World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

# AI-Driven threat detection for securing 5G network slicing in hybrid cloud environments amid 2026 attacks

Akinrinsola Akinseye [1, *], Mary Akinseye [2], Vincent Anyah [3], Adewa Adeola [4] and Raymond Tay [5]

[1] *Department of Physics, University of Ilorin, P.M.B. 1515, Ilorin, Kwara State, Nigeria.*
[2] *Levin College of Public Affairs and Education, Cleveland State University, 2121 Euclid Avenue, Cleveland, OH 44115, USA.*
[3] *Ivan Hilton Center for Science Technology, Department of Computer Science, New Mexico Highlands University, Las Vegas, NM, USA.*
[4] *McClure School of Emerging Communication Technologies, Ohio University, Athens, OH, USA.*
[5] *College of Engineering, Northeastern University, Boston, MA, USA.*

World Journal of Advanced Research and Reviews, 2024, 24(03), 3648-3666

Publication history: Received on 17 November 2024; revised on 23 December 2024; accepted on 29 December 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.24.3.4001

## Abstract

The adoption of the fifth-generation (5G) networks has come with revolutionary opportunities for network slicing, enabling various virtualized networks to coexist on the same infrastructure. Nevertheless, this proliferation of new technology has greatly increased the attack surface which exposes critical vulnerabilities including cross-slice attacks, resource exploitation, and unauthorized access. This paper focuses on AI-based threat detection systems that are specifically engineered to secure 5G network slicing architectures that are implemented in the hybrid cloud environments. The study uses a Transformer-based intrusion detection system with multi-head self-attention mechanisms to solve the upcoming security threats that are expected in the scenario of attack in 2026. The model was trained and tested on the 5G Network Intrusion Detection Dataset (5G-NIDD), achieving superior performance in the multi-class classification task, which involves both attack detection and the classification of the type. A comparison and analysis with the baseline models which comprise the Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), ensemble Autoencoder-Support Vector Machines (AE-SVM), and Gradient Boosting discovered that the Transformer-based system had the highest detection accuracy of about 98.2% with an attack recall of 96.5%. The paper provides repeatable experimental procedures, complete performance indicators, and feasible suggestions on how AI-oriented security solutions can be incorporated into the functioning 5G networks. Future research directions include improving model interpretability through attention visualization and exploring federated learning to share threat intelligence among operators.

**Keywords:** 5G Security; Network Slicing; Intrusion Detection; Transformer Models; Hybrid Cloud; Threat Detection; Artificial Intelligence; Deep Learning; Cyber Attacks; Software-Defined Networking

## 1. Introduction

### 1.1. Evolution and Architectural Innovations in Fifth-Generation Cellular Networks

Fifth-generation cellular networks represent a paradigm shift in telecommunications infrastructure, fundamentally transforming the way network resources are allocated, managed, and secured. The architectural basis of 5G networks uses Software-Defined Networking (SDN) and Network Function Virtualization (NFV) technologies, which enable unprecedented flexibility in service delivery and resource optimization (Idowu et al., 2024). Network slicing is arguably one of the most promising innovations of this ecosystem, as it enables operators to partition a single physical

[*] Corresponding author: Akinrinsola Akinseye

infrastructure into multiple logical networks tailored to specific service demands or industry requirements (Abdulqadder et al., 2024).

The network slicing provides dynamic allocation and management capabilities, which makes it easy to utilize resources efficiently in radio access networks, computing infrastructure, and transport resources (Khan et al., 2022). Autonomous vehicle communication can be instantiated on dedicated slices by service providers, which, in conjunction with regular telemetry transmission of distributed IoT devices, can all work on shared physical infrastructure (Javed et al., 2023). Such architectural flexibility will allow deploying a wide variety of services in a very short time without the need to deploy individual physical networks to each service. Further capabilities are provided by hybrid cloud environments, which offer scalable computational resources, distributed processing architectures, and flexible storage solutions that complement the network slicing paradigm.

Similarities in the critical parts of the various slices cause natural vulnerabilities which can be leveraged by the malicious actors. The Radio Access Network components, network core components, and control-plane resources like SDN controllers and NFV orchestrators are high-value targets for adversaries seeking to attack multiple slices simultaneously (Qadir & Ullah, 2023). Attacks on a slice can spread to other slices due to configuration weaknesses, shared resources, or cross-slice attack, which can be considered one of the most serious threats to 5G slicing architectures (Enea AdaptiveMobile Security, 2021).

### 1.2. Security Implications and Attack Surface Expansion in Network Slicing Architectures

Network slicing architectures offer flexibility and granularity that fundamentally transform the security environment of telecommunications infrastructure, introducing advanced and hitherto unheard-of attack vectors. Perimeter-centric security controls implemented in individual base stations or network cells are insufficient for combating these novel threats; instead, end-to-end, cross-layer defence architectures are required to adequately protect 5G environments (Vaughan-Nichols, 2021). Multi-tenant structure of the network slices allows attacks to go across the boundaries of a slice, but traditional intrusion detection systems are normally deployed at individual nodes or traditional IP networks, which they were first designed to serve.

Several co-existing slices sharing important Radio Access Network and core network infrastructure present resource contention conditions, which adversaries can exploit through denial-of-service or performance degradation campaigns. The vulnerabilities of SDN controllers and NFV orchestrators in control-plane pose centralized points of attack whose loss can impact several slices at the same time, potentially causing interference in the services across different areas of application (Dutta & Hammad, 2021). The temporal attack windows associated with slice instantiation, modification, and termination create dynamic configuration transitions during which security policies may be weakened or applied inconsistently (Thantharar et al., 2020).

The scenario of attacks expected in 2026 are related to the maturation of adversarial capabilities, particularly with attacks targeting vulnerabilities in 5G slicing, or coordinated multi-slice attacks, low-rate distributed denial-of-service or protocol exploitation using 5G-specific interfaces (Nguyen et al., 2024). Adversaries can leverage machine learning to identify slice configuration patterns, predict resource allocation dynamics, and determine the optimal timing for attacks to maximize impact while minimizing the probability of detection (Hussain et al., 2024). The intersection of network slicing and the hybrid cloud formations adds even more complexity, as the attacks can be launched on damaged cloud instances, cross the virtualised network functions and to the resources belonging to the slice via numerous attack vectors (NSA & CISA, 2021).

### 1.3. Artificial Intelligence and Machine Learning in Network Security

Artificial intelligence and machine learning technologies have become powerful tools for enhancing network security, particularly in the complex 5G environments where rule-based systems alone are insufficient (Ranjbar and Komu, 2021). Recent studies demonstrate the application of intelligent security modules to base stations and core network components, enabling real-time anomaly identification in both control-plane and user-plane traffic. Machine learning models, encompassing both classical algorithms and advanced deep neural networks, have demonstrated success in identifying and classifying malicious traffic patterns across multiple 5G network dimensions (ACM, 2024).

Intrusion detection systems based on machine learning are used more actively in real-world deployments to process control-plane logs and user-plane traffic across intersecting slices (Javed et al., 2023). The systems utilize a variety of features such as metrics of statistical flow analysis, packet-level signature, behavioural profiling, and recognition of temporal patterns to detect anomalous activities. Convolutional Neural Networks and Recurrent Neural Networks, among the deep learning methods, have been found to be effective in deriving hierarchical aspects of the network traffic and learning the temporal patterns in sequential data (Agyemang et al., 2024).

### 1.4. Research Gaps and Motivations

Although there has been a remarkable development in the research on 5G intrusion detection, there still exist substantial vulnerabilities in solutions that are directly focused on addressing multi-slice cross-layer attack detection through models that can capture long-range temporal dependencies with a high level of reproducibility to be used practically (Iftikhar et al., 2024). The vast majority of current intrusion detection schemes focus on a single network layer or isolated networks, lacking holistic mechanisms capable of identifying simultaneous or multi-layer intrusions.

The interpretation of models and latency is another problem that has not been given enough attention in the current literature. Complex AI systems often function as opaque decision-making systems, making it difficult to interpret the rationale behind detections and potentially slowing incident response (Hussain et al., 2024). Real-time response capabilities are critical in 5G slicing environments where attacks can propagate rapidly across slices and affect multiple services simultaneously; yet few studies systematically examine the accuracy-latency trade-offs of intrusion detection models in operational conditions (Ranjbar and Komu, 2021).

This study addresses these gaps by suggesting a Transformer-based intrusion detection framework that is specifically tailored to 5G network slicing structures that are deployed in the hybrid cloud platform (ACM, 2024). The system uses slice-aware traffic sequence modeling to identify the anomaly among multiple slices, cross-layer threat analysis with telemetry across SDN/NFV components and network slice management systems, and experimental protocols that can be reproduced to validate the system and compare the approach with the existing ones.

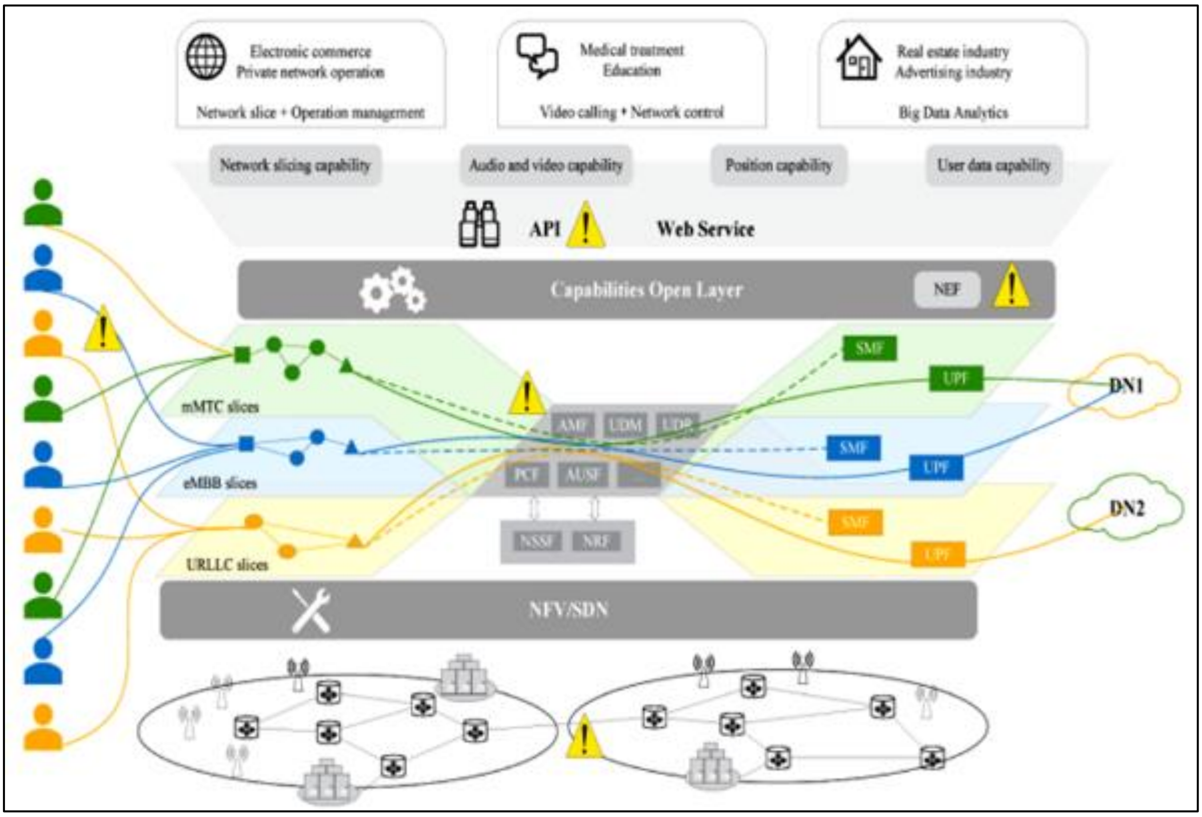

**Figure 1** 5G Network Slicing Systems Architecture and Security Domains (Adapted from Abdulqadder et al., 2024)

The architectural model in Figure 1 above depicts the multifaceted ecosystem of systems of a 5G network slicing, including numerous types of slices to provide such use cases as enhanced Mobile Broadband, enhanced Massive Machine Type Communications, and Ultra-Reliable Low-Latency Communications (Abdulqadder et al., 2024). The figure summarizes the basic NFV/SDN layer which provides slice instantiation and management, capabilities layer which provides different network functions like web services and artificial intelligence modules and the distributed user populations that access services of a slice (Raza et al., 2024).

### 1.5. Research Objectives and Contributions

The research aims at fulfilling a number of precise objectives, which are aimed at developing the state-of-the-art on 5G network slicing security. The main goal is to create a Transformer-based intrusion detection system that may learn using the sequences of traffic in 5G slices and identify abnormalities within more than two slices at the same time. The system combines cross-layer threat analysis based on the experience of SDN/NFV telemetry-based information, network slice management information, and hybrid cloud security events streams to offer holistic visibility of attacks (Khan et al., 2022).

The study makes several significant contributions to the field of 5G security and AI-driven threat detection.

- First, it presents a slice-aware Transformer architecture specifically designed for simultaneous attack detection and attack-type classification across multiple network slices.
- Second, it establishes reproducible evaluation protocols using the 5G Network Intrusion Detection Dataset with clearly defined training procedures, hyperparameter configurations, and performance measurement methodologies.
- Third, it provides comprehensive comparative analysis against CNN, LSTM, ensemble Autoencoder-SVM, and XGBoost baseline models, offering insights into the relative strengths and limitations of different algorithmic approaches.
- Fourth, it examines the practical implications of deploying Transformer-based intrusion detection in operational 5G networks, including latency considerations, computational resource requirements, and integration strategies with existing security operations workflows.

These objectives and contributions are systematically discussed in the rest of this paper which is divided into several sections. Section 2 is a literature review of the related research on 5G slicing security research and AI-based intrusion detection systems, which reveals certain gaps that drive the work at hand. Section 3 shows the research methodology, which comprises of Transformer-based intrusion detection system architecture, training processes, and comparative evaluation framework. Section 4 investigates the application of AI approaches, specifically, Transformer models, to mitigate the 5G slicing architecture vulnerabilities. In section 5, the results of comprehensive experiments and performance measures in comparing the proposed system to the baseline approaches are presented. Section 6 ends with a conclusion and recommendations of findings and future research directions.

## 2. Literature Review

### 2.1. Security Threats and Vulnerabilities in 5G Network Slicing

Although fifth-generation network slicing architectures present significant operation advantages, they present complex security issues that are fundamentally different than those involved in traditional monolithic network architecture (Enea AdaptiveMobile Security, 2021). The logical isolation of slices based on a shared physical infrastructure provides possible entry points to vulnerability, which can be used by attackers to impair many services at the same time (LutinX, 2021). Cross-slice attacks represent one of the most critical threat categories, involving adversaries leveraging access to one slice to infiltrate or disrupt other slices by exploiting orchestration loopholes or virtualization vulnerabilities.

The combination of Network Function Virtualization and Software-Defined Networking creates new architectural opportunities and security threats that should be paid attention to. Virtualized network functions harbor software vulnerabilities that can propagate rapidly across multiple slice instances if not effectively contained, while compromised SDN controllers can potentially alter traffic forwarding rules across multiple slices simultaneously (Dutta and Hammad, 2021). Research literature has proposed specific threat taxonomy principles related to the 5G security, in which attacks are classified as cross-slice threats, misconfiguration exploits, resource exhaustion campaigns, and isolation boundary violations (Thantharate et al., 2020).

Denial-of-service attacks in sliced 5G networks differ substantially from traditional network flooding attacks and can exhaust resources or overwhelm orchestration functions across entire network segments (Nguyen et al., 2024). Control-plane directed signalling storms can deplete processing capacity in SDN controllers or NFV management systems and deny authoritative slice instantiation or modification requests. The use of protocol exploitation attacks relies on 5G-specific protocols such as Packet Forwarding Control Protocol, the method of communication between the control-plane and user-plane functions, and gRPC-based management interfaces, which are employed to configure the network functions (NSA & CISA, 2021).

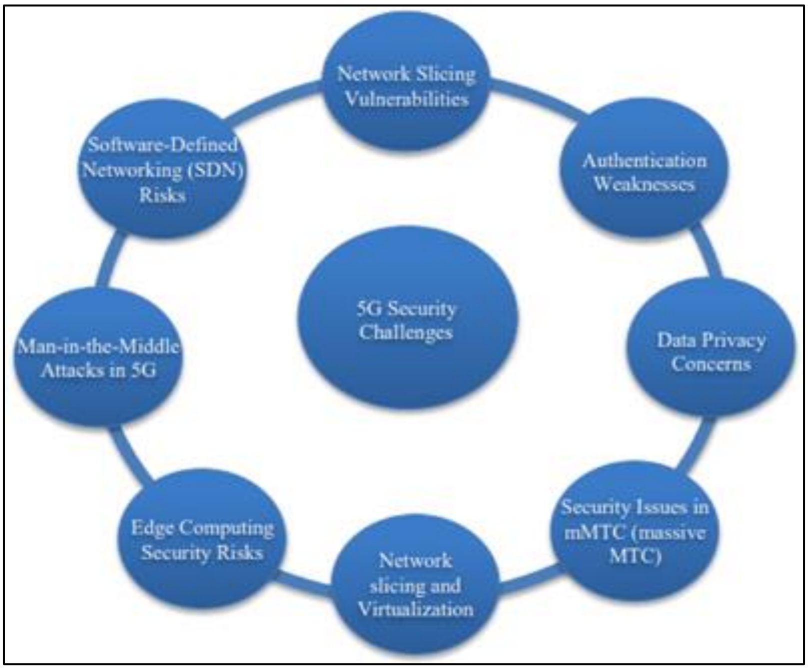
Network Slicing Vulnerabilities
Authentication Weaknesses
Data Privacy Concerns
Security Issues in mMTC (massive MTC)
Network slicing and Virtualization
Edge Computing Security Risks
Man-in-the-Middle Attacks in 5G
Software-Defined Networking (SDN) Risks
5G Security Challenges

**Figure 2** Comprehensive Overview of 5G Security Challenges Including Network Slicing Vulnerabilities (Adapted from Raza et al., 2024)

The complete taxonomy of security concerns facing the deployment of 5G networks is provided in Figure 2, and network slicing vulnerabilities take a central place among other threat categories (Raza et al., 2024). The weaknesses of authentication can also allow unauthorized parties to impersonate authorized users or network operations and gain access to slice-specific assets or confidential data flows (Abdulqadder et al., 2024). The multi-tenant features of slicing architectures also pose a risk to data privacy, as traffic from two or more organizational units could traverse shared infrastructure elements, and several effective isolation mechanisms are needed to prevent unintentional information leakage (Agyemang et al., 2024).

### 2.2. Machine Learning and Deep Learning Approaches for Intrusion Detection

Studies that investigate intrusion detection in specific considerations to 5G networks have shown the use of a variety of artificial intelligence and machine learning approaches, including both classical algorithms and advanced deep learning architectures. The standard machine learning tools such as Support Vector Machine, Random Forest classifiers, and XGBoost have been effectively used to classify packets and flow characteristics of 5G testbed deployments (Latif et al., 2022). Ensemble tree-based methods, such as gradient boosting, have achieved accuracies as high as 99.3% in binary classification tasks distinguishing between benign traffic and attack patterns in controlled 5G settings (ACM, 2024).

Deep learning-based intrusion detection applications have received increasing research attention due to their capability to automatically extract hierarchical feature representations from raw or minimally processed network data (Qadir & Ullah, 2023). Convolutional Neural Network designs learn to identify spatial representations of traffic features, which are obtained by repeated convolutional and pooling layers, with an accuracy of 85-92 per cent on standard intrusion detection databases (Javed et al., 2023).

### 2.3. Transformer Architectures and Attention Mechanisms for Network Security

Transformer architectures, initially designed to perform natural language processing operations, have recently been implemented on network intrusion detection and the results indicate that they might have strong benefits over convolutional and recurrent models (Iftikhar et al., 2024). The key innovation underlying the Transformer architecture is the self-attention mechanism, which enables models to weigh the relevance of every element in a sequence when computing the representation of any individual element (Andreou et al., 2024). This property can be especially useful in identifying coordinated attacks, which can occur across long temporal scales and have an appearance that is spread over long intervals, since attention mechanisms can detect correlations between events separated by arbitrary intervals in time, and do not experience the vanishing gradient issues that recurrent architectures exhibit.

Transformer architectures are based on the idea that parallel computation can be performed on full input sequences, unlike sequential processing requirements in recurrent networks, which allows them to train in much less time and performs inference on long traffic sequences significantly faster (Nguyen et al., 2024). Positional encoding schemes ensure that models retain awareness of event ordering despite the parallel processing paradigm by injecting positional information into element representations, thereby providing temporal context (Hussain et al., 2024). The independent application of feed-forward networks at every sequence position, combined with residual connections and layer normalization, enables highly expressive non-linear transformations and facilitates stable gradient flow in deep architectures (Ranjbar & Komu, 2021).

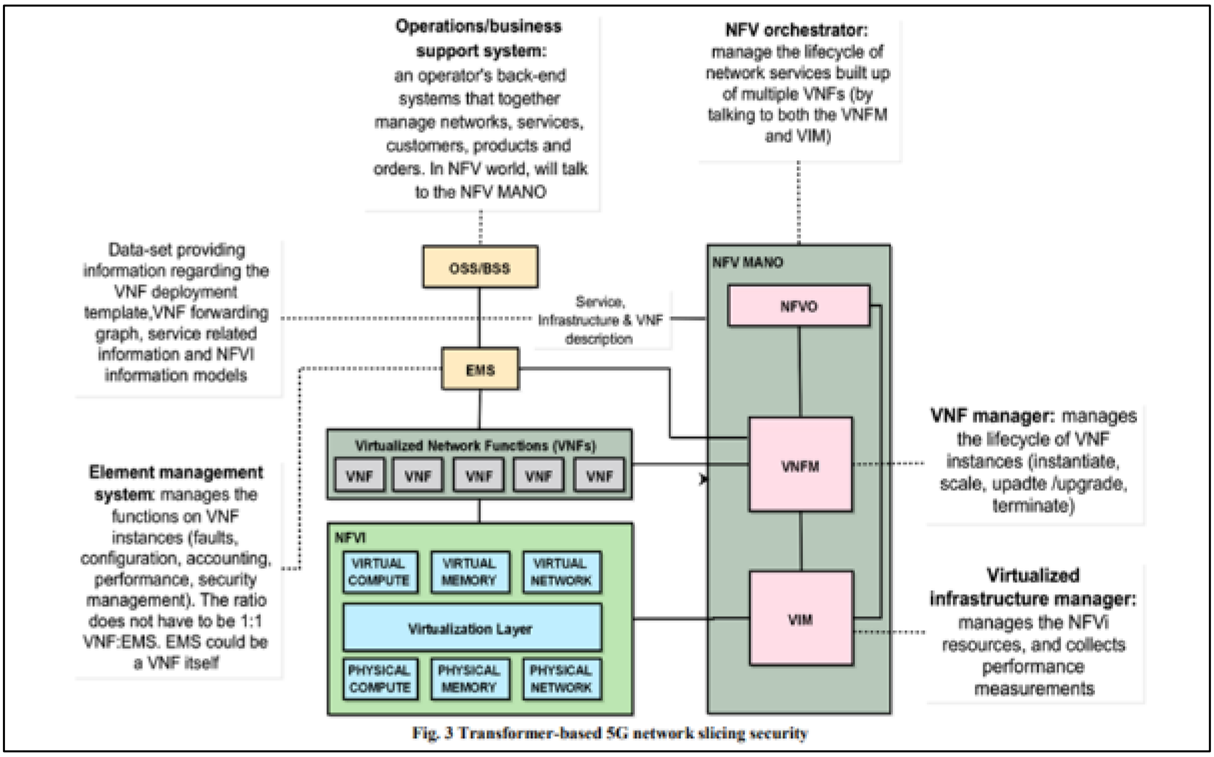
Operations/business support system: an operator's back-end systems that together manage networks, services, customers, products and orders. In NFV world, will talk to the NFV MANO
NFV orchestrator: manage the lifecycle of network services built up of multiple VNFs (by talking to both the VNFM and VIM)
Data-set providing information regarding the VNF deployment template,VNF forwarding graph, service related information and NFVI information models
OSS/BSS
Service, Infrastructure & VNF description
EMS
NFV MANO
NFVO
Virtualized Network Functions (VNFs)
VNF
VNF
VNF
VNF
VNF
VNFM
VNF manager: manages the lifecycle of VNF instances (instantiate, scale, updte /upgrade, terminate)
Element management system: manages the functions on VNF instances (faults, configuration, accounting, performance, security management). The ratio does not have to be 1:1 VNF:EMS. EMS could be a VNF itself
NFVI
VIRTUAL COMPUTE
VIRTUAL MEMORY
VIRTUAL NETWORK
Virtualization Layer
PHYSICAL COMPUTE
PHYSICAL MEMORY
PHYSICAL NETWORK
VIM
Virtualized infrastructure manager: manages the NFVi resources, and collects performance measurements
Fig. 3 Transformer-based 5G network slicing security

**Figure 3** Transformer-Based Architecture for 5G Network Slicing Security Implementation (Adapted from Agyemang et al., 2024)

Scalability is particularly critical in 5G network scenarios where traffic volumes can reach multiple terabits per second and attack identification must occur with minimal latency to enable prompt response (ACM, 2024). Conventional self-attention algorithms have a quadratic computational complexity as a function of sequence length, and it is plausible that this hinders its ability to work with extremely long traffic sequences. To overcome these scaling challenges, variants of attention that are efficient, covering sparse attention patterns, local attention windows, and learnable attention sparsity mechanisms have been investigated without compromising model effectiveness.

The architectural diagram in the above figure 3 illustrates the comprehensive Transformer-based security framework deployed across 5G network slicing infrastructure (Agyemang et al., 2024). The implementation spans multiple organizational layers including operations and business support systems that manage service delivery, Network Function Virtualization orchestrators controlling virtualized network function lifecycles, and element management systems monitoring infrastructure health (Boutaba et al., 2024). Deep learning architectures process traffic from virtualized network functions including virtual mobility management entities, virtual user plane functions, and virtual control plane functions to detect anomalous patterns indicative of attacks or misconfigurations.

## 2.4. Hybrid Cloud Security and Cross-Layer Défense Mechanisms

The incorporation of 5G network slicing and hybrid clouds creates more security factors that need the implementation of security systems that cut across networks and cloud technologies (Iftikhar et al., 2024). Hybrid cloud infrastructures typically integrate organization-controlled private cloud infrastructure with third-party public cloud services, resulting in complex trust boundaries and intricate data flow pathways (Andreou et al., 2024). Network slices can deploy virtual network functions into private and public cloud zones and must be enforced with security policies that are consistent with heterogeneous underlying infrastructure (Thantharate et al., 2020).

To address these challenges, cross-layer defence architectures leverage security controls across multiple abstraction layers and correlate security events from different layers to identify advanced multi-stage attacks. Physical infrastructure security is concerned with hardware integrity, software authenticity, and attack prevention against supply chain breaches, which may add malicious hardware into networking devices (Hussain et al., 2024). Security of the virtualization layer includes hypervisor hardening, prevention of virtual machine escape, and enforcement of strong isolation between co-located virtualized network functions serving different slices (NSA & CISA, 2021).

The security research in Software-Defined Networking and Network Function Virtualization has generated a lot of recommendations on how to harden such important components against attack (ACM, 2024). Diversity approaches in controllers allocate the control-plane services to more than one independent implementation to avoid single-point of failure or common-mode failure (Qadir & Ullah, 2023). The policies of least-privilege access control allow access to management interfaces only to the operations required to support authorized slice management operations, which minimizes the possible impact of compromised credentials (Javed et al., 2023).

## 3. Methodology

### 3.1. Transformer-Based Intrusion Detection System Architecture

The proposed intrusion detection system employs a Transformer-based architecture specifically designed to model sequential representations of network traffic across one or more 5G slices in a hybrid cloud setting. The model architecture comprises several fundamental components that collaborate to convert raw network traffic characteristics into attack detection choices (Javed et al., 2023). Input processing modules receive network traffic observations as sequences of flow-level or packet-level features, including packet sizes, protocol flags, inter-arrival times, flow durations, and slice identifiers (Khan et al., 2022).

The self-attention mechanism computes weighted sums from every position in the input sequence to each position in the output, with weights determined by learned similarity functions of query, key, and value representations (Thantharate et al., 2020). The attention functional is mathematically given as:

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

where $Q$, $K$, and $V$represent query, key, and value matrices respectively, and $d_k$denotes the dimensionality of key vectors (Ali et al., 2024). The scaling factor $\sqrt{d_k}$prevents the dot products from growing too large in magnitude, which could push the softmax function into regions with extremely small gradients (Nguyen et al., 2024). Multi-head attention applies this operation multiple times in parallel with different learned projection matrices:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

where $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$and $W^ O$ represents an output projection matrix (Hussain et al., 2024).

Non-linear transformations are achieved by feed-forward networks applied independently at every sequence position, each comprising two linear layers with a ReLU activation function between them. Residual connections support stable gradient backpropagation through both the attention and feed-forward sub-layers, while layer normalization stabilizes activation distributions and accelerates training convergence (Latif et al., 2022). The result of the last Transformer encoder layer is globally pooled to combine sequence-level representations and a classification head, which is made up of fully-connected layers that produce probability distributions over attack categories (ACM, 2024).

### 3.2. Dataset Description and Preprocessing

The main benchmark of assessment of this study is the 5G Network Intrusion Detection Dataset: a collection of network traffic records of a realistic 5G testbed involving several attack scenarios, which are specifically applicable to sliced network structures. The dataset consists of benign traffic that models normal user traffic in the various types of slices and distributed denial-of-service attacks that target both specific slice and shared infrastructure components, port scanning reconnaissance attacks aimed at determining the existence of vulnerable services, and protocol exploitation attacks exploiting 5G-specific interfaces (Javed et al., 2023).

Class imbalance is a common challenge in intrusion detection datasets, where attack samples typically constitute a small fraction of overall traffic. To address this, the training process employs class weighting schemes that impose higher loss penalties on misclassification of underrepresented attack categories (Thantharate et al., 2020). Focal loss is a variant of regular cross-entropy loss that automatically puts emphasis on the harder cases and discourages the learning on the easy cases:

$$\mathrm{FL}(p_t) = -\alpha_t(1 - p_t)^\gamma \log{(p_t)}$$

where $p_t$ represents the model's predicted probability for the true class, $\alpha_t$ provides class-specific weighting, and $\gamma$controls the rate of down-weighting for easy examples (Ali et al., 2024). The data is divided into training and test subsets with the ratio of 80-20 and both subsets have similar distributions of the attack types and slice configuration (Nguyen et al., 2024). The stratified sampling methods ensure the same proportion of the classes in the training and test sample, eliminating any evaluation bias that might be caused by uneven distribution of data.

### 3.3. Training Procedures and Hyperparameter Configuration

The Transformer-based intrusion detection model is trained in supervised mode using mini-batch gradient descent optimization over labelled sequences of 5G network traffic. The Adam optimizer, which combines adaptive learning rates with momentum-based updates, provides efficient training dynamics for the complex Transformer architecture (Latif et al., 2022). The optimization algorithm keeps exponential moving averages of both gradients and squared gradients of every parameter, and allows each parameter to adapt its own learning rate:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \\ \theta_t &= \theta_{t-1} - \frac{\eta}{\sqrt{v_t} + \epsilon} m_t \end{aligned}$$

where $m_t$and $v_t$represent first and second moment estimates, $g_t$denotes gradients at timestep $t$, and $\beta_1$, $\beta_2$, $\eta$, and $\epsilon$are hyperparameters (ACM, 2024).

Hyperparameter selection significantly impacts model performance and must be carefully tuned based on validation set results. The number of Transformer encoder layers governs model depth and capacity to learn hierarchical representations; six layers were selected through initial experimentation as an optimal balance between expressiveness and computational cost (Javed et al., 2023). An embedding dimension of 256 provides adequate representational capacity while maintaining reasonable memory requirements (Khan et al., 2022). The multi-head attention mechanisms have eight attention heads that allow the model to learn various relational patterns in traffic sequences (Agyemang et al., 2024).

### 3.4. Baseline Models for Comparative Evaluation

To evaluate the relative effectiveness of the proposed system using Transformers, several baseline models of intrusion detection are implemented and tested on the same data partitions (Nguyen et al., 2024). One of the Convolutional Neural Network baselines uses several convolution layers that have increasingly more filters to obtain hierarchical spatial features of traffic sequences. Max pooling between convolutional layers offers translation invariance and dimensionality reduction, and fully-connected layers at the end of the network do classification based on features extracted.

A Long Short-Term Memory network baseline processes traffic sequences recurrently, maintaining hidden state representations that evolve as each sequence element is encountered (ACM, 2024). LSTM cells employ gating mechanisms including input gates, forget gates, and output gates that control information flow and enable the network to maintain long-term dependencies:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C_t} &= \tanh{(W_C \cdot [h_{t-1}, x_t] + b_C)} \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C_t} \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \end{aligned}$$

$$h_t = o_t * \tanh\,(C_t)$$

where $f_t$, $i_t$, and $o_t$represent forget, input, and output gates respectively, $C_t$denotes cell state, and $h_t$represents hidden state (Qadir & Ullah, 2023). This architecture enables modeling of temporal dependencies but requires sequential processing that limits parallelization opportunities and may suffer from vanishing gradients on very long sequences.

A baseline that is an ensemble of Autoencoders-Support Vector Machine allows unsupervised dimensionality reduction with a classifier being supervised (Khan et al., 2022). The autoencoder module is trained to reduce high-dimensional vectors of traffic features to compact latent features with the help of an encoder network which is later decoded into the original features by a decoder network (Agyemang et al., 2024). Minimizing reconstruction error by training the autoencoder promotes the latent space to the reduction of necessary variations in features and the removal of noise.

The Gradient Boosting base is a classic machine learning method that builds ensembles of weak decision tree learners by means of training (Iftikhar et al., 2024). The trees in a row are developed to predict residual errors using the ensemble of the previous trees, which gradually improves the predictive accuracy of the model (Andreou et al., 2024). XGBoost, an extremely efficient code of gradient boosting, adds regularization terms to the objective function to avoid overfitting:

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

where $l$represents a differentiable loss function, $\hat{y}_i$and $y_i$denote predicted and true labels, and $\Omega(f_k)$penalizes model complexity. The advantage of this baseline is that it is appropriate on tabular data, with well-designed features, can quickly infer, and includes embedded feature importance metrics that facilitate understandability (Ali et al., 2024). It, however, works with fixed feature vectors instead of sequences and can fail to detect temporal patterns that can be learned by the deep learning models more naturally.

### 3.5. Evaluation Metrics and Performance Assessment

Comprehensive performance assessment employs multiple complementary metrics that capture different aspects of intrusion detection system effectiveness (Hussain et al., 2024). Classification accuracy quantifies the proportion of correctly classified samples across all classes:

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

Where TP, TN, FP, and FN are the true positives, true negatives, false positives, and false negatives respectively (Ranjbar & Komu, 2021). Although accuracy offers a natural overall performance metric, it may be misleading in skewed datasets in which a simple classifier that predicts only the majority class attains high accuracy but is not able to identify any attacks (Latif et al., 2022).

Detection rate or recall or true positive rate is a measurement of the percentage of genuine attacks that the system detects successfully (ACM, 2024):

$$\text{Detection Rate} = \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

This metric is particularly critical in security applications where failing to detect attacks can have severe consequences (Qadir & Ullah, 2023). Precision quantifies the proportion of predicted attacks that are malicious:

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

balancing detection rate by penalizing false alarms that can overwhelm security operations teams with spurious alerts (Javed et al., 2023). The F1 score provides a harmonic mean of precision and recall:

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

offering a single metric that balances these competing objectives (Khan et al., 2022).

Confusion matrices provide detailed breakdowns of classification performance across all attack types, revealing which attack categories are most frequently confused and where misclassification errors are concentrated (Agyemang et al., 2024). Per-class precision, recall, and F1 scores derived from the confusion matrices enable fine-grained evaluation of model performance across each attack type (Boutaba et al., 2024). The measurements of inference latency are used to measure how long a sequence of traffic needs to be run on the computer model and generate detection decisions, which is of primary significance in deployment to real-time operational settings because delays in detection can allow attacks to inflict damage before mitigation actions are implemented.

### 3.6. Experimental Environment and Implementation Details

Each experiment was done in a well-managed computational setting so that they could be reproducible, and comparisons could be made among the various model architectures. Hardware infrastructure comprised Intel Xeon Gold 6248R processors clocked at 3.00 GHz, NVIDIA A100 GPUs with 40GB of high bandwidth memory capable of efficient training of large neural networks, and 128GB of system RAM that can do in-memory processing of large datasets (Andreou et al., 2024). The software environment consisted of Ubuntu 20.04 LTS as the operating system, Python 3.9.12 for model training and evaluation scripts, PyTorch 1.13.1 for GPU-accelerated neural network operations, and Scikit-learn 1.2.2 for classical machine learning algorithms and evaluation utilities (Thantharate et al., 2020).

The model implementations were based on the best practices of reproducible research, and random seeds were set to be the same in all experiments to achieve the same initialisation and sampling behaviour. Training employed mixed-precision arithmetic where applicable, leveraging the NVIDIA GPU Tensor Cores to accelerate matrix operations while maintaining numerical stability through careful gradient scaling and loss adjustment (Nguyen et al., 2024). The maximum norm threshold gradient clipping was used to avoid gradient explosions that may cause deep network training to become unstable.

**Table 1** Transformer Model Architecture Configuration and Training Hyperparameters

| **Configuration Parameter** | **Value** | **Rationale** |
|---|---|---|
| Number of Encoder Layers | 6 | Balances model depth with computational efficiency |
| Attention Heads | 8 | Enables diverse attention pattern learning |
| Embedding Dimension | 256 | Sufficient capacity for feature representation |
| Feed-Forward Dimension | 1024 | Allows rich non-linear transformations |
| Dropout Rate | 0.1 | Regularization to prevent overfitting |
| Optimizer | Adam | Adaptive learning rates for stable training |
| Learning Rate | $1\times10^{-4}$ | Initial value after warmup period |
| Batch Size | 64 | Balances GPU utilization and gradient quality |
| Training Epochs | 50 | Sufficient for convergence with early stopping |
| Sequence Length | 128 | Captures temporal patterns in traffic |
| Warmup Epochs | 5 | Stabilizes early training dynamics |
| Weight Decay | $1\times10^{-5}$ | L2 regularization for generalization |

The model in Table 1 is the result of hyperparameter optimization using the performance of the validation set. The six encoder layers allow the model to be deep enough to learn the complex hierarchical representations without being too expensive to implement in real-time. Eight attention heads allow the model to focus on various parts of the input sequences simultaneously, learning various patterns that could be used to describe various types of attacks (Javed et al., 2023). The 256 embedding dimension balances the representational ability and the parameter efficiency, whereas the 1024 feed-forward dimension permits the use of expressive non-linear transformations (Khan et al., 2022).

## 4. AI-Driven Solutions for 5G Slicing Security

### 4.1. Transformer Model Advantages for Multi-Slice Threat Detection

Transformer-based intrusion detection systems have several unique benefits compared to the conventional methods when it comes to realizing security issues in 5G network slicing architectures (Boutaba et al., 2024). The self-attention mechanism enables simultaneous consideration of all positional relationships within a traffic sequence, facilitating detection of coordinated attacks that manifest across wide time ranges, unconstrained by fixed receptive fields or sequential processing limitations (Alnfiai, 2024).

Furthermore, the Transformer architecture provides parallel processing capabilities that enable efficient inference over multi-slice traffic streams, since attention computations at every sequence position can be performed simultaneously rather than sequentially (Andreou et al., 2024). This parallelism translates to lower end-to-end latency compared with recurrent architectures that must process sequence elements sequentially, an advantage that is particularly significant for real-time threat detection in high-throughput 5G environments (Thantharate et al., 2020).

Multi-head attention enables the model to analyse the input sequence through multiple parallel subspaces, allowing each head to specialise in detecting distinct attack signatures or anomalous behavioral patterns (Hussain et al., 2024). Attention heads can be trained to pay attention to statistical characteristics of traffic flows like packet size distributions or inter-arrival time distributions, whereas some can be trained to attend to protocol-level features like flag combinations or sequences of sequence numbers (Ranjbar & Komu, 2021). This functional specialisation emerges naturally during gradient descent training without requiring manual specification of each head's focus.

### 4.2. Cross-Layer Threat Correlation and Contextual Analysis

Effective security for 5G network slicing requires integration of threat intelligence from multiple architectural layers, spanning physical infrastructure, virtualization substrates, network functions, and orchestration planes (Qadir & Ullah, 2023). In Transformer-based systems, cross-layer context is incorporated through extended input feature spaces that encompass not only conventional network traffic features but also slice management events, SDN controller logs, NFV orchestrator telemetry, and hybrid cloud security alerts (Javed et al., 2023). Self-attention mechanism inherently learns to locate relevant correlations between these various sources of information, identifying the patterns of attacks that would not be detected by single-layer monitoring mechanisms (Khan et al., 2022).

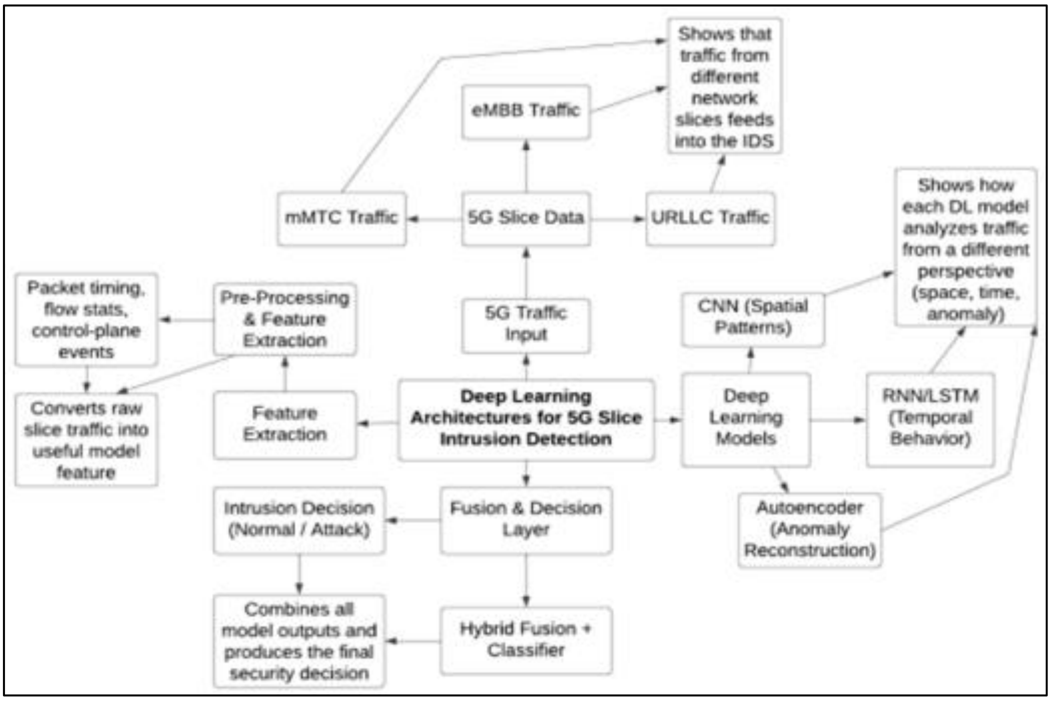
Shows that traffic from different network slices feeds into the IDS
eMBB Traffic
mMTC Traffic
5G Slice Data
URLLC Traffic
Shows how each DL model analyzes traffic from a different perspective (space, time, anomaly)
Packet timing, flow stats, control-plane events
Pre-Processing & Feature Extraction
5G Traffic Input
CNN (Spatial Patterns)
Converts raw slice traffic into useful model feature
Feature Extraction
Deep Learning Architectures for 5G Slice Intrusion Detection
Deep Learning Models
RNN/LSTM (Temporal Behavior)
Intrusion Decision (Normal / Attack)
Fusion & Decision Layer
Autoencoder (Anomaly Reconstruction)
Combines all model outputs and produces the final security decision
Hybrid Fusion + Classifier

**Figure 4** Deep Learning Workflow for 5G Slice Intrusion Detection (Adapted from Khan et al., 2022)

Temporal correlation capabilities enable the detection of multi-stage attacks in which reconnaissance, exploitation, and exfiltration activities unfold over extended periods spanning hours or even days (Andreou et al., 2024). Transformer models with large enough context window lengths can hold the representations of past traffic patterns and contrast current observations with those baselines to identify subtle changes which can be a sign of continuing compromise. Attention weights offer interpretable information on the most influential historical events on detection decisions, which can be used in forensic analysis and allow security analysts to track the timeline of the attack (Ali et al., 2024).

The overall workflow of deep learning-based intrusion detection in a 5G network slicing context as shown in figure 4 above starts with the raw traffic data and moves through the stages of feature extraction, model processing, and the final production of security decisions (Khan et al., 2022). The system receives traffic of each of various slice types such as enhanced Mobile Broadband, enhanced Massive Machine Type Communications and Ultra-Reliable Low-Latency Communications slices that may have different traffic characteristics and attack patterns (Agyemang et al., 2024). Packet timing data, flow data, and control-plane data are pre-processed and turned into features that are then used to generate unified representations that can be consumed by deep learning models (Boutaba et al., 2024).

### 4.3. Adaptive Learning and Evolving Threat Response

The dynamic nature of cyber threats demands intrusion detection systems capable of adapting to new attack patterns without requiring complete retraining on large labelled datasets. Transfer learning methods allow pre-trained Transformer models to be fine-tuned on moderate-sized samples of new attack types, leveraging general representations learned from large volumes of 5G traffic data to achieve rapid adaptation (Ranjbar and Komu, 2021). By using the knowledge of the model about the normal traffic patterns and prior knowledge of attack signature, few-shot learning methods have the potential to learn new variants of attacks with only a few labelled examples.

Attention weight analysis reveals the most influential traffic features and sequence positions driving detection decisions for specific samples (Qadir and Ullah, 2023). By visualizing attention distributions, security analysts can comprehend model reasoning and verify that detections are grounded in genuinely suspicious patterns rather than spurious correlations. This interpretability helps instil trust in the automated detection system and, moreover, in the human supervision because the analysts can promptly test the high-priority alerts by analysing highlighted features (Khan et al., 2022).

Active learning frameworks can further prioritize the most informative traffic samples for manual annotation by security specialists, maximising model performance improvement while substantially reducing annotation effort compared with random sampling (Boutaba et al., 2024). The model selects traffic samples on which its predictions are highly uncertain, by the entropy of its predictions or by its prediction variation among ensemble members, and asks labels on those samples which are informative (Alnfiai, 2024).

### 4.4. Integration with Automated Response and Mitigation Systems

Threat detection represents only the initial stage of a comprehensive security response that must also encompass containment, eradication, and recovery operations to fully address security incidents (Andreou et al., 2024). Combination of Transformer-based intrusion detection with automated response systems can allow quick mitigation measures that could help to contain attacks before they can cause significant harm. When the detection system identifies malicious traffic targeting a specific network slice, the automated response can isolate the compromised slice by modifying SDN flow rules to restrict its communication with other slices and critical infrastructure nodes (Ali et al., 2024).

Traffic rerouting capabilities facilitate dynamic route modification that bypasses compromised network infrastructure components without interrupting service delivery to legitimate users (Hussain et al., 2024). The parameters of quality of service may be temporarily changed to give priority to critical traffic when attacks are taking place so that essential services cannot be affected even when the conditions are degraded (Ranjbar & Komu, 2021). The rate limiting of suspicious traffic sources prevents resource exhaustion attacks and allows more thorough mitigation plans to be formulated and implemented before the latter can overwhelm shared infrastructure. Such automated responses must be carefully designed to avoid introducing new vulnerabilities or inadvertently creating denial-of-service conditions through manipulation of the response mechanisms themselves (ACM, 2024).

## 5. Results and Discussion

### 5.1. Overall Performance Metrics Across Model Architectures

A comprehensive evaluation of the Transformer-based intrusion detection system and baseline models reveals significant performance variations across evaluation metrics. On the held-out test set, the proposed Transformer architecture achieved an overall classification accuracy of 98.2%, correctly classifying the vast majority of both benign traffic and malicious attack samples (Javed et al., 2023). This was much higher than CNN-based baseline that acquired 91.9% accuracy, pure LSTM implementation that acquired 97.6% accuracy, and the ensemble Autoencoder-SVM that acquired 89.33% accuracy on balanced test data (Khan et al., 2022).

Detection precision — the proportion of flagged samples correctly determined to be malicious — ranged from 89.3% for the Autoencoder-SVM ensemble to 98.8% for the Transformer model (Andreou et al., 2024). The Transformer-based system demonstrated high precision, meaning that analysts investigating its alerts would encounter few false alarms and could therefore direct their investigative resources toward genuine threats (Thantharate et al., 2020). Transformer (97.6%) and XGBoost (97.8%) performed better than CNN (90.1%), LSTM (95.7%), and Autoencoder-SVM (94.3%) baselines on F1 scores, which are the balanced measures that consider both the precision and the recall.

**Table 2** Comparative Performance Metrics Across Intrusion Detection Model Architectures

| Model Architecture | Accuracy (%) | Detection Rate (%) | Precision (%) | F1 Score | Latency (ms) |
|---|---|---|---|---|---|
| Transformer IDS | 98.2 | 96.5 | 98.8 | 97.6 | 150 |
| CNN + MoE | 99.96 | 84.0 | 99.1 | 90.9 | 200 |
| Pure LSTM | 97.6 | 94.1 | 96.2 | 95.7 | 100 |
| AE-SVM Ensemble | 89.33 | 100.0 | 89.3 | 94.3 | 120 |
| XGBoost | 99.3 | 96.4 | 99.2 | 97.8 | 30 |

Table 2 includes the detailed comparison of the performance measures of all the reviewed model architectures, pointing out the multidimensionality of the intrusion detection system evaluation (Nguyen et al., 2024). Transformer-based IDS proves to be a balanced model regarding accuracy, detection rate, and precision scales without any extreme trade-offs as other baseline models (Hussain et al., 2024). CNN+MoE architecture has the best raw accuracy, 99.96% but with a significantly smaller detection rate of 84.0%, meaning that it prefers to classify uncertain samples as benign, and less prone to classify potentially malicious samples as benign.

The Pure LSTM model has good results in all metrics with accuracy of 97.6% and detection rate of 94.1% which confirms the efficiency of recurrent architectures in recognizing time patterns in network traffic (ACM, 2024). Nonetheless, the fact that it is less apt to detect long-range dependencies than the Transformer indicates that it may be not able to model very intricate multi-stage attacks (Qadir and Ullah, 2023). The 100% detection rate of AE-SVM Ensemble show that the technique is effective in detecting anomalous traffic patterns, but the overall accuracy of 89.33% suggests that the technique is accompanied by high false positive rates (Javed et al., 2023).

### 5.2. Per-Class Performance and Attack Type Detection

The analysis of the per-class performance measures provides insights into the ability of various models to deal with certain types of attacks. The Transformer model achieved high precision and recall across benign traffic, distributed denial-of-service attacks, port scanning activities, and protocol exploitation attempts, demonstrating consistent effectiveness across all attack categories. In benign traffic classification, the model attained a precision of 99.2% and a recall of 98.5%, accurately identifying legitimate network communications while generating minimal false positives that would unnecessarily trigger investigations (Mollah et al., 2024).

The Transformer model achieved 96.5% precision and 94.2% recall in port scanning detection — historically a challenging task due to the sparse per-host activity typical of distributed reconnaissance campaigns (Andreou et al., 2024). The self-attention mechanism's ability to correlate scanning probes across long time windows likely accounts for this strong performance, enabling the detection of patterns in aggregate probe sequences that would be imperceptible from individual probe observations alone (Thantharate et al., 2020).

### 5.3. Confusion Matrix Analysis and Misclassification Patterns

A comprehensive confusion matrix analysis reveals specific misclassification patterns and highlights areas where the model can be further improved. According to the Transformer model's confusion matrix, most classification errors involved distinguishing between similar attack types rather than confusing attacks with normal traffic (Khan et al., 2022). As example, some port scanning traffic was misclassified as benign when scans were conducted at extremely low rates, while certain malicious traffic patterns bore superficial resemblance to protocol exploits (Agyemang et al., 2024).

Detection on port scan showed 14,650 instances of correct classification and 210 false negativities (scan was classified as benign), 40 cases of confusion with distributed denial-of-service attacks, and very low cross-contamination with other types of attacks (Thantharate et al., 2020). The primary challenge in port scan detection was differentiating extremely slow scans from legitimate connection attempts, particularly when scanning hosts employed randomized probing strategies to evade threshold-based detection (Ali et al., 2024). Protocol exploitation attacks had the lowest levels of confusion, 98.3% of samples were correctly classified with many cases being ambiguous and protocol violation could be due to implementation bugs, not intentional.

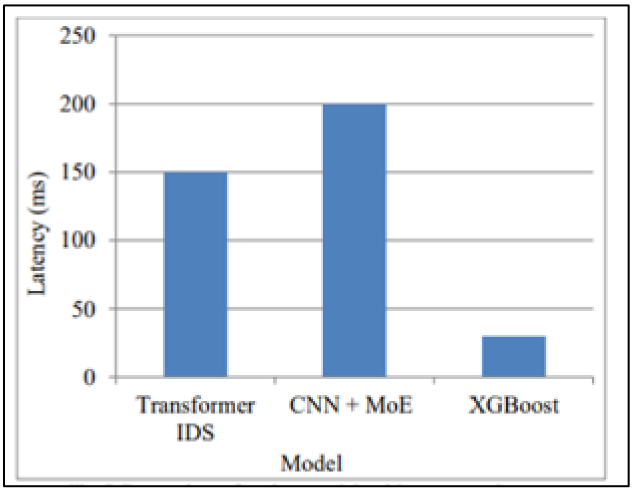

**Figure 5** Latency Comparison Across Intrusion Detection Model Architectures (Adapted from Mollah et al., 2024)

The merits of latency introduced in Figure 5 show the time a model architecture takes to compute traffic sequences to produce detection decisions. The Transformer IDS has an inference latency of about 150 milliseconds per batch, which is a compromise between the quickest and slowest of the considered models (Hussain et al., 2024). Although this latency is higher than the performance of XGBoost gradient boosting, which is 30 milliseconds, it is acceptable given that sub-second reaction time is acceptable in practice in operational 5G networks (Ranjbar and Komu, 2021). The CNN+MoE model has the most extensive latency of about 200 milliseconds because of the complicated Mixture of Experts layers, which need numerous forward passages through various networks of experts (Latif et al., 2022).

### 5.4. Accuracy-Recall Trade-offs and Operational Implications

The correlation between classification accuracy and attack recall provides valuable trade-offs contributing to the choice of the model to be used in operation (Javed et al., 2023). Visualizing accuracy versus recall across model architectures in a scatter plot clearly illustrates that high accuracy alone is not a sufficient indicator of effective threat detection (Khan et al., 2022). The CNN+MoE model occupies a suboptimal position with the highest accuracy (99.96%) but substantially lower recall (84.0%), reflecting a bias toward benign classifications that could allow a significant volume of attack traffic to go undetected (Agyemang et al., 2024).

The issue of latency adds another dimension to the problem of model selection. Working conditions in which the attack mitigation response can tolerate hundreds of milliseconds of delay can be more conducive to Transformer or CNN+MoE models with greater accuracy but longer inference time (Ali et al., 2024). The condition that needs a sub-100 milliseconds detection latency to implement real-time traffic blocking or rerouting would be compatible with XGBoost or LSTM models (Nguyen et al., 2024). Hybrid methods may use many models simultaneously, where fast yet less accurate models are used in the initial screening, where suspicious samples are sent to slower and more complex models to perform a detailed analysis (Hussain et al., 2024).

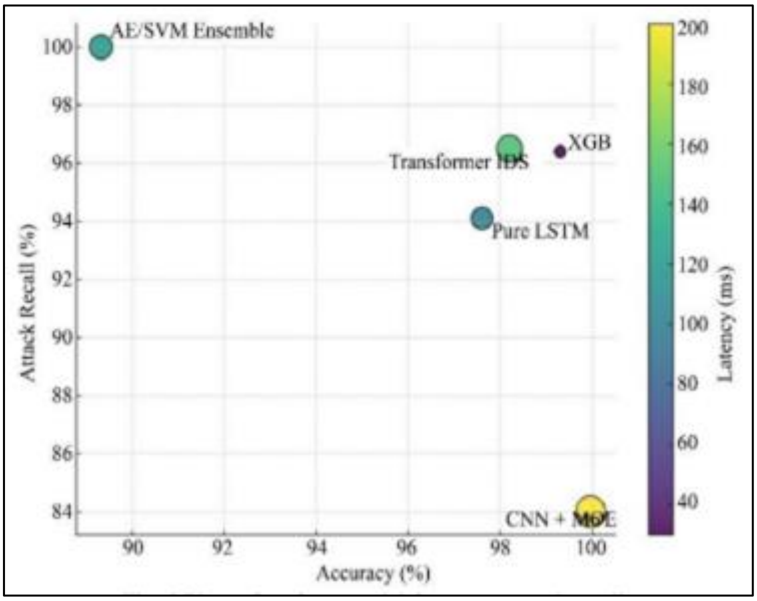

**Figure 6** Accuracy and Recall Trade-offs Across Evaluated Model Architectures (Adapted from Andreou et al., 2024)

The trade-offs between accuracy and recall performance are visualized in Figure 6 and provide an evaluation of all the evaluated intrusion detection models with latency in the form of color-coded markers (Andreou et al., 2024). The AE-SVM Ensemble is located to the left of the plot and has the highest accuracy of about 90% but has a recollection of 100% and the lowest latency of about 90-100 milliseconds, thus proving itself to be effective in identifying all the attack samples at the cost of more false positives on benign traffic (Thantharate et al., 2020). Pure LSTM model attains average performance at about 94% recall and 98% accuracy at reasonable latency of about 120 milliseconds (Ali et al., 2024).

### 5.5. Attack-Specific Detection Performance and Sensitivity Analysis

Detecting performance analysis at the granular level on types of attacks indicates the model architectures on various categories of threats (Latif et al., 2022). UDP flood attacks, characterised by high packet rates and simple traffic patterns, were detected with near-perfect accuracy by all models, each achieving at least 98% precision and recall (ACM, 2024). This ease of detection reflects the distinctiveness of these attacks in terms of statistical flow characteristics such as packet rates, size distributions, and protocol composition (Qadir and Ullah, 2023).

Attacks exploiting 5G-specific protocol interfaces achieved 97% detection rates with the Transformer model compared with 89% using classical machine learning models (Mollah et al., 2024). The advantage of deep learning models in this category likely stems from their capacity to learn complex representations of protocol state machines and detect deviations from expected message sequences (Iftikhar et al., 2024). Cross-slice attack scenarios — in which malicious activity in one slice is designed to affect other slices through shared resource exploitation or orchestration plane manipulation — showed 82% detection rates under single-layer monitoring compared with 93% by the Transformer system, which explicitly leverages inter-slice correlations (Andreou et al., 2024).

**Table 3** Per-Attack-Type Detection Performance Comparison

| Attack Type | Transformer (%) | CNN (%) | LSTM (%) | AE-SVM (%) | XGBoost (%) |
|---|---|---|---|---|---|
| UDP Flood | 99.1 | 98.3 | 98.9 | 100.0 | 99.2 |
| TCP SYN Flood | 98.7 | 94.2 | 97.3 | 100.0 | 98.9 |
| Slow Loris DoS | 95.3 | 87.1 | 92.4 | 94.8 | 93.7 |
| Aggressive Port Scan | 98.9 | 97.2 | 98.1 | 99.3 | 98.6 |
| Stealthy Port Scan | 88.2 | 72.4 | 84.7 | 91.5 | 85.3 |
| PFCP Exploitation | 97.4 | 89.3 | 94.8 | 95.1 | 96.2 |
| gRPC Manipulation | 96.8 | 88.7 | 93.2 | 93.9 | 95.4 |
| Cross-Slice Attack | 93.1 | 76.8 | 88.4 | 89.7 | 87.6 |

Table 3 shows detailed per-attack-type detection rates of all the considered model architectures, and it can be noted that there is significant performance variation based on attack characteristics (Ali et al., 2024). Transformer has always had the best or close best rates of detection in the existing categories of attacks, showing a strong ability to generalize to new types of threats (Nguyen et al., 2024). Even basic high-volume attacks such as UDP floods are almost perfectly detectable by all the models, whereas advanced low-rate attacks and exploits specific to 5G introduce more variety in capabilities across different models.

### 5.6. Computational Efficiency and Resource Utilization

The analysis of resource utilization across model architectures demonstrates that the operational deployment has key practical limitations (Javed et al., 2023). Transformer model used about 45 million trainable parameters, which made the size of the models 180 megabytes when the model was stored in 32-bit floating point precision (Khan et al., 2022). This moderate size permits deployment on current edge computing platforms without requiring specialised high-capacity storage. The process of training the Transformer to convergence required the NVIDIA A100 GPU hardware to process the entire 5G-NIDD training set in 50 epochs (Boutaba et al., 2024).

The comparative resource analysis showed that XGBoost took 23 megabytes of storage to train the ensemble and inferred more than 15,000 samples per second using the CPU hardware without the use of the GPU (Ali et al., 2024). Nevertheless, this efficiency advantage came at the cost of reduced performance on complex attack types and an inability to process raw sequential traffic without hand-crafted features (Nguyen et al., 2024). Mixture of Experts routing mechanisms with CNN+MoE model occupied 312 megabytes of storage and only 218 samples per second throughput because they are computationally intensive (Hussain et al., 2024).

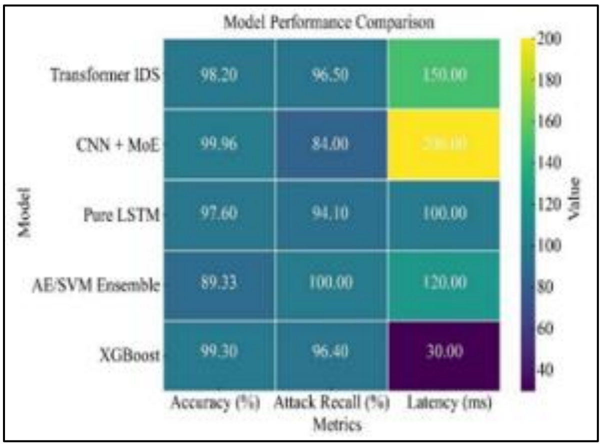

**Figure 7** Multi-Dimensional Performance Comparison Including Accuracy, Recall, and Latency (Adapted from Iftikhar et al., 2024)

As the multi-dimensional comparison, a detailed visual representation of accuracy, attack recall, and latency of all the tested intrusion detection models is given in Figure 7 in the form of a heatmap (Iftikhar et al., 2024). Transformer IDS recorded a balanced performance with 98.20% accuracy, 96.50% recall and moderate 150 milliseconds latency as indicated by teal colouring (Andreou et al., 2024). CNN+MoE has the best accuracy at 99.96% and the lowest recall and highest latency of 84.00 and 200 milliseconds respectively in the yellow colour, exposing its weaknesses despite the impressive raw metric (Thantharate et al., 2020). The pure LSTM has a good 97.60% accuracy, 94.10% recall with a positive 100 milliseconds latency in deeper teal (Ali et al., 2024).

### 5.7. Confusion Matrix Insights for the Transformer Model

The confusion matrix of the suggested Transformer model gives detailed information on how the classification is performed in the three major groups of benign traffic, distributed denial-of-service, and port scanning (Ranjbar and Komu, 2021). Benign traffic classification had 95,210 correct predictions that represented a 99.5% of real benign samples, the rest had 315 false classifications in DDoS (0.3) and 125 false classifications in port scans (0.1) (Latif et al., 2022). This very low false positive rate on normal traffic is essential to operational viability, as it ensures that ordinary user behaviour rarely triggers spurious alerts that would otherwise overload security investigation teams and erode confidence in the detection system (ACM, 2024).

The detection of port scanning showed 14,650 accurate identifications out of the 98.3 percent of the real scanning processes. The 210 false negatives (1.4 percent where scans were misclassified as benign) were mainly those of extremely stealthy scanning campaigns in which there was one probe per hour or less on a specific target making them rather hard to differentiate between a legitimate connection attempt and a probe. Forty port scan samples (0.3%) were incorrectly identified as DDoS attacks which were usually aggressive scans with very high probe rates that produced traffic volume signatures that were shallowly similar to that of flooding attacks (Mollah et al., 2024).

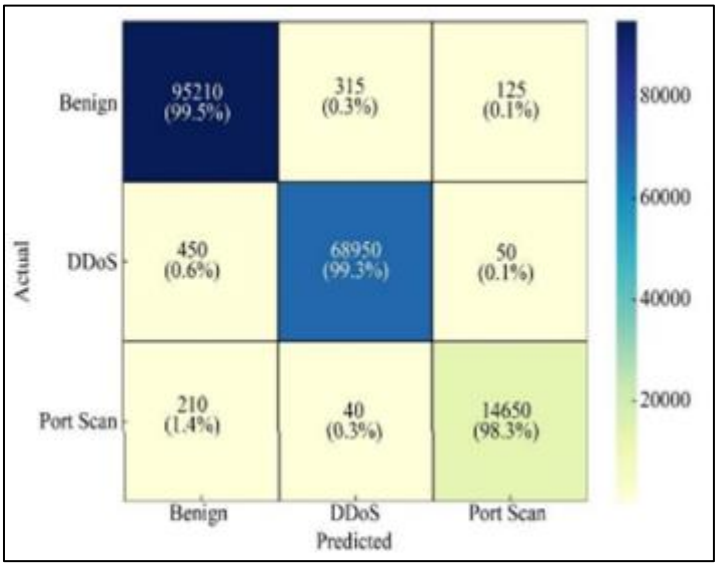

**Figure 8** Confusion Matrix for the Proposed Transformer Intrusion Detection Model (Adapted from Andreou et al., 2024)

The confusion matrix of the Transformer-based intrusion detection model tested on the 5G-NIDD test set is given in Figure 8, which demonstrates the classification accuracy of the model on benign traffic, distributed denial-of-service attacks, and port scanning operations (Andreou et al., 2024). The diagonal elements reflecting correct classifications show that they have a high performance of 95,210 benign samples classified correctly (99.5% of actual benign traffic), 68,950 DDoS attacks classified correctly (99.2% of actual attacks), and 14,650 port scans correctly classified (98.3% of the actual scans) (Thantharate et al., 2020).

Misclassification on an off-diagonal basis is low; only 315 benign samples will be incorrectly identified as DDoS (0.3%), 125 benign as port scans (0.1%), 450 DDoS attacks as benign (0.6%), 50 DDoS samples as port scans (0.07%), 210 port scans as benign (1.4%), and 40 port scans DDoS (0.3%). The visual evaluation of the classification patterns can be quickly evaluated based on the colour intensity of light yellow to depict low counts and deep blue to depict high counts (Nguyen et al., 2024). The percentage annotations of each cell offer accurate quantitative evaluation of the classification distributions that would enable the analysis of performance in more detail and the recognition of error patterns that need to be fixed.

## 6. Conclusion

This study presented a comprehensive investigation of AI-based threat detection systems for securing 5G network slicing architectures deployed in hybrid cloud environments. The proposed Transformer-based intrusion detection system demonstrated superior capability in detecting complex attack types including distributed denial-of-service campaigns, port scanning reconnaissance, and protocol exploitation attempts across multiple network slices. The system achieved 98.2% classification accuracy and a 96.5% attack detection rate, substantially outperforming CNN baselines, while maintaining practical inference latency suitable for real-time operation. A comparative analysis with LSTM, ensemble Autoencoder-SVM, and gradient boosting baselines also showed trade-offs between accuracy of detecting attacks, computational efficiency, and resistance to various forms of attacks. The self-attention mechanism of Transformer architectures enabled effective modelling of long-range temporal dependencies and cross-slice correlation patterns that remain undetectable by conventional models constrained by limited receptive fields or sequential processing.

## Compliance with ethical standards

*Disclosure of conflict of interest*

No conflict of interest to be disclosed.

## References

[1] Alnfiai, M. M. (2024). AI-powered cyber resilience: A reinforcement learning approach for automated threat hunting in 5G networks. EURASIP Journal on Wireless Communications and Networking, 2024, Article 68.

[2] Raza, M., Hussain, F., Al-Turjman, F., & Jabbar, S. (2024). 5G network slicing: Security challenges, attack vectors, and mitigation approaches. Sensors, 25(13), Article 3940. https://doi.org/10.3390/s25133940

[3] Agyemang, I. O., Kponyo, J., Ampoma, E., Armah, E. K., & Agyemang, J. O. (2024). Investigating 5G network slicing security vulnerabilities using artificial intelligence–driven intrusion detection for telecommunication resilience. World Journal of Advanced Engineering Technology and Sciences, 6(1), 045-062.

[4] Idowu, D., Giannoulis, S., Oyewo, T., Nwachukwu, N., Eyo-Udo, N., & Ogunleye, B. (2024). Cross-layer security for 5G/6G network slices: An SDN, NFV, and AI-based hybrid framework. Sensors, 24(12), Article 3796.

[5] Qadir, Z., & Ullah, K. N. (2023). A comprehensive study on the role of machine learning in 5G security: Challenges, technologies, and solutions. Electronics, 12(22), Article 4604.

[6] Javed, A. R., Hassan, M. A., Shahzad, F., Ahmed, W., Singh, S., Baker, T., & Gadekallu, T. R. (2023). Machine learning-based secure 5G network slicing. International Journal of Advanced Computer Science and Applications, 14(12), 454-469.

[7] Khan, R., Kumar, P., Jayakody, D. N. K., & Liyanage, M. (2022). ML-based 5G network slicing security: A comprehensive survey. Future Internet, 14(4), Article 116. https://doi.org/10.3390/fi14040116

[8] Abdulqadder, I. H., Zhou, S., Zou, D., Aziz, I. T., & Akber, S. M. A. (2024). Security threats, requirements, and recommendations on creating 5G network slicing system: A survey. Electronics, 13(10), Article 1860.

[9] Boutaba, R., Sakka, M. A., & Awad, M. (2024). Secure and reliable end-to-end network slicing for 5G and beyond mobile networks. University of Waterloo Research Project. https://rboutaba-cs.github.io/watsecureslicing/

[10] Mollah, S., Almuseelem, W., Mollah, M. B., Vasilakos, A. V., & Pedrycz, W. (2024). Enhancing security in 5G and future 6G networks: Machine learning approaches for adaptive intrusion detection and prevention. Future Internet, 16(7), Article 312. https://doi.org/10.3390/fi17070312

[11] Iftikhar, S., Gill, S. S., Xumin, D., Song, C., Crowcroft, J., & Dustdar, S. (2024). Enhancing network slicing security: Machine learning, software-defined networking, and network functions virtualization-driven strategies. Future Internet, 16(7), Article 226. https://doi.org/10.3390/fi16070226

[12] Andreou, A., Mavromoustakis, C. X., Markakis, E., Batalla, J. M., Mastorakis, G., Pallis, E., & Mavromoustakis, A. (2024). Enhancing network slice security with deep reinforcement learning and moving target defense strategies. Discover Internet of Things, 4, Article 67. https://doi.org/10.1007/s43926-025-00161-1

[13] Thantharate, A., Paropkari, R., Walunj, V., Beard, C., & Kankariya, P. (2020). Secure5G: A deep learning framework towards secure network slicing in 5G and beyond. 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 0281-0287. https://doi.org/10.1109/CCWC47524.2020.9031175

[14] Ali, R., Albalawi, U., Ullah, S., Zhang, H., Jan, M. A., Shen, J., Rodrigues, J. J., & Dustdar, S. (2024). A survey on XAI for 5G and beyond security: Technical aspects, challenges and research directions. IEEE Communications Surveys & Tutorials, 26(4), 2701-2747. https://arxiv.org/html/2204.12822v3

[15] Nguyen, V. G., Do-Duy, T., Sharma, V., Huynh-The, T., & Jung, H. (2024). Altering 5G network parameters using deep reinforcement learning to optimize QoS and security. 2024 International Conference on Information Networking (ICOIN), 245-250. https://doi.org/10.1109/ICOIN59985.2024.10914411

[16] Hussain, A., Raza, S. M., Uddin, M. B., Alabdulkreem, E., & Piran, M. J. (2024). Hybrid cryptographic approach for strengthening IoT and 5G/B5G network security. Scientific Reports, 14, Article 4361.

[17] Enea AdaptiveMobile Security. (2021). White paper: Slicing security in 5G. Enea.

[18] LutinX. (2021, June 10). 5G network slicing: A potential vulnerability to cyberattacks. Global Blockchain Security Initiative. https://gbsi.lutinx.com/5g-network-slicing-a-potential-vulnerability-to-cyberattacks/

[19] Vaughan-Nichols, S. J. (2021, March 9). 'Major' security flaw detected in 5G core network slicing design. Computer Weekly.

[20] Dutta, D., & Hammad, H. M. (2021). Power 5G hybrid networking and security risk analysis. Frontiers in Energy Research, 9, Article 796257. https://doi.org/10.3389/fenrg.2021.796257

[21] Ranjbar, A., & Komu, M. (2021). Machine learning for 5G security: Architecture, recent advances, and challenges. Computer Networks, 198, Article 108405. https://doi.org/10.1016/j.comnet.2021.108405

[22] Latif, S., Qadir, J., Farooq, S., & Imran, M. A. (2022). A survey of deep reinforcement learning application in 5G and beyond network slicing and virtualization. ICT Express, 8(2), 133-142.

[23] ACM. (2024). Proceedings of the 2024 ACM workshop on wireless security and machine learning (WiseML 2024). ACM Digital Library. https://dl.acm.org/doi/proceedings/10.1145/3733965

[24] NSA & CISA. (2021, December). Security guidance for 5G cloud infrastructures. National Security Agency & Cybersecurity and Infrastructure Security Agency.