Computer Modeling in Engineering & Sciences

Tech Science Press

Doi:10.32604/cmes.2025.067756

ARTICLE

ScalaDetect-5G: Ultra High-Precision Highly Elastic Deep Intrusion Detection System for 5G Network

Shengjia Chang, Baojiang Cui* and Shaocong Feng

School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing, 100876, China

*Corresponding Author: Baojiang Cui. Email: cuibj@bupt.edu.cn

Received: 12 May 2025; Accepted: 15 July 2025; Published: 30 September 2025

ABSTRACT: With the rapid advancement of mobile communication networks, key technologies such as Multi-access Edge Computing (MEC) and Network Function Virtualization (NFV) have enhanced the quality of service for 5G users but have also significantly increased the complexity of network threats. Traditional static defense mechanisms are inadequate for addressing the dynamic and heterogeneous nature of modern attack vectors. To overcome these challenges, this paper presents a novel algorithmic framework, ScalaDetect-5G, designed for high-precision intrusion detection in 5G environments. ScalaDetect-5G adopts a three-stage architecture comprising traffic feature extraction, elastic representation, and adaptive classification. Specifically, an enhanced Concrete Autoencoder (CAE) is employed to reconstruct and compress high-dimensional network traffic features, producing compact and expressive representations suitable for large-scale 5G deployments. To further improve accuracy in ambiguous traffic classification, a Residual Convolutional Long Short-Term Memory model with an attention mechanism (ResCLA) is introduced, enabling multi-level modeling of spatial–temporal dependencies and effective detection of subtle anomalies. Extensive experiments on benchmark datasets—including 5G-NIDD, CIC-IDS2017, ToN-IoT, and BoT-IoT—demonstrate that ScalaDetect-5G consistently achieves F1 scores exceeding 99.19% across diverse network environments, indicating strong generalization and real-time deployment capabilities. Overall, ScalaDetect-5G achieves a balance between detection accuracy and deployment efficiency, offering a scalable, flexible, and effective solution for intrusion detection in 5G and next-generation networks.

KEYWORDS: 5G security; network intrusion detection; feature engineering; deep learning

1 Introduction

Over the past four decades, mobile communication systems have experienced a profound transformation, transitioning from the first-generation analog systems to the highly sophisticated fifth-generation (5G) digital infrastructures [1]. This evolution has signified a paradigm shift from primarily voice-centric services to a broad spectrum of data-driven applications, encompassing diverse functionalities such as video streaming and real-time cloud-based services [2,3]. These developments have not only redefined the user experience but also played a pivotal role in the exponential growth of the global information and communication technology (ICT) industry. Mobile communication networks are characterized by distinctive features such as dynamic flexibility, user mobility, adaptive bandwidth allocation, and real-time responsiveness. These attributes fundamentally differentiate them from traditional wired internet technologies [4], significantly enhancing accessibility, responsiveness, and scalability in data transmission and information exchange. With the recent global deployment of 5G technology, mobile networks are on

the brink of a significant leap in capabilities, offering transformative benefits in terms of performance, connectivity, and application scope [5]. Specifically, 5G networks are engineered to provide ultra-high data rates (up to 10 Gbps), immense system capacity (supporting up to 10^6 devices per km^2), and ultra-low latency (as low as 1 ms) with enhanced reliability. These technological advances facilitate the seamless integration of heterogeneous devices and services, including smart terminals, edge computing platforms, and industrial Internet of Things (IoT) systems. Consequently, 5G is not merely an upgrade in communication technology but also a foundational technology that enables intelligent applications such as autonomous driving, remote surgery environments. Thus, 5G plays a crucial role in driving innovation and economic development across various sectors.

The advent of 5G mobile communication technology has precipitated a surge of transformative innovations, notably Multi-Access Edge Computing (MEC) and Network Function Virtualization (NFV) [6,7]. These technologies enhance the agility, efficiency, and scalability of mobile networks through the decentralization of computation and the virtualization of network functions. However, they concurrently increase architectural complexity, diversify services, and expand system openness, thereby enlarging the network's attack surface and escalating its vulnerability to cyber threats. Looking forward, sixth-generation (6G) networks are poised to further revolutionize the communications landscape by delivering even faster transmission speeds (e.g., in the terabits-per-second range), near-instantaneous end-to-end latency (as low as 0.1 ms), and pervasive ultra-dense connectivity. Despite these anticipated advancements, such progress will inevitably be accompanied by increasingly intricate and dynamic security challenges. The conventional security mechanisms employed in 4G networks, which are primarily based on perimeter defense and static rule matching, are insufficient to counter the sophisticated and evolving threat landscape presented by 5G and forthcoming 6G systems. Attackers can exploit vulnerabilities in advanced components such as baseband processing, network slicing, service orchestration, and protocol implementations [8], potentially leading to critical service disruptions, breaches of user privacy, and extensive impacts on the economic and social stability of nations. Therefore, it is imperative to develop and implement proactive, adaptive, and high-performance intrusion detection and mitigation systems tailored specifically for the unique threat vectors introduced by next-generation mobile networks. Such systems must incorporate real-time monitoring, context-aware analysis, and AI-enhanced decision-making to effectively safeguard network infrastructures in the 5G and 6G eras.

The potent pattern recognition capabilities of deep learning are increasingly being harnessed for network intrusion detection, particularly within the realms of traffic characterization and classification [9–11]. Traffic characterization entails the transformation of raw network data into meaningful representations, employing both low-level features and high-level semantic abstractions. Conversely, traffic classification involves the categorization of network flows into benign or malicious groups, utilizing both conventional neural models and hierarchical deep learning architectures that incorporate temporal and spatial dependencies. However, the inherent complexity, heterogeneity, and dynamic nature of 5G traffic demand highly flexible and adaptive representational models to ensure precise and robust traffic characterization. With the increasing integration of 5G in critical infrastructural sectors such as smart cities, intelligent transportation systems (ITS), and the industrial Internet of Things (IIoT), the repercussions of subpar detection accuracy in intrusion detection systems (IDSs) can be severe, leading to service disruptions, suboptimal resource utilization, and potential safety hazards. The principal limitations of existing IDSs are as follows: (1) The feature extraction capabilities of low-level traffic descriptors, such as packet size, inter-arrival time, and flow duration, are insufficient for capturing the multidimensional patterns inherent in 5G traffic. For instance, 5G networks introduce new traffic types, such as ultra-reliable low-latency communications (URLLC) and massive machine-type communications (mMTC), which have distinct traffic behaviors that

traditional low-level features cannot adequately represent. While high-level features (e.g., aggregated traffic patterns or machine learning-based representations) can provide more abstract representations, they often suffer from scalability issues. As 5G mobile network architectures evolve and diversify, such as in the case of network slicing and dynamic spectrum allocation, high-level features may fail to capture the complex and evolving relationships between traffic patterns and network topologies, leading to reduced effectiveness. (2) Conventional classification methods, such as decision trees and support vector machines, often lead to low detection precision in 5G environments. These methods struggle to meet the real-time and stringent security requirements of next-generation mobile communication systems, especially when dealing with high volumes of data and high-speed traffic. In particular, traditional methods exhibit significant delays in classifying traffic patterns, causing them to miss short-lived attack signals in high-speed, low-latency 5G environments. Although hierarchical classification strategies, which organize traffic into multiple levels of classification (e.g., benign vs. malicious and further subcategories), can help reduce false alarm rates, they still face challenges in accurately identifying malicious behaviors within ambiguous, stealthy, or obfuscated traffic. These issues are exacerbated in 5G, where attack vectors are increasingly sophisticated and harder to distinguish from legitimate traffic due to the use of techniques like encryption and traffic obfuscation.

To achieve effective and discriminative representation learning of complex 5G network traffic characteristics and to significantly improve the accuracy of traffic classification, this study introduces an innovative intrusion detection framework termed ScalaDetect-5G. Initially, advanced features are extracted from network traffic using CICFLOWMETER. This framework incorporates an enhanced Concrete Autoencoder (CAE) for executing deep compression and latent-space reconstruction of high-dimensional traffic features, thereby facilitating the extraction of compact yet information-rich feature representations. Unlike traditional autoencoders, the augmented CAE utilizes stochastic regularization and differentiable feature selection to preserve only the most salient dimensions for intrusion detection. Furthermore, we propose a customized Residual Convolutional-Long Short-Term Memory with Attention (ResCLA) classification model that effectively captures both local and global dependencies in sequential traffic data through the integration of residual CLA blocks. This three-tier architecture establishes an end-to-end IDS specifically designed for the stringent security demands of 5G networks. It not only allows for flexible adaptation to variable traffic patterns but also achieves high precision in identifying ambiguous or adversarial attack vectors. Experimental results demonstrate that the proposed framework surpasses baseline models in both detection accuracy and computational efficiency across various 5G scenarios. The key innovations of this study are outlined as follows:

- We introduce ScalaDetect-5G, a novel system designed to effectively detect malicious traffic in 5G networks. This system operationalizes malicious traffic detection through a three-stage process: feature extraction, reconstruction, and detection for anomalous traffic in 5G networks.
- To tackle the scalability challenges associated with high-dimensional network features, we employ the enhanced CAE. This model dynamically reconstructs and compresses traffic characteristics, thus offering a highly elastic and efficient representation that is well-suited for large-scale 5G environments.
- Addressing the issue of low detection accuracy in ambiguous or borderline traffic scenarios, the ResCLA model is applied to extract and integrate multi-granular spatiotemporal features. This approach allows the system to identify subtle behavioral anomalies across network layers, thereby ensuring robust and precise anomaly detection in the complex and heterogeneous contexts of 5G networks.

2 Related Works

As indicated in Table 1, contemporary research in the domain of deep learning-based IDSs for 5G networks primarily concentrates on two pivotal areas: traffic characterization and traffic classification. Traffic

characterization entails both low-level and high-level feature extraction methods. Xiao et al. [12] introduced a Convolutional Neural Network (CNN)-based intrusion detection model by converting traffic data into images after dimensionality reduction, which helps in extracting discriminative features. Similarly, Kim et al. [13] focused on Denial of Service (DoS) attack detection by training CNNs on image-transformed traffic data derived from CSE-CIC-IDS2018 datasets. Expanding on these approaches, Kim and Pak [14] proposed an optimized method that converts traffic data into RGB image representations using multiple transformers, demonstrating superior detection performance over grayscale-based and non-image-based methods. These visual methods enable convolutional neural networks to directly learn spatial patterns from the traffic matrices. Conversely, high-level characterization is achieved through the extraction of protocol-specific and statistical features, utilizing sophisticated tools such as CICFLOWMETER, Zeek, Argus, and nProbe [15–17]. These tools provide enriched semantic representations that are particularly effective for behavior-based intrusion detection. In terms of traffic classification, the methodologies can broadly be grouped into traditional and hierarchical model-based approaches. Traditional classification generally employs deep learning models such as CNNs [18], LSTMs [19], and hybrid architectures like CNN+LSTM [20], which analyze either the temporal or spatial dependencies of network flows. More recently, hierarchical models such as LuNet [21] and Pelican [22] have been developed. These models exhibit a hierarchical structure, facilitating fine-grained classification across multiple traffic categories. This capability significantly enhances generalization and scalability in dynamic 5G environments. The progressive evolution of IDS methodologies underscores a trend toward more interpretable and adaptive systems, optimized for the challenges of next-generation network environments.

Table 1: Comparison of deep learning-based IDS methods for 5G traffic characterization and classification

Type	Method	Description	Problems
Characterization	Low-level feature	Converting normalized traffic (with fixed size) into grayscale or RGB images	Potential information loss or redundancy due to normalization of traffic
	High-level feature	Generating high-level features using tools such as Argus and nProbe	Poor generalizability of features due to the close association with specific traffic types
Classification	Traditional model	Employing deep learning models such as CNN, LSTM, and CNN+LSTM for detecting abnormal traffic	Low detection accuracy, insufficient for meeting the security demands of 5G networks
	Hierarchical model	Utilizing neural networks with hierarchical structures, such as LuNet and Pelican, for traffic anomaly detection	Ineffectiveness in identifying ambiguous traffic, leading to erroneous evaluations

2.1 Traffic Characterization

Low-level features of network traffic are extracted directly from raw byte streams and are considered shallow representations due to their limited semantic abstraction. This straightforward and computationally efficient approach is well-suited for large-scale preprocessing pipelines. In research conducted

by Xiao et al., network traffic data were first processed using dimensionality reduction techniques and then transformed into two-dimensional image matrices. These visual representations enabled the effective training of a convolutional neural network (CNN)-based intrusion detection system, thereby demonstrating the effectiveness of shallow feature representations in deep learning-based IDSs [12]. Similarly, Kim et al. utilized image transformation techniques on CSE-CIC-IDS2018 datasets, converting raw traffic into grayscale images. Despite achieving moderate detection performance, the study highlighted significant limitations of grayscale-based traffic encoding, particularly in capturing temporal or protocol-specific nuances [14]. To enhance representation fidelity, Husnain et al. transformed network flows from the CICDDoS2019 dataset into RGB images and employed a ResNet-18 model for intrusion detection [23]. Although RGB encoding better preserved structural and spatial features than grayscale, the approach still faced performance limitations due to the lossy nature of image transformations. Currently, the predominant method for generating low-level traffic features remains image-based conversion, either to grayscale or RGB. However, this approach necessitates truncation and scaling operations to fit network data into fixed-size image formats, potentially leading to the loss of essential packet-level or session-specific information. Such compromises can significantly undermine the accuracy of anomaly detection, especially in detecting fine-grained or stealthy attacks typical in 5G networks. Therefore, while low-level visual representations continue to be a focus of research, future efforts should aim to mitigate inherent information losses through enhanced encoding techniques or hybrid feature fusion frameworks.

The role of high-level features in network traffic analysis is fundamentally rooted in the rich domain knowledge and empirical insights that have been accumulated by cybersecurity experts over the years. These features play an indispensable role in enhancing the semantic understanding of complex traffic patterns. When compared to their low-level counterparts, high-level features demonstrate superior representational capabilities, particularly in distinguishing the behavioral characteristics of different types of cyber attacks. Several leading research entities, including the Canadian Institute for Cybersecurity, the University of New South Wales, the Cyber Range Lab, and Queensland University, have employed advanced tools such as CICFLOWMETER, Zeek, Argus, and nProbe to extract these high-level features from network traffic datasets that cover a broad spectrum of attack scenarios [15–17]. These tools facilitate the derivation of important metrics such as flow-level statistics, protocol flags, temporal metrics, and aggregated behavioral indicators, which are crucial for detailed intrusion detection. Building upon the aforementioned high-level feature representations, recent studies have increasingly emphasized the critical role of feature selection in optimizing detection performance while minimizing computational overhead. Notably, Sapna et al. [24] adopted a filter-based feature selection approach driven by the F1-score. The F1-score, serving as a statistically robust and unsupervised evaluation metric, ranks features based on their ability to maximize inter-class variance while minimizing intra-class variance. Despite the effectiveness of the high-level features developed in these studies within certain security contexts, they often exhibit a lack of adaptability and generalizability when applied across diverse network environments, including heterogeneous 5G infrastructures. This rigidity restricts their scalability and cross-domain applicability. To overcome this limitation, Balin et al. introduced a pioneering methodology employing a *CAE* to facilitate the flexible reconstruction of high-level features. This method enables the system to dynamically adjust feature representations in response to evolving traffic patterns [25]. The approach integrates an interpretable and differentiable selection mechanism that retains only the most informative features while discarding those that are redundant or specific to certain contexts. In the subsequent sections of this study, we will conduct a thorough analysis of the practical feasibility and implementation strategies for deploying the CAE in the reconstruction of 5G traffic representations. Additionally, we will discuss potential challenges and corresponding mitigation techniques to ensure robustness in real-world deployment scenarios.

2.2 Traffic Classification

CNNs and Recurrent Neural Networks (RNNs) represent two pivotal deep learning architectures that have shown substantial efficacy in identifying intrusive activities within network environments. CNNs utilize convolutional kernels to discern spatial correlations and localized patterns in network traffic, rendering them particularly suitable for processing structured traffic matrices or image-like representations. Conversely, RNNs are engineered to manage sequential data through the use of logic gates and memory cells—such as Long Short-Term Memory (LSTM) units or Gated Recurrent Units (GRUs)—to capture temporal dependencies and event sequences in network flows. The amalgamation of these architectures has given rise to hybrid models that are adept at concurrently learning spatial and temporal characteristics of traffic data. A notable example, the CNN-LSTM model, has been effectively deployed on the WSN-DS dataset, where it achieved high levels of intrusion detection accuracy by delineating the complex spatiotemporal dynamics within wireless sensor network traffic, as demonstrated by Halbouni [26]. Despite these promising outcomes, conventional models encounter several limitations in real-world 5G network scenarios, particularly when faced with encrypted traffic, evolving attack patterns, or imbalanced data distributions typically seen in operational environments. Furthermore, these traditional models frequently necessitate extensive manual feature tuning and exhibit limited robustness to adversarial perturbations. Hence, while CNN and RNN-based architectures establish a robust foundation for intrusion detection, there is an escalating necessity to devise more adaptable and resilient models that can sustain high detection precision in the intricate, high-throughput, and low-latency contexts of 5G networks.

In pursuit of enhancing the classification efficacy of IDSs, researchers are increasingly investigating neural network architectures with hierarchical structures, especially for managing complex and heterogeneous network traffic. Hierarchical models excel in learning multi-resolution representations by processing features at various levels of abstraction, thereby proficiently capturing intricate patterns within spatiotemporal data. Wu et al. introduced two hierarchical models, LuNet and Pelican, which were assessed using the UNSW-NB15 datasets. These models utilize deep hierarchical frameworks to extract multi-scale spatial and temporal features from network traffic, achieving notable predictive accuracies on the UNSW-NB15 dataset [21,22]. Although hierarchical models demonstrate superior performance in detecting well-defined intrusion behaviors, the accurate identification of ambiguous or borderline traffic continues to pose a significant challenge, particularly when data exhibits overlapping distributions or low inter-class separability. To mitigate this challenge, Guo et al. proposed an innovative approach by integrating an *external attention mechanism* that includes shared memory modules, implemented via linear transformation and normalization layers [27]. This architecture empowers the model to discern global contextual relationships and latent correlations between traffic instances, thereby augmenting its capability to distinguish complex intrusion behaviors. Experimental evaluations have affirmed the efficacy of this model in enhancing classification performance in challenging scenarios. Nevertheless, the generalizability and robustness of external attention mechanisms in real-world 5G environments—characterized by ultra-low latency, massive connectivity, and highly dynamic traffic patterns—remain to be thoroughly validated. The subsequent chapters of this study will therefore delve into a comprehensive exploration of the applicability, advantages, and limitations of external attention frameworks within the context of 5G intrusion detection, alongside proposed strategies to bolster their deployment feasibility in operational settings. Meanwhile, Ali et al. [28] proposed a Hybrid Deep Reinforcement Learning-based IDS (HDRL-IDS) using an actor–critic architecture to enable autonomous detection of evolving threats. By combining network- and host-level feature analysis, their system integrates the strengths of Network-based Intrusion Detection System (NIDS) and Host-based Intrusion Detection System (HIDS). While this work is significant for its focus on decision-making under uncertainty, it primarily emphasizes reinforcement learning and policy optimization, whereas the present study focuses on enhancing

the representational and discriminative capacity of hierarchical deep learning models, particularly within 5G environments.

3 Proposed Method

Within the 5G access network, we have implemented a rigorous security framework by deploying the ScalaDetect-5G IDS, which plays an integral role in recognizing and mitigating potential cybersecurity threats. This system continuously monitors high-volume, heterogeneous 5G traffic in real time, utilizing advanced analytics to detect anomalies indicative of malicious activity. Specifically, the ScalaDetect-5G IDS is tailored to address the dynamic and decentralized characteristics of 5G environments, thereby ensuring elevated detection accuracy. The detection process begins with the *feature extraction module*, which extracts key features from raw traffic using statistical and deep learning methods. These features are subsequently processed by the *feature reconstruction module*, which dynamically extracts key features informed by the preceding module's extractions. The final stage, the *intrusion detection module*, employs machine learning classifiers to assess the likelihood of intrusions based on the detected deviations. This tripartite architecture not only augments detection precision but also supports scalability and real-time responsiveness, rendering it highly suitable for intricate 5G scenarios. The overall structure and data flow of the ScalaDetect-5G IDS are delineated in Fig. 1.

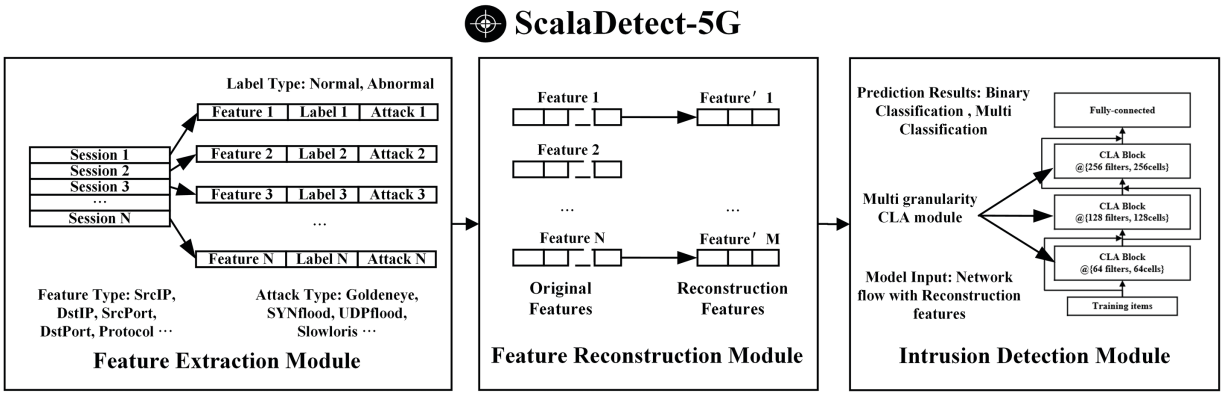


Figure 1: Architecture and data flow of the ScalaDetect-5G IDS

Initially, the *feature extraction module* collects and aggregates raw network traffic data based on five-tuple information (i.e., source IP, destination IP, source port, destination port, and protocol type), which serves as a fundamental identifier for individual network flows. This module then implements a comprehensive data preprocessing pipeline that encompasses normalization, noise filtering, and temporal segmentation to enhance feature clarity. High-level statistical and temporal features are subsequently gleaned using methodologies such as entropy calculation, packet inter-arrival time analysis, and flow duration statistics [17]. Following this, the *feature reconstruction module* utilizes an advanced CAE architecture, which capitalizes on the feature correlation matrix derived from the Concrete distribution to discern latent dependencies among features. This module performs elastic and unsupervised reconstruction of traffic patterns, thereby adeptly capturing nuanced deviations from normative behavior, even in scenarios involving zero-day attacks. Finally, the *intrusion detection module* establishes the CLA block, integrating both temporal and spatial dependencies within the traffic data. This underpins the ResCLA model, a deep learning-based detection framework that amalgamates residual learning with attention mechanisms to fortify detection robustness and interpretability. In subsequent sections, we will provide a detailed

exposition of the underlying *traffic characterization* and *traffic classification* techniques employed by the ScalaDetect-5G system.

3.1 Traffic Feature Extraction Using CICFLOWMETER

The tool `CICFLOWMETER` [29] is employed within the *feature extraction module* to derive a comprehensive set of 82 statistical and flow-based features from raw 5G network traffic. These features encompass a broad spectrum of traffic attributes, including but not limited to packet size distributions, inter-arrival times, bidirectional byte counts, average flow durations, header-based flags, and flow-based entropy metrics. This high-dimensional feature set provides a robust and informative initial representation of 5G traffic behavior, facilitating various analytical tasks such as flow classification, anomaly detection, and behavioral profiling.

Nevertheless, the inherent heterogeneity, rapid evolution, and encryption characteristics of contemporary 5G traffic pose significant challenges. Predefined handcrafted features often exhibit considerable redundancy and may lack sufficient discriminative power in the face of stealthy attack scenarios, polymorphic malware, or obfuscated benign traffic patterns. These challenges highlight the necessity for adaptive and dynamic feature extraction methodologies that can effectively respond to the complex and evolving nature of network traffic.

The 82 network traffic features generated by CICFlowMeter are meticulously categorized into eight distinct groups, encompassing a wide array of essential traffic characteristics. These categories include basic flow information, time-related features, packet counts, packet sizes, traffic rates, TCP flag bits, active and idle features, and advanced statistical metrics. As shown in Table 2, each category is accompanied by a comprehensive description and representative examples. This classification not only aids in a deeper understanding of the functional roles of these features but also underscores their significance in the broader context of network traffic analysis.

Table 2: Classification of CICFlowMeter network traffic features

Category	Description
Basic flow information	Attributes that identify the flow.
Time-related features	Statistics on packet inter-arrival times and flow duration.
Packet count features	Counts of packets in forward and backward directions.
Packet size features	Statistics on packet lengths in both directions.
Flow rate features	Measures of byte and packet transmission rates.
TCP flag features	Counts of TCP flags used in the flow.
Active and idle features	Statistics on active and idle time periods in the flow.
Advanced statistical features	Subflow information and average segment sizes.

These features prove instrumental in a variety of applications, particularly in the fields of network intrusion detection and anomaly detection. They play a crucial role in identifying subtle deviations in network behavior that may signal malicious activities or system vulnerabilities. The classification schema outlined in the table further enables researchers and network security professionals to gain a thorough understanding of how these features can enhance the security posture of modern networks. In the context of emerging technologies such as 5G, leveraging these features allows security practitioners to develop more sophisticated and effective detection systems capable of contending with increasingly complex cyber threats in dynamic network environments.

3.2 Traffic Characterization Technique Based on the Concrete Autoencoder

To enhance the extraction of pivotal features from high-dimensional network traffic data, we have developed a *feature reconstruction module* that operates subsequent to the initial feature extraction phase. This module is meticulously designed to compress, reorganize, and enrich the original feature space by leveraging a synergy of learned abstraction and data-driven selection mechanisms. Such architectural refinement significantly elevates the generalizability, robustness, and interpretability of downstream classification tasks. At the core of this module lies the *Concrete selector layer*, which exploits the Concrete distribution (also known as the Gumbel-Softmax distribution) to perform soft and differentiable feature selection. In contrast to conventional binary masking methods, this approach introduces a stochastic sampling process over a continuous relaxation of the categorical distribution, thereby enabling end-to-end gradient-based optimization during training while approximating discrete selection behavior.

Building upon this foundation, we introduce an enhanced *Concrete Autoencoder* (CAE), wherein the selector layer is seamlessly integrated into a symmetric encoder-decoder architecture. As shown in Fig. 2, the encoder incorporates a series of learnable Concrete selector nodes that compress the high-dimensional input vector into a latent representation by selectively engaging the most informative features. The decoder then endeavors to reconstruct the original input vector from this compressed embedding, thereby implicitly guiding the selector nodes to identify and preserve salient and non-redundant features. To further elevate the discriminative power and representation diversity of the selected features, we incorporate a correlation-aware regularization strategy during training. This regularization term penalizes the selection of highly correlated features across different selector nodes, thereby promoting orthogonality and encouraging each node to capture distinct aspects of the input, which ultimately augments the expressive capacity of the latent space.

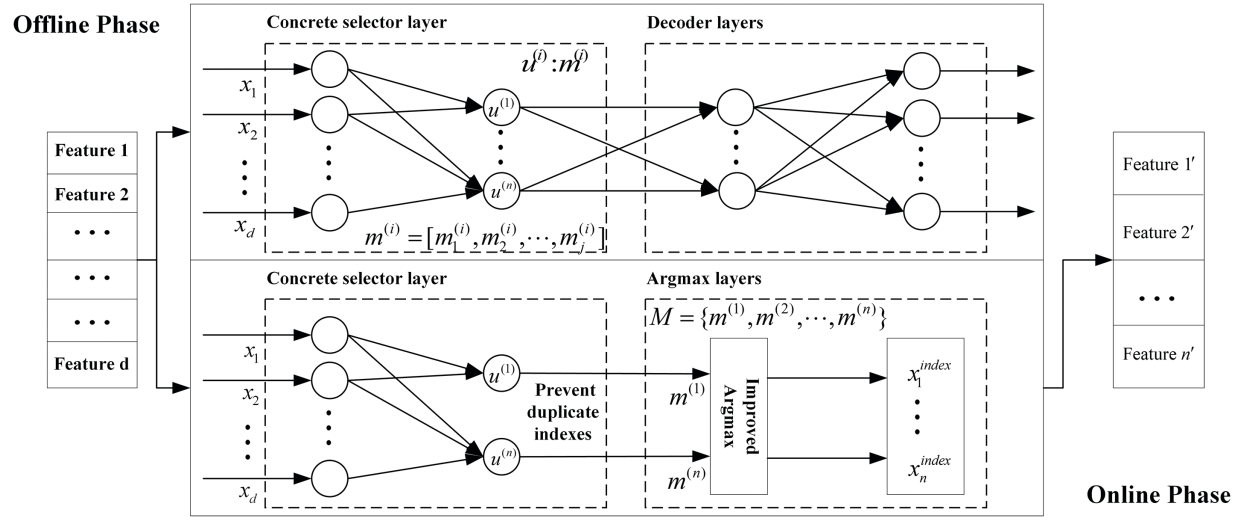


Figure 2: Schematic of training and inference phases in the concrete autoencoder

The Concrete distribution [30] enables a continuous relaxation of categorical random variables, thereby facilitating differentiable sampling within stochastic neural architectures. It defines a probabilistic sample vector $m \in \mathbb{R}^d$ as:

$$m_j = \frac{\exp\left((\log \alpha_j + g_j)/T\right)}{\sum_{k=1}^{d} \exp\left((\log \alpha_k + g_k)/T\right)} \tag{1}$$

In this equation, m_j denotes the probability assigned to feature j, where α_j is its importance score (logit), and g_j is a random sample drawn from the standard Gumbel distribution. The temperature parameter T controls how smooth or sharp the output distribution is. As T decreases, the distribution becomes more discrete, approaching a one-hot vector that emphasizes a single feature. This reparameterized formulation ensures the sampling process remains differentiable, allowing the feature selector to be trained efficiently with backpropagation. By encouraging sparsity in the output vector, this mechanism enables the model to automatically select the most informative features while suppressing redundant ones.

In the proposed architecture, each selector node $u^{(i)}$ generates a d-dimensional Concrete vector $m^{(i)}$ and performs a linear projection $x \cdot m^{(i)}$, where x denotes the input feature vector. As T decreases, each $m^{(i)}$ increasingly resembles a one-hot vector, effectively driving the node to focus on a single feature with high probability. During inference, the soft probabilistic sampling of the Concrete vector is replaced by a deterministic selection mechanism using the arg max operation, ensuring that each selector outputs the feature with the highest selection confidence. Formally, the output of each selector node is given by $x_{\text{index}(i)}$, where index(i) denotes the position of the feature with the maximum probability in $m^{(i)}$. Fundamentally, each Concrete variable is designed to isolate a single, salient feature from the input space, thereby promoting interpretability and sparsity in the reconstructed representation.

In this reconstruction framework, each selector independently learns to represent and reconstruct distinct semantic components of the input feature space, allowing for highly parallelized and scalable learning. However, such independence introduces the risk of *feature collisions*, wherein multiple selector nodes converge upon the same input dimension, thereby leading to redundant selections. This is especially problematic in the context of 5G traffic analysis, where vital features are often sparse, temporally entangled, or obscured by noise. To mitigate this redundancy, we introduce an improved CAE algorithm as detailed in Algorithm 1. This algorithm initializes a selector weight matrix and iteratively performs a greedy arg max operation to extract the most salient feature per selector node. After each extraction, the corresponding feature is masked out from the remaining selectors to prevent overlap, thereby ensuring that each selected feature is unique, highly relevant, and non-redundant across all selector nodes.

Algorithm 1: Improved CAE

Input: Feature weight matrix $M = \{m^{(1)}, m^{(2)}, \ldots, m^{(n)}\}$;
k is the number of feature selectors; $m^{(i)}$ is a d-dimensional column vector.
for $i \in \{1 \ldots \text{num\_feature}\}$ **do**
$row = M[i, :]$
$index(i) = \arg\max(row)$
$M[:, index(i)] = 0$ **end for**
Output: Non-redundant feature index set *index*

In summary, our enhanced CAE architecture transforms an initial d-dimensional statistical feature vector into a compact n-dimensional representation via n selector nodes, each deterministically selecting the most informative feature. This reduced vector maintains high informational density and reflects a refined integration of expert-curated statistical descriptors with data-adaptive deep learning abstractions. While the choice of n directly affects reconstruction quality and classification performance, its optimal value is explored and validated through subsequent experiments. Additionally, the temperature parameter$T$ in the Concrete distribution plays a critical role in balancing soft exploration and hard selection; lower values of T (e.g., $T < 0.2$) yield sharper selections but may hinder gradient smoothness during training. With appropriate hyperparameter tuning, the model achieves effective dimensionality reduction, suppresses overfitting, and

enhances generalization across evolving and heterogeneous 5G traffic patterns—thereby establishing a strong foundation for downstream intrusion detection tasks.

3.3 Traffic Classification Technique Based on the ResCLA Model

Hierarchical models generally exhibit superior performance over traditional flat architectures in terms of detection accuracy, particularly within environments characterized by high-dimensional, non-stationary data such as 5G networks. However, the effective classification of ambiguous, encrypted, or deliberately obfuscated traffic patterns remains a persistent and unresolved challenge. These traffic flows frequently display intricate spatial-temporal correlations, exhibit dynamically evolving behaviors, and are susceptible to adversarial manipulations, which collectively undermine the robustness and generalization capabilities of conventional deep learning models. To address these deficiencies, we propose a novel hybrid classification framework, termed the ResCLA model. This architecture enhances the representational power of deep neural networks through the integration of external attention mechanisms, hierarchical fusion of heterogeneous modules, and residual shortcut connections, enabling the capture of fine-grained dependencies across multiple abstraction levels. The ResCLA model is meticulously designed to operate under the stringent demands of real-world network conditions, where rapid traffic fluctuations and adversarial concealment techniques are increasingly prevalent.

The proposed ResCLA model takes as input the traffic features reconstructed by the feature characterization module and outputs a predicted traffic class label (e.g., *Normal*, *Goldeneye*, *ICMP-flood*). It is constructed in three principal stages. First, we design modular units—referred to as CLA blocks—that integrate CNNs, LSTM networks, and External Attention layers. Each module within a CLA block is specialized to extract a specific aspect of network traffic features, including spatial, temporal, and contextual information. Next, multiple CLA blocks are hierarchically stacked to form a deep architecture that enables the model to encode both low-level signal characteristics and high-level behavioral semantics. To facilitate effective gradient propagation and mitigate issues such as vanishing gradients, we incorporate non-linear residual (shortcut) connections across non-contiguous layers. This architecture allows ResCLA to learn robust and expressive representations from noisy or obfuscated traffic data, while maintaining stable convergence and strong generalization to previously unseen traffic types. The overall structure is illustrated in Fig. 3.

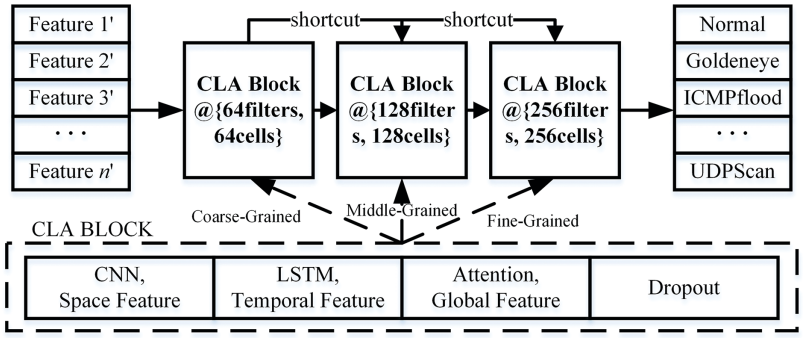


Figure 3: Schematic diagram of the proposed ResCLA model architecture

CNNs: CNNs are deep feedforward models that utilize convolutional filters and pooling to extract multi-scale spatial features. Leveraging local connectivity, weight sharing, and translation invariance, CNNs are well-suited for modeling fine-grained patterns in network traffic. In our design, each CLA block integrates one-dimensional convolutions tailored for 5G traffic sequences, effectively detecting localized

anomalies such as bursts, jitter, or latency spikes. Stacked layers expand the receptive field, enabling the capture of compound patterns over longer time windows, which may indicate multi-stage attacks or protocol-specific behaviors.

LSTMs: LSTMs, a variant of RNNs, use input, forget, and output gates to capture long-range temporal dependencies while addressing the vanishing gradient problem. In the ResCLA model, each CLA block incorporates LSTM modules to learn sequential patterns in traffic data, such as periodic behaviors, session dependencies, or stealthy low-rate attacks. This temporal modeling is essential for distinguishing flows with similar short-term characteristics but divergent long-term behaviors, enhancing the model's sensitivity to complex time-based threats like slow DoS attacks and distributed probes.

External Attention Mechanism [27]: Traditional self-attention mechanisms compute attention scores by projecting input features into query (Q), key (K), and value (V) matrices, with the output given by:

$$A = (\alpha)_{i,j} = \text{softmax}(QK^{\top}) \tag{2}$$

$$F_{\text{out}} = AV \tag{3}$$

where $F \in \mathbb{R}^{N\times d}$ denotes the input feature matrix, $A \in \mathbb{R}^{N\times N}$ represents the pairwise attention matrix, and $F_{\text{out}} \in \mathbb{R}^{N\times d}$ is the output feature representation. Although effective for capturing intra-sample dependencies, this approach incurs quadratic complexity with respect to sequence length N, and may struggle with capturing inter-sample dependencies across large-scale traffic datasets. To alleviate these limitations, we incorporate an *external attention mechanism* that decouples attention score computation from input-specific key-value generation by utilizing global memory units. The modified attention mechanism is defined as:

$$A = \text{Norm}(FM_k^{\top}) \tag{4}$$

$$F_{\text{out}} = AM_v \tag{5}$$

where $M_k, M_v \in \mathbb{R}^{S\times d}$ are learnable global memory matrices that store key and value embeddings, respectively, and $A \in \mathbb{R}^{N\times S}$ represents the normalized attention scores. This configuration enables the model to incorporate global contextual information learned from the entire training set, thus overcoming the locality constraints of self-attention. In the context of 5G traffic analysis—where correlations among distant flows may reveal the presence of distributed or temporally disjoint intrusions—external attention significantly improves the model's capacity to capture broad relational patterns, thereby enhancing both classification accuracy and operational scalability.

ResCLA Model Summary: The ResCLA model begins with input vectors reconstructed from the CAE, as outlined in Section 3.2. These feature vectors are first processed by convolutional layers that extract spatial attributes, followed by LSTM layers that capture temporal dependencies, and finally by external attention layers that uncover global relational patterns across flows. The hierarchical stacking of CLA blocks at three distinct abstraction levels—namely, CLA Block@64 filters, 64 cells (coarse-grained), CLA Block@128 filters, 128 cells (middle-grained), and CLA Block@256 filters, 256 cells (fine-grained)—facilitates progressive, multi-resolution feature abstraction. The coarse-grained CLA block focuses on capturing broad spatial distributions and flow-level statistical patterns, serving as an initial encoder of low-level traffic signatures. The middle-grained block further refines temporal dependencies and transitional dynamics that span over moderate-length sequences, enabling the model to detect session-level behaviors and sustained attack patterns. The fine-grained block emphasizes complex inter-flow correlations and high-resolution feature interactions, extracting subtle temporal anomalies and semantic variations that may indicate stealthy or adversarial activities. Embedded residual connections between these non-contiguous CLA blocks ensure the preservation and fusion of both shallow and deep representations, promoting feature reuse, enhancing

gradient flow, and mitigating degradation in deep architectures. This synergy between hierarchical depth and modular granularity not only stabilizes training but also improves interpretability and transparency by maintaining the traceability of discriminative features throughout the network. Furthermore, the granularity-aware modular design of ResCLA enables targeted attribution analysis at each abstraction level, thereby supporting explainable AI (XAI) practices in security-critical 5G environments.

In conclusion, the ResCLA model embodies a robust, scalable, and interpretable hybrid architecture that synthesizes spatial, temporal, and contextual representations for high-precision traffic classification. Through its multi-path design enriched with attention-driven global memory and deep residual linkages, the model demonstrates strong capability in recognizing ambiguous, evolving, and previously unseen traffic patterns. These qualities render it particularly suitable for real-time intrusion detection and anomaly identification in complex, high-throughput 5G network environments. Furthermore, the model's effectiveness has been validated under challenging conditions, including imbalanced class distributions and adversarial traffic scenarios. By leveraging attention mechanisms and deep residual encoding, ResCLA demonstrates resilience against class skew and maintains high detection fidelity even in the presence of obfuscated or adversarially crafted inputs. In addition, its lightweight inference modules ensure fast processing throughput, confirming its practicality for real-time deployment in mission-critical, latency-sensitive 5G systems. As further evidenced by the ablation study results, only the synergistic integration of spatial encoding, temporal modeling, and contextual attention mechanisms enables robust and accurate intrusion detection in diverse and complex 5G network traffic environments.

4 Evaluation and Discussion

This section presents a comprehensive evaluation of the intrusion detection system proposed in this study, denoted as ScalaDetect-5G. The experiments were carried out using both publicly accessible and proprietary datasets that are pertinent to 5G and associated network environments. These datasets include a diverse array of traffic types and attack vectors, such as Distributed Denial of Service (DDoS), DoS, spoofing attacks, and signalling-based anomalies, which are prevalent in 5G network architectures. The experimental setup was conducted on a high-performance computing platform equipped with the following specifications: a 64-bit Windows 10 operating system, an Intel Core i7-10900 processor featuring 10 cores and 20 threads with a base frequency of 2.8 GHz, and an Nvidia GTX1080 GPU with 8 GB of video RAM. This configuration provided the necessary computational power to handle the parallel processing demands of deep learning algorithms efficiently.

In addition to evaluating detection performance, we also assessed the computational efficiency and deployment feasibility of ScalaDetect-5G from a system resource perspective. Despite its deep, granularity-aware architecture featuring hierarchically stacked CLA blocks—each composed of convolutional, recurrent, and external attention modules—the final trained model is highly compact, with a disk size of approximately 12 MB. This lightweight design enables seamless deployment on edge device, where storage and computational resources are constrained. During inference, the model maintains a modest memory footprint of less than 1.5 GB, making it suitable for real-time traffic analysis even under resource-limited conditions. The complete training pipeline, which includes both feature reconstruction via CAE and classification via ResCLA, completes within approximately 2.3 h over 100 epochs on the specified hardware configuration. These results highlight the favorable trade-off achieved by ScalaDetect-5G between representational expressiveness and computational efficiency, thereby supporting its scalability and practicality in high-throughput, latency-sensitive 5G intrusion detection deployments.

During the implementation phase, cutting-edge technologies were utilized to optimize the training efficiency and stability of convergence of the models. Notably, the `PyTorch Lightning` framework was

implemented to facilitate the development and modularization of deep learning models. Furthermore, the Adaptive Moment Estimation optimizer was employed to effectively minimize the training loss function, expressed as:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \ell(f(x_i;\theta), y_i) \tag{6}$$

In this equation, θ represents the model parameters, while x_i and y_i denote the input-output pairs. The function $\ell(\cdot)$ refers to the loss function utilized, such as cross-entropy for classification tasks. The integration of these advanced techniques has enabled rapid prototyping and precise fine-tuning of the neural network architectures, which are specifically designed to address the unique challenges posed by 5G traffic patterns. To further improve generalization and mitigate overfitting, regularization strategies such as dropout, weight decay, and early stopping were also adopted. These techniques help constrain model complexity and ensure stable performance when exposed to noisy or previously unseen traffic data.

4.1 Datasets and Metrics

Owing to their inherent architectural complexity, heterogeneous service requirements, and dynamic topology, 5G networks are markedly more vulnerable to sophisticated multi-vector security threats than their legacy mobile network predecessors. These threats range from protocol-specific vulnerabilities to AI-driven evasion techniques. In response to this increased security risk, we conducted extensive experiments on a novel IDS specifically designed for 5G environments. We assessed its performance using over 2 million raw network traffic records from the 5G-NIDD dataset [31], which includes a diverse mix of signaling and user-plane data representative of realistic 5G use cases. Given the potential for adversaries to exploit vulnerabilities in legacy or transitional infrastructures to compromise 5G domains, it is crucial to develop IDS solutions that can generalize from traditional attack patterns. In this vein, we also analyzed over 16 million raw traffic instances from the CIC-IDS2017 dataset [17], which covers a broad spectrum of conventional network intrusions such as brute-force attacks, botnets, and port scanning.

Moreover, while 5G networks play a pivotal role in facilitating low-latency, high-density IoT connectivity, they also inherently inherit IoT-related security risks, which may propagate within the 5G core. To evaluate this aspect, we examined over 5 million and 7 million raw traffic samples from the ToN-IoT [15] and BoT-IoT [16] datasets, respectively. These datasets enabled us to assess the IDS's capability to detect and classify IoT-induced anomalies within a 5G-integrated context. All traffic data were preprocessed by aggregating raw packets into communication sessions and extracting high-level statistical and protocol-specific features. Specifically, each session was characterized by 82 fine-grained features, systematically categorized into seven principal groups: *Network Identifiers*, *Flow Descriptors*, *Interval Times*, *Flag Features*, *Sub-flow Descriptors*, *Header Descriptors*, and *Flow Time Metrics*. This categorization facilitates comprehensive profiling of traffic behavior suitable for both supervised and unsupervised learning models. This multi-perspective feature engineering process significantly enhances the system's capability to discern benign from malicious behaviors under the complex operational conditions characteristic of 5G ecosystems. To ensure robust evaluation, 80% of the data was allocated to the training set and the remaining 20% to the testing set. Additionally, 5-fold cross-validation was conducted to assess generalizability and reduce the risk of overfitting. Nonetheless, it is important to acknowledge the inherent limitations of the employed datasets. Both ToN-IoT and BoT-IoT are partially synthetic and may not fully reflect the diversity and unpredictability of real-world 5G deployments, especially in the presence of encrypted communications or multi-protocol traffic. As such, future work should consider incorporating more representative, real-world 5G traffic data to further evaluate

the generalizability and robustness of the proposed detection framework across heterogeneous and evolving network environments.

To conduct a thorough and rigorous performance assessment of the proposed intrusion detection framework ScalaDetect-5G, we evaluated the classification efficacy of its central security perception module, the ResCLA model, employing four widely recognized evaluation metrics: accuracy, precision, recall, and F1-score. These metrics provide complementary insights into the model's behavior, particularly within the context of complex and imbalanced network traffic distributions characteristic of realistic 5G scenarios. Specifically, **accuracy** measures the overall ratio of correctly classified instances and serves as a straightforward gauge of the model's overall performance across diverse categories. **Precision** determines the fraction of correctly identified positive instances among all instances predicted as positive, thus gauging the model's efficacy in reducing false positives. Conversely, **recall** quantifies the model's capability to detect all actual positive instances, underscoring its thoroughness in intrusion detection. The **F1-score**, representing the harmonic mean of precision and recall, offers a balanced measure that proves particularly valuable in scenarios with imbalanced datasets, where the predominance of benign samples markedly surpasses that of malicious ones.

The mathematical definitions of these evaluation metrics are presented below:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

where *TP*, *TN*, *FP*, and *FN* represent the number of true positives, true negatives, false positives, and false negatives, respectively. Collectively, these metrics establish a comprehensive and balanced framework for evaluating the ResCLA model's capability to differentiate between benign and malicious network traffic within a 5G security context.

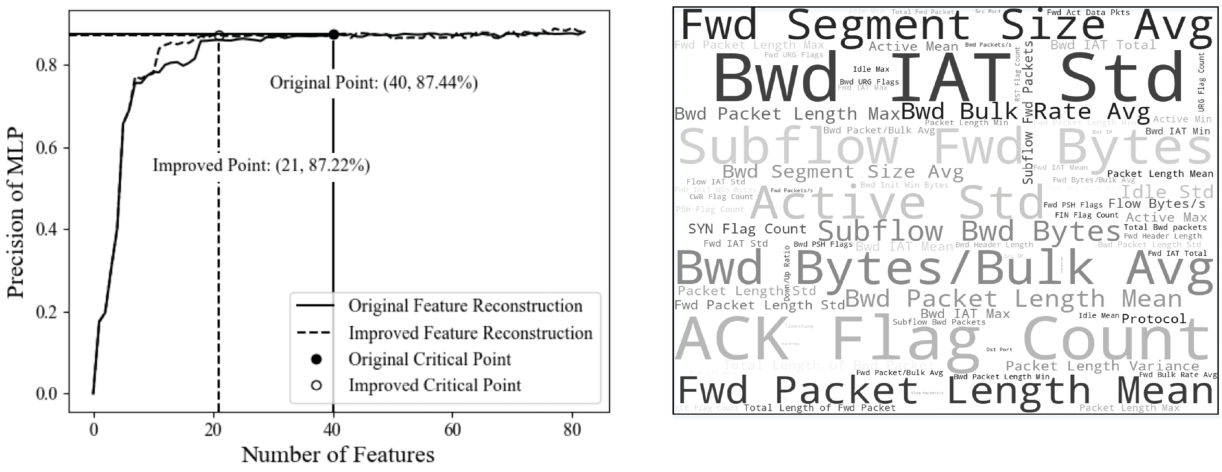
4.2 Experimental Results and Discussion

A rigorous series of feature reconstruction experiments was conducted across multiple benchmark datasets to assess the efficacy of the proposed ScalaDetect-5G intrusion detection framework. Specifically, traffic-level features and multi-class labels were extracted from raw packet-level data within four representative datasets: 5G-NIDD, CIC-IDS2017, ToN-IoT, and BoT-IoT. These datasets collectively cover a broad spectrum of attack types and network behaviors pertinent to 5G and IoT environments. For feature selection and dimensionality reduction, both the original CAE and its enhanced version were employed. The enhanced CAE was specifically developed to more effectively capture non-linear dependencies and attenuate noise in high-dimensional feature spaces. These methods facilitated the generation of two reconstructed datasets: the *original reconstructed dataset* and the *enhanced reconstructed dataset*.

To evaluate the discriminative capability of the selected features, the intrusion detection performance was analyzed across various feature subsets derived from the reconstructed datasets. The number of features in these subsets, denoted as `num_feature`, ranged from 1 to 82. Each subset comprised the top-k features, which were ranked according to their learned importance weights from the reconstruction model. Subsequently, each subset was utilized to train a multi-layer perceptron (MLP)-based classifier. The detection accuracy was then measured to compare the representational quality of the features selected by

the two autoencoders. It is worth noting that this experiment primarily focuses on evaluating the feature reconstruction module of ScalaDetect-5G, and thus employs MLP—a widely used neural network—as the baseline classifier for traffic classification. The detection accuracy was then measured to compare the representational quality of the features selected by the two autoencoders.

As depicted in Fig. 4a, both reconstruction approaches exhibit a similar trend in classification performance: there is a rapid increase in accuracy up to a critical point, followed by fluctuations or a plateau. These critical points, marked by solid black dots, are reached at `num_feature` = 21 for the original CAE and `num_feature` = 40 for the enhanced CAE, achieving peak detection accuracies of 88.22% and 88.44%, respectively. Fig. 4b visualizes the reconstructed feature space using a word cloud, where font size indicates each feature's relative importance as determined by the CAE. Prominent features such as `ACK Flag Count`, `Bwd IAT Std`, and `Fwd Segment Size Avg` highlight the importance of time- and flag-based descriptors in capturing traffic burstiness, flow control behavior, and inter-packet timing—key indicators of 5G network anomalies. Additional high-impact features, including `Fwd Packet Length Mean`, `Bwd Bytes/Bulk Avg`, and `Idle Std`, further reveal insights into packet size variability, directional byte flow, and inactivity patterns. The presence of both direction-specific and statistical features suggests that effective intrusion detection in 5G requires modeling not only instantaneous traffic attributes but also temporal dependencies and session-level dynamics. These findings confirm that both CAE variants reduce dimensionality effectively by discarding noisy or redundant features, thereby improving convergence and accuracy. Notably, the enhanced CAE yields superior performance by extracting a broader range of semantically meaningful traffic characteristics, particularly those related to subtle or low-rate attack patterns, underscoring its value in 5G intrusion detection.



(a) (b)

Figure 4: (**a**) Intrusion detection precision of the 5G-NIDD dataset (left), and (**b**) feature cloud of the 5G-NIDD dataset (right)

Table 3 summarizes the outcomes of the feature reconstruction experiments conducted on four representative intrusion detection datasets: 5G-NIDD, CIC-IDS, ToN-IoT, and BoT-IoT, utilizing the enhanced CAE. This table details the number of features at the critical performance points, as well as variations in detection precision and processing time before and after feature reconstruction. Several key observations emerge from these results: 1. The critical features identified by the enhanced CAE demonstrate dataset-specific patterns. For example, the 5G-NIDD dataset includes 45 significant features such as `ACK Flag Count` and

`Subflow Fwd Bytes`; CIC-IDS highlights 54 important features including `Fwd Pkt Len Mean` and `Fwd Act Data Pkts`; ToN-IoT features 58 key attributes like `Subflow Bwd Byts` and `Subflow Fwd Byts`; BoT-IoT contains 52 prominent features, with `Bwd Seg Size Avg` and `Bwd IAT Std` being notable examples. 2. In terms of accuracy improvement, post-reconstruction detection precision increased by 0.77%, 0.58%, 0.85%, and 0.88% for 5G-NIDD, CIC-IDS, ToN-IoT, and BoT-IoT, respectively. These improvements reflect the CAE's ability to retain essential discriminative information while eliminating redundant dimensions. 3. Regarding time efficiency, the application of reconstructed features resulted in consistent reductions in detection latency, with average time savings of 3.57, 4.38, 4.59, and 4.55 s for the respective datasets. This enhancement in performance demonstrates the CAE's advantage in computational efficiency without compromising detection fidelity. Collectively, these experimental findings substantiate the enhanced CAE as an effective and scalable feature reconstruction mechanism for heterogeneous network traffic, facilitating adaptive and efficient intrusion detection across various 5G-related data environments.

Table 3: Intrusion detection performance of the improved CAE

Datasets	Number of critical point features	Precision variation	Time variation
5G-NIDD	21 (ACK Flag Count, . . .)	−0.22%	−3.57 s
CIC-IDS	25 (Fwd Pkt Len Mean, . . .)	+0.38%	−4.38 s
ToN-IoT	28 (Subflow Bwd Byts, . . .)	−0.25%	−4.59 s
BoT-IoT	24 (Bwd Seg Size Avg, . . .)	+0.32%	−4.55 s

To rigorously assess the intrusion detection effectiveness of the proposed framework ScalaDetect-5G, we conducted a comprehensive set of experiments across four widely adopted benchmark datasets: 5G-NIDD, CIC-IDS, ToN-IoT, and BoT-IoT. The initial phase of our evaluation involved the application of the enhanced CAE to each dataset. This process facilitated the extraction of critical features, which were used to construct optimized reconstructed feature subsets. These subsets were designed to retain the most informative dimensions while minimizing redundancy. Subsequently, we deployed both the baseline intrusion detection models *LuNet Pelican* and the proposed deep attention-based model ResCLA to perform classification tasks on each reconstructed dataset. To ensure a robust comparison, we evaluated the models using multiple performance metrics, including the confusion matrix, F1 score, false positive rate (FPR), and area under the ROC curve (AUC). While F1 score provides a balanced assessment of precision and recall, FPR highlights the models' susceptibility to false alarms, which is critical in intrusion detection scenarios. AUC, on the other hand, offers a holistic view of classification robustness across varying threshold settings.

Table 4 presents the comparative false positive rates (FPRs) of three representative intrusion detection models—LuNet, Pelican, and ScalaDetect-5G—across four widely used benchmark datasets: 5G-NIDD, CIC-IDS, ToN-IoT, and BoT-IoT. The results indicate that ScalaDetect-5G consistently achieves the lowest FPR among all models across every dataset. For instance, on the 5G-NIDD dataset, ScalaDetect-5G records an exceptionally low FPR of 0.35%, compared to 1.82% for LuNet and 1.17% for Pelican. Similar trends are observed in the CIC-IDS and BoT-IoT datasets, where ScalaDetect-5G maintains FPRs below 0.4%. Even in the more challenging ToN-IoT dataset, which features a diverse range of attack scenarios, ScalaDetect-5G maintains a favorable FPR of just 0.41%. These results suggest that ScalaDetect-5G is highly effective at reducing false alarms, a critical requirement for practical deployment in real-world intrusion detection systems.

Table 4: Comparison of FPR and AUC for different models evaluated on multiple benchmark datasets

Dataset	FPR			AUC		
	LuNet	Pelican	ScalaDetect-5G	LuNet	Pelican	ScalaDetect-5G
5G-NIDD	1.82%	1.17%	0.35%	0.975	0.983	0.997
CIC-IDS	1.79%	1.03%	0.38%	0.972	0.985	0.995
ToN-IoT	1.72%	1.15%	0.41%	0.979	0.982	0.991
BoT-IoT	1.82%	1.17%	0.35%	0.976	0.983	0.993

In addition to minimizing FPR, ScalaDetect-5G also demonstrates superior classification performance in terms of the area under the ROC curve (AUC). As shown in Table 4, ScalaDetect-5G consistently achieves the highest AUC values across all datasets. On the 5G-NIDD dataset, it achieves a near-optimal AUC of 0.997, outperforming both LuNet (0.975) and Pelican (0.983). This trend persists across the CIC-IDS and BoT-IoT datasets, where ScalaDetect-5G reaches AUC values above 0.99. Even under the diverse attack conditions in ToN-IoT, the model maintains a high AUC of 0.991. These high AUC scores indicate that ScalaDetect-5G possesses strong discriminatory capability across varying decision thresholds, further validating its robustness and generalizability in detecting a wide range of network intrusions.

As depicted in Fig. 5, the confusion matrices offer a visual representation of the classification outcomes for various traffic categories within the 5G-NIDD dataset using both the Pelican and ResCLA models. Two significant observations can be noted: (1) The Pelican model achieves high F1 scores for well-defined and frequent traffic categories such as *Goldeneye*, *ICMP-flood*, and *Normal* traffic; however, it underperforms in more ambiguous and less frequent classes, such as *SYNScan* and *UDPScan*. This underperformance indicates a limitation in generalizing to subtle intrusion patterns. (2) Conversely, the ResCLA model consistently secures high F1 scores across all traffic categories, an achievement attributable to its external attention mechanism and enhanced feature interaction. This capability allows for more accurate modeling of complex patterns in 5G traffic. These findings underscore that although the Pelican model is effective for conventional threat detection, it encounters difficulties with edge-case anomalies. In contrast, ResCLA demonstrates robust performance in managing both common and ambiguous network traffic within the high-dimensional environments of 5G networks. However, it is important to note that the current experiments did not include more challenging adversarial traffic samples, the detection of which will be a key focus of future work.

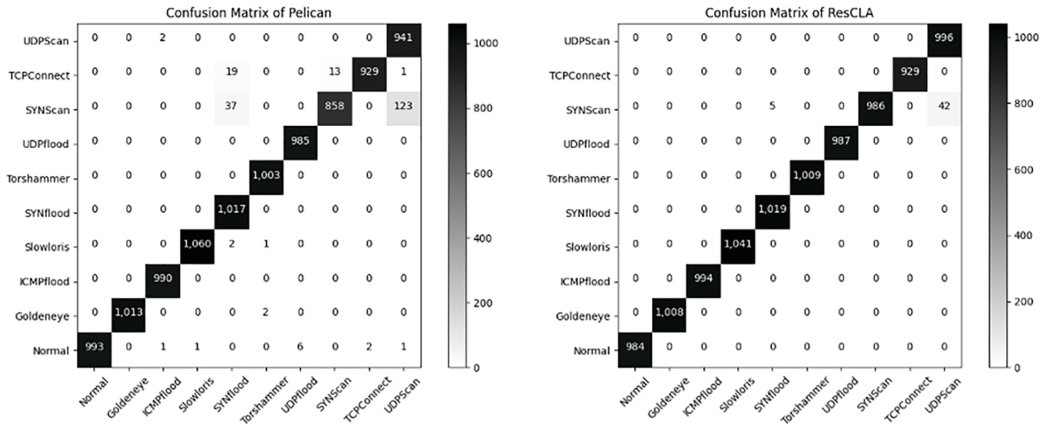


Figure 5: Confusion matrix of 5G-NIDD with different models

Table 5 presents a quantitative comparison between the Pelican and ResCLA models regarding their capability to identify ambiguous intrusion patterns across four distinct datasets. Each dataset encapsulates unique classes of ambiguous traffic, often characterized by temporal irregularities, low frequency, or overlapping signatures with benign flows. Notable examples include *SYNScan* and *UDPScan* in 5G-NIDD, *BruteForce* and *SlowHTTPTest* in CIC-IDS, *DoS* and *DDoS* in ToN-IoT, and *Password* and *Injection* in BoT-IoT. The Pelican model achieves average F1 scores of 97.84%, 98.30%, 98.11%, and 97.92% across these datasets, respectively. In contrast, the ResCLA model demonstrates superior performance with F1 scores of 99.51%, 99.43%, 99.19%, and 99.24%. These comparative gains of 1.67%, 1.13%, 1.08%, and 1.32% underscore ResCLA's enhanced capability in general classification and its significant proficiency in detecting subtle, ambiguous, and low-frequency attack vectors. These findings affirm the effectiveness of integrating attention-enhanced architectures within intrusion detection frameworks tailored for 5G and heterogeneous network environments, thereby ensuring high accuracy and robustness in practical applications.

Table 5: Comparison of intrusion detection performance between different models

Datasets	Ambiguous network traffic	F1 of Pelican	F1 of ResCLA	Comparison
5G-NIDD	SYNScan, UDPScan	97.84%	99.51%	+1.67%
CIC-IDS	BruteForce, SlowHTTPTest	98.30%	99.43%	+1.13%
ToN-IoT	DoS, DDoS	98.11%	99.19%	+1.08%
BoT-IoT	Password, Injection	97.92%	99.24%	+1.32%

To rigorously evaluate the individual contributions of each architectural component within the proposed ResCLA framework, we conducted a comprehensive series of **ablation experiments**. These experiments entailed systematically removing key functional modules—namely, the CNN, the LSTM unit, and the External Attention mechanism—either individually or in various combinations. The objective of this empirical analysis was to quantify the marginal and joint impacts of these modules on model performance, particularly in the context of detecting ambiguous, low-frequency, and evolving intrusion traffic patterns prevalent in heterogeneous 5G-enabled IoT networks.

The ablation study utilized four widely adopted and diverse intrusion detection datasets: **5G-NIDD**, **CIC-IDS2017**, **ToN-IoT**, and **BoT-IoT**. Each dataset presents unique challenges in terms of traffic distribution, attack sparsity, and temporal complexity. By assessing the F1 score—a balanced metric that combines precision and recall—we provided a robust and interpretable measure of detection fidelity under various architectural constraints.

As depicted in Table 6, the exclusion of any individual architectural component significantly affects the F1 score across various datasets, thus underscoring the essential role of each module within the ResCLA model. For example, the removal of the External Attention mechanism (referenced in the third row) resulted in an average decrease in the F1 score of approximately 1.5%. This observation highlights its pivotal role in modeling long-range dependencies and contextual relationships in high-dimensional 5G traffic streams. This mechanism is particularly adept at capturing cross-feature interactions and non-local semantics, which traditional convolutional or recurrent architectures might overlook.

Moreover, the absence of the CNN module has a pronounced impact on the model's performance. Eliminating this module leads to an average decline in the F1 score of more than 1.1%, underscoring its critical function in extracting localized spatial features and refining hierarchical traffic patterns from raw

data. These spatial features are crucial for identifying subtle intrusion signatures that may manifest within specific temporal or protocol-bound contexts.

Table 6: Ablation study on ResCLA model across different datasets

Configuration	F1 of 5G-NIDD	F1 of CIC-IDS	F1 of ToN-IoT	F1 of BoT-IoT
Full ResCLA	99.51%	99.43%	99.19%	99.24%
No CNN	97.88%	98.55%	98.32%	98.40%
No LSTM	98.21%	98.74%	98.59%	98.61%
No Attention	97.46%	98.33%	98.10%	98.02%
No LSTM + Attention	96.92%	97.88%	97.73%	97.61%
No CNN + Attention	96.65%	97.60%	97.42%	97.50%

The LSTM unit, which is designed to capture temporal dependencies and sequential patterns, also demonstrates high efficacy. Removing this module results in a consistent decrease in F1 scores, particularly noticeable in the ToN-IoT and BoT-IoT datasets, where time-evolving attacks (e.g., DoS, reconnaissance) are common. This indicates the model's diminid capability to detect long-term behavioral correlations in time-series traffic data without the LSTM framework.

Further insights are revealed under dual ablation scenarios. The simultaneous removal of both the LSTM and Attention modules leads to a substantial drop in the F1 score—up to 2.59% on the 5G-NIDD dataset—illustrating the synergistic benefits of combining temporal modeling with global attention. Similarly, the concurrent exclusion of the CNN and Attention modules (referenced in the last row) results in the most significant performance degradation across all datasets, confirming the complementary and interdependent nature of spatial feature extraction and global context encoding.

5 Conclusion

In this study, we introduced the ScalaDetect-5G framework, which has proven to be a robust, high-precision, and scalable deep intrusion detection system, specifically designed to address the complex and dynamic characteristics of 5G networks. Through the utilization of an enhanced CAE for adaptive feature reconstruction, coupled with the ResCLA for fine-grained classification, ScalaDetect-5G efficiently discerns the intricate patterns of both benign and malicious network traffic. Our experimental evaluation, conducted across several benchmark datasets, including 5G-NIDD, CIC-IDS, ToN-IoT, and BoT-IoT, demonstrates that ScalaDetect-5G consistently achieves exceptional detection performance. Notably, it attains mean F1 scores exceeding 99.19% with low standard deviation (e.g., $\leq$0.21%) across multiple runs, even under scenarios involving ambiguous, low-frequency, or obfuscated attack traffic. Furthermore, 95% confidence intervals indicate the statistical robustness of the observed performance gains. The system's elasticity allows it to scale with evolving traffic patterns and sustain high detection accuracy under fluctuating network loads, reinforcing its suitability for real-time deployment in mission-critical 5G infrastructures.

Additionally, the modular architecture and model-agnostic design of the system provide the flexibility to extend its application to cross-domain traffic analysis and accommodate emerging network paradigms. This adaptability positions ScalaDetect-5G as a forward-compatible security solution adept at managing the heterogeneous and high-dimensional nature of next-generation network environments. The results underscore its potential to serve as a foundational component in mobile network defense strategies, offering both practical deployability and theoretical innovation.

Looking to the future, as 6G communication architectures begin to integrate terrestrial, aerial, maritime, and space-based communication domains, the complexity and intelligence of the threat landscape are anticipated to increase significantly. To meet these upcoming challenges, future research will focus on enhancing the robustness of the CAE against adversarial samples, aiming to ensure reliable feature extraction under evasion or obfuscation attacks. Additionally, reinforcement learning-based approaches will be investigated to further improve the system's adaptability and decision-making accuracy in highly dynamic and uncertain 6G environments. These advancements are expected to propel ScalaDetect-5G beyond its current capabilities, ensuring its continued efficacy and relevance in safeguarding both current 5G deployments and future 6G ecosystems.

Acknowledgement: The authors would like to express their sincere appreciation to the colleagues at the School of Cyberspace Security, Beijing University of Posts and Telecommunications, for their valuable discussions, technical insights, and continuous support throughout the research process. Their constructive input greatly contributed to shaping the direction of this study. The authors are also grateful to the anonymous reviewers for their thoughtful comments and critical suggestions, which significantly improved the clarity, rigor, and overall quality of the manuscript.

Funding Statement: This research was conducted without any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors did not receive any financial support for the design, execution, analysis, or publication of this study.

Author Contributions: The authors confirm that each individual made substantial contributions to the work reported in this manuscript. Specifically, Shengjia Chang was responsible for the original draft preparation and early-stage writing. Baojiang Cui provided comprehensive revisions, critical feedback, and final polishing of the manuscript. Shaocong Feng contributed to the methodological design, data processing, and experimental validation. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to the sensitivity and privacy concerns of the data used in this research, participants did not consent to the public sharing of their information. Consequently, the supporting data are not publicly available. Researchers interested in further collaboration or validation may contact the corresponding author under specific data-use agreements, subject to ethical review and institutional approval.

Ethics Approval: This study did not involve any human subjects, animal experiments, or interventions requiring ethical approval. Therefore, ethical review and approval were not applicable to this work.

Conflicts of Interest: The authors declare that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Wang CX, You X, Gao X, Zhu X, Li Z, Zhang C, et al. On the road to 6G: visions, requirements, key technologies, and testbeds. IEEE Commun Surv Tutor. 2023;25(2):905–74. doi:10.1109/COMST.2023.3249835.
2. Shen LH, Feng KT, Hanzo L. Five facets of 6G: research challenges and opportunities. ACM Comput Surv. 2023;55(11):1–39.
3. Banafaa M, Shayea I, Din J, Hadri Azmi M, Alashbi A, Ibrahim Daradkeh Y, et al. 6G mobile communication technology: requirements, targets, applications, challenges, advantages, and opportunities. Alex Eng J. 2023;64:245–74. doi:10.1016/j.aej.2022.08.017.
4. Hakak S, Gadekallu TR, Maddikunta PKR, Ramu SP, Parimala M, De Alwis C, et al. Autonomous vehicles in 5G and beyond: a survey. Veh Commun. 2023;39(2):100551. doi:10.1016/j.vehcom.2022.100551.
5. Attaran M. The impact of 5G on the evolution of intelligent automation and industry digitization. J Ambient Intell Humaniz Comput. 2023;14(5):5977–93. doi:10.1007/s12652-020-02521-x.

6. Singh R, Sukapuram R, Chakraborty S. A survey of mobility-aware multi-access edge computing: challenges, use cases and future directions. Ad Hoc Netw. 2023;140(8):103044. doi:10.1016/j.adhoc.2022.103044.
7. Attaoui W, Sabir E, Elbiaze H, Guizani M. VNF and CNF placement in 5G: recent advances and future trends. IEEE Trans Netw Serv Manag. 2023;20(4):4698–733. doi:10.1109/tnsm.2023.3264005.
8. Cui Z, Cui B, Su L, Du H, Xu J, Fu J. A formal security analysis of the fast authentication procedure based on the security context in 5G networks. Soft Comput. 2024;28(3):1865–81. doi:10.1007/s00500-023-09486-x.
9. Kasongo SM. A deep learning technique for intrusion detection system using a Recurrent Neural Networks based framework. Comput Commun. 2023;199(1):113–25. doi:10.1016/j.comcom.2022.12.010.
10. Henry A, Gautam S, Khanna S, Rabie K, Shongwe T, Bhattacharya P, et al. Composition of hybrid deep learning model and feature optimization for intrusion detection system. Sensors. 2023;23(2):890. doi:10.3390/s23020890.
11. Hnamte V, Hussain J. DCNNBiLSTM: an efficient hybrid deep learning-based intrusion detection system. Telematics Inform Rep. 2023;10(1):100053. doi:10.1016/j.teler.2023.100053.
12. Xiao Y, Xing C, Zhang T, Zhao Z. An intrusion detection model based on feature reduction and convolutional neural networks. IEEE Access. 2019;7:42210–9. doi:10.1109/access.2019.2904620.
13. Kim J, Kim J, Kim H, Shim M, Choi E. CNN-based network intrusion detection against denial-of-service attacks. Electronics. 2020;9(6):916. doi:10.3390/electronics9060916.
14. Kim T, Pak W. Deep learning-based network intrusion detection using multiple image transformers. Appl Sci. 2023;13(5):2754. doi:10.3390/app13052754.
15. Koroniotis N, Moustafa N, Sitnikova E, Turnbull B. Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: bot-IoT dataset. Future Gener Comput Syst. 2019;100(7):779–96. doi:10.1016/j.future.2019.05.041.
16. Moustafa N. Ton-IoT Datasets; 2019. doi:10.21227/fesz-dm97.
17. Sharafaldin I, Habibi Lashkari A, Ghorbani AA. Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: Proceedings of the 4th International Conference on Information Systems Security and Privacy; 2018 Jan 22—24; Funchal, Madeira, Portugal: SciTePress—Science and Technology Publications; 2018. p. 108–16. doi:10.5220/0006639801080116.
18. Okey OD, Melgarejo DC, Saadi M, Rosa RL, Kleinschmidt JH, Rodríguez DZ. Transfer learning approach to IDS on cloud IoT devices using optimized CNN. IEEE Access. 2023;11(4):1023–38. doi:10.1109/access.2022.3233775.
19. Hnamte V, Nhung-Nguyen H, Hussain J, Hwa-Kim Y. A novel two-stage deep learning model for network intrusion detection: lstm-AE. IEEE Access. 2023;11:37131–48. doi:10.1109/access.2023.3266979.
20. Faruqui N, Abu Yousuf M, Whaiduzzaman M, Azad A, Alyami SA, Liò P, et al. SafetyMed: a novel IoMT intrusion detection system using CNN-LSTM hybridization. Electronics. 2023;12(17):3541. doi:10.3390/electronics12173541.
21. Wu P, Guo H. LuNet: a deep neural network for network intrusion detection. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI); 2019 Dec 6–9; Xiamen, China: IEEE; 2019. p. 617–24. doi:10.1109/ssci44817.2019.9003126.
22. Wu P, Guo H, Moustafa N. Pelican: a deep residual network for network intrusion detection. In: 2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W); 2020 Jun 29–Jul 2; Valencia, Spain: IEEE; 2020. p. 55–62. doi:10.1109/dsn-w50199.2020.00018.
23. Hussain F, Abbas SG, Husnain M, Fayyaz UU, Shahzad F, Shah GA. IoT DoS and DDoS attack detection using ResNet. In: 2020 IEEE 23rd International Multitopic Conference (INMIC); 2020 Nov 5–7; Bahawalpur, Pakistan: IEEE; 2020. p. 1–6. doi:10.1109/inmic50486.2020.9318216.
24. Sadhwani S, Mathur A, Muthalagu R, Pawar PM. 5G-SIID: an intelligent hybrid DDoS intrusion detector for 5G IoT networks. Int J Mach Learn Cybern. 2025;16(2):1243–63. doi:10.1007/s13042-024-02332-y.
25. Balın MF, Abid A, Zou J. Concrete autoencoders: differentiable feature selection and reconstruction. In: International Conference on Machine Learning; Long Beach, CA, USA: PMLR; 2019. p. 444–53.
26. Halbouni A, Gunawan TS, Habaebi MH, Halbouni M, Kartiwi M, Ahmad R. CNN-LSTM: hybrid deep neural network for network intrusion detection system. IEEE Access. 2022;10(11):99837–49. doi:10.1109/access.2022.3206425.

27. Guo MH, Liu ZN, Mu TJ, Hu SM. Beyond self-attention: external attention using two linear layers for visual tasks. IEEE Trans Pattern Anal Mach Intell. 2022;45(5):5436–47. doi:10.1109/tpami.2022.3211006.
28. Ghubaish A, Yang Z, Jain R. HDRL-IDS: a hybrid deep reinforcement learning intrusion detection system for enhancing the security of medical applications in 5G networks. In: 2024 International Conference on Smart Applications, Communications and Networking (SmartNets); 2024 May 28–30; Harrisonburg, VA, USA: IEEE; 2024. p. 1–6. doi:10.1109/SmartNets61466.2024.10577692.
29. Hsupeng B, Lee KW, Wei TE, Wang SH. Explainable malware detection using predefined network flow. In: 2022 24th International Conference on Advanced Communication Technology (ICACT); 2022 Feb 13–16; Pyeongchang, Kwangwoon-do, Republic of Korea: IEEE; 2022. p. 27–33.
30. Maddison CJ, Mnih A, Teh YW. The concrete distribution: a continuous relaxation of discrete random variables. In: International Conference on Learning Representations; San Juan, PR, USA; 2016.
31. Samarakoon S, Siriwardhana Y, Porambage P, Liyanage M, Chang SY, Kim J, et al. 5G-NIDD: a comprehensive network intrusion detection dataset generated over 5G wireless network. arXiv:2212.01298. 2022.