengths of unsupervised learning for anomaly detection with supervised classifiers for accurate attack classification and enhancing intrusion detection performance [19, 20, 21]. Even though hybrid intrusion frameworks obtained strong performance, existing studies pay limited attention to the development of hierarchical frameworks where anomaly detection is first implemented followed by the type of identification attack.

To bridge this gap, this paper proposes a hierarchical two-stage intrusion detection framework based on the integration of the Asymmetric Stacked AutoEncoder (ASAE) with XGBoost for binary anomaly detection and multi-class attack classification. As an initial stage, the ASAE learns only from normal traffic to identify abnormal traffic using reconstruction errors. In the second stage, only the traffic instances detected as abnormal by the ASAE are forwarded to the XGBoost classifier, which assigns them to their corresponding attack categories. The proposed framework is evaluated in two real- world datasets, TON_IoT Network and IOTID20, using a standardized experimental protocol. The experimental results demonstrate its effectiveness in both binary and multi-class attack classification, while maintaining stable performance across different intrusion detection scenarios.

The key contributions of the work are listed as follows:

- A hierarchical two-stage intrusion detection framework is proposed by integrating the ASAE with the XGBoost classifier for binary anomaly detection and multi-class attack classification in IoT environments.
- The ASAE learns only from normal traffic to identify abnormal traffic using reconstruction errors, while XGBoost identifies the corresponding attack types using the latent representations learned by the ASAE for the traffic instances detected as abnormal.
- The proposed framework is evaluated in two real-world datasets using a standardized experimental protocol.

- Experimental results demonstrate its effectiveness in both binary and multi class attack classification by using two datasets, TON_IoT and IoTID20.

The remainder of this paper is organized as follows: Section 2 explores the current literature. Section 3 presents the proposed framework. Section 4 focuses on the experimental results and discussion. Section 5 outlines the limitations of the proposed framework, offering insights and suggesting future research directions. Finally, section 6 concludes the paper.

2. Related work

A wide range of studies have explored machine learning, deep learning and hybrid framework to enhance intrusion detection systems. A critical review of recent studies is presented below, focusing on the strengths and weaknesses of the existing approaches.

In [22], the authors introduced an XGBoost-based network intrusion detection system for IoT environments to enhance multi-class attack detection while enhancing interpretability and computational efficiency. The proposed framework integrates feature dimensionality reduction, class balancing and SHAP-based interpretability analysis to tackle the challenges of high-dimensional network traffic and class imbalance. This model was evaluated on ToN-IoT dataset and achieved high multi- class detection performance while ensuring interpretability and computational efficiency and offering transparent decision-making through SHAP analysis. However, this framework employs XGBoost directly on the preprocessed and balanced labeled dataset without first detecting abnormal traffic through an anomaly detection stage.

Based on the concept of the deep learning approaches, Ayad et al. [23] present an efficient and lightweight intrusion detection model that merges an asymmetric stacked autoencoder with a Deep Neural Network (DNN) employing one-class learning for real-time anomaly detection in IoT environments. The ASAE is employed for dimensionality reduction of network traffic and to learn compact latent representations from normal traffic, while the OC-DNN learns normal behavior patterns to detect abnormal traffic. The authors used the Data Driven Threshold Determination (DDTD) method to enhance anomaly detection performance and they evaluated their model on the BoT-IoT dataset and achieved a high detection rate with few false positives and ensured fast response times. These results confirm that this model is suitable for deployment in resource-constrained IoT environments. Nevertheless, the proposed framework is limited to binary anomaly detection, restricting the final decision on normal and abnormal traffic and does not extend a supplement stage to attack category classification.

To enhance binary intrusion detection in IoT environments, this study [24] introduces a hybrid CNN-LSTM framework that integrates preprocessing pipeline with deep learning-based classification. The CNN extracts representative spatial features from network traffic, while the LSTM models temporal dependencies to enhance intrusion detection performance. The proposed framework is evaluated on two benchmark datasets namely, ToN_IoT and UNSW-NB15, and demonstrated strong binary detection performance across both datasets. However, this framework is limited to binary intrusion detection and lacks a multi-class attack stage after anomaly detection.

A hybrid intrusion detection framework is presented in [20]. This framework combines an Autoencoder-Gaussian Mixture Model (AE-GMM) with two supervised learning techniques, XGBoost and Logistic Regression to enhance intrusion detection performance. Firstly, the autoencoder is designed to learn compressed latent representations of normal network traffic, while the GMM was then applied to the latent representations of the normal data to model their probability distribution. This framework uses the extracted latent features as input to XGBoost and Logistic Regression for supervised attack classification. The proposed framework's performance was evaluated using the NSL-KDD and UNSW-NB15 datasets and achieved the best detection performance across both datasets and among the three evaluated models. While the proposed framework is evaluated only in a binary intrusion detection, it does not identify more information of attack categories after anomaly detection.

The authors at [25] proposed a robust hybrid intrusion detection system that employs an advanced hybrid architecture integrating convolutional neural networks (CNN) with multilayer perceptron (MLP) especially tailored for IoT networks which enables accurate detection and classification across both binary and multi-class IoT network traffic. The CNN component extracts and enhances features from network traffic data, while the MLP

the final classification task. Regarding class imbalance, this model incorporates the enhanced hybrid adaptive synthetic sampling-synthetic minority oversampling technique (ADASYN-SMOTE) for binary classification, an advanced ADASYN for multi-class classification and applies edited nearest neighbors (ENN) with class weights. The proposed framework was tested on the IoT-23 and NF-BoT-IoT-v2 datasets and achieved high accuracy across both binary and multi-class classification, which demonstrates its effectiveness as a robust solution for enhancing IoT network security. However, the proposed framework is based on a two-stage supervised classification and does not explore unsupervised anomaly detection methods.

Motivated by the need to enhance anomaly-based intrusion detection performance, Rahman et al. [26] introduced a Stacked Intrusion Detection System (SIDS) that relies on multiple machine learning and deep learning models including Stochastic Gradient Descent (SGD), Logistic Regression (LR), Random Forest (RF), and Deep Neural Network (DNN). The proposed framework relies on a stacked ensemble strategy that combines the predictions of the different models to enhance intrusion detection performance. The proposed framework's performance was evaluated using the NSL-KDD dataset and achieved high detection performance in terms of accuracy, precision, recall, and F1-score. Nevertheless, the proposed framework combines multiple models within a stacking framework for intrusion detection rather than separating anomaly detection from the following attack classification into two sequential stages.

To enhance both intrusion detection performance and model interpretability, the authors at [27] proposed an explainable Deep Neural Network (DNN) framework integrated with the Shapley Additive Explanations (SHAP) method. In the proposed framework, the DNN performs network intrusion detection, while SHAP provides both global and local explanations of how network traffic features contribute to the model's predictions. The proposed framework was tested on the CIC-IDS2017 dataset and achieved high intrusion detection performance while maintaining computational efficiency and model interpretability. However, this framework is limited to binary classification between normal and abnormal traffic and does not extend a supplement stage to classify different attacks categories.

Inspired by these insights, the following section describes the proposed hierarchical ASAE-XGBoost framework along with its overall architecture.

3. The proposed framework

As mentioned above, we have proposed a hierarchical ASAE-XGBoost intrusion detection framework that integrates the ASAE with the XGBoost classifier for binary anomaly detection and multi-class attack classification in IoT environments. The process starts with data preprocessing, including data cleaning, one hot encoding, and label encoding for the attack labels, followed by a train-test split. Then, the training split is divided into training (80%) and validation (20%) to train the model and determine the optimal threshold. In addition, before training our ASAE, we apply a standard scaler, which is fitted only on the normal training samples and then applied to the validation split and test split to maintain a standardized uniform feature representation while avoiding data leakage. As a first stage, the ASAE is trained only on the normal training samples, while the Data Driven Threshold Determination (DDTD) method is applied to determine the optimal threshold on the validation split. We employ, then, the selected threshold on the unseen test split to identify normal and abnormal traffic. Finally, as a second stage, the XGBoost classifier is trained on the latent space learned by the ASAE encoder, using the multi-class labels and only the traffic samples identified by the ASAE as abnormal are forwarded to the XGBoost to classify attack types. Figure 1 below depicts the overall architecture of the proposed framework. In the following, we will discuss each step shown in Figure 1, starting by the first stage, the ASAE-based anomaly detection.

3.1. ASAE-based binary anomaly detection

As an initial stage, we designed an Asymmetric Stacked Autoencoder (ASAE) for binary anomaly detection. Unlike traditional supervised intrusion detection methods, which require labeled data to classify network traffic as normal or abnormal. However, acquiring labeled data for IoT environments remains challenging and difficult to accomplish in practice. Our ASAE addresses this limitation by relying on an unsupervised autoencoder to learn

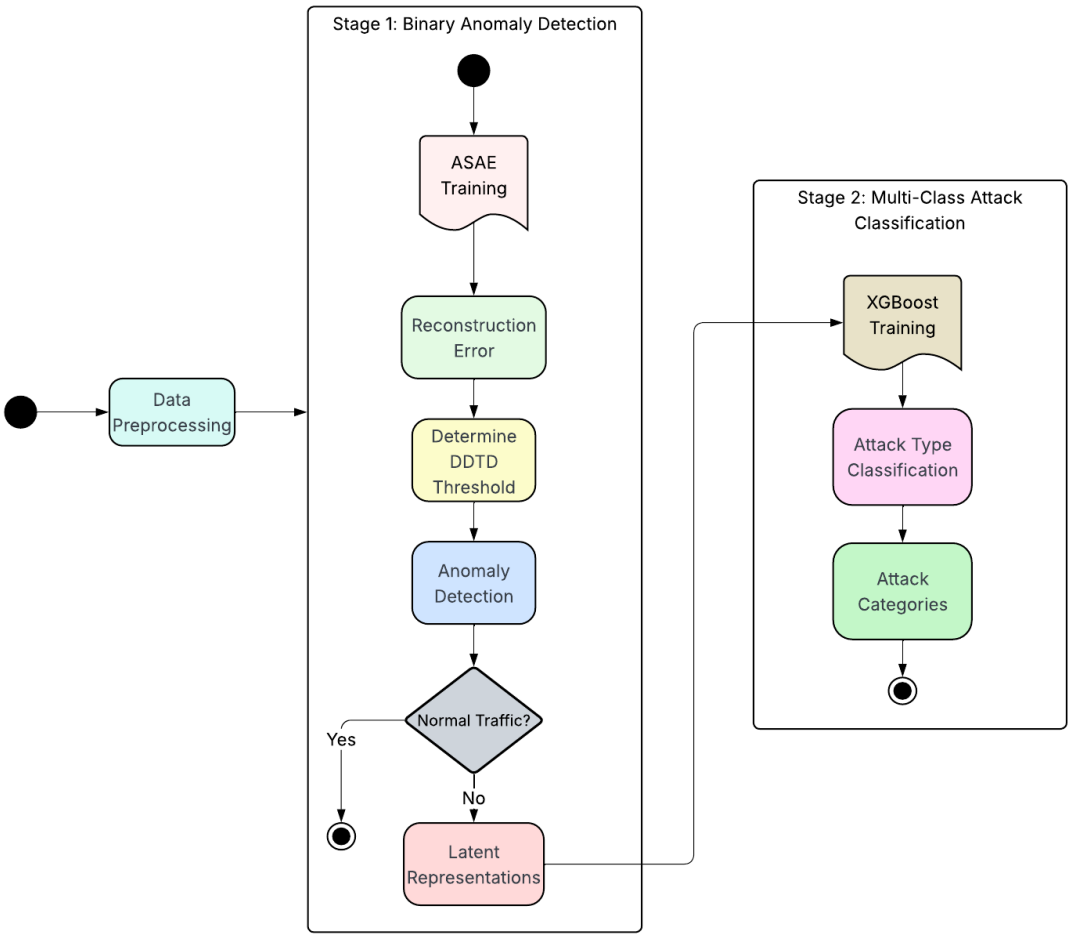


Figure 1. Architecture of the proposed framework.

the representative patterns of normal network traffic. In addition, the proposed ASAE employs a one-class learning strategy trained only on normal traffic to bridge the challenges associated with class imbalance in IoT intrusion detection.

Table 1 and Table 2 present the architecture and training configuration of the proposed ASAE. It relies on an encoder-decoder architecture that learns compact latent representation of normal network traffic while minimizing the reconstruction error. The encoder learns a compressed latent representation of the input feature vector, while the decoder reconstructs the original input from this learned representation. This asymmetric architecture allows the model to capture the more discriminative features of normal network traffic while enhancing its reconstruction capability. This ASAE employs ReLu activations functions across all hidden layers to learn discriminative latent representations, while the output layers employ a linear activation function to reconstruct the original features.Then, the encoder employs L1 activity regularization to generate sparse latent representations, while enhancing the model's ability to detect anomalies. The ASAE is optimized through the Adam optimizer to minimize the Mean Squared Error (MSE) loss function. Moreover, we employ early stopping based on the validation loss to avoid overfitting and restore the best-performing model weights.

The ASAE is trained only on normal traffic employing a one-class learning strategy. It learns the representative patterns on normal network traffic. As a result, this model builds a compact representation of normal network

Table 1. Model Architecture

Parameters	Encoder	Decoder
Layers	2	4
Neurons per layer	Input_dim, 64, 20	64, 128, 256, 512, input_dim
Activation function	ReLU	ReLU
Output activation	–	Linear
L1 Activity regularization	1×10^{-5}	–

Table 2. Model Training

Parameters	Value
Epochs	100
Optimizer	Adam
Loss function	MSE
Batch Size	32

behavior, which serves as a reference for anomaly detection. Each traffic sample is forwarded to the trained ASAE for reconstruction. As it is trained only on normal traffic, normal instances are reconstructed along with low reconstruction errors. Unlike normal traffic, anomalous samples differ from those learned by the model, thereby producing higher reconstruction errors. Therefore, this reconstruction error serves as an indicator to discriminate between normal and abnormal traffic. The reconstruction error is determined using the MSE between the input sample and the corresponding reconstructed output as follows:

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(x_j - \hat{x}_j)^2 \tag{1}$$

Whereby x denotes the input traffic sample, $\hat{x}_j$ denotes the corresponding reconstructed output and N presents the number of input features.

As mentioned above, the ASAE is trained to reconstruct normal traffic with low reconstruction error. In the validation split, we determined an optimal threshold to separate between normal and abnormal traffic using the Data Driven Threshold Determination (DDTD) method [23]. This method starts by evaluating the reconstruction error for each sample and then generates the associated Precision-Recall curve. After that, the thresholds associated with the intersection between the precision the optimal reconstruction error threshold [23]. Finally, the selected threshold is applied on the test split to classify traffic samples. Traffic is classified as normal when its reconstruction error is below the threshold, while those exceeding the threshold are selected as abnormal.

3.2. XGBoost-based attack classification

As a second stage, we forwarded only the traffic samples selected as anomalous by the ASAE to the XGBoost classifier for multi-class attack classification. While the ASAE is designed for one-class anomaly detection, XGBoost relies on supervised learning with multi-class attack labels. This classifier is trained on the latent space generated by the encoder rather than the original feature space, thereby reducing the feature dimensionality while enhancing computational efficiency and facilitating attack classification.

Following ASAE training, latent representations of the training samples are extracted from the encoder, which are then applied as an input to the XGBoost classifier. The hyperparameters of XGBoost are optimized using the Optuna framework. In each Optuna trial, an objective function sampled a set of candidate hyperparameters, including the number of trees n_estimators (50–200), the maximum tree depth max_depth (3–7), the learning rate learning_rate (0.01-0.3, log-scaled), the subsampling ratio subsample (0.6–1.0), the column subsampling ratio

colsample_bytree (0.6–1.0), the minimum child weight min_child_weight (1–10), the splitting penalty gamma (0.0–5.0), and the L1/L2 regularization terms reg_alpha and reg_lambda (0.0–5.0). For each hyperparameter configuration, the latent training samples were further split into an internal 80/20 train-validation partition (stratified, random_state = 42), whereby 80% of the training data was used for training the corresponding XGBoost model and the remaining 20% was used for validation set to test its performance when optimizing hyperparameters with Optuna. The weighted F1 score on this validation split served as the optimization objective and was maximized by Optuna over 50 trials. After the optimization process converged, the best hyperparameter configuration (retrieved from study.best_trial.params) was combined with fixed settings: objective = "multi:softprob", eval_metric = "logloss", and random_state = 42 for training the best XGBoost classifier. Finally, only the traffic samples detected as anomalous by the ASAE are encoded into the latent space and classified by the XGBoost classifier into their corresponding attack types.

3.3. Hierarchical ASAE-XGBoost framework

Lastly, as illustrated in Figure 1, we propose a hierarchical two-stage intrusion detection framework based on the integration of the ASAE with XGBoost for binary anomaly detection and multi-class attack classification. At the first stage, the ASAE classifies network traffic as normal or abnormal based on the reconstruction error and threshold determined through the DDTD method. Following anomaly detection, only the traffic samples detected as anomalous by the ASAE are encoded into the latent space and classified by the XGBoost classifier into their corresponding attack types. This proposed hierarchical architecture combines the strengths of one-class anomaly detection with the effectiveness of supervised multi-class classification to identify different attack categories. In addition, the proposed hierarchical framework enhances computational efficiency by forwarding only the traffic samples detected as abnormal to the second stage. Furthermore, the compact latent representations learned by the ASAE reduce feature dimensionality while enhancing attack classification.

4. Experimental results and discussion

In this section, we introduce the experimental results of the proposed hierarchical ASAE-XGBoost framework. Firstly, we describe the benchmark datasets followed by the associated data preprocessing steps. Secondly, we present the experimental results for both binary anomaly detection and multi-class attack classification on each dataset. Finally, we analyze the computational cost of the proposed framework followed by a benchmarking against recent frameworks.

4.1. Benchmark datasets

The simulation validation of the proposed approach is evaluated on two real-world datasets, TON_IoT Network and IOTID20, using a standardized experimental protocol. The TON_IoT Network dataset [28] is a modern dataset built to test intrusion detection systems in authentic IoT and IIoT settings. It brings together a mix of data drawn from IoT/IIoT sensors, Linux and Windows hosts, and network traces captured in a large scale industrial IoT testbed. In this work, we concentrate on the Train_Test_Network subset, featuring labeled network flows categorized into two classes: normal and attack (abnormal), and covering multiple attack types including backdoor, DDoS, DoS, injection, MiTM, password, ransomware, scanning, and XSS. This dataset contains 211043 network flow records that include 50000 normal traffic and 161043 malicious traffic records resulting in a 76.31% attack ratio. The malicious traffic data spreads across different attack types with 20000 examples for each major attack category (backdoor, DDoS, DoS, injection, password, ransomware, scanning, and XSS) while MiTM represents teh most imbalanced category with 1043 instances. The dataset contains 44 original network-flow features, including source and destination IP addresses and ports, protocol information, traffic size, connection-state indicators, HTTP metadata, and a label identifying normal or malicious behavior. In the present work, anomaly detection is conducted using binary labels, while multi-class labels are used for attack classification.
The IOTID20 dataset [29] is a modern dataset built to test intrusion-detection systems in realistic IoT settings. It was built from a realistic samrt-home environment including IoT devices and networked devices. This dataset

includes network traffic obtained from both normal network activities and malicious atatck scenarios. In this work, we concentrate on the binary classification setting, where each network traffic is grouped into two categories: normal traffic and anomalous traffic. The anomalous traffic covers multiple attack types including DoS, MiTM ARP Spoofing, Mirai, and Scan. This dataset contains 625783 network flow records featuring 86 original network-flow features including source and destination IP addresses and ports, protocol information, traffic statistics and labeling attributes identifying normal and anomalous behavior. Similarly to the TON_IoT Network dataset anomaly detection is conducted using binary labels, while multi-class labels are used for attack classification.

4.2. Data preprocessing

Following the datasets descriptions, a number of preprocessing steps were applied before model training to enhance data quality while maintaining a standardized representation of network traffic across both benchmark datasets. Firstly, we applied a data cleaning process to remove duplicate records, handle missing and infinite values, and eliminate high-cardinality features with limited contribution to model learning. Next, we encoded the remaining categorical features using one-hot encoding. After that, binary labels were used for anomaly detection, while the multi-class labels were maintained for attack classification. Then, the datasets were partitioned into a stratified training and testing split to maintain the class distribution. Following this, the training split is divided into training and validation sets based on the same stratified strategy. Finally, we applied Standard Scaler to normalize the features using only the normal training samples for fitting, and applying the same transformation to the validation and test sets to avoid data leakage.

4.3. Performance on TON_Iot Network Dataset

The performance of the proposed hierarchical ASAE-XGBoost is first evaluated on the TON_IoT Network dataset. This evaluation reports the performance of both the ASAE for binary anomaly detection and the XGBoost classifier for multi-class attack classification.

4.3.1. Binary anomaly detection

Firstly, we evaluate the binary anomaly detection performance of the ASAE using various metrics including accuracy, precision, recall, F1-score, ROC-AUC and PR AUC. Table 3 below summarizes the performance achieved during the binary anomaly detection stage.

Table 3. Binary anomaly detection results on TON_IoT Network Dataset

Model	**ACC (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**FPR (%)**	**FNR (%)**	**ROC-AUC (%)**	**PR-AUC (%)**
ASAE	98.35	98.97	98.92	98.94	3.65	1.08	98.24	98.67

As illustrated in Table 3, these results demonstrate that the ASAE identifies successfully abnormal traffic reaching an accuracy of 98.35%, along with a precision of 98.97%, a recall of 98.92% and a F1-score of 98.94%. In addition, ROC-AUC and PR-AUC scores of 98.24% and 98.67% confirm the model's effectiveness discriminative ability. These low false positive (3.65%) and false negative (1.08%) rates confirm the ability of the ASAE to detect the majority of attacks while generating only a limited number of false alarms.

Figure 2 below shows the confusion matrix of the ASAE and confirms its effectiveness. It shows that the model correctly classified 44048 attack samples and 12152 normal samples, with only 460 false positives and 483 false negatives. These results demonstrate the effectiveness of the ASAE in separating normal from abnormal traffic with only a few samples being misclassified.

Figure 3 hereafter presents the Receiver Operating Characteristic (ROC) curve of the ASAE, demonstrating its effectiveness in discriminating between normal and abnormal traffic across various decision thresholds. The ASAE achieved a ROC-AUC of 98.24%, while the ROC curve remained near the upper-left corner reflecting its strong detection capability.

Finally, Figure 4 illustrates the Precision-Recall (PR) curve of the ASAE, demonstrating its effectiveness in maintaining high precision and recall across various decision thresholds. The ASAE achieved a PR-AUC of

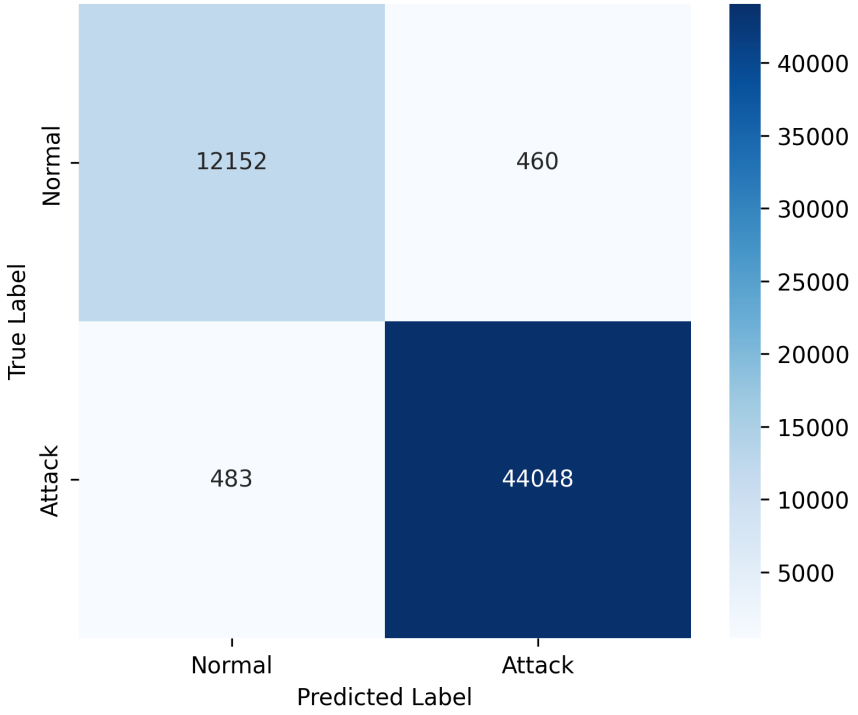


Figure 2. ASAE Confusion Matrix on TON_IoT Network Dataset.

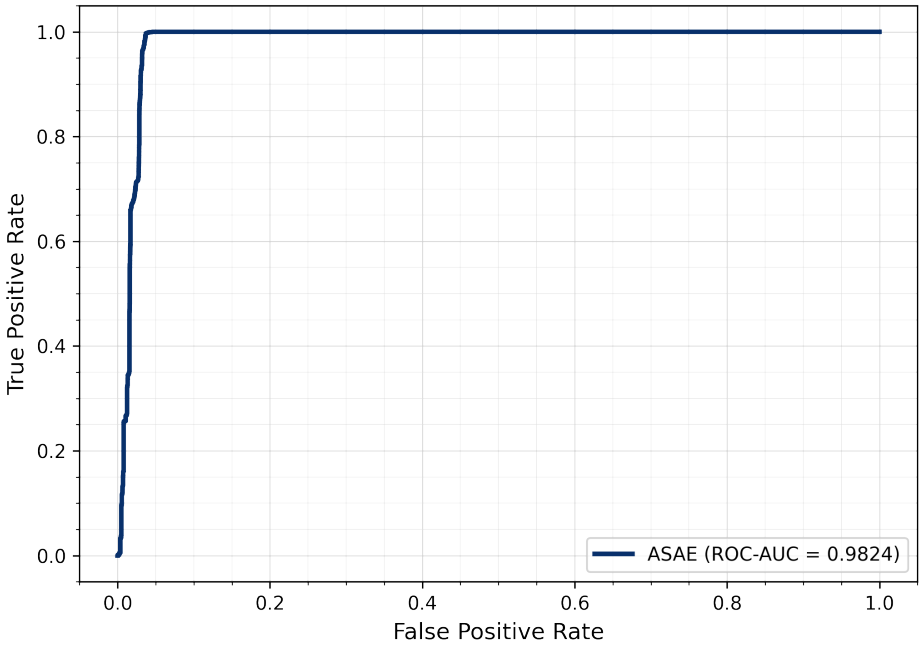


Figure 3. ASAE ROC curve on TON_IoT Network Dataset.

98.67%, while the PR curve remained near the upper-right corner reflecting its strong balance between precision and recall in anomaly detection.

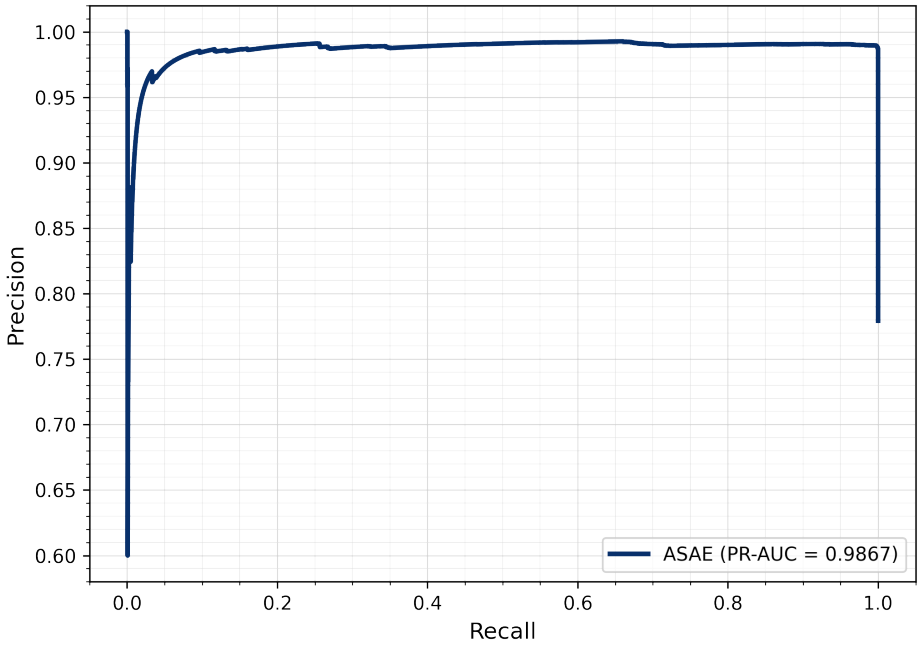


Figure 4. ASAE PR curve on TON_IoT Network Dataset.

4.3.2. Multi-class attack classification

Once abnormal traffic has been detected, the XGBoost classifier identifies the attack category associated with each anomalous sample. Table 4 below summarizes the performance achieved for each attack category during the multi-class attack classification stage.

Table 4. Multi-class attack classification results on TON_IoT Network Dataset

Class	Metrics							
	ACC (%)	Precision (%)	Recall (%)	F1-score (%)	FPR (%)	FNR (%)	ROC-AUC (%)	PR-AUC (%)
Backdoor	99.96	99.98	99.72	99.85	0.00	0.28	100.00	99.99
DDoS	99.07	96.37	96.66	96.52	0.56	3.34	99.83	98.98
DoS	99.32	97.32	97.22	97.27	0.38	2.78	99.88	99.61
Injection	98.80	95.25	95.77	95.51	0.73	4.23	99.80	98.98
MiTM	99.64	76.13	73.52	74.80	0.17	26.48	99.84	84.85
Normal	99.80	92.63	87.39	89.93	0.07	12.61	99.93	96.42
Password	99.42	97.79	97.83	97.81	0.34	2.17	99.92	99.57
Ransomware	99.96	99.79	99.81	99.80	0.02	0.19	100.00	99.98
Scanning	99.35	97.48	97.63	97.55	0.39	2.37	99.92	99.58
XSS	99.65	98.34	98.25	98.30	0.19	1.75	99.98	99.86

As shown in Table 4, these results demonstrate that the XGBoost classifier achieved strong multi-class

classification performance. We note that all attack categories accuracy above 98%, while precision, recall and F1-score maintained high values across most attack categories. Particularly, we observed that Backdoor, DDoS, DoS, Injection, Password, Ransomware, Scanning and XSS classes achieved F1-score above 95%, indicating that the latent representations learned by the ASAE are well suited for identifying different attack categories. However, the MiTM class achieved lower precision, recall and F1-score values, reflecting the difficulty of discriminating it from the other classes, which may be explained by similarities between its latent represnetations and those associated with others categories. Regarding the Normal class, its lower precision, recall and F1-score can be explained by the fact that it represents normal instances identified as abnormal by the ASAE in the first stage. It is included in the second stage to allow XGBoost to correclty reclassify these instances and consequently reduce false positives, while enhancing the final classification decision. In addition, ROC-AUC and PR-AUC values nearly reaching 100% confirms the strong discriminative capability of the learned latent representations. While the MiTM class achieved lower PR-AUC than the other attack categories, the corresponding ROC-AUC remained strong demonstrating the strong discriminative capability of the XGBoost classifier. Finally, we can observe that most attack categories have low false positive and false negative rates reflecting the reliable classification performance of the XGBoost classifier.

Figure 5 displays the multi-class confusion matrix of the XGBoost classifier. It exhibits high values along the main diagonal indicating that traffic samples are correctly classified to their respective categories. The MiTM presents more misclassifications than the other attack categories, reflecting its lower precision, recall, F1-score reported in Table 4. Despite these misclassifications, most MiTM samples are still correctly identified. Finally, the confusion matrix demonstrates the effectiveness of the XGBoost classifier in multi-class attack classification.

Figure 6 illustrates the one-vs-rest confusion matrices of the XGBoost classifier for each attack category. The classifier correctly classifies most traffic samples across all categories, which is illustrated by the main diagonal of each confusion matrix. This indicates that the learned latent representations allow the XGBoost classifier to accurately identify each attack category. We observed that the Backdoor and Ransomware classes are classified with the highest accuracy, with only a limited number of false positives and false negatives. In the same way, the DDoS, DoS, Injection, Password, Scanning and XSS are accurately identified, demonstrating the effectiveness of the XGBoost classifier in identifying different attack categories. In addition, the MiTM and Normal classes show more false negatives than the other categories, indicating that these classes are more challenging for classification. Finally, these one-vs-rest confusion matrices confirm the effectiveness of the XGBoost classifier to accurately classify each attack category.

Figure 7 shows the one-vs-rest ROC curves of the XGBoost classifier for each attack category. We observed that all curves remained near the upper-left corner demonstrating an excellent discrimination capability between each traffic category and the other classes. For all traffic categories, ROC-AUC scores vary from 99.80% to 100% demonstrating the excellent discriminative capability of the XGBoost classifier across all categories. Specifically, the Backdoor and Ransomware classes achieved a best ROC-AUC of 100%, while the other categories obtained scores over 99.80%. Finally, this indicates that the learned latent representations offer effective discriminative features allowing the XGBoost classifier to accurately identify each attack category.

Finally, Figure 8 below illustrates the one-vs-rest PR curves of the XGBoost classifier for each attack category. As we can see, most curves remained near the upper right corner demonstrating an effectiveness in maintaining high precision and recall across most traffic categories. For all traffic categories, PR-AUC scores vary from 98.98% to 99.99% for the Backdoor, DDoS, DoS, Injection, Password, Ransomware Scanning, and XSS classes demonstrating an excellent multi-class classification performance. Conversely, the MiTM and Normal classes achieved lower PR-AUC scores of 84.85% and 96.42%, respectively, indicating that these classes are more challenging to identify accurately. Finally, PR-AUC scores maintain high, indicating that this classifier achieves a strong balance between precision and recall across most traffic categories.

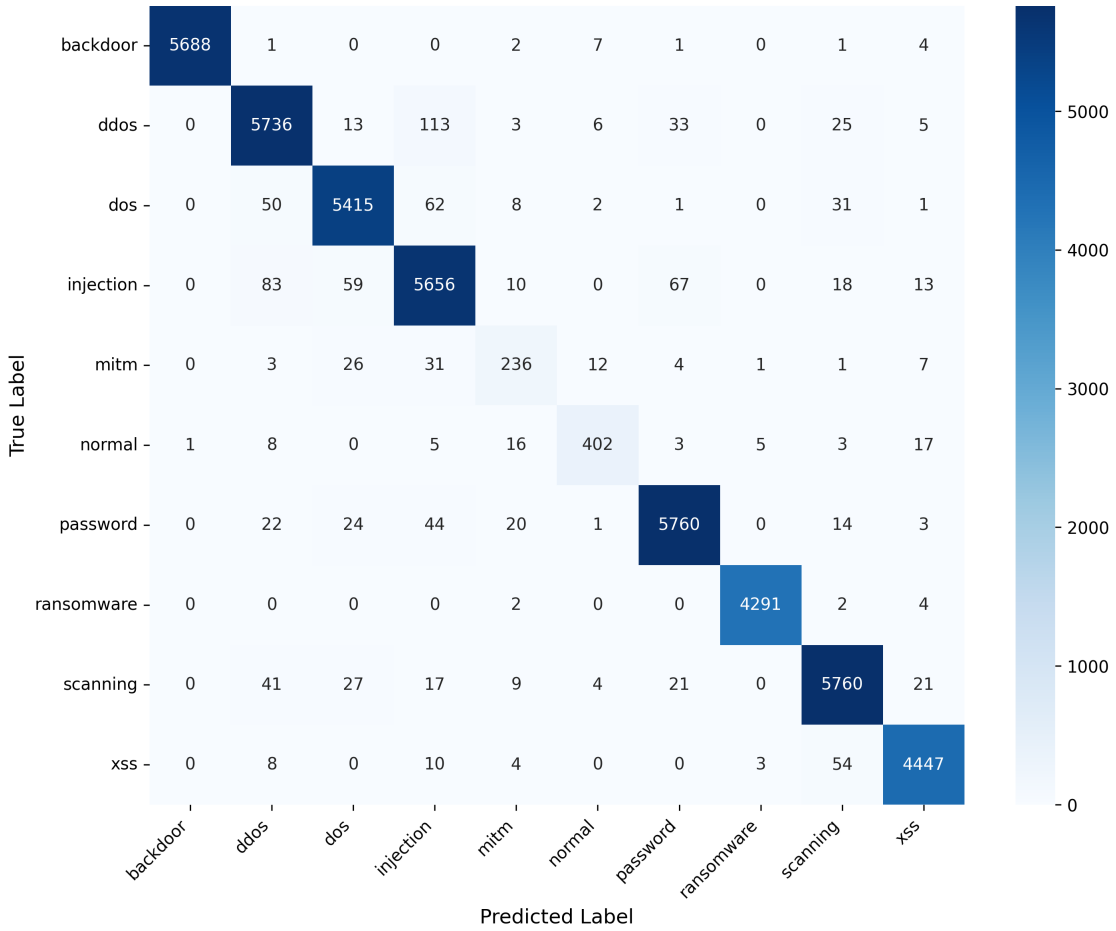


Figure 5. Multi-class Confusion Matrix on TON_IoT Network Dataset.

4.4. Performance on IOTID20 Dataset

The performance of the proposed hierarchical ASAE-XGBoost is then evaluated on the IOTID20 dataset. This evaluation reports the performance of both the ASAE for binary anomaly detection and the XGBoost classifier for multi-class attack classification.

4.4.1. Binary anomaly detection

Firstly, we evaluate the binary anomaly detection performance of the ASAE using various metrics including accuracy, precision, recall, F1-score, ROC-AUC and PR AUC. Table 5 below summarizes the performance achieved during the binary anomaly detection stage.

Table 5. Binary anomaly detection results on IOTID20 Dataset

Model	**ACC (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**FPR (%)**	**FNR (%)**	**ROC-AUC (%)**	**PR-AUC (%)**
ASAE	99.86	99.88	99.97	99.92	1.30	0.03	99.72	99.97

As illustrated in Table 5, these results demonstrate that the ASAE identifies successfully abnormal traffic reaching an accuracy of 99.86%, along with a precision of 99.88%, a recall of 99.97% and a F1-score of 99.92%. In addition, ROC-AUC and PR-AUC scores of 99.72% and 99.97% confirm the model's effectiveness discriminative

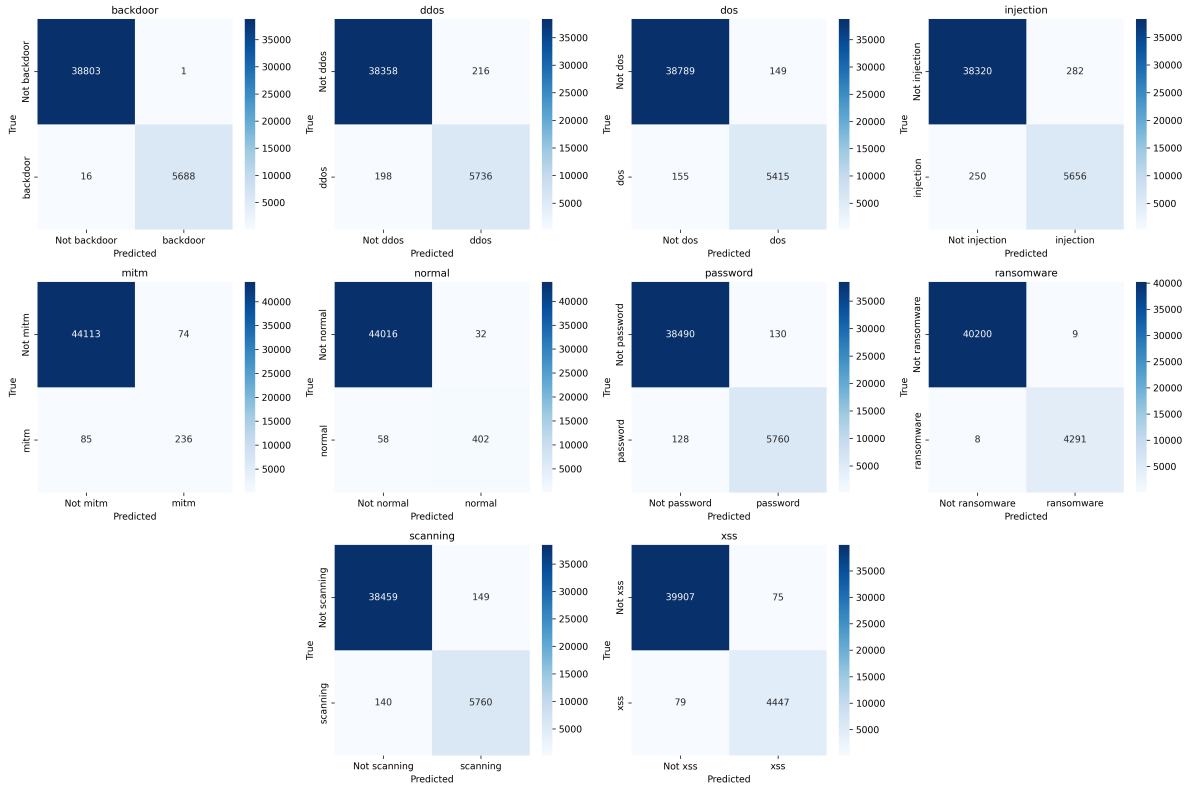


Figure 6. One-vs-Rest Confusion Matrices on TON_IoT Network Dataset.

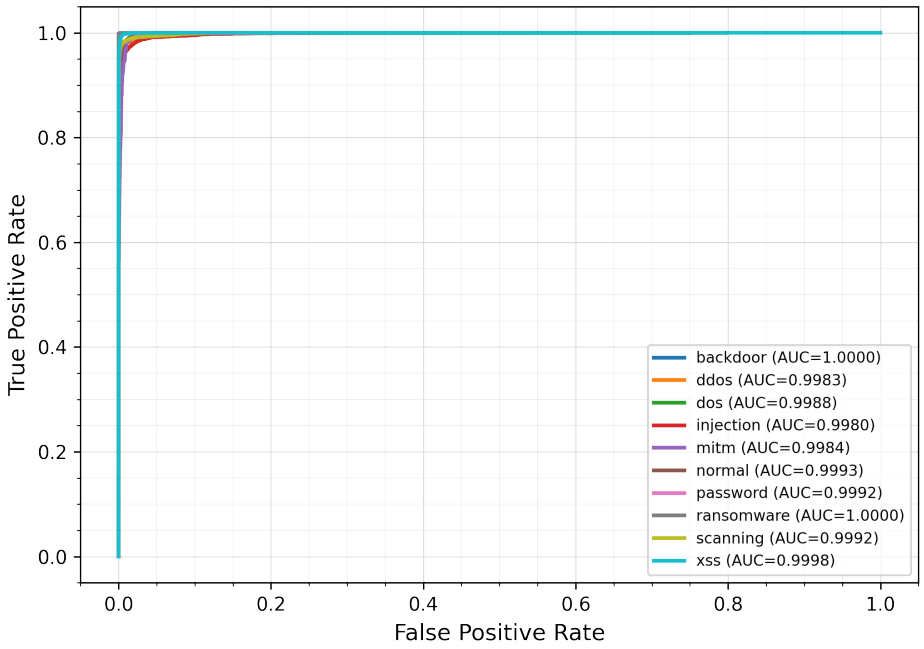


Figure 7. One-vs-Rest ROC Curves on TON_IoT Network Dataset.

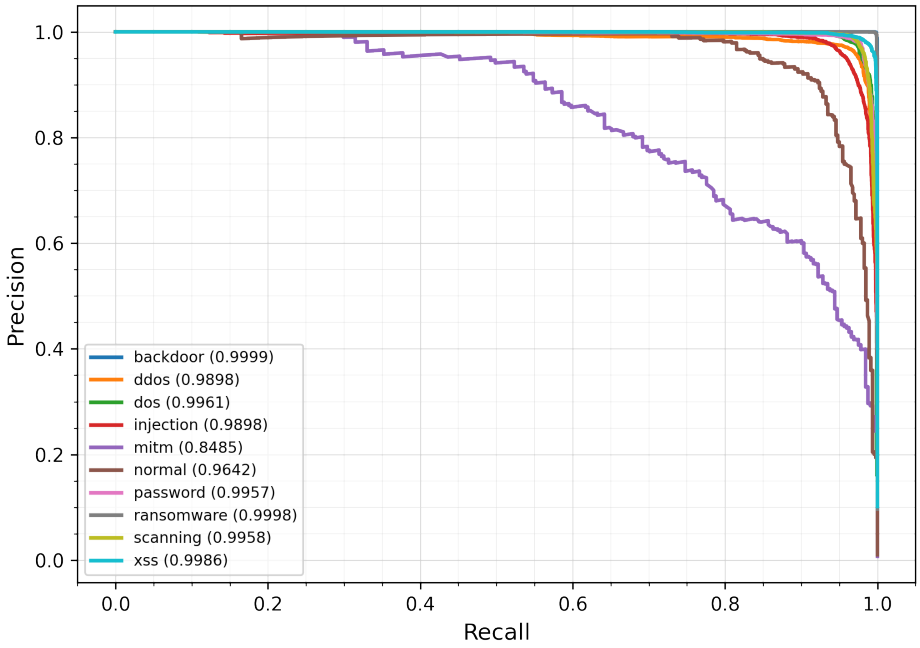


Figure 8. One-vs-Rest PR Curves on TON_IoT Network Dataset.

ability. These low false positive (1.30%) and false negative (0.03%) rates confirm the ability of the ASAE to detect the majority of attacks while generating only a limited number of false alarms.

Figure 9 below shows the confusion matrix of the ASAE and confirms its effectiveness. It shows that the model correctly classified 126887 attack samples and 11428 normal samples, with only 151 false positives and 43 false negatives. These results demonstrate the effectiveness of the ASAE in separating normal from abnormal traffic with only a few samples being misclassified.

Figure 10 hereafter presents the Receiver Operating Characteristic (ROC) curve of the ASAE, demonstrating its effectiveness in discriminating between normal and abnormal traffic across various decision thresholds. The ASAE achieved a ROC-AUC of 99.72%, while the ROC curve remained near the upper-left corner reflecting its strong detection capability.

Finally, Figure 11 illustrates the Precision-Recall (PR) curve of the ASAE, demonstrating its effectiveness in maintaining high precision and recall across various decision thresholds. The ASAE achieved a PR-AUC of 99.97%, while the PR curve remained near the upper-right corner reflecting its strong balance between precision and recall in anomaly detection.

4.4.2. Multi-class attack classification

Once abnormal traffic has been detected, the XGBoost classifier identifies the attack category associated with each anomalous sample. Table 6 below summarizes the performance achieved for each attack category during the multi-class attack classification stage.

As shown in Table 6, these results demonstrate that the XGBoost classifier achieved strong multi-class classification performance. We note that all attack categories accuracy above 99%, while precision, recall and F1- score maintained high values across most attack categories. Particularly, we observed that DoS, MiTM ARP Spoofing, Mirai, and Scan classes achieved F1-score above 98%, indicating that the latent representations learned by the ASAE are well suited for identifying different attack categories. However, the Normal class achieved lower precision, recall and F1 score values than the others categories, reflecting the difficulty of discriminating it from

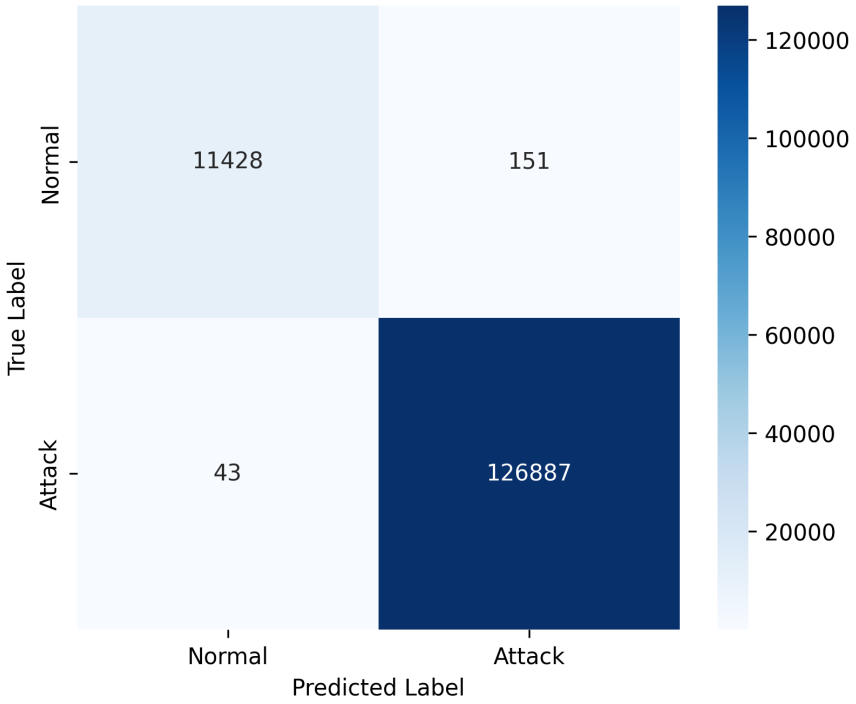


Figure 9. ASAE Confusion Matrix on IOTID20 Dataset.

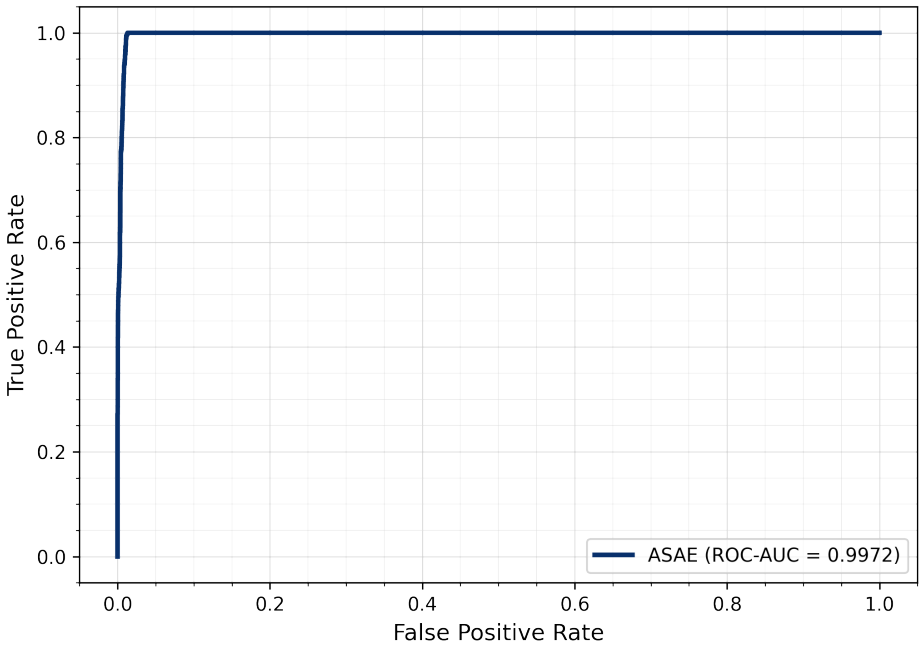


Figure 10. ASAE ROC curve on IOTID20 Dataset.

the other classes, which may be explained by similarities between its latent representations and those associated

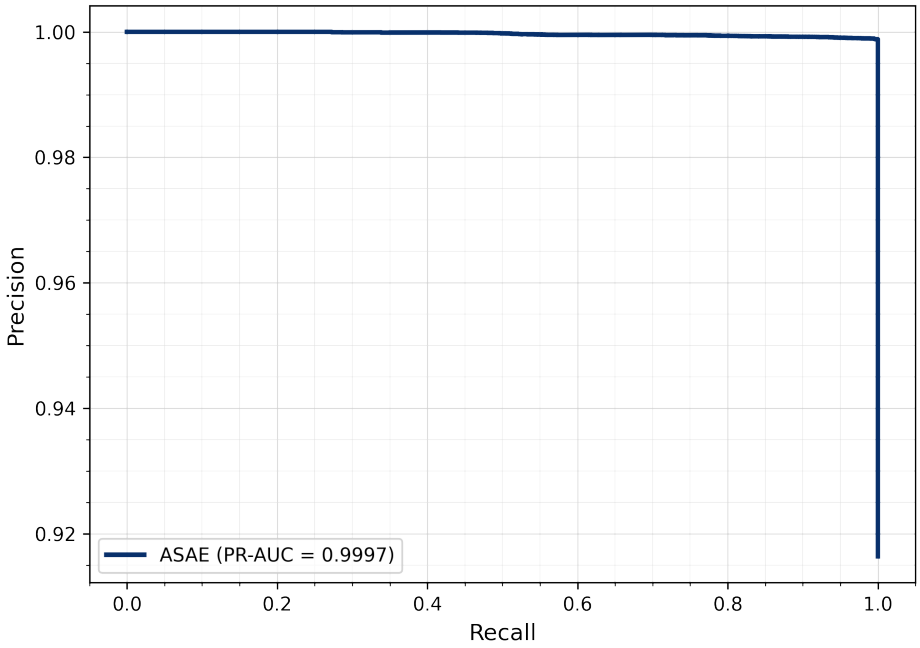


Figure 11. ASAE PR curve on IOTID20 Dataset.

Table 6. Multi-class attack classification results on IOTID20 Dataset

Class	Metrics							
	ACC (%)	Precision (%)	Recall (%)	F1-score (%)	FPR (%)	FNR (%)	ROC-AUC (%)	PR-AUC (%)
DoS	99.97	99.98	99.84	99.91	0.00	0.16	100.00	99.99
MiTM ARP Spoofing	99.86	98.97	98.68	98.82	0.07	1.32	99.99	99.88
Mirai	99.69	99.66	99.87	99.76	0.68	0.13	99.99	100.00
Normal	99.94	73.68	83.44	78.26	0.04	16.56	99.91	84.54
Scan	99.74	99.49	98.60	99.04	0.08	1.40	99.99	99.93

with others categories. Regarding the Normal classes, it is included in the second stage to allow XGBoost to correctly reclassify normal instances identified as abnormal by the ASAE in the first stage and consequently reduce false positives, while enhancing the final classification decision. In addition, ROC-AUC and PR-AUC values nearly reaching 100% confirms the strong discriminative capability of the learned latent representations. While the Normal class achieved lower PR-AUC than the other categories, the corresponding ROC-AUC remained strong demonstrating the strong discriminative capability of the XGBoost classifier. Finally, we can observe that most categories have low false positive and false negative rates, in contrast to the Normal class which showed a higher false negative rate indicating that normal traffic is more challenging than the other categories. Finally, the XGBoost classifier maintains an excellent multi-class classification performance across most categories.

Figure 12 displays the multi-class confusion matrix of the XGBoost classifier. It exhibits high values along the main diagonal indicating that the most of traffic samples are correctly classified to their respective categories. We observed that misclassifications are limited across all different traffic categories, while the Normal class shows more misclassifications than the other categories, reflecting the lower precision, recall and F1-score as reported in Table 6. Even with these misclassifications, most normal samples are correctly identified. Finally, the confusion matrix demonstrates the effectiveness of the XGBoost classifier in multi-class attack classification.

Figure 13 illustrates the one-vs-rest confusion matrices of the XGBoost classifier for each attack category. The

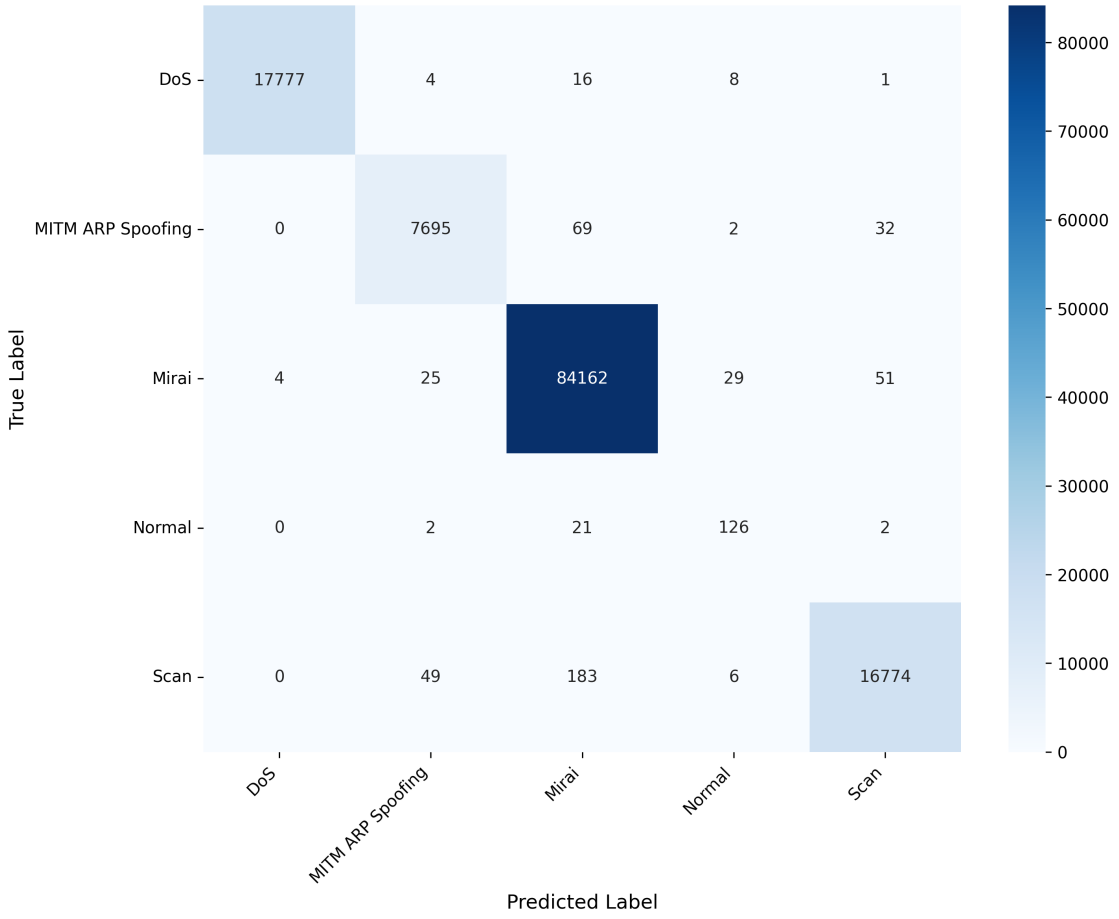


Figure 12. Multi-class Confusion Matrix on IOTID20 Dataset.

classifier correctly classifies most traffic samples across all categories, which is illustrated by the main diagonal of each confusion matrix. This indicates that the learned latent representations allow the XGBoost classifier to accurately identify each attack category. Finally, false positives and false negatives are low across different categories confirming the effectiveness of the XGBoost classifier to accurately classify each attack category.

Figure 14 shows the one-vs-rest ROC curves of the XGBoost classifier for each attack category. We observed that all curves remained near the upper-left corner demonstrating an excellent discrimination capability between each traffic category and the other classes. For all traffic categories, ROC-AUC scores vary from 99.91% to 100% demonstrating the excellent discriminative capability of the XGBoost classifier across all categories. Specifically, the DoS class achieved a best ROC-AUC of 100%, while the other categories obtained scores over 99.90%. Finally, this indicates that the learned latent representations offer effective discriminative features allowing the XGBoost classifier to accurately identify each attack category.

Finally, Figure 15 below illustrates the one-vs-rest PR curves of the XGBoost classifier for each attack category. As we can see, most curves remained near the upper-right corner demonstrating an effectiveness in maintaining high precision and recall across most traffic categories. For all traffic categories, PR-AUC scores vary from 99.88% to 100% for the DoS, MiTM ARP Spoofing, Mirai, and Scan classes demonstrating an excellent multi-class classification performance. Conversely, the Normal classes achieved lower PR-AUC scores of 84.54% indicating that this class is more challenging to identify accurately. Finally, PR-AUC scores maintain high, indicating that this classifier achieves a strong balance between precision and recall across most traffic categories.

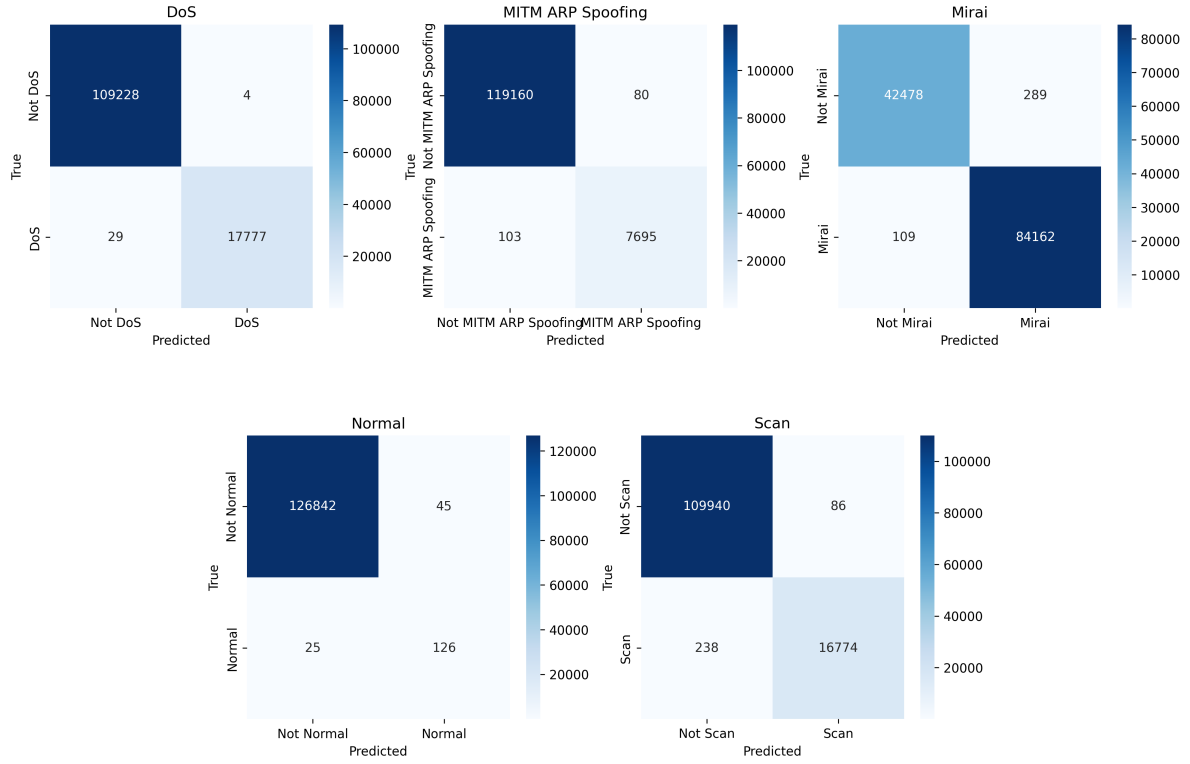


Figure 13. One-vs-Rest Confusion Matrices on IOTID20 Dataset.

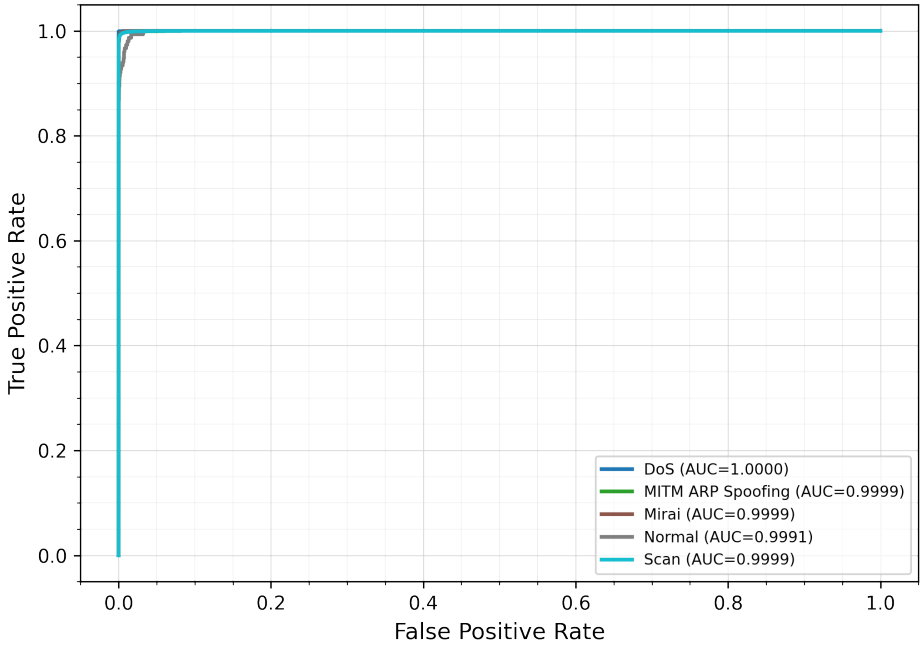


Figure 14. One-vs-Rest ROC Curves on IOTID20 Dataset.

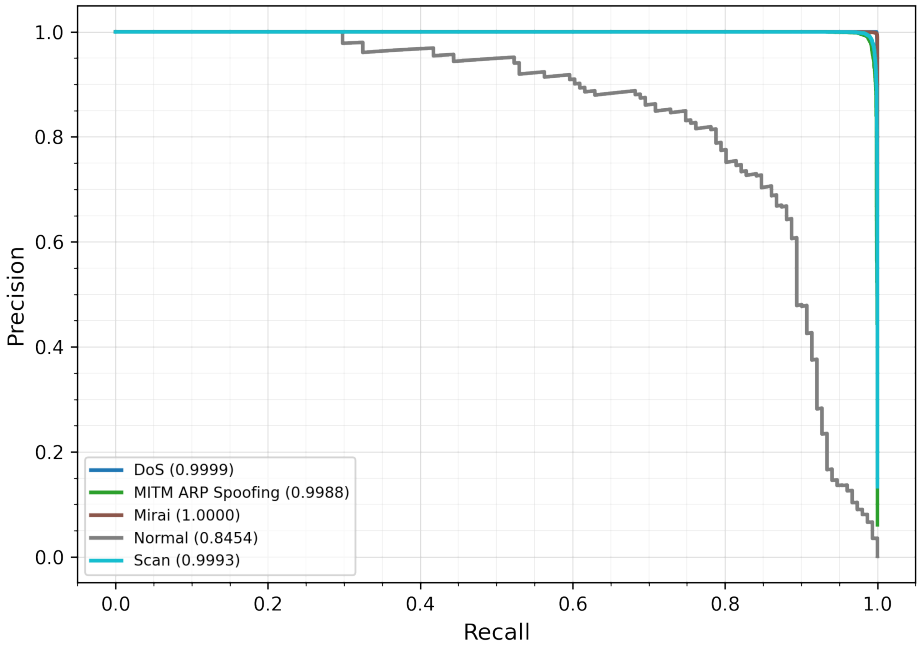


Figure 15. One-vs-Rest PR Curves on IOTID20 Dataset.

4.5. Comparison with Previous Research

This section presents a comparison of the proposed hierarchical ASAE XGBoost framework with several intrusion detection approaches evaluated on the TON_IoT Network and IOTID20 datasets. Based on the intrusion detection task, the comparison includes both binary anomaly detection and multi-class attack classification in terms of F1-score. We selected the F1-score as the main comparison metric as it provides a balanced assessment of precision and recall. Table 7 summarizes the comparative results for binary anomaly detection, while Tables 8 and 9 present the comparative results for multi-class attack classification on TON_IoT Network and IOTID20 Network datasets, respectively.

Table 7. Comparison of binary anomaly detection with previous research

Dataset	Research	Technique	Year	F1-score (%)
TON_IoT Network	[30]	J48	2024	99.10
	[31]	DistillGuard	2025	96.40
	Our research	**Proposed**	2026	98.94
IoTID20	[32]	3BD-KNN	2025	99.16
	[33]	CNN-GRU	2025	99.83
	Our research	**Proposed**	2026	99.92

As shown in Tables 7, 8, and 9, these results demonstrate that the proposed ASAE-XGBoost framework achieved strong performance compared with previous approaches on both TON_IoT Network and IOTID20 datasets. Regarding binary anomaly detection, the proposed framework achieved 98.94% and 99.92% on TON_IoT Network

Table 8. Comparison of multi-class attack classification with previous research on TON_IoT Network Dataset

Class	F1-score (Proposed)	F1-score [30]	F1-score [34]
Backdoor	99.85%	99.90%	98.81%
DDoS	96.52%	97.50%	97.81%
DoS	97.27%	98.50%	92.40%
Injection	95.51%	96.90%	82.21%
MiTM	74.80%	68.60%	23.75%
Normal	89.93%	99.30%	99.14%
Password	97.81%	98.20%	94.78%
Ransomware	99.80%	88.60%	82.85%
Scanning	97.55%	98.90%	99.08%
XSS	98.30%	92.10%	96.35%

Table 9. Comparison of multi-class attack classification with previous research on IOTID20 Dataset

Class	F1-score (Proposed)	F1-score [35]	F1-score [36]
DoS	99.91%	99.70%	99.89%
MiTM ARP Spoofing	98.82%	45.53%	99.60%
Mirai	99.76%	87.38%	99.86%
Normal	78.26%	94.72%	99.05%
Scan	99.04%	53.74%	99.51%

and IOTID20 datasets, respectively. Especially, the results achieved on IOTID20 exceeded that reported in [32] and [33], while the results obtained on TON_IoT Network remained comparable to [30] and surpassed [31]. In terms of multi-class attack classification, most attack categories achieved high F1-scores across both datasets. On TON_IoT Network dataset, the proposed framework obtained F1-scores above 95% across eight of the ten classes, while Backdoor (99.85%), Ransomware (99.80%) and XSS (98.30%) achieved the highest F1-score. Against the results reported in [30], the proposed framework achieved the best F1-scores for the MiTM, Ransomware and XSS and also exceeded [34] for Backdoor, DoS, Injection, MiTM, Password, Ransomware and XSS. On IOTID20 dataset, the proposed framework obtained F1-scores above 98% for DoS, MiTM ARP Spoofing, Mirai and Scan. Across these four attack classes, the proposed framework exceeded results reported in [35], while remaining comparable to [36] for DoS, MiTM ARP Spoofing, Mirai and Scan. Regarding the Normal class, the relatively lower F1-score is explained by its specific role in the second stage, where this class represents normal instances identified as abnormal in the first stage by the ASAE rather than the overall normal traffic. Even with the challenging instances, the proposed framework maintained high F1-scores across all attack categories, reflecting reliable performance to identify the corresponding attack categories.

4.6. Computational Cost Analysis

In this section, we evaluate the computational cost of the hierarchical ASAE XGBoost framework. We analyze the preprocessing, training, inference and per samples detection times on the TON_IoT Network and IOTID20 datasets to examine its computational efficiency. The preprocessing process required 133.27 seconds for TON_IoT Network and 94.08 seconds for IOTID20. The ASAE was trained in less than 90 s on both datasets, while the respective inference times were 11.29 seconds and 9.68 seconds, respectively. Moreover, the average ASAE detection time

per sample was maintained in the microsecond level, which required only 3.42×10^{-6} seconds/per sample on TON_IoT Network and 9.80×10^{-7} seconds/per sample on IOTID20. In the same way, the XGBoost classifier required a low computational cost. The training process required 9.40 seconds on TON_IoT Network and 27.50 seconds on IOTID20, while the respective inference times were 0.47 seconds and 0.87 seconds, respectively. Moreover, the average XGBoost classifier detection time per sample was maintained in the microsecond level, which was below 1.1×10^{-5} seconds/per sample on both datasets. Finally, these results indicate that the proposed framework achieves fast training and inference with very low per-sample detection times on both datasets. This indicates that it is a lightweight and computationally inexpensive framework in real time diverse IoT intrusion detection systems. Table 10 summarizes the computational cost of the proposed framework on both datasets.

Table 10. Cross Dataset Computational Cost of ASAE and XGBoost

Model	Dataset	Training time (s)	Inference time (s)	Detection time (s) / sample
ASAE	**TON_IoT Network**	89.48	11.29	3.42×10^{-6}
	IOTID20	84.95	9.68	9.80×10^{-7}
XGBoost classifier	**TON_IoT Network**	9.40	0.47	1.06×10^{-5}
	IOTID20	27.50	0.87	6.83×10^{-6}

4.7. Robustness and Generalization Across Datasets

Results obtained on TON_IoT Network and IOTID20 datasets are analyzed to evaluate the performance stability and the robustness of the proposed hierarchical ASAE-XGBoost framework. Given the differences in traffic patterns and attack distributions between the two datasets, the proposed framework achieves reliable performance across both datasets, maintaining its stable performance in diverse IoT network environments. As we observed above, the ASAE maintains strong binary anomaly detection performance achieved by high accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC values obtained across both datasets with different traffic patterns and attack distributions. In the same way, the XGBoost classifier maintains strong multi-attack classification performance by accurately identifying the different attack categories across both datasets with different traffic patterns and attack distributions. The performance metrics show only minor differences between the two datasets, which mainly result from their specific features and not from the proposed framework. In terms of computational cost, the proposed framework achieves fast training and inference with very low per-sample detection times across both datasets, which indicate that this framework maintains its lightweight and computationally inexpensive in real-time diverse IoT intrusion detection systems. In general, the proposed hierarchical ASAE-XGBoost framework maintains stable performance across both datasets, reflecting its reliable binary anomaly detection performance, accurate attack classification along with a low computational cost. Finally, these results demonstrate the reliable performance of this framework across diverse IoT network environments, while maintaining its high detection performance, remaining lightweight and computationally inexpensive in real-time diverse IoT intrusion detection systems.

5. Limitations and future work

While the proposed hierarchical ASAE-XGBoost framework demonstrates strong performances across both datasets, each dataset was used to train and test the framework separately. As a result, training the framework in one dataset and testing it in another dataset remains an interesting direction of our future perspectives. Several focused future directions can strengthen practical impact and generalization capability of the proposed hierarchical ASAE-XGBoost framework. Firstly, we will focus on cross-dataset evaluation by training the framework on one dataset and testing it on another. This will also confirm the framework's cross-dataset generalization capability across datasets with diverse traffic patterns and attack distributions. Secondly, we will evaluate the effectiveness of the proposed framework against unknown (zero-day) attacks to also confirm its robustness in diverse IoT network environments.

6. Conclusion

In this work, we have proposed a hierarchical two-stage intrusion detection framework based on the integration of the ASAE with XGBoost for binary anomaly detection and multi-class attack classification. It integrates the ASAE, which is learned only from normal traffic to identify abnormal traffic, along with the XGBoost classifier that classifies the traffic instances detected as abnormal by the ASAE into their corresponding categories using the learned latent representations. The proposed framework is evaluated in two real-world datasets, TON_IoT Network and IOTID20, using a standardized experimental protocol and demonstrating its effectiveness. Therefore, the experimental results demonstrate reliable binary anomaly detection performance, accurate attack classification along with a low computational cost across both datasets. Finally, the proposed framework constitutes a lightweight, robust and effective intrusion detection system in real-time diverse IoT network environments.

Acknowledgement

This work was supported by the National Center for Scientific and Technical Research (CNRST), Morocco, under the PhD Associate Scholarship (PASS).

REFERENCES

1. M. A. Khan et al., *A Deep Learning-Based Intrusion Detection System for MQTT Enabled IoT*, Sensors, vol. 21, no. 21, p. 7016, Oct. 2021, doi: 10.3390/s21217016.
2. K. Dubey, R. Dubey, S. Panedy, and S. Kumar, *A Review of IoT Security: Machine Learning and Deep Learning Perspective*, Procedia Computer Science, vol. 235, pp. 335–346, 2024, doi: 10.1016/j.procs.2024.04.034.
3. J. Li, M. S. Othman, H. Chen, and L. M. Yusuf, *'Optimizing IoT intrusion detection system: feature selection versus feature extraction in machine learning*, J Big Data, vol. 11, no. 1, p. 36, Feb. 2024, doi: 10.1186/s40537-024-00892-y.
4. J. Sengupta, S. Ruj, and S. Das Bit, *A Comprehensive Survey on Attacks, Security Issues and Blockchain Solutions for IoT and IIoT*, Journal of Network and Computer Applications, vol. 149, p. 102481, Jan. 2020, doi: 10.1016/j.jnca.2019.102481.
5. H. Pourrahmani, A. Yavarinasab, A. M. H. Monazzah, and J. Van Herle, *A review of the security vulnerabilities and countermeasures in the Internet of Things solutions: A bright future for the Blockchain*, Internet of Things, vol. 23, p. 100888, Oct. 2023, doi: 10.1016/j.iot.2023.100888.
6. H. Fereidouni, O. Fadeitcheva, and M. Zalai, *IoT and Man-in-the-Middle Attacks*, Security and Privacy, vol. 8, no. 2, p. e70016, Mar. 2025, doi: 10.1002/spy2.70016.
7. O. Zenzoum, A. Elmounadi, and H. Kharraz Aroussi, *Protecting Wireless Sensor Networks Against Sybil Attacks*, Stat., optim. inf. comput., vol. 15, no. 4, pp. 2686–2695, Dec. 2025, doi: 10.19139/soic-2310-5070-3007.
8. King Faisal University and A. A. Almuqren, *Cybersecurity threats, countermeasures and mitigation techniques on the IoT: Future research directions*, JCSRA, vol. 1, no. 1, pp. 1–11, Jan. 2025, doi: 10.63180/jcsra.thestap.2025.1.1.
9. A. Pinto, L.-C. Herrera, Y. Donoso, and J. A. Gutierrez, *Survey on Intrusion Detection Systems Based on Machine Learning Techniques for the Protection of Critical Infrastructure*, Sensors, vol. 23, no. 5, p. 2415, Feb. 2023, doi: 10.3390/s23052415.
10. N. Berbiche and J. El Alami, *Enhancing Anomaly-Based Intrusion Detection Systems: A Hybrid Approach Integrating Feature Selection and Bayesian Hyperparameter Optimization*, ISI, vol. 28, no. 5, Oct. 2023, doi: 10.18280/isi.280506.
11. A. Khraisat and A. Alazab, *A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges*, Cybersecur, vol. 4, no. 1, p. 18, Mar. 2021, doi: 10.1186/s42400-021-00077-7.
12. A. L. Buczak and E. Guven, *A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection*, IEEE Commun. Surv. Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016, doi: 10.1109/COMST.2015.2494502.
13. A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, *Survey of intrusion detection systems: techniques, datasets and challenges*, Cybersecur, vol. 2, no. 1, p. 20, Dec. 2019, doi: 10.1186/s42400-019-0038-7.
14. M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, *Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study*, Journal of Information Security and Applications, vol. 50, p. 102419, Feb. 2020, doi: 10.1016/j.jisa.2019.102419.
15. A. Nejari, A. Badaoui, A. Elmounadi, and N. Berbiche, *AI-Driven IoT Security: A Comprehensive Review on Current Trends and Future Perspectives*, in 2025 International Conference on Circuit, Systems and Communication (ICCSC), Fez, Morocco: IEEE, Jun. 2025, pp. 1–6. doi: 10.1109/ICCSC66714.2025.11135299.
16. A. A. Neloy and M. Turgeon, *A comprehensive study of auto-encoders for anomaly detection: Efficiency and trade-offs*, Machine Learning with Applications, vol. 17, p. 100572, Sep. 2024, doi: 10.1016/j.mlwa.2024.100572.
17. E. Rodríguez et al., *Transfer-Learning-Based Intrusion Detection Framework in IoT Networks*, Sensors, vol. 22, no. 15, p. 5621, Jul. 2022, doi: 10.3390/s22155621.
18. A. H. Ali et al., *Unveiling machine learning strategies and considerations in intrusion detection systems: a comprehensive survey*, Front. Comput. Sci., vol. 6, p. 1387354, Jun. 2024, doi: 10.3389/fcomp.2024.1387354.

19. H. V. M.V. and J. T., *A hybrid model for detecting intrusions using stacked autoencoders and extreme gradient boosting*, Computers & Security, vol. 150, p. 104212, Mar. 2025, doi: 10.1016/j.cose.2024.104212.
20. S. A. Hussein and S. R. Répás, *A Hybrid Intrusion Detection Framework Using Deep Autoencoder and Machine Learning Models*, AI, vol. 7, no. 2, p. 39, Jan. 2026, doi: 10.3390/ai7020039.
21. A. Villafranca and M.-D. Cano, *A hybrid inference pipeline for IDS: Combining DNNs and XGBoost through stacking for real-world intrusion detection*, Ad Hoc Networks, vol. 188, p. 104227, Jul. 2026, doi: 10.1016/j.adhoc.2026.104227.
22. Y. Hu, K. Xiao, L. Luo, and L. Chen, *An XGBoost-Based Intrusion Detection Framework with Interpretability Analysis for IoT Networks*, Applied Sciences, vol. 16, no. 2, p. 980, Jan. 2026, doi: 10.3390/app16020980.
23. A. G. Ayad, M. M. El-Gayar, N. A. Hikal, and N. A. Sakr, *Efficient Real-Time Anomaly Detection in IoT Networks Using One-Class Autoencoder and Deep Neural Network*, Electronics, vol. 14, no. 1, p. 104, Dec. 2024, doi: 10.3390/electronics14010104.
24. A. M. H. Othman, M. F. Ab Razak, A. Firdaus, H. Tahaei, and M. Gheisari, *A Hybrid CNN–LSTM Model for IoT Intrusion Detection: A Robustness Analysis Across Datasets*, Future Internet, vol. 18, no. 7, p. 345, Jun. 2026, doi: 10.3390/fi18070345.
25. H. Kamal and M. Mashaly, *Robust Intrusion Detection System Using an Improved Hybrid Deep Learning Model for Binary and Multi-Class Classification in IoT Networks*, Technologies, vol. 13, no. 3, p. 102, Mar. 2025, doi: 10.3390/technologies13030102.
26. A. Rahman et al., *Stacked Ensemble Method: An Advanced Machine Learning Approach for Anomaly-based Intrusion Detection System*, Stat., optim. inf. comput., vol. 14, no. 1, pp. 434–453, May 2025, doi: 10.19139/soic-2310-5070-2352.
27. F. M. Setiawan, D. Devianto, and M. Almuhayar, *Explainable Deep Neural Network for a Reliable Intrusion Detection System with Shapley Additive Explanation Method*, Stat., optim. inf. comput., vol. 15, no. 5, pp. 4083–4099, Feb. 2026, doi: 10.19139/soic-2310-5070-3141.
28. A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, *TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems*, IEEE Access, vol. 8, pp. 165130–165150, 2020, doi: 10.1109/ACCESS.2020.3022862.
29. I. Ullah and Q. H. Mahmoud, *IoTID20: IoT Network Intrusion Dataset [625K Flows, 4 Attack Types, 83 Features]*, Springer, 2020. Accessed: Jun. 01, 2026. [Online]. Available: https://www.kaggle.com/datasets/rohulaminlabid/iotid20-dataset
30. J. García, J. Entrena, and Á. Alesanco, *Empirical evaluation of feature selection methods for machine learning based intrusion detection in IoT scenarios*, Internet of Things, vol. 28, p. 101367, Dec. 2024, doi: 10.1016/j.iot.2024.101367.
31. N. AL-Nomasy, A. Alamri, A. Aljuhani, and P. Kumar, *Transformer-based knowledge distillation for explainable intrusion detection system*, Computers & Security, vol. 154, p. 104417, Jul. 2025, doi: 10.1016/j.cose.2025.104417.
32. H. Bao and J. Gao, *Network intrusion detection based on improved KNN algorithm*, Sci Rep, vol. 15, no. 1, p. 29842, Aug. 2025, doi: 10.1038/s41598-025-14199-2.
33. K. O. Adefemi, M. B. Mutanga, and O. A. Alimi, *A Hybrid CNN–GRU Deep Learning Model for IoT Network Intrusion Detection*, JSAN, vol. 14, no. 5, p. 96, Sep. 2025, doi: 10.3390/jsan14050096.
34. C. Du, Y. Guo, and Y. Zhang, *A Deep Learning-Based Intrusion Detection Model Integrating Convolutional Neural Network and Vision Transformer for Network Traffic Attack in the Internet of Things*, Electronics, vol. 13, no. 14, p. 2685, Jul. 2024, doi: 10.3390/electronics13142685.
35. B. M. Kouassi, A. B. Ballo, K. J. Ayikpa, D. Mamadou, and Y. Diabagate, *Optimized Intrusion Detection in the IoT Through Statistical Selection and Classification with CatBoost and SNN*, Technologies, vol. 13, no. 10, p. 441, Sep. 2025, doi: 10.3390/technologies13100441.
36. H. Karamollaoğlu, İ. A. Doğru, and İ. Yücedağ, *An Efficient Deep Learningbased Intrusion Detection System for Internet of Things Networks with Hybrid Feature Reduction and Data Balancing Techniques*, ITC, vol. 53, no. 1, pp. 243–261, Mar. 2024, doi: 10.5755/j01.itc.53.1.34933.