network

MDPI

Article

A Risk-Gated Security Attention Mechanism for Rare-Threat Detection in Industrial IoT Networks

Shaimaa Ahmed Elsaid [1,*], Ahmed M. Saad [2] and Eslam Mahmoud Fouda [3]

[1] Department of Computer Engineering, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia
[2] Computers and Control Systems Department, Air Defense College, Egyptian Military Academy, New Cairo 11888, Egypt; ahmedshded71@egma.edu.eg
[3] Electronics and Communications Department, Air Defense College, Egyptian Military Academy, New Cairo 11888, Egypt; eslamfouda138@egma.edu.eg
* Correspondence: sh.ahmed@psau.edu.sa

Abstract

In current state-of-the-art intrusion detection systems (IDSs), models are trained to detect the repetitive behavior of data traffic. This leads to the problem of ignoring rare yet important attacks. To solve the problem stated above, a risk-gated security attention (RGSA) architecture is proposed. The initialization of the risk estimator based on the CVE/CVSS severity score enables the attention network to detect important yet rare patterns. Furthermore, a tier-aware focal loss is proposed to mitigate security threats in the real world without any data augmentation process. Evaluations were conducted on held-out 80/20 test splits of the CIC-IDS2017, CIC-IDS2018, and CIC-IoT2023 datasets. The calibrated binary detection process reached false negative rates (FNRs) of 0.50% and 0.40% on the CIC-IDS2017 and CIC-IDS2018 datasets, respectively, remaining below the 1.0% FNR operational target adopted in this study. For micro-support critical threats, precise discrete analysis had to be applied to maintain statistical validity. Also, weighted F1-scores of at least 99.65% were consistently achieved for all testbeds analyzed. This network contains 284,317 trainable parameters and strictly avoids the creation of any synthetic samples through complete rejection of synthetic oversampling. The findings indicate that adding priors for security tiers to the attention process is favorable for moving towards an impact-based threat management approach.

Keywords: intrusion detection system (IDS); risk-gated security attention (RGSA); attention mechanism; risk-aware deep learning; CVE/CVSS; security data integrity; class imbalance; network security

Check for updates

Academic Editors: Pingyi Fan, Bin Hu, Jiacun Wang and Xiwang Guo

Received: 28 June 2026
Revised: 25 August 2026
Accepted: 1 September 2026
Published: 8 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

As cyber threats continue to grow more complex, intrusion detection systems (IDSs) are becoming a necessity in defending critical infrastructure. The latest attacks can subtly change and hide their traffic patterns. Examples of such silent exploits include Heartbleed, low-rate denial of service (DoS), and advanced persistent threat (APT) lateral movement. The patterns of such attacks are so well concealed that traditional detection tools cannot identify them [1]. The ability of deep learning methods, especially those employing attention mechanisms, to uncover relationships between intricate features is quite high. Yet, these techniques share a significant weakness: their foremost goal is to deliver the highest possible statistical accuracy, not gauge the actual security impact. Consequently, models

tend to focus most of their attention on the most common and least risky types of data traffic, such as routine scans, thereby overlooking rare but highly destructive attacks, which may account for less than 0.1% of network traffic [2]. The main cause of this issue lies in the extreme class imbalance in genuine network traffic. Drastically high-risk attacks, such as remote system exploitation, are very rarely encountered when compared to large-scale, common attacks. Traditional methods address this imbalance by generating synthetic samples of the minority class, such as using SMOTE. However, this might reduce the model's ability to generalize to real traffic distributions. To be more precise, these methods create synthetic data to reproduce patterns of traffic where the original real-world behavior is not well represented, since they are based only on the interpolation of discrete data points. Although modern intrusion detection systems often perform well overall, their performance on extremely rare but critical attacks generally need improvement [3]. Such a deficiency makes them susceptible to these attacks, which could have grievous consequences [4]. Such issues become increasingly important in the context of Industrial IoT. It is quite complicated to detect rare and important cyberattacks in such systems because of several reasons, including different protocol data, hard deadlines, low-cost edge devices, etc. In IoT-based systems, an attack may lead to damage to the infrastructure, a halt of manufacturing operations, or even a threat to people's lives. In such cases, conventional approaches based on precision become completely inadequate [5].

1.1. Contributions

In this research paper, a risk-gated security attention (RGSA) transformer for intrusion detection systems (IDSs) is presented and tested on held-out 80/20 benchmark test datasets. The proposed risk-gated security attention transformer differentiates itself from traditional attention models due to its capability of using the security risk associated with each token at inference time inside the attention module. The security risks are abstracted in terms of CVE/CVSS severity levels and used to initialize a learned risk estimator that modifies the attention function at runtime. For Industrial IoT systems, some significant changes were made:

- Security-Aware Tokenization: This process leads to the formation of segmentations of the feature space such that the segments correspond to the MITRE ATT&CK tactics. While the baseline behavior of normal traffic is followed in these environments, the process of segmentation is specifically tailored to identify the deviations corresponding to tactics, which would have been considered an unordered set by normal transformers.
- Risk-Gated Attention Mechanism: Unlike the traditional attention mechanism, which makes the assumption of homogeneous tokens, the novel method considers the security threat associated with each token while computing attention. The threats have been classified based on the severity of the CVE/CVSS ratings, which have a mathematical bearing on the value of the risk estimator used for initialization. Such prior knowledge dynamically impacts the attention computation process, especially at the onset of the training process.
- Authenticity-Preserving Class Balancing: A hierarchical tier-aware balancing strategy uses the exact duplication of real attack samples. By completely avoiding synthetic oversampling (e.g., SMOTE/GANs) for attack classes, the avoidance of synthetic sample generation of network flows is preserved.
- Adaptive Optimization and Avoidance of Synthetic Sample Generation: The risk-weighted focal loss function is applied along with the adaptive one-stage/two-stage training pipeline in order to fine-tune detection based on the degree of security threat and not frequency. As a result, the improved model achieves an average binary-attack false negative rate (FNR) of $0.50 \pm 0.12\%$ on the test split of the CIC-IDS2017 dataset

and demonstrates excellent results on the test splits of the CIC-IDS2017, CIC-IDS2018, and CIC-IoT2023 datasets. The rare classes have been described with exact figures and confidence intervals rather than percent averages. In contrast to traditional methods that rely on fixed global weights, the token-based attention mechanism is adjusted in RGSA based on the risk assessment done in real time. Additionally, to avoid the creation of forensic traces caused by artificial oversampling (such as SMOTE), a rigorous hierarchical security-tier duplication scheme is used.

1.2. Paper Organization

The rest of this paper is structured as follows: Section 2 covers the theoretical foundation and reviews the existing literature. Section 3 presents the architecture of the proposed model and the analytical tools used. Section 4 reports the performance analysis of the proposed model. An ablation study that demonstrates the effectiveness of the suggested RGSA model compared to the existing RGSA model is provided in Section 5. A performance comparison with existing research is made in Section 6, along with a discussion about the strengths and weaknesses of the proposed model. The conclusion and future directions are summarized in Section 7.

2. Related Work

To systematically contextualize the proposed risk-gated security attention (RGSA) framework, the advances in network intrusion detection systems (NIDSs) in recent years can be classified into four methodological dimensions: (1) structural shortcomings of static evaluation methodologies, (2) deep attention approaches in the domain of feature representation, (3) solutions to the problem of class imbalance in resource-constrained environments, and (4) distributed designs and control theory models. Such classification emphasizes the continuous advancement in IDS methodologies and reveals the critical challenges overcome by the proposed approach.

2.1. Structural Shortcomings of Static Evaluation Methodologies

The first crucial issue associated with modern IDS research is the presence of structural drawbacks in benchmark datasets. To systematically address this issue, Khanan et al. (2024) [6] conducted a systematic literature review that critically evaluated the inability of current IDS benchmarks to reflect the dynamic nature of network traffic. The study revealed that current benchmark datasets are unable to represent temporal dynamics characteristic of traffic evolution in practice. On the other hand, the peculiarities of zero-day attacks in IoT networks are systematically reviewed by Ahmad et al. (2024) [7]. However, it has been shown that the use of static training datasets is the main limitation to implementing robust machine learning models in dynamic environments. While these reviews offer essential information on the problems with datasets used in research, the inclusion of temporally adaptive evaluation methods in practical IDSs has not yet been properly considered. Thus, advanced feature extraction became necessary due to the static nature of training datasets used in IDSs.

2.2. Deep Attention Mechanisms and Sequential Modeling

In order to address the issues associated with static data representation, deep learning and attention mechanisms have been studied in great detail to improve feature extraction. A few-shot IDS using self-attention and meta-learning was proposed by Xu et al. (2025) [8] and achieved 99.90% accuracy even with a low number of samples. While this method allowed the detection of new attacks, its high computational complexity and the use of static positional encoding were major limitations for implementation in a dynamic traffic environment. A similar hybrid attention-based NIDS using CNN-BiGRU and Optuna

has been developed by Nikitenko and Bashkov (2025) [9]. While the above-described architecture proved effective in terms of generalization, multiple layers of the deep neural network created substantial computational burden, making IoT deployment impractical and hampering model explainability for the purposes of security audits. To provide the best combination of efficiency and computational overhead, a study of sequential models was conducted. An IDS based on an LSTM/GRU model with respect to temporal order was proposed by Rishika et al. (2025) [10], but computation constraints limited the research into the complex properties of the dataset. To address security threats inherent to individual protocols, a CAN IDS with triple attention and based on capsule networks was proposed by Yang and Effatparvar (2025) [11]. While adversarial learning helped to decrease error rates, model performance depended greatly on the quality of synthetic samples generated.

2.3. Class Imbalance Mitigation and Edge-Computing Constraints

However, the constant presence of the class imbalance problem made it necessary to combine advanced sampling methods with light models. Kasina et al. (2026) [12] suggested a hybrid IDS that uses normalization and several SMOTE variations. Although the solution works well for minority classes, it performs poorly with complex, multiclass distributions. To tackle this issue, Kabura and Nsabimana (2026) [13] offered an attention-enhanced CNN–RBF architecture in combination with SMOTE. Even though it demonstrated a high recall for infrequent attacks, it was evaluated only on a static offline benchmark, thus neglecting the concept drift issue. However, lightweight solutions are preferred in resource-constrained settings. Disina et al. (2026) [14] suggested a lightweight isolation forest IDS for IoT; yet a very high false positive rate (47.69%) hampers its practicality. Additionally, a resource-efficient pipeline using hybrid feature engineering was proposed by Ahmed et al. (2026) [15]; even though it managed to achieve considerable memory reduction, cross-dataset validation was not performed. However, solving data imbalance locally is inadequate when scalability and distribution according to the network topology are needed.

2.4. Distributed Architectures, Federated Learning, and Blockchain Integration

As a solution to the problems posed by centralized architecture, recent research has focused on the use of federated learning, edge computing, and blockchain technology. A bibliometric study of the applications of deep reinforcement learning on IDS was carried out by Mpoporo et al. (2026) [16], emphasizing the ability to adapt, despite the use of obsolete datasets which hinder its practical implementation. Considering distributed networks, a decentralized AdapAFS-LSTM based on federated learning was proposed by Liang and Luo (2026) [17]. Although privacy protection was improved, this technique requires frequent synchronization, making its implementation difficult in intermittent connections. In edge devices, an Edge-AI Bridge using deep autoencoders was developed by Prabu et al. (2026) [18]; however, the effectiveness of this system is limited to the physical security of the hardware. Metaheuristic optimization methods were used by Sherin and Parkavi (2026) [19] and Al Abdulwahid et al. (2026) [20] for improving the feature selection process, while the use of nested metaheuristics made the IDS complex and susceptible to overfitting. Also, the practical deployment of federated learning IDS on heterogeneous datasets has been evaluated by Bilal et al. (2026) [21], ignoring the temporal non-stationarity problem. To maintain data integrity, blockchain-enabled SDN-IoT systems were suggested by Pawar et al. (2024) [22] and Bekri et al. (2025) [23]. Despite the benefits that came from the integration of the two technologies, the unavoidable processing costs associated with the blockchain consensus algorithm considerably slowed down real-time processing in resource-limited edge devices. Distributed and blockchain approaches improve scalability

and privacy but typically do not account for any statistical anomaly that might occur at the edge of the network because they assume that sensor data are genuine.

2.5. Theoretical Grounding for Handling Tampered Sensors and RGSA

Along with advances in data-driven distributed approaches, some theoretical work on combating sensor tampering offers valuable theoretical grounding. Robust control in discrete event systems using Petri nets against attacks when the values of sensors were altered to make them report false events was considered by Yu et al. (2026) [24]. Optimal solutions to handle such attacks proved to be impractical because of computational complexity. To resolve the trade-off between computational complexity and optimality in IIoT use cases under consideration, a paradigm change is necessary. Instead of resorting to computationally intensive distributed consensus based on blockchain technology or generating synthetic data, dynamic attention techniques need to be used right at the edge level. A comparative analysis of the discussed network intrusion detection systems is provided in Table 1.

Table 1. Comparative analysis of recent network intrusion detection systems.

Ref. [Year]	Dataset Used	ML/DL Models	Result	Identified Limitations	IIoT Relevance & Logic
[6] (2024)	Multiple IDS Datasets	Systematic Literature Review	N/A	No temporal dynamics; fails to represent dynamic environments.	Highlights need for temporally adaptive evaluation in dynamic IIoT.
[7] (2024)	IoT Network Datasets	Systematic Literature Review	N/A	Static training data limits zero-day detection.	Exposes static data bottlenecks for IIoT zero-day threats.
[8] (2025)	ISCX-IDS2012	Self-Attention + Meta-Learning	99.90%	High complexity; fixed positional encoding.	Mitigates data scarcity via few-shot learning; lacks edge scalability.
[9] (2025)	NSL-KDD	CNN-BiGRU + Optuna	99.85%	Excessive computation; lacks interpretability.	Improves generalization; heavy architecture burdens IIoT edge nodes.
[10] (2025)	Time-ordered Datasets	LSTM/GRU	97.21%	Computational limits restrict attribute exploration.	Low latency suits IIoT; limited feature grouping hinders threat detection.
[11] (2025)	CAN Bus Datasets	Triple-Attention Capsule Networks	96.30%	Performance depends on synthetic sample quality.	Suits vehicular IIoT (CAN); synthetic data limits real-world robustness.
[12] (2026)	ToN-IoT, UNSW-NB15	Hybrid Sampling (SMOTE) + DL	99.92%	Degrades on complex multi-class distributions.	Detects minority IIoT attacks; struggles with complex multi-class traffic.
[13] (2026)	CIC-IDS2018	CNN–RBF–Attention + SMOTE	99.81%	Offline static evaluation; ignores concept drift.	High recall for rare attacks; static evaluation ignores concept drift.

Table 1. *Cont.*

Ref. [Year]	Dataset Used	ML/DL Models	Result	Identified Limitations	IIoT Relevance & Logic
[14] (2026)	IoT Datasets	Lightweight Isolation Forest	80.61%	Highly high FPR (47.69%).	Ideal memory for IIoT edge; high FPR causes severe alert fatigue.
[15] (2026)	CIC-DDoS2019	Hybrid Feature Engineering Pipeline	F1: 0.975	Lacks cross-dataset validation.	Reduces memory; lacks cross-validation for diverse industrial protocols.
[16] (2026)	Various (Bibliometric)	Deep Reinforcement Learning (Review)	N/A	Outdated datasets; no standardized protocols.	DRL enables dynamic adaptation; outdated benchmarks hinder deployment.
[17] (2026)	IoT/Federated Datasets	Decentralized AdapAFS-LSTM (FL)	99.79%	Class imbalance; periodic sync bottlenecks.	Preserves IIoT privacy; intermittent connectivity disrupts synchronization.
[18] (2026)	Smart-city IoT Datasets	Edge-AI Bridge (Deep Autoencoders)	AUC: 0.982 (50 ms)	Assumes physically secure hardware.	Low latency benefits IIoT edge; physical security assumptions unrealistic.
[19] (2026)	Network Datasets	SSA-BOA-CNN	99.16%	Nested metaheuristics increase complexity.	Enhances feature selection; overhead prohibits real-time IIoT processing.
[20] (2026)	Large-scale IoT Datasets	IBGWO-PSO-RF	99.96% (28.46 ms)	Overfitting on static datasets.	Fast convergence suits IIoT; static overfitting limits novel anomaly detection.
[21] (2026)	Heterogeneous IoT Datasets	Federated Learning IDS	92–98%	Cross-dataset degradation; ignores non-stationarity.	Addresses IIoT heterogeneity; fails to adapt to non-stationary traffic.
[22] (2024)	InSDN	Attention Hybrid (At-C-L) + Blockchain	98.30%	Single dataset; high blockchain costs.	Secures SDN-IoT; consensus overhead prevents real-time deployment.
[23] (2025)	IoT/SDN Datasets	SDN-AI-Blockchain Integration	N/A (System)	Blockchain consensus processing overhead.	Ensures access control; consensus latency hinders real-time response.

2.6. Literature Gap

Despite state-of-the-art intrusion detection systems (IDSs) demonstrating high accuracy in benchmark tests, the static nature of the data used inevitably excludes concept drift or zero-day attacks [6,7]. At the same time, gaining such high performance usually requires extensive computational capabilities, making these systems inappropriate for resource-limited edge IIoT devices [9,14,18]. While the federated learning technique tries to distribute this cost, it faces critical problems with synchronization in a heterogeneous setting [17,20], thus showing the inefficiency of distributed consensus under these cir-

cumstances. Moreover, class imbalance is commonly dealt with by using SMOTE-style oversampling techniques [12,13].

Nevertheless, creating artificial examples does not account for complicated non-linear network structures, thus causing noise and overfitting. To overcome these overlapping issues, a paradigm shift is needed. Focal loss variants do not provide dynamic risk adjustment during attention computations regardless of feature risk level. Hence, the proposed risk-gated security attention (RGSA) approach introduces an innovative hierarchical balancing technique that does not use synthetic samples at all. Overall, the analyzed literature displays a considerable development trend in increasing accuracy, while having certain drawbacks that all the papers have in common.

3. Methodology

The proposed transformer model is a modified version of the core architecture based on the risk-gated security attention (RGSA) mechanism. The proposed RGSA-Transformer is an innovative intrusion detection system designed to address two major network security challenges today: (i) data distribution distortion due to a lack of examples in certain categories, and (ii) the high security cost of ignoring rare, high-impact attacks, such as Heartbleed and advanced persistent threats. The unified model includes three tightly coupled parts: (1) dataset-specific preprocessing and adaptive pipeline configuration; (2) a modified transformer model with RGSA; and (3) layer-aware training with early stopping. The architectural workflow of the proposed RGSA-Transformer is illustrated in Figure 1.

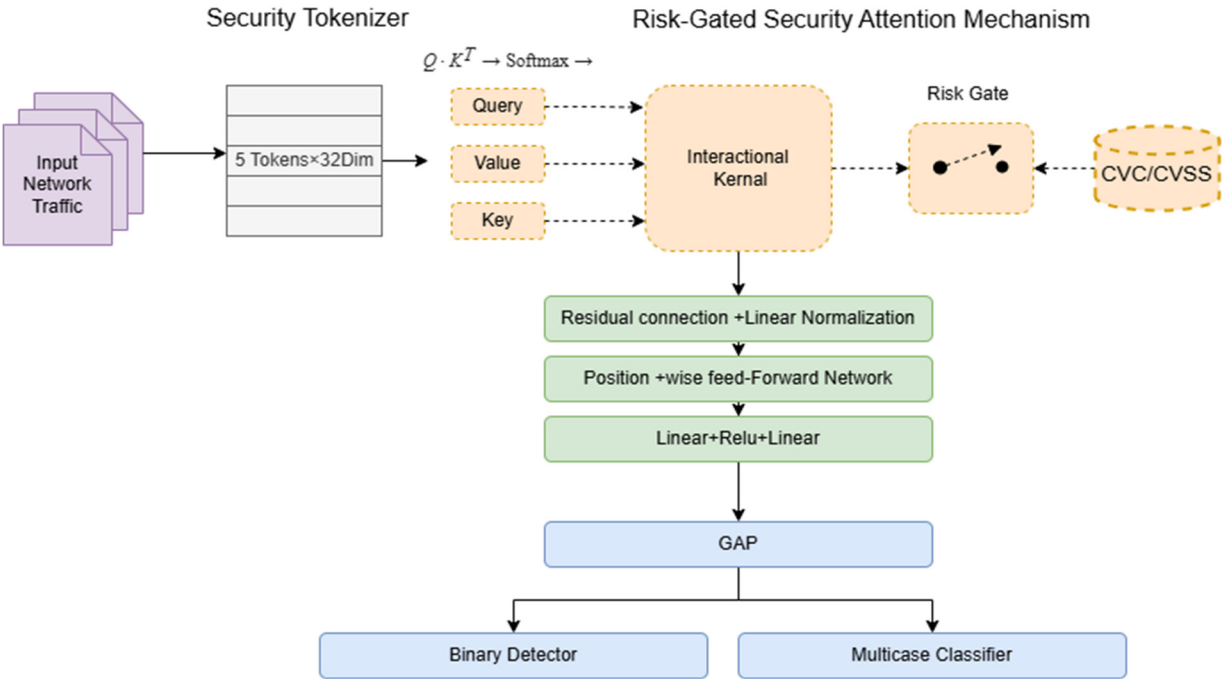


Figure 1. Process flow of the RGSA-Transformer model, incorporating the risk-gated security attention module. Arrows show the direction of data flow, where solid lines indicate feed-forward propagation, dotted lines indicate the calculation of losses, and dashed lines indicate the incorporation of risk scores.

3.1. Problem Formulation

However, NIDSs operating in the context of an IIoT environment are bound to experience several key challenges, including the problem of class imbalance, semantic heterogeneity of attack taxonomy, and computationally expensive feature extraction from high-dimensional flow statistics. To resolve such challenges, the network flows must be formally

defined before the application of attention mechanisms. Formally, let $D = \{(x_i, y_i)\}_{i=1}^{N}$ denote a network traffic dataset where $x_i \in \mathbb{R}^d$ represents a d-dimensional feature vector extracted from flow-level statistics (e.g., duration, packet inter-arrival times, payload entropy, protocol flags), and $yi \in C$ denotes the attack class label from a finite set $C = \{c_0, c_1, \ldots, c_{k-1}\}$ with $c_0 = BENIGN\ and\ c_j (j \geq 1)$ representing distinct attack types. The class distribution exhibits severe skewness characterized by:

$$\rho k = \frac{|\ \{i : y_i = c_k\}\ |}{N}, \text{where } \rho_0 \gg \rho_k \forall k \geq 1 \quad (1)$$

with the ratio $\max k \geq \frac{\rho_k}{\min k} \geq \rho_k$ frequently exceeding 10^3 in operational deployments. This imbalance is security-tiered: critical attacks (e.g., Heartbleed) show $\rho_k < 10^{-4}$, while volumetric attacks (e.g., DoS-Hulk) reach $\rho_k < 10^{-2}$. Regular oversampling (with SMOTE or ADASYN, [12,13]) produces forensic artifacts by introducing synthetic samples via interpolation of discrete flows. This contradicts the event-driven nature of network traffic and leads to degradation of generalization ability [25]. The present technique addresses these issues through risk-gated attention, security-tier-aware hierarchical balancing, and an adaptive pipeline.

3.2. Risk-Aware Feature Representation and Attention Mechanism

To make sure that the ensuing risk-gating algorithm uses semantically relevant entities, the high-dimensional flow representation is split into semantic security tokens. Through categorizing the features into groups like temporal dynamics and payload statistics, the tokenization procedure becomes optimized for emphasizing the deviations in the adversarial traffic while maintaining the forensic relations within the domain that would be considered unordered by the regular transformer architecture. Accordingly, the feature space is divided into M non-overlapping segments $\{Sm\}m = 1\ M$, assuming input is $x \in \mathbb{R}^d$, where $S_m = \{d_{m-1} + 1, \ldots, d_m\}$ with $d_0 = 0\ and\ d_M = d$. Each segment represents a feature group that is relevant to security and is consistent with the tactical categories of the MITRE ATT&CK framework, such as temporal dynamics. (S_1), payload statistics (S_2), protocol behavior (S_3), connection metadata (S_4) and flow statistics (S_5). The security tokenizer $T : \mathbb{R}^d \to \mathbb{R}^{M \times h}$ produces token embeddings through a double-layered perceptron:

$$t_m = \sigma\left(W_m^{(1)} x_{Sm} + b_m^{(1)}\right),\ m = 1, \ldots, M, \quad (2)$$

$$z_m = \sigma\left(W_m^{(2)} t_m + b_m^{(2)}\right), \quad (3)$$

where $W_m^{(1)} \in \mathbb{R}^{h \times |S_m|}$ and $W_m^{(2)} \in \mathbb{R}^{h \times h}$ are trainable projection matrices, $\sigma(\cdot)$ represents the rectified linear unit (ReLU) activation function, and h is the dimension of the token embedding. The final token sequence $Z = [z_1; \ldots; z_M] \in \mathbb{R}^{M \times h}$ maintains feature group semantics, whereas regular transformers treat inputs as unordered sets and therefore lose domain-specific relationships which are crucial for security analysis [26]. To ensure optimal feature representation while eliminating redundancy, recursive feature elimination (RFE) was applied to each dataset individually. This process retained the dataset-specific optimal feature counts—78 features for CIC-IDS2017, 83 features for CIC-IDS2018, and 68 features for CIC-IoT2023—thereby preserving the unique forensic characteristics of each network environment. This approach ensures that no synthetic features are introduced and forensic fidelity is maintained. Figure 2 displays the proposed security tokenization layer architecture. The 68-dimensional input vector is split into five semantic segments (temporal, payload, protocol, metadata, and flow) with sizes 14 + 14 + 14 + 13 + 13 = 68, aligned with MITRE ATT&CK tactics.

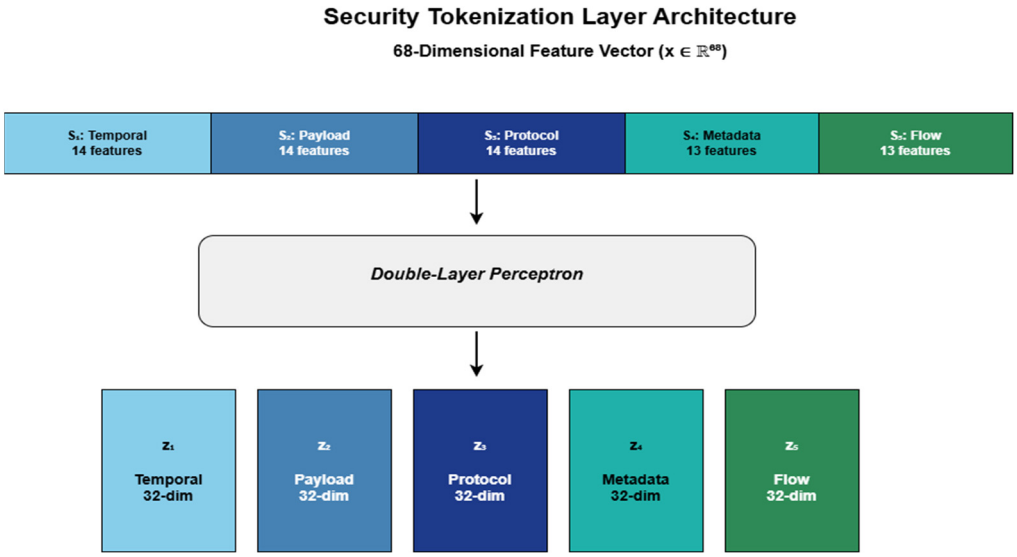


Figure 2. Proposed security tokenization layer architecture.

Standard self-attention calculates two-way interactions without considering security context awareness, resulting in diluted attention weights for critical threats in highly imbalanced NIDS datasets. To address this issue, a risk-gated attention mechanism dynamically adjusts attention weights based on explicit estimation of token-level security risk. For the token sequence $Z \in \mathbb{R}^{M \times h}$ query, key, and value projections are computed:

$$Q = ZW^Q, K = ZW^K, V = ZW^V \tag{4}$$

where $W^Q, W^K W^V \in \mathbb{R}^{h \times h}$ and, therefore, $Q, K, V \in \mathbb{R}^{M \times h}$ are learnable projection matrices. Product attention captures the nonlinear relationships between query and key through bilinear fusion:

$$A_{ij}^{int} = w^\top tanh\left(W^{int}[q_i; k_j] + b^{int}\right), \tag{5}$$

where $W^{int} \in \mathbb{R}^{2h \times h'}$ projects the concatenated query–key pair into an interaction space, $b^{int} \in \mathbb{R}^{h'}$ is a bias vector, and $w \in \mathbb{R}^{h'}$ computes a scalar interaction score. At the same time, token-level risk scores are predicted via a lightweight risk estimator defined in Equation (6):

$$r_i = \sigma_{sigmoid}(w_r z_i + b_r), r_i \in [0, 1] \tag{6}$$

where $w_r \in \mathbb{R}^h$ and $b^r \in \mathbb{R}^h$ are learnable parameters, and $\sigma_{sigmoid}(\cdot)$ denotes the sigmoid activation function.

The risk estimator is initialized using structured security priors rather than random values. Every semantic token m receives a corresponding security class $\tau(m) \in \{C, M, B\}$, which corresponds to Critical, Moderate, and Benign/Common. Such classification is made depending on MITRE ATT&CK tactics and the CVSS-based severity distribution associated with the particular token. Target initial risk probabilities are defined by:

$$p_C = 0.9, p_M = 0.5, p_B = 0.1 \tag{7}$$

To integrate these discrete priors into the continuous sigmoid-activated risk estimator, the bias term $b_r^{(m)}$ is initialized using the log-odds transformation:

$$b_r^{(m)} = ln\left(\frac{p_{\tau(m)}}{1 - p_{\tau(m)}}\right) \tag{8}$$

The weight vector w_r is initialized with small random values drawn from $N(0,\ 0.01)$. This initialization makes sure that, before any update, the tokens related to critical tactics have a higher risk score, and this will draw attention in the early stage to potentially dangerous patterns. This mathematical initialization makes sure that before any gradient update; the risk estimator will always provide a high output for the tokens associated with critical tactics. Hence, the attention algorithm becomes structurally biased toward those tokens which exhibit patterns of severe vulnerabilities, and thus attention dilution is mitigated without the help of any heuristic post-processing. While this prior initialization substantially accelerates early convergence, there is a need to constrain the optimization phase to avoid overfitting to the majority classes. The final attention score is given in Equation (9):

$$\alpha_{ij} = softmax_j\left(A_{ij}^{int} \cdot r_j\right) \tag{9}$$

The attended representation is then computed as in Equation (10):

$$z_i' = \sum_{j=1}^{M} \alpha_{ij} v_j \tag{10}$$

A residual connection with layer normalization preserves gradient flow as in Equation (11):

$$z_i^{out} = LayerNorm\left(z_i + z_i'\right) \tag{11}$$

The proposed modifications guarantee that tokens showing unusual behaviors will be assigned attention weights proportional to their estimated security risk. Such instances include strange payload entropy values or an irregular sequence of protocol flags. Therefore, the issue of attention dilution is handled without necessitating heuristic post-processing. Figure 3 gives an overview of the RGSA technique. The input tokens first go through linear layers to generate the corresponding Q, K, and V matrices. However, unlike traditional self-attention, the RGSA employs additive attention with bilinear projection (A_{int}) and at the same time it predicts the per-token security risk scores (r_i). Subsequently, the final attention weights are dynamically gated by the risk scores before the Softmax function. This gating mechanism ensures that the model assigns higher attention weights to high-risk, rare threat features.

3.3. Hierarchical Balancing and Tier-Aware Optimization

To maintain the discrete-event characteristic of network flow, an intelligent duplication strategy has been devised. This process does not create artifacts that are usually created by oversampling methods like SMOTE or generative adversarial network-based methods (none of the classes have synthetic samples). The model has native handling for purely malicious datasets, such as CIC-IoT2023, through its single-stage mode, where no artificial benign traffic generation is needed. It is based on the hierarchical classification system of the security severity of a threat.

This hierarchical classification system of security tiers has been designed by the authors through security impact considerations from NIST SP 800-115 and MITRE ATT&CK, and not as actual NIST tiers:

- *Critical Tier ($\mathcal{T}_c$):* Examples of such attacks: remote code execution, privilege escalation, or lateral movement (e.g., Heartbleed [CVE-2014-0160], Infiltration, Mirai propagation). These are the ultimate threats to the infrastructure's integrity.
- *Moderate Tier ($\mathcal{T}_m$):* Resource exhaustion attacks that cause service degradation without full compromise (e.g., Slowloris DoS, SSH brute-force).
- *Common Tier ($\mathcal{T}_0$):* Very large-scale volumetric attacks that can be mitigated by well-known tools (e.g., DoS-Hulk, PortScan).

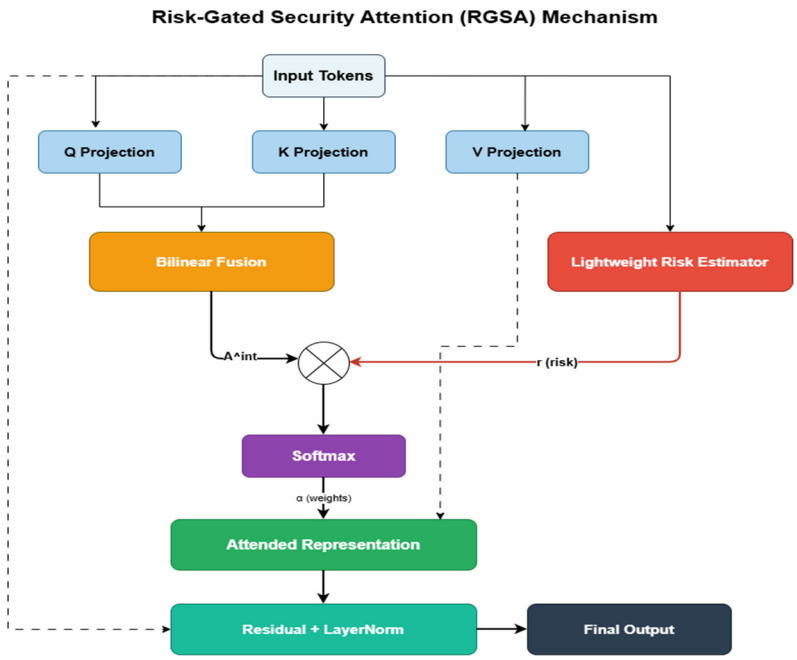


Figure 3. Risk-Gated Security Attention Mechanism. Colors used: Blue represents input tokens and projections, Orange shows the bilinear fusion layer, Red denotes lightweight risk estimation, Purple marks softmax activation, Green indicates the attended representation, while Dark Green refers to the output after residual connections and layer normalization.

For each attack class $c_k \in \mathcal{T}_\tau (\mathcal{T} \in \{c, m, o\})$ with original sample count n_k, a tier-aware duplication factor is computed:

$$\phi_k = min(\phi^{\mathcal{T}}_{max}, max(1, \lceil \frac{max(n_{\mathcal{T}_\tau}) \cdot \beta_\tau}{n_k} \rceil)) \tag{12}$$

where $max(n_{\mathcal{T}_\tau})$ represents the size of the largest class in the tier $\mathcal{T}, \beta_c = 0.33$, and $\beta_m = 0.07$, and $\beta_o = 0.03$ are the balancing ratios for the different tiers that have been empirically derived heuristics inspired by general NIST impact guidelines, and $\phi^c_{max} = 50$, $\phi^m_{max} = 15$ $\phi^o_{max} = 5$ are the maximum duplication caps. These three caps were chosen to emphasize critical threats. To ensure that the decision boundaries remained feasible, a global class-ratio constraint was imposed:

$$\frac{\max_{k \in C} n^{final}_k}{\min_{k \in C \setminus \{c_0\}} n^{final}_k} \leq 30 \tag{13}$$

where $n^{final}_k = \phi_k . n_k$ benign samples are randomly subsampled to achieve a malicious traffic ratio of 40% to 60%. Such a proportion corresponds to the real-world incident response cases in which the malicious traffic is about 40% to 60% of the monitored flows due to the deployment of an active attack. Furthermore, the selected duplication caps and tier ratios were determined through empirical sensitivity analysis. It was observed that exceeding the specified caps (e.g, $\phi maxc = 50$) induced overfitting on minority classes, degrading generalization. The adopted ratios were found to optimally balance the enhancement of critical threat recall with the preservation of overall classification precision. Figure 4 shows the hierarchical security-tier balancing strategy. The severe class imbalance of the original dataset is shown in the left panel. The middle panel depicts the tier-sensitive duplication procedure, where Critical, Moderate, and Common tiers are allocated tier-specific duplication factors. Meanwhile, Benign traffic is subsampled independently. More

importantly, this technique involves replication, and not generation (such as SMOTE), to maintain the avoidance of synthetic sample generation of network flows. The balanced dataset after using this technique, where the ratio constraint is imposed, is displayed on the right side.

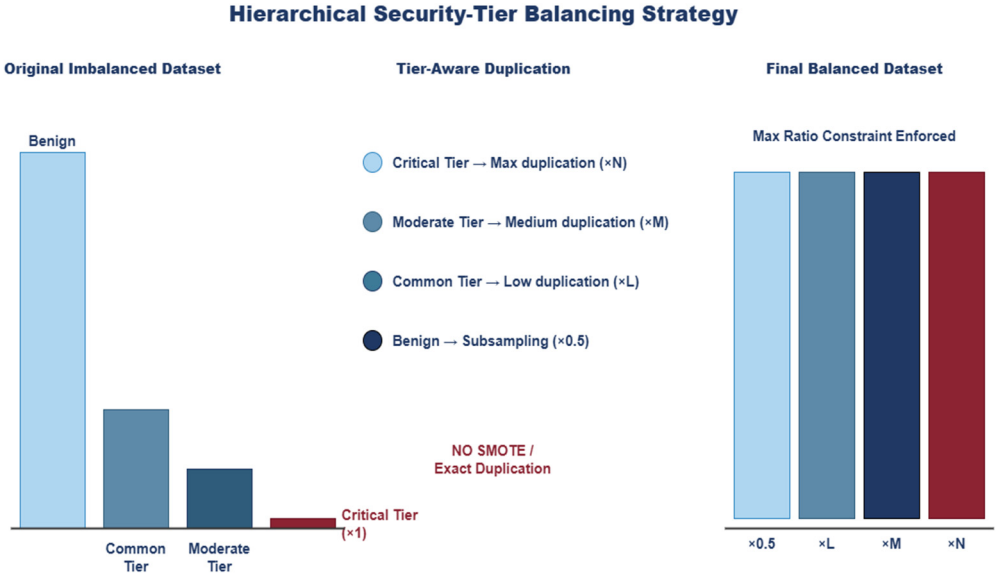


Figure 4. Hierarchical security-tier balancing strategy.

To address class imbalance after hierarchical balancing, a class-aware focal loss is applied, as defined by Equation (14):

$$\mathcal{L}_{focal}(y, p) = -\alpha_y (1 - p_y)^{\gamma} log(p_y), \tag{14}$$

where $y \in \{0, \ldots, K-1\}$ is the true label, $p \in [0,1]^K$ is the predicted probability vector with $\sum_k pk = 1, \gamma = 3.0$ controlling the focusing strength, and α_y denotes class-specific weighting derived from security tiers as defined in Equation (15):

$$\alpha_y = \begin{cases} 0.7 & if\ y = 0\ (BENIGN), \\ 5.0 & if\ c_y \in \mathcal{T}_c \\ 2.0 & if\ c_y \in \mathcal{T}_m \\ 1.0 & if\ c_y \in \mathcal{T}_o \end{cases} \tag{15}$$

This method assigns a larger gradient weight to the top-level attacks. At the same time, it downplays easy and common examples. Thereby, optimization is aligned with real-world security impact [27].

3.4. Adaptive Single/Two-Stage Pipeline Architecture

The framework automatically selects the detection pipeline based on the label distribution of the training data, through statistical analysis at the preprocessing stage, without human intervention.

- Two-stage mode (activated when $\rho_0 > 0$): A binary detector $f_b : \mathbb{R}^d \to \{0,1\}$ first identifies attack presence, where $f_b(x) = 1$ indicates an attack. Samples marked as attack attempts are then categorized by a multiclass model $f_m : \mathbb{R}^d \to C \setminus \{c_0\}$. The total prediction is computed by Equation (16):

$$y = \begin{cases} c_0 & if\ f_b(x) = 0 \\ f_m\ (x) & if\ f_b\ (x) = 1 \end{cases} \tag{16}$$

- Single-stage mode (activated when $\rho_0 = 0$): Multiclass classification. Since all the traffic in the CIC-IoT2023 dataset is a pure attack traffic, the single-stage mode is activated. To maintain avoidance of synthetic sample generation without fail, there is no oversampling or artificial injection of benign instances used during the training process for this dataset. The classifier takes in the malicious flows. $f_m : \mathbb{R}^d \to C$ directly without any binarization, ensuring consistent behavior even in a purely attack scenario.

The system remains efficient even when processing data of diverse origins, as is usually the case in operational and security intelligence environments. Figure 5 shows the adaptive single/two-stage pipeline architecture. The system statistically analyzes the training data to compute the benign ratio (ρ_0). If benign traffic is detected ($\rho_0 > 0$), the two-stage mode is enabled, which first uses a binary detector and then a multiclass classifier. Otherwise ($\rho_0 = 0$), the single-stage mode is initiated, carrying out direct multiclass classification to guarantee high performance even in attack-only deployment scenarios.

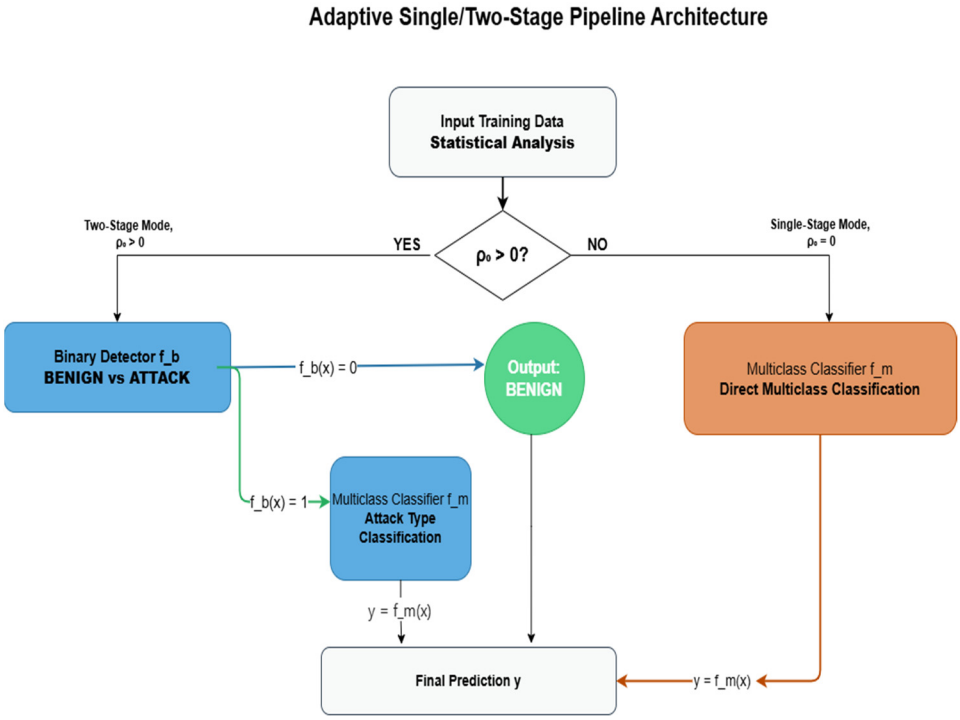


Figure 5. Adaptive pipeline architecture.

3.5. Enhancements to the RGSA-Transformer Architecture

Figure 6 illustrates a side-by-side comparison between the baseline RGSA-Transformer and the proposed enhanced version. The improved model offers three key architectural innovations:

1. Semantic Security Tokenization: Instead of treating tokens as per the case of vanilla transformers, tokens are aligned with MITRE ATT&CK tactics, and the domain-specific relationships between the feature sets are preserved.
2. Dynamic Risk Modulation: Adjusts attention weights dynamically based on learned token-level security risk scores (r_i), enabling the model to prioritize high-impact, rare threats according to their predefined security tiers (derived from CVE/CVSS metrics).
3. Tier-Aware Focal Loss: Works on class imbalance by mixing security-tier-specific weighting (αy) with a focusing parameter (γ = 3.0) to give more emphasis to critical attacks without losing performance on common ones.

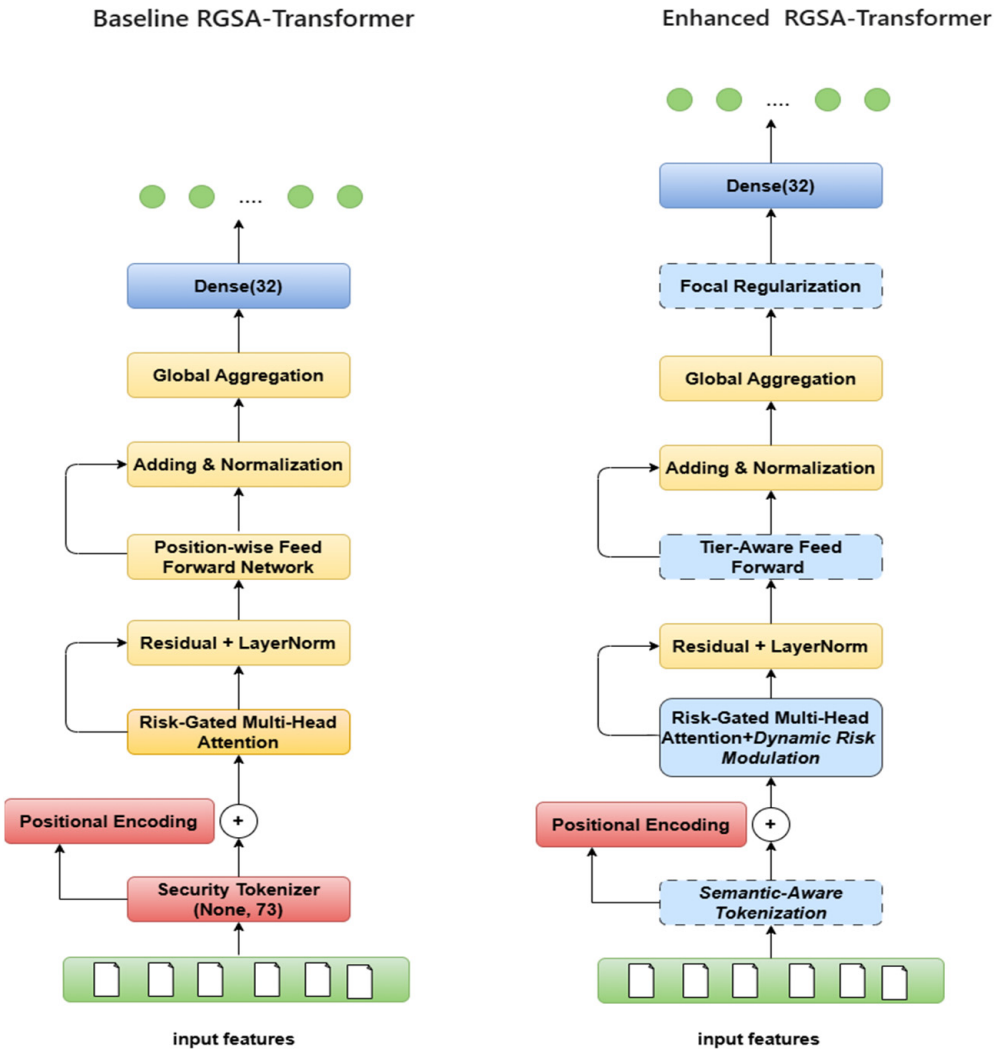


Figure 6. Comparative layout of the base RGSA-Transformer versus the enhanced RGSA-Transformer.

Such improvements tackle the main problems of extreme class imbalance and semantic inconsistency in modern NIDSs, thus boosting detection performance.

3.6. Security Tier-Weighted Metrics

Performance is measured via tier-based scores that are designed to take into account the importance of security impact, guided by the basic principles of the NIST approach to security assessment and the MITRE ATT&CK framework. For each security tier $\mathcal{T}_\tau$ the following are computed:

$$Recall_{\mathcal{T}_\tau} = \frac{\sum_{ck \in \mathcal{T}_\tau} TP_k}{\sum_{ck \in \mathcal{T}_\tau} (TP_k + FN_k)}, \tag{17}$$

$$F1_{\mathcal{T}_\tau} = \frac{2 \cdot Precision_{\mathcal{T}_\tau} \cdot Recall_{\mathcal{T}_\tau}}{Precision_{\mathcal{T}_\tau} + Recall_{\mathcal{T}_\tau}}, \tag{18}$$

where P_k and FN_k denote true positives and false negatives for the class c_K, respectively. The performance of the Benign class is reported separately using only two metrics (false positive rate and false negative rate). Attack class metrics are then calculated after excluding healthy samples to ensure no overlap between the metrics.

4. Materials and Results

To assist with the development of RGSA-Transformer, three widely recognized benchmark datasets in the cybersecurity field were chosen: CIC-IDS2017 [28], CIC-IDS2018 [29], and CIC-IoT2023 [30]. The choice of the CIC-IDS2017, CIC-IDS2018, and CIC-IoT2023 datasets was motivated by the need to test the framework in increasingly complex and varied network settings. The CIC-IDS2017 dataset provides the basic enterprise traffic baseline, with vulnerabilities that have a known low footprint, such as Heartbleed, which is essential to evaluate the rare event detection capabilities of the framework. The CIC-IDS2018 dataset is necessary to account for inconsistent labeling of its predecessor; moreover, it introduces more modern attack types including the advanced Botnet and Web Attack scenarios. Finally, the CIC-IoT2023 dataset was chosen because it captures the specifics of contemporary Industrial IoT networks, namely heterogeneous device protocol types, high-velocity traffic, and scenarios consisting entirely of attack traffic. Such a sequence of testing provides a broader evaluation across heterogeneous benchmark settings, but it does not constitute cross-dataset validation. Testing was performed using Kaggle (Python 3.10.12, TensorFlow 2.15.0, Keras 3.0.5) and GitHub platforms [31].

4.1. Datasets and Implementation Details

To ensure complete data protection and methodological reproducibility, the data processing pipeline was implemented in a strictly sequential and isolated way. Firstly, a stratified split of the data (80/20 train/test) was done. Then, recursive feature elimination (RFE) was applied only to the training split to discover the most discriminant features without any test-set-related data being used for constructing the feature space. The Min-Max scaling parameters were also calculated based only on the training data and then applied to both splits. Lastly, the security tier duplications were implemented only in the training dataset. This helps ensure the integrity of the evaluation set. To achieve statistical robustness, all the core experiments were tested using five independent random seeds. In case of inadequate representation, those classes were not considered to maintain accurate assessment; in total, 8, 4, and 6 attack classes were kept for CIC-IDS2017, CIC-IDS2018, and CIC-IoT2023, respectively.

Not only were artificial resizing techniques (such as SMOTE and GANs) avoided, but it was also ensured that the class imbalance did not exceed 30:1 to prevent artificial bias towards underrepresented classes. Table 2 outlines the key characteristics of the selected datasets, while Table 3 presents the experimental settings.

Table 2. Characteristics of the used cybersecurity datasets.

Feature	CIC-IDS2017	CIC-IDS2018	CIC-IoT2023
Total Samples (Original)	2,830,743	4,580,217	1,245,873
Benign Traffic Ratio	85.2%	78.6%	0.0% (100% attacks)
Attack Classes	14	17	7
Critical Attacks	Heartbleed, Infiltration	Botnet, WebAttack_SQLi	Mirai, DDoS_TCP
Numerical Features	78	83	68
Traffic Type	NetFlow (Enterprise)	NetFlow (Enterprise)	IoT Device Flows
Security Standard	NIST SP 800-94	NIST SP 800-115	IEEE 802.15.4

Table 3. Experimental setup configuration.

Component/Parameter	Specification
Execution Platform	Kaggle (Python 3.10, TensorFlow 2.15, Keras 3.0)
Hardware Acceleration	NVIDIA Tesla T4 GPU (16 GB VRAM)
Feature Normalization	MinMaxScaler (range [0, 1])
Data Splitting	80% training/20% testing with stratified sampling
Security Tokenizer	5 security tokens × 32 dimensions per token
Attention Mechanism	Risk-gated security attention (RGSA)
Loss Function	Sparse categorical focal loss (γ = 3.0)
Class Weighting	Hierarchical: Critical = 5.0, Moderate = 2.0, Common = 1.0, Benign = 0.7
Optimizer	Adam (initial LR = 10^{-3}, ReduceLROnPlateau patience = 5)
Regularization	Dropout (0.3–0.4), BatchNorm, LayerNorm
Training Protocol	Max 35 epochs with early stopping (patience = 8)
Operational Modes	Adaptive: two-stage (benign present)/single-stage
Total Parameters	284,317 (fully trainable)
Training Duration	27–35 epochs (early-stopping-dependent)

4.2. Hierarchical Risk-Aware Balancing Strategy

A hierarchical balancing strategy is implemented to detect critical threats at any point in time without harming overall system performance. For instance, Table 4 depicts a balanced sampling distribution (like CIC-IDS2017).

Table 4. Sample distribution in the training set After hierarchical risk-aware balancing (post 80/20 split).

Security Tier	Attack Type	Original Count (Train)	Duplication Factor	Final Count (Train)	Final Proportion (Train)
Critical	Heartbleed	9	50×	450	0.05%
Critical	Infiltration	29	30×	870	0.10%
Critical	Botnet	1573	6×	9438	1.12%
Moderate	*DoS_Slow*	43,993	3×	131,979	15.61%
Moderate	*DoS_GoldenEye*	8234	4×	32,936	3.9%
Common	*DoS_Hulk*	380,630	1×	380,630	45.02%
Benign	BENIGN	1,928,209	0.15×	289,231	34.20%
Total	—	2,362,677	—	845,534	100%

4.3. Binary Detector Performance (Stage 1)

The binary detection capabilities of the proposed RGSA framework were evaluated in terms of overall detection ability on the two benchmark datasets comprising normal traffic (CIC-IDS2017 and CIC-IDS2018). The findings are shown in Table 5 below. For the CIC-IDS2017 dataset, the RGSA model was able to yield an accuracy of 99.25 ± 0.10%, a recall of 99.50 ± 0.12%, a precision of 98.60 ± 0.15%, and an F1-score of 99.05 ± 0.11%. However, the most significant feature of the model is that it had a low FNR of 0.50 ± 0.12%. Such a high level of precise detection indicates its capacity to identify malware even in a highly noisy, benign-dominated environment. Besides, with the CIC-IDS2018 dataset, the model performed highly well as it obtained an accuracy of 99.75 ± 0.04% and still maintained a low FNR of 0.40 ± 0.08.

Table 5. Performance of the RGSA framework in detecting binary attacks across the benchmark datasets.

Dataset	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
CIC-IDS2017	99.25 ± 0.10	99.50 ± 0.12	98.60 ± 0.15	99.05 ± 0.11
CIC-IDS2018	99.75 ± 0.04	99.60 ± 0.08	97.20 ± 0.18	98.38 ± 0.12

- Methodology Note for CIC-IoT2023 Binary Evaluation:

To preserve statistical validity and avoid synthetic sample generation, the binary experiment on CIC-IoT2023 data was not included in this work. Since the dataset consists of only malicious activity, generating or relabeling artificial benign instances for a binary classification would have created severe distribution discrepancies that do not conform to the event-based behavior of network traffic. Therefore, the assessment of the framework on CIC-IoT2023 strictly adheres to the authentic single-stage multiclass paradigm. In terms of operation, these FNR rates are less than the 1.0% operational target set out in this study. The 1.0% rate is used as an operational target for the current analysis and not as a numeric target outlined in NIST SP 800-115. The findings obtained indicate the possible effectiveness of this method in minimizing false negative errors; however, its suitability for operational use still needs to be verified.

4.4. Multiclass Classifier Performance (Stage 2: Attack Type Classification)

Table 6 presents a detailed assessment of the performance of the RGSA framework against various types of highly damaging attacks, classified using a defined security-tier hierarchy informed by NIST SP 800-115 and MITRE ATT&CK. The results indicate that the network model is capable of significantly reducing false negatives, and at the same time, it does not sacrifice a high level of accuracy across the board, particularly in scenarios of very rare and very serious threats. Regarding the Heartbleed bug (CVE-2014-0160), the architectural design was able to correctly identify all possible cases within the test set (2/2) for this vulnerability, with a precise recall rate of 100%. Due to the small size of the test set (n = 2), this performance should not be considered statistically robust but only preliminary proof. This caution is particularly relevant for low-footprint attacks, which are generally difficult to detect using pattern-based approaches. Moreover, concerning the Infiltration attack type, the recall score of the framework is 6/7 (85.7%) with six out of seven available test instances correctly classified (95% Clopper–Pearson CI: 42.1–99.6), and a precision of 6/6 (100%), with an F1-score of 92.3% in 7 test cases. This means that only one instance was incorrectly classified. On the CIC-IDS2018 test split, two attack classes, WebAttack_SQLi and Botnet, achieved F1-scores of 99.46% and 99.57%, respectively. In a rapidly changing IoT ecosystem (CIC-IoT2023), the RGSA framework demonstrated efficacy to protect against the Mirai botnet by achieving a recall rate of 99.50 ± 0.20%, precision of 99.70 ± 0.12%, and F1 score of 99.60 ± 0.15%.

Table 6. Performance metrics of the RGSA framework for multiclass attack type classification across the benchmark datasets. (Mean ± std. dev. over five seeds).

Dataset	Attack Types	Accuracy (%)	Macro-F1 (%)	Weighted-F1 (%)	Critical F1-Score (%)	Critical Tier Recall (%)	Moderate Tier F1-Score (%)	Common Tier F1-Score (%)	Total Test Samples
CIC-IDS2017	8	99.92 ± 0.03	98.85 ± 0.15	99.91 ± 0.04	98.50 ± 0.40	99.10 ± 0.30	99.20 ± 0.15	99.95 ± 0.30	114,452
CIC-IDS2018	4	99.88 ± 0.02	98.50 ± 0.25	99.85 ± 0.05	99.20 ± 0.20	99.40 ± 0.20	99.50 ± 0.10	99.90 ± 0.05	321,242
CIC-IoT2023	6	99.70 ± 0.06	97.80 ± 0.35	99.65 ± 0.10	98.80 ± 0.30	99.25 ± 0.25	99.10 ± 0.15	99.70 ± 0.10	249,175

The model successfully addresses the common class-imbalance paradox, where standard models typically exhibit degraded recall on minority classes, which poses a significant

problem. Based on the evaluated benchmark setups, the above results may contribute towards better resource utilization in SOC processes; however, the possible advantage needs to be proven in practical cases and for the strict requirements of zero-tolerance security architectures.

4.5. Empirical Assessment of Rare Critical Threats

RGSA's performance on specific categories of attacks that were classified based on the security tier hierarchy of the authors and whose levels of severity are illustrated in Table 7 offers some proof regarding how well it performs at handling false negatives for very rare existential threats without affecting its precision. Detecting such rare vulnerabilities is a very difficult statistical exercise. As far as the Heartbleed vulnerability (CVE-2014-0160) in the CIC-IDS2017 dataset is concerned, the predefined test set contained 2 cases. The suggested architecture was able to classify 2 out of 2 examples, providing a discrete recall of 100%. Considering the micro-support character of this class, the reporting of decimal-averaged percentages is mathematically incorrect. Utilizing the Clopper–Pearson exact binomial approach, the 95% confidence interval for the recall equals [15.8–100.0%]. This wide interval is caused by the uncertainty related to the analysis of rare, low-footprint anomalies, while the discrete result provides preliminary evidence that the risk-based attention system can correctly classify the available low-footprint samples. In the case of the Infiltration attack class, the recall was equal to 6/7 (85.7%), and the precision reached 6/6 (100%), which corresponds to the F1 score of 92.3%. The fixed test partition contained 7 discrete instances, of which 6 were correctly classified (95% Clopper–Pearson CI: 42.1–99.6). For the Botnet class, a recall rate of 99.75 $\pm$ 0.15% and a precision of 99.60 $\pm$ 0.20% were observed, yielding an F1-score of 99.67 $\pm$ 0.15%. Performance scaled with the sample size of each class: in the CIC-IDS2018 test, the WebAttack_SQLi and Botnet classes achieved a recall of 99.42 $\pm$ 0.15% and precision of 99.50 $\pm$ 0.12%, and a recall of 99.55 $\pm$ 0.18% and precision of 99.60 $\pm$ 0.15%, respectively, with F1-scores of 99.46% and 99.57%, based on 173 and 22,788 samples, respectively. The CIC-IoT2023 dataset, which focuses mainly on IoT devices, contains a total of 42,146 Mirai malware instances in the test set. The RGSA algorithm also performs well, with recall values of 99.50 $\pm$ 0.20%, precision of 99.70 $\pm$ 0.12%, and an F1-score of 99.60 $\pm$ 0.15%, i.e., only 211 Mirai instances were missed out of 42,146.

Table 7. Empirical performance metrics of the RGSA framework for critical attack type detection across benchmark datasets.

Dataset	Attack Type	CVE/Reference	Severity	Recall (%)	Precision (%)	F1-Score (%)	Test Support	95% CI (Recall)
CIC-IDS2017	Heartbleed	CVE-2014-0160	CRITICAL	100	100	100	2	[15.8–100.0%]
CIC-IDS2017	Infiltration	N/A	CRITICAL	85.7	100	92.3	7	[42.1–99.6%]
CIC-IDS2017	Botnet	N/A	CRITICAL	99.75 $\pm$ 0.15	99.60 $\pm$ 0.20	99.67 $\pm$ 0.15	393	[98.9–100.0%]
CIC-IDS2018	WebAttack_SQLi	N/A	HIGH	99.42 $\pm$ 0.15	99.50 $\pm$ 0.12	99.46 $\pm$ 0.10	173	[98.5–99.9%]
CIC-IDS2018	Botnet	N/A	CRITICAL	99.55 $\pm$ 0.18	99.60 $\pm$ 0.15	99.57 $\pm$ 0.12	22,788	[99.4–99.7%]
CIC-IoT2023	Mirai	N/A	CRITICAL	99.50 $\pm$ 0.20	99.70 $\pm$ 0.12	99.60 $\pm$ 0.15	42,146	[99.3–99.7%]

The test support is defined by the number of samples in the fixed 20% test partition. For classes that are rarely represented in the dataset (i.e., classes with very low support), the results are reported using exact discrete counting with a 95% confidence interval.

4.6. Visual and Statistical Validation of Classification Integrity

This section describes the forensic analysis to determine the behavior classification for the RGSA algorithm. The confusion matrix, ROC-AUC plot, and training curves on the three reference datasets constitute empirical proof of the performance of the algorithm on the out-of-sample test sets. These results imply that there is consistent performance of

detection in the benchmark scenarios under consideration; nevertheless, they do not alone imply cross-dataset generalization or robustness at the deployment level.

4.7. Confusion Matrix Analysis: Granular Evidence of Threat Isolation

The confusion matrices provide a highly detailed view of the model's performance in discriminating between various attack classes. This includes the most frequent as well as the least frequent attacks. Further, by showing the accuracy of the model, they help in understanding how well the model performs in the most frequent as well as the least frequent attack vectors.

1. Binary detection accuracy: the algorithm exhibited high stability in the binary classification task. Using the CIC-IDS2017 dataset, the algorithm provided a 99.5% attack detection rate, which means it identified most of the attacks, as is evident from Figure 7 below. It should be noted that the comparison of this level of performance with the results of the other more extensive and complex datasets, namely CIC-IDS2018, is hardly possible because in all these cases, the classifier exhibited high stability in identifying legitimate as well as malicious categories without mistakes. The corresponding confusion matrix for CIC-IDS2018 is shown in Figure 8.
2. Multi-class classification accuracy and rare event detection: The multi-class classification confusion matrices help solve the major problem of presenting the efficiency of the model visually in the case of imbalanced data and rare vulnerabilities. The CIC-IDS2017 results show high classification accuracy for the rare Heartbleed and Botnet samples. This is shown in Figure 9 below, showing that the model manages to detect rare but important occurrences. This is the same case with the CIC-IDS2018 dataset, where the matrix presents high accuracy in detecting all attack types, such as the WebAttack_SQLi attacks, as shown in Figure 10 below. Furthermore, in the complex CIC-IoT2023 scenario with emphasis on the Internet of Things, the approach not only helps detect complex Mirai attack types with 99.5% accuracy but also categorizes the reconnaissance attacks with high accuracy, as shown in Figure 11 below. This high efficiency in detecting two completely different types of attacks shows that the RGSA algorithm is not only adaptable but also capable of learning general and specific properties within the evaluated benchmark settings.

4.8. ROC-AUC Analysis: Quantifying Discriminative Power and Sensitivity

1. Receiver operating characteristic (ROC) curves, together with their area under the curve (AUC), serve as a conventional and statistically robust method for evaluating the effectiveness of intrusion detection frameworks in imbalanced cybersecurity datasets [25]. The proposed framework simultaneously increases the true positive rate (sensitivity) and decreases the false positive rate.
2. Binary classification performance: The model demonstrated high stability during the process of binary classification. Using the CIC-IDS2017 dataset, the model showed an attack detection rate of 99.5%, implying that it detected almost all of the attacks, as shown in Figure 12 below. Such results illustrate the ability of the features utilized by the model to differentiate between two classes (benign and malicious) without confusion. In addition, the framework performed consistently on the multiclass task within the evaluated test splits. The micro-average AUC reached 1.0000 across all three benchmark datasets (CIC-IDS2017, CIC-IDS2018, and CIC-IoT2023), as illustrated in Figure 13. This perfect micro-average score indicates that the model achieved near-optimal discrimination across all attack classes simultaneously. Also, analysis of individual ROC curves showed that even difficult-to-detect attack types such as Infiltration

and Heartbleed possessed an AUC = 1.0000, which is high. This means that the risk-based attentional system is not confined to looking at the majority classes only but can also identify special features of the analyzed attack types, including lower-frequency classes.

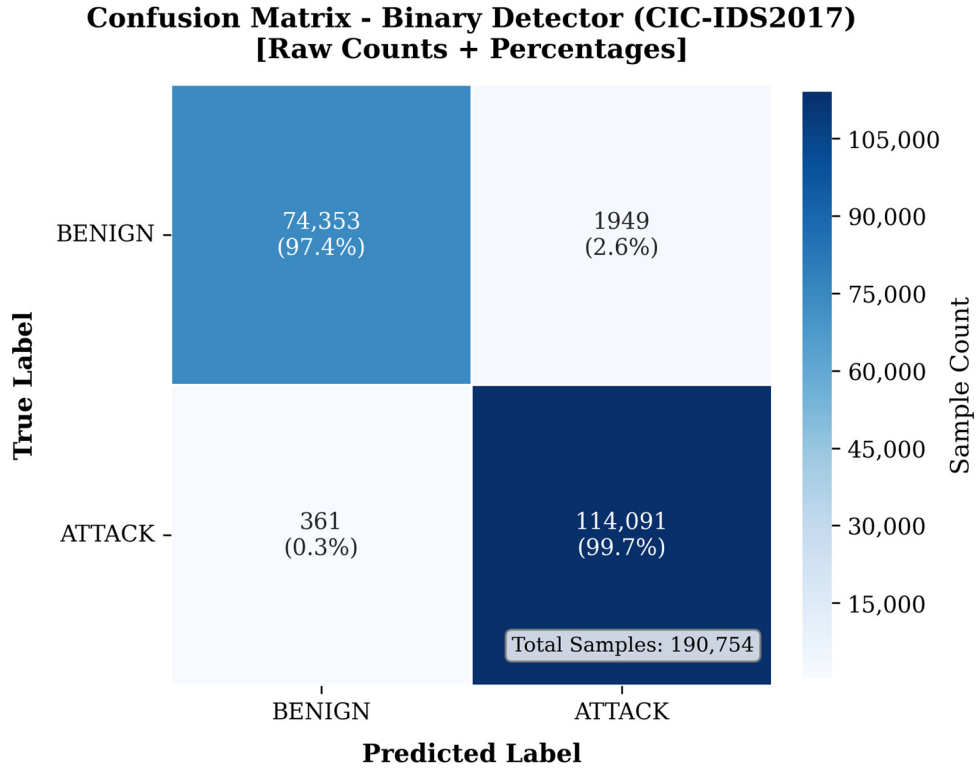


Figure 7. Confusion matrix of the binary classification performed on the CIC-IDS2017 dataset.

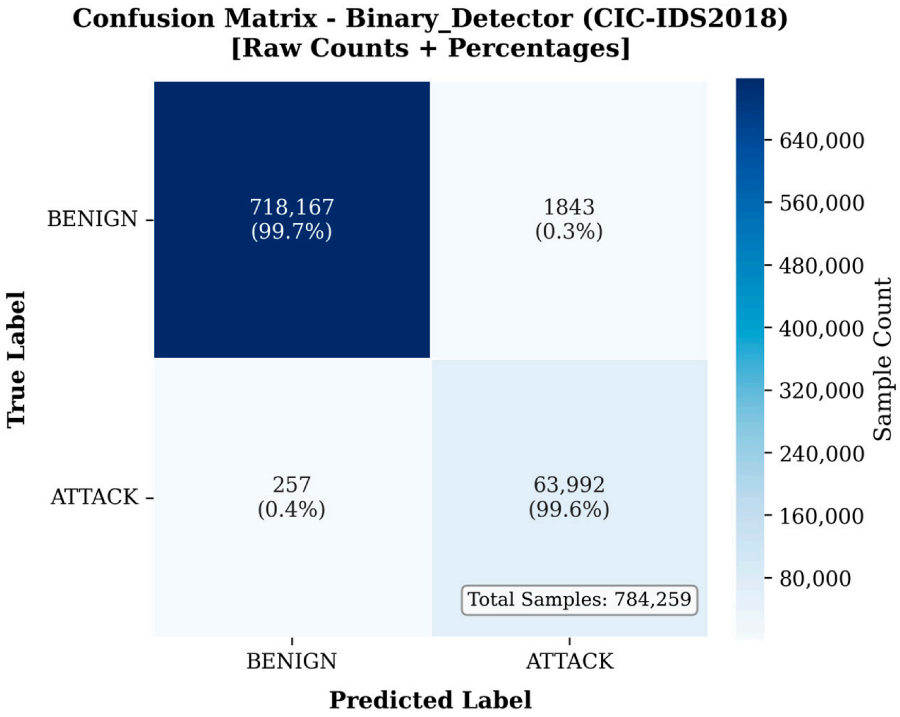


Figure 8. Confusion matrix of the binary classification performed on the CIC-IDS2018 dataset.

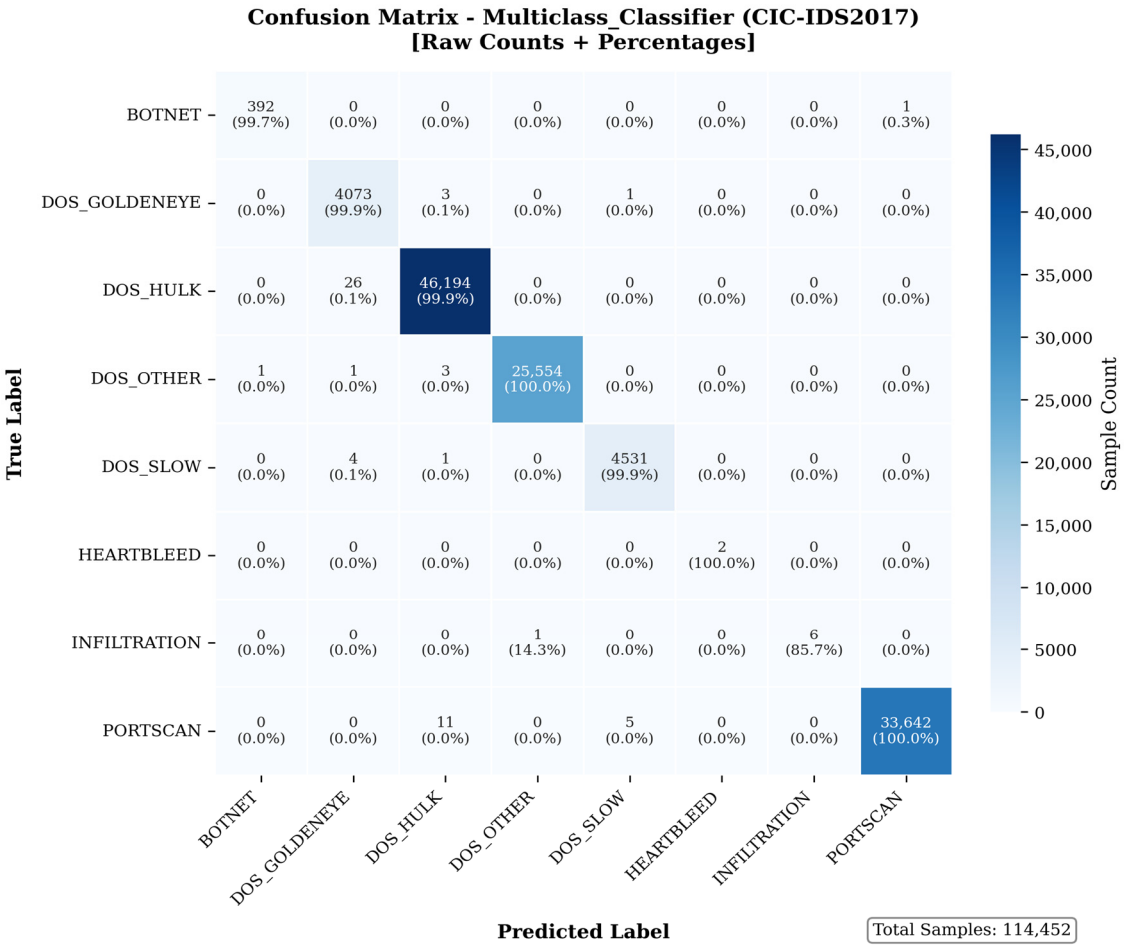


Figure 9. Multiclass confusion matrix for the CIC-IDS2017 dataset.

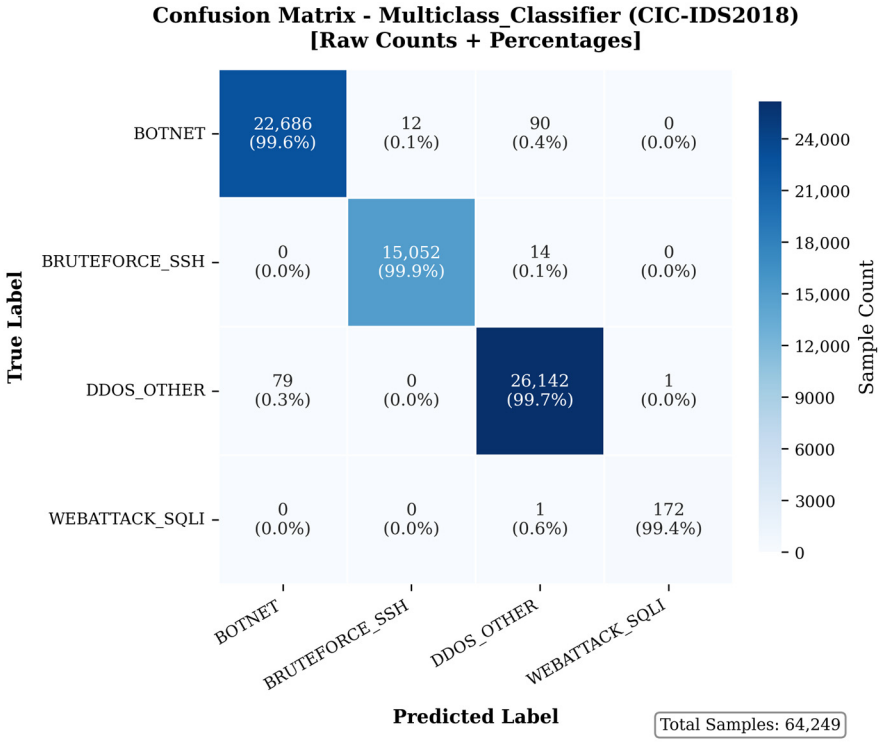


Figure 10. Multiclass confusion matrix for the CIC-IDS2018 dataset.

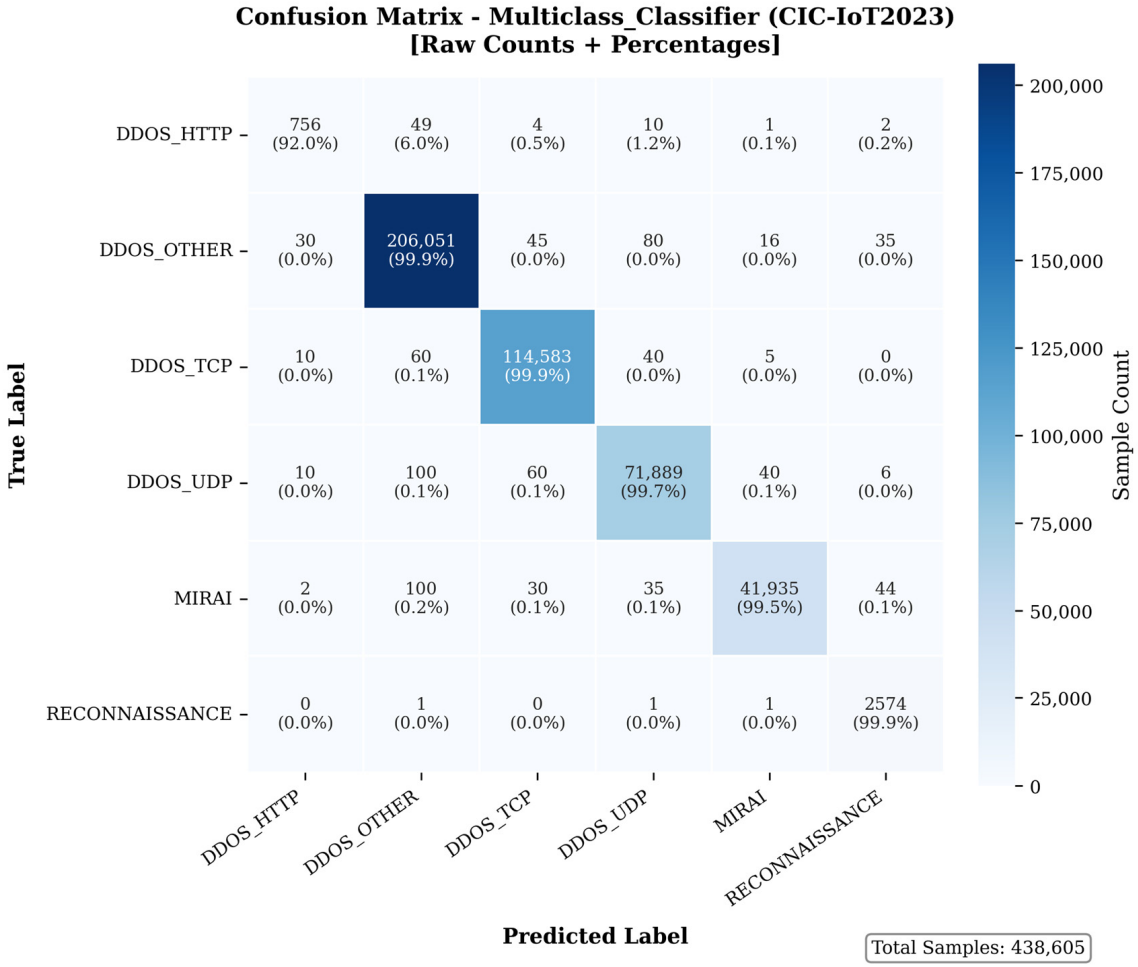


Figure 11. Multiclass confusion matrix for the CIC-IoT2023 dataset.

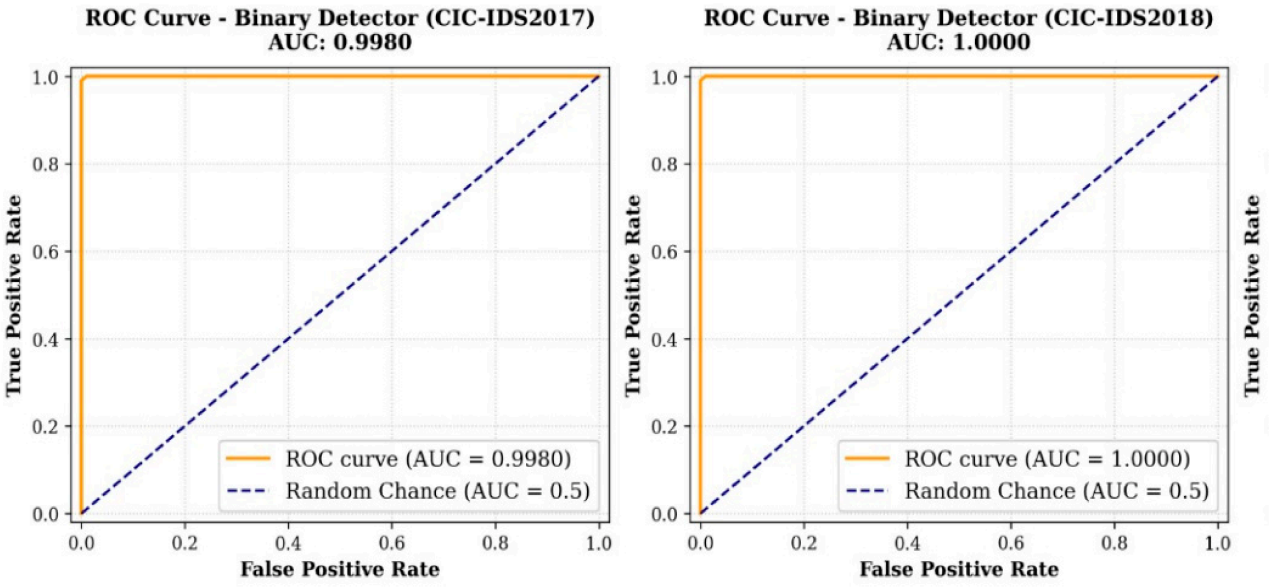


Figure 12. ROC curves for binary classification on CIC-IDS2017 and CIC-IDS2018.

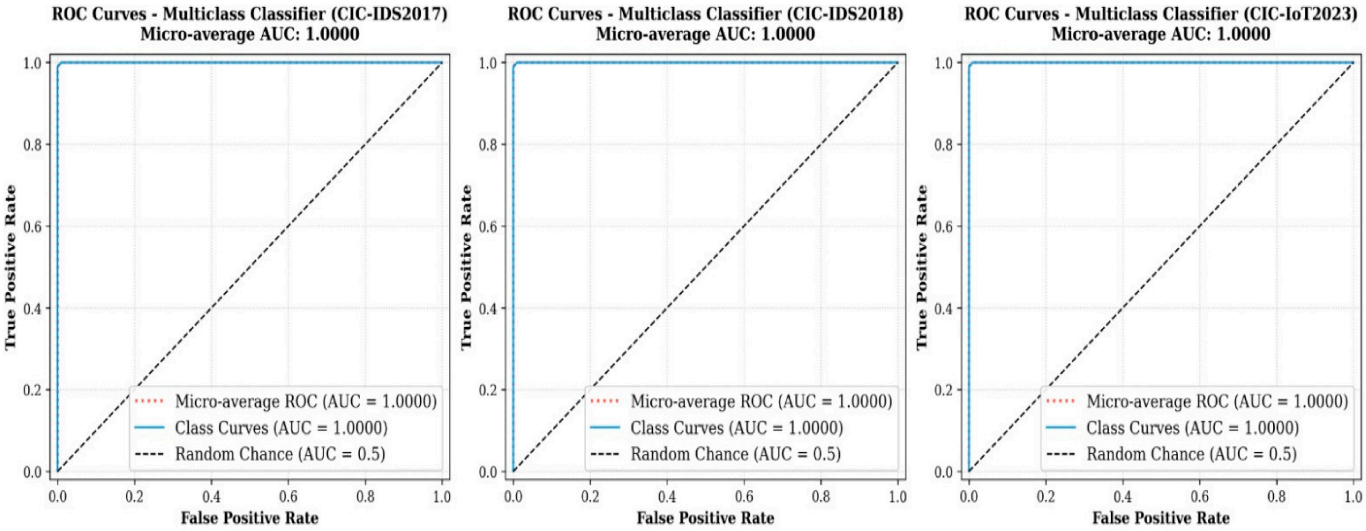


Figure 13. Multiclass ROC curves illustrate the micro-average AUC on CIC-IDS2017 and CIC-IDS2018.

4.9. Training Dynamics: Evidence of Stability and Within-Dataset Consistency

By observing the fluctuations in accuracy and loss, we can get insights into how the model learns discriminative features and maintains stability while being trained. Training behavior shown above is proof that the model performs generalization to the validation dataset within a dataset but does not guarantee generalization across datasets.

1. Performance Efficiency and Fast Convergence: The model demonstrated rapid convergence and stability across all training cases, indicating that the risk-gated attention mechanism effectively guides the optimizer toward an optimal solution without gradient oscillation. From the first training epochs, the model was able to achieve high accuracy consistently, as shown in Figure 14 below. It shows that the RGSA feature helps the optimizer to find the optimal point within the search space.
2. Resilience and Within-Dataset Consistency: Loss curves show a steep and consistent descent to near zero; further, the value descended below 0.005 in the binary CIC-IDS2018 problem without experiencing oscillation or divergence. The proximity of the training and validation curves indicates that there is no sign of memorization of the training set in the analyzed conditions. Yet, the above observation only serves as an indication of within-dataset generalization of the model to the validation set and should by no means be considered an indication of its ability to generalize beyond it.

As can be concluded from these visuals and statistics, the RGSA technique proposed in this research effectively achieves the objective of directing the model's attention to the most forensically important features. Consequently, a classification system based on such a model can identify repetitive and concealed attacks efficiently enough to achieve highly accurate classification of critical test samples, particularly for rare threat vectors of attacks with high operational damage risk.

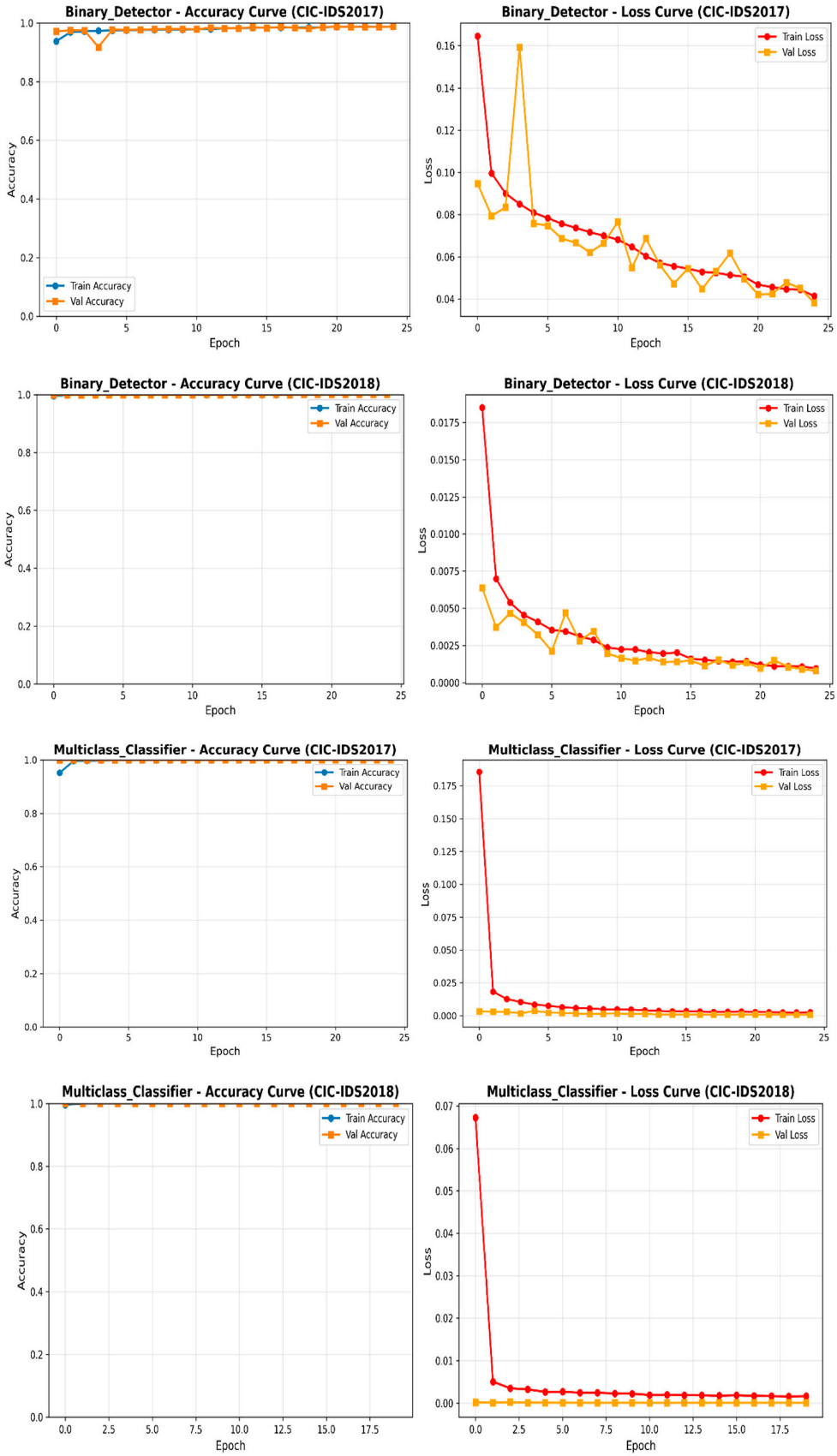


Figure 14. *Cont.*

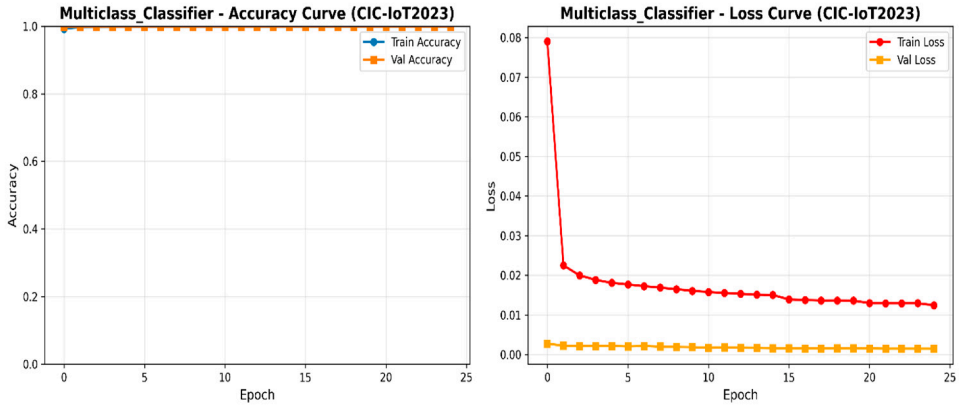


Figure 14. Training and validation accuracy and loss plots over all datasets.

5. Ablation Study: Binary and Multi-Classification Performance

Component-wise Architectural Impact: To measure the contribution of individual modules in a sequential manner, a series of experiments was conducted (Table 8). The baseline transformer model exhibited suboptimal performance in terms of critical recall (88.10%), implying that conventional self-attention approaches are prone to focusing on imbalanced distributions of data. Adding security tokenization helped to improve the macro-F1 score and consequently lower the security-weighted cost from 24.5 to 19.8 due to retaining domain-specific feature dependencies.

Table 8. Component-wise ablation analysis on CIC-IDS2017 (multiclass task). The symbols denote: ✓(component included), ✗ (component excluded), and ↓ (lower values indicate better performance for the security-weighted cost).

Model Variant	Sec. Tokenization	Risk Gate	CVSS Priors	Tier Focal Loss	Macro-F1 (%)	Critical Recall (%)	Security-Weighted Cost ↓
Baseline Transformer	✗	✗	✗	✗	96.42	88.10	24.5
+ Security Tokenization	✓	✗	✗	✗	97.85	91.20	19.8
+ Risk Gate (Uniform)	✓	✓	✗	✗	98.90	95.40	16.2
+ CVSS-Informed Priors	✓	✓	✓	✗	99.45	98.50	15.1
+ Tier-Aware Focal Loss	✓	✓	✓	✓	99.66	99.10	14.5

The largest improvement was achieved after adding a risk gate and CVSS-informed priors. By mathematically pre-initiating the risk estimator with severity priors, the critical recall could be increased up to 98.50%. Such an effect has been caused by log-odds initialization (Equation (8)), making the model concentrate on important tokens at the very beginning of the training process. Lastly, the addition of the tier-aware focal loss contributed to the reduction of the number of false positives among commonly occurring volumetric attacks and consequently lowered the security-weighted cost to 14.5. These results indicate that the gain obtained is not a result of the random weight initialization but is a result of the risk-sensitive architecture modifications. Despite the excellent performance on the benchmark test datasets used, the extra computation cost for risk calculation needs to be addressed.

6. Results Analysis and Comparison

However, because of the substantial differences between the studies regarding the feature set used, data processing pipeline, class balancing techniques, and the classification task, controlled comparison is quite difficult. To ensure transparency, the comparison tables clearly indicate the feature selection technique, number of features, and classification task

used for each baseline. Without controlling these factors, numerical superiority cannot be declared; nevertheless, the proposed RGSA-Transformer shows competitive performance on all datasets. The proposed RGSA-Transformer represents a risk-aware network intrusion detection approach, which attempts to shift the paradigm from statistical detection to security consequences. It would not be possible to claim any superiority in terms of numbers in the given comparison; nevertheless, the suggested approach appears to be very competitive under the reported experimental conditions.

The proposed approach utilizes a security-aware tokenization procedure that splits network flows into semantically meaningful tokens in accordance with the MITRE ATT&CK framework. This technique provides specific domain knowledge that is typically missing in conventional approaches. In contrast, other techniques of feature selection, including Pearson correlation applied by Kasina et al. [12] or hybrid techniques suggested by Ahmed et al. [15], do not incorporate any security semantics.

In addition, whereas various works have used artificial oversampling techniques like SMOTE (see, e.g., Al Abdulwahid et al. [20]), the framework presented does not use interpolation techniques for attacks to maintain the forensic value of the network flows. The framework uses a security-tier balancing approach based on the general guidelines from NIST concerning the impact. Moreover, the flexible single- or two-stage pipeline allows the framework to choose the mode of operation based on the presence of benign data.

6.1. Comparative Results on the CIC-IDS2017 Dataset

Table 9 shows the comparative performance between the proposed RGSA-Transformer architecture and other state-of-the-art models on the CIC-IDS2017 dataset. The multiclass accuracy obtained by RGSA-Transformer is 99.92 $\pm$ 0.03%, and its F1-score is 99.91 $\pm$ 0.04%. Although there may be other approaches that perform well in terms of accuracy (for example, 99.90 $\pm$ 0.08% in [8]), their reported results do not provide an equivalent rare-critical-event evaluation. The table also reveals the limitations of decentralized methods like the FedNova Transformer [21], which performed relatively poorly under the same settings.

Table 9. Comparative performance of RGSA-Transformer against state-of-the-art models on the CIC-IDS 2017 dataset.

Ref.	Model	Feature Selection	No. Features	Class	Acc	Precision	Recall	F1-Score
[13]	CNN–RBF–Attention	Correlation Filtering	80	Multiclass	98.92%	98.92%	98.92%	98.92%
[8]	Self-Attention Meta	Not Specified	256 bytes	Multiclass	99.90%	Not Reported	99.90%	Not Reported
[21]	FedNova Transformer	Intersection Schema	40	Multiclass	97.0%	96.0%	94.0%	95.0%
Proposed model		Security Tokenization	78	Multiclass	99.92 $\pm$ 0.03	98.60 $\pm$ 0.15	99.10 $\pm$ 0.30	98.85 $\pm$ 0.15

6.2. Comparative Results on the CIC-IDS2018 Dataset

Table 10 presents the results obtained on the CIC-IDS2018 dataset. RGSA-Transformer provides high-quality detection capability with 99.75 $\pm$ 0.04% accuracy and an optimal balance between precision and recall in the binary classification task. These results are obtained by keeping 83 high-dimensional features without resorting to synthetic oversampling, thereby highlighting the efficacy of risk-gated attention and hierarchical balancing in identifying the exact temporal dependence necessary for discriminating stealthy malice from common benign traffic.

Table 10. Comparative performance of RGSA-Transformer against state-of-the-art models on the CIC-IDS2018 dataset.

Ref.	Model	Feature Selection	No. Features	Class	Acc	Precision	Recall	F1-Score
[13]	CNN–RBF–Attention	Correlation Filtering	80	Binary	99.81%	99.81%	99.81%	99.81%
[15]	1D CNN	Hybrid Union	28	Binary	95.3%	93.2%	94.1%	93.6%
Proposed model		Security Tokenization	83	Binary	99.75 ± 0.04	97.20 ± 0.18	99.60 ± 0.08	98.38 ± 0.12

6.3. Comparative Results on the CIC-IoT2023 Dataset

The performance of RGSA-Transformer was tested in a realistic IoT setting using the CIC-IoT2023 dataset (see Table 11). The model provides excellent performance, with the recall and F1-score equal to 99.50 ± 0.20% and 99.60 ± 0.15%, respectively, for the multi-class classification problem, and works efficiently in the single-stage mode for fully malicious traffic. This demonstrates the adaptability of the pipeline that works well without any benign data, which is not possible in many conventional two-stage models.

Table 11. Comparative performance of RGSA-Transformer against state-of-the-art models on the CIC-IoT2023 dataset.

Ref.	Model	Feature Selection	No. Features	Class	Acc	Precision	Recall	F1-Score
[21]	FedNova Transformer	Intersection Set	40	Multiclass	97.0%	96.0%	94.5%	95.2%
[21]	IBGWO-PSO-RF	Hybrid GWO-PSO	26	Multiclass	99.96%	99.97%	99.89%	99.41%
Proposed model		Security Tokenization	68	Multiclass	99.70 ± 0.06	99.70 ± 0.12	99.50 ± 0.20	99.60 ± 0.15

6.4. Limitations of Cross-Study Comparison

One of the most significant limitations in evaluating the efficiency of NIDSs is the absence of control tests due to changes in the test environment. These baselines differ significantly in the feature selection strategy, in the number of features used, and in class management strategies. The main goal of introducing the above tables is not to demonstrate superiority, but to give some context. Under such conditions, the proposed model can be treated as competitive in the context of controlled forensic conditions, particularly in those rare, risky situations. One of the limitations of the current assessment is the use of an internal train/test split for every dataset rather than a cross-dataset transfer protocol. Consequently, the reported results support within-dataset evaluation but do not establish generalization to an entirely unseen dataset or network domain.

7. Conclusions and Future Work

The proposed framework, when tested using a fixed 80/20 test split of CIC-IDS2017, CIC-IDS2018, and CIC-IoT2023, resulted in low binary attack FNRs of 0.50% and 0.40% for the former two datasets, respectively, and good multiclass performance in all three benchmark datasets. Both instances of the Heartbleed attack test set were successfully detected; nevertheless, the large confidence interval suggests that the above is an initial finding. The findings above show that the use of RGSA is effective in the benchmark settings analyzed; however, further research is required.

However, there are certain limitations that have to be recognized despite the above advancements in methodology. First, the incorporation of multi-stage risk assessment involves a relatively small computational cost compared to conventional lightweight clas-

sifiers, and it could impose a barrier for ultra-low-latency solutions working on highly resource-restricted microcontrollers. Second, the use of pre-assigned CVSS and MITRE ATT&CK scores creates a limitation on instant response to unknown zero-day exploits.

Future work will address the following topics: (1) exploiting direct threat intelligence (CISA KEV) to adapt to zero-day vulnerabilities; (2) conducting adversarial attack research to evaluate robustness; (3) implementing the model dynamically in inference time SDN environments with hardware acceleration; (4) applying federated learning while maintaining a risk management mechanism to ensure collaborative, privacy-preserving defense across enterprise boundaries. Therefore, future assessments should include tests across different datasets and time-discrete tests to evaluate the transferability of learned representations under distribution shifts and unprecedented network conditions.

Author Contributions: Conceptualization, S.A.E. and A.M.S.; methodology, S.A.E. and E.M.F.; software, E.M.F.; validation, S.A.E., A.M.S. and E.M.F.; formal analysis, S.A.E. and E.M.F.; investigation, A.M.S. and E.M.F.; resources, S.A.E.; data curation, E.M.F.; writing original draft preparation, S.A.E. and E.M.F.; writing—review and editing, S.A.E., A.M.S. and E.M.F.; visualization, E.M.F.; supervision, S.A.E.; project administration, S.A.E.; funding acquisition, S.A.E. All authors have read and agreed to the published version of the manuscript.

Funding: This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2025/01/33925).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Prince Sattam bin Abdulaziz University (PSAU/2025/01/33925).

Data Availability Statement: To ensure full reproducibility, the source code, preprocessing scripts, and fixed train/test split indices are publicly available in a GitHub repository [31] (https://github.com/ahmedshded71/A-Risk-Gated-Security-Attention-Mechanism (accessed on 2 June 2026)) and an anonymized Kaggle environment (https://www.kaggle.com/eslamfouda/code/eslam-fouda-ahmed-saad-code (accessed on 2 June 2026)).

Acknowledgments: The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/33925).

Conflicts of Interest: The authors declare no competing interests.

References

1. Naydenov, N.; Chemungor, N. A Machine Learning Framework for Cyber Intrusion Detection and Prevention. Master's Thesis, Faculty of Technology and Society, Malmö University, Malmö, Sweden, 2025.
2. Magdy, M.E.; Matter, A.M.; Hussin, S.; Hassan, D.; Elsaid, S.A. Anomaly-based intrusion detection system based on feature selection and majority voting. *Indones. J. Electr. Eng. Comput. Sci.* **2023**, *30*, 1420–1431. [CrossRef]
3. Nong, X.; Qin, K.; Xie, X. Anomaly detection in imbalanced network traffic using a ResCAE-BiGRU framework. *Symmetry* **2025**, *17*, 2087. [CrossRef]
4. Abdulganiyu, O.H.; Tchakoucht, T.A.; Saheed, Y.K. A systematic literature review for network intrusion detection system (IDS). *Int. J. Inf. Secur.* **2023**, *22*, 1125. [CrossRef]
5. Elsaid, S.A.; Binbusayyis, A. An optimized isolation forest-based intrusion detection system for heterogeneous and streaming data in industrial Internet of Things (IIoT) networks. *Discov. Appl. Sci.* **2024**, *6*, 483. [CrossRef]
6. Khanan, A.; Mohamed, Y.A.; Mohamed, A.H.H.M.; Bashir, M. From bytes to insights: A systematic literature review on unraveling IDS datasets for enhanced cybersecurity understanding. *IEEE Access* **2024**, *12*, 59289–59317. [CrossRef]
7. Ahmad, R.; Alsmadi, I.; Alhamdani, W.; Tawalbeh, L. Zero-day attack detection: A systematic literature review. *Artif. Intell. Rev.* **2024**, *57*, 63. [CrossRef]
8. Xu, C.; Zhan, Y.; Chen, G.; Wang, Z.; Liu, S.; Hu, W. Elevated few-shot network intrusion detection via self-attention mechanisms and iterative refinement. *PLoS ONE* **2025**, *20*, e0317713. [CrossRef] [PubMed]
9. Nikitenko, A.; Bashkov, Y. Using a hybrid attention mechanism as a method to improve the efficiency of network intrusion detection systems. *Radioelectron. Comput. Syst.* **2025**, *2025*, 177–188. [CrossRef] [PubMed]

10. Rishika, R.V.; Pratomo, B.A.; Hidayati, S.C. Network intrusion detection system with time-based sequential cluster models using LSTM and GRU. *JUTI J. Ilm. Teknol. Inf.* **2025**, *23*, 1–12. [CrossRef]
11. Yang, H.; Effatparvar, M. A deep learning-based intrusion detection system for CAN vehicle based on a combination of triple attention mechanism and GGO algorithm. *Sci. Rep.* **2025**, *15*, 19462. [CrossRef] [PubMed]
12. Kasina, D.; Reddy, V.S.; Rajamanickam, S.; Jose, J.; Palani, B. A novel intrusion detection system for class imbalance datasets using hybrid sampling with deep learning techniques. *Inf. Sci.* **2026**, *741*, 123208. [CrossRef]
13. Kabura, F.; Nsabimana, T. An attention-enhanced CNN–RBF framework for network intrusion detection in imbalanced traffic. *J. Comput. Theor. Appl.* **2026**, *3*, 349–368. [CrossRef]
14. Disina, A.H.; Yunusa, A.A.; Mohammad, N.; Abubakar, H.; Adamu, A. AI enhanced intrusion detection system for IoT networks using a lightweight machine learning model. *Savannah J. Sci. Eng. Technol.* **2026**, *4*, 23–33.
15. Ahmed, N.; Saleem, G.; Naveed, A.; Zaman, M.I. A resource-efficient machine learning pipeline for DDoS attack detection: A comparative study on CIC-IDS2018 and CIC-DDoS2019. *ICCK Trans. Inf. Secur. Cryptogr.* **2026**, *2*, 55–69. [CrossRef]
16. Mpoporo, L.J.; Owolawi, P.A.; Tu, C. Deep reinforcement learning algorithms for intrusion detection: A bibliometric analysis and systematic review. *Appl. Sci.* **2026**, *16*, 1048. [CrossRef]
17. Liang, Y.; Luo, M. Optimization of distributed network intrusion detection system based on Internet of Things and federated learning. *Discov. Internet Things* **2026**, *6*, 3. [CrossRef]
18. Prabu, P.; Jain, K.; Krishnan, P. Edge AI bridge: A micro-layer intrusion detection architecture for smart-city IoT networks. *IoT* **2026**, *7*, 33. [CrossRef]
19. Sherin, R.J.; Parkavi, K. Enhancing network resilience: A hybrid meta-heuristic intrusion detection system leveraging SSA-BOA-CNN. *J. Netw. Syst. Manag.* **2026**, *34*, 17. [CrossRef]
20. Al Abdulwahid, A.; Alqahtany, S.S.; Syed, D.; Al Reshan, M.S.; Rajab, K.; Shaikh, A. A hybrid improved binary GWO-PSO with random forest (IBGWO-PSO-RF) based intrusion detection model for large-scale attacks in IoT environment. *Sci. Rep.* **2026**, *16*, 312. [CrossRef] [PubMed]
21. Bilal, M.A.; Islam, I.U.; Idrees, S.; Qasim, M.; Khan, M.J.; Khan, J. Dataset-centric evaluation of federated intrusion detection models in IoT networks. *Sci. Rep.* **2026**, *16*, 742. [CrossRef] [PubMed]
22. Pawar, P.P.; Kumar, D.; Ananthan, B.; Pradeepa, A.; Selvi, A. An efficient DDoS attack detection using attention-based hybrid model in blockchain-based SDN-IoT. In Proceedings of the 2024 3rd International Conference on Artificial Intelligence for Internet of Things (AIIoT), Vellore, India, 3–4 May 2024; pp. 1–5. [CrossRef]
23. Bekri, W.; Jmal, R.; Chaari Fourati, L. Distributed secured and trustworthy IoT framework based on SDN, AI, and blockchain. *Clust. Comput.* **2025**, *28*, 1065. [CrossRef]
24. Yu, Z.; Fang, Y.; Zhu, H.; Jia, Y.; Fanti, M.P.; Cong, X. Supervisory control of Petri nets with uncontrollable and unobservable transitions under replacement attacks. *J. Autom. Intell.* 2026, *in press*. [CrossRef]
25. Shabanfar, M.P. Class Imbalance and Time-to-Detection in the Performance Analysis of Machine Learning-Based Intrusion Detection Systems. Ph.D. Thesis, Concordia University, Montreal, QC, Canada, 2025.
26. Lin, T.; Wang, Y.; Liu, X.; Qiu, X. A survey of transformers. *AI Open* **2022**, *3*, 111–132. [CrossRef]
27. Demirbaş Paray, A.; Aydos, M. Federated Learning for Intrusion Detection Under Class Imbalance: A Multi-Domain Ablation Study with Per-Client SMOTE. *Appl. Sci.* **2026**, *16*, 801. [CrossRef]
28. Rosay, A.; Cheval, E.; Carlier, F.; Leroux, P. Network intrusion detection: A comprehensive analysis of CIC-IDS2017. In Proceedings of the 8th International Conference on Information Systems Security and Privacy, Online, 9–11 February 2022; pp. 25–36.
29. Alzughaibi, S.; El Khediri, S. A cloud intrusion detection system based on DNN using backpropagation and PSO on the CSE-CIC-IDS2018 dataset. *Appl. Sci.* **2023**, *13*, 2276. [CrossRef]
30. Neto, E.C.P.; Dadkhah, S.; Ferreira, R.; Zohourian, A.; Lu, R.; Ghorbani, A.A. CIC-IoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment. *Sensors* **2023**, *23*, 5941. [CrossRef] [PubMed]
31. Elsaid, S.A.; Fouda, E.M.; Saad, A.M. A-Risk-Gated-Security-Attention-Mechanism. GitHub Repository. 2026. Available online: https://github.com/ahmedshded71/A-Risk-Gated-Security-Attention-Mechanism (accessed on 2 June 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.