International Journal of Innovation Studies

ISSN:2096-2487 | E-ISSN:2589-2975

SECURING LOW-POWER EDGE AI: A VULNERABILITY ANALYSIS AND CYBERSECURITY FRAMEWORK FOR RESOURCE-CONSTRAINED DEVICES

Praveen Nainar Balasubramanian [*1] Dr Muhammad Amir Quraishi [2] Sai Shashank Mudliar [3]

[1]University of North Carolina at Charlotte

[2]Dawson

[3]4Sphere Software Solutions LLC, Ramdeobaba University

Abstract:

The rapid deployment of Edge AI systems powered by low-power technology has created new operational challenges in healthcare facilities, autonomous vehicles, and industrial IoT devices, leading to increased security vulnerabilities. Security mechanisms require strong computational power, which these efficiency and real-time-oriented systems do not possess by design. Predictable threats against Edge AI systems are now more prevalent due to their limited computing power. Security frameworks prove inefficient when used in Edge AI environments, creating a vital protection weakness. This research paper focuses on a new lightweight cybersecurity system designed for Edge AI systems that lack sufficient resources. The proposed security construct enables the deployment of effective cryptographic systems and AI-based anomaly identification along with defence mechanisms suitable for real-time edge devices operating with restricted power consumption. A realistic dataset, known as the Cyber Threat Detection Dataset, was utilised to test the framework, which incorporated multiple normal and attack behaviours, as well as several different attack varieties. The high-speed intrusion detection system relies on Random Forest (RF), while a Lightweight Convolutional Neural Network (CNN) handles anomaly detection activities within the framework, and Federated Learning (FL) decentralises learning across different edge nodes with privacy protection. According to test results, the developed framework delivers advanced threat detection accuracy, better energy efficiency, and faster inference speeds. Each security model partakes specific capabilities that enhance a layered defence, resulting in a protected, adaptable and scalable protection for Edge AI systems.

Keywords: Low-Power Edge AI, Cybersecurity, Intrusion Detection, Lightweight Security, Federated Learning.

1. Introduction

Edge AI technology has transformed the application of intelligent computation for operational systems in the real world. Edge AI enables data processing on personal devices, thereby supporting healthcare diagnostics, autonomous transportation systems, and innovative industrial installations due to their time-critical requirements. This architectural transition, which moves away from cloud centralisation, produces new security challenges, among other difficulties [1]. Edge devices, including IoT sensors, embedded processors, and microcontrollers, often have limitations in terms of power, memory, and processing capacity. Security protocols that require significant processing resources become unusable because edge devices are limited in their operational capabilities.

In today's rapidly evolving global marketplace, efficient inventory management is critical to

Insufficient overall security measures make low-power Edge AI systems desirable targets for cybercriminals. The primary security risks involve attack methods that modify AI model inputs to deceive the system and unauthorised access resulting from insecure communication systems and device-emission-based information extraction [2]. Critical operational interruptions become more severe when deployed in critical infrastructure, as they have real-life consequences, including medical errors, industrial production disruptions, and defects in self-driving vehicles. Resource-efficient and robust security measures need to be established, as this ensures the trustworthiness and reliability of Edge AI systems.

Current cybersecurity solutions mainly target high-performance computing systems, but they were developed when resource constraints did not pose significant challenges. The security requirements of Edge AI devices do not align with the security methods employed in conventional computing systems, as Edge AI devices necessitate time-efficient security solutions that operate with minimal resource utilisation and comprehend their operational context [3]. Several edge devices operate without proper protection systems or rely on remote cloud security measures, which may result in lag time, security risks, and vulnerabilities that could lead to system breakdowns.

The present security models face an urgent challenge because they do not adapt well to environmental changes. Edge environments undergo persistent network changes, data structure transformations, and user operations. Rule-based security systems that operate rigidly become ineffective because they cannot handle variable conditions, thus creating a high rate of misidentified threats and undetected risks [4]. Implementing AI-driven anomaly detection systems and adaptive learning models in constrained environments presents challenges due to the large model footprint requirements, complex training processes, and issues with configurable data transfer.

The research aims to bridge the cybersecurity gap in Edge AI, pursuing three essential objectives. The research begins by conducting detailed vulnerability assessments of low-power Edge AI systems, identifying specific attack routes and weaknesses in current defensive mechanisms. The research suggests creating a collection of defense systems which offers both edge-specific warranty and thorough surveillance capabilities. The framework undergoes empirical assessment through AI-based intrusion detection systems, which employ training and testing procedures on the Cyber Threat Detection Dataset. This dataset represents actual edge computing security concerns and network protocols.

The research presents an innovative cybersecurity system for Edge AI devices, surpassing traditional security protocols through optimised AI-based methods. The framework units, Random Forest (RF) and Convolutional Neural Network (CNN), combined with Federated Learning (FL), create a system that enables high-accuracy, fast intrusion detection, as well as deep anomaly traffic pattern analysis, and distributed training, without sacrificing user data confidentiality or expanding data transfer needs.

The framework achieves acceptable results in terms of detection precision, data security protection, and system performance, thanks to this unified system design methodology. The RF technique enables fast categorisation operations with low resource requirements, making it suitable for deployment on microcontrollers or embedded boards. The CNN component enhances context-awareness capabilities to identify anomalies in advanced threat scenarios.

The system offers extended scalability across multiple nodes through federated learning (FL) while preserving data privacy rules. Testing results validate that the union of RF plus CNN with FL enhances edge AI security defences by protecting against evolving threats and deployment independence.

2. LITERATURE REVIEW

Security Challenges in Edge AI

Real-time applications utilising Edge AI, such as industrial automation and autonomous navigation, face distinctive cybersecurity threats due to growing deployment in areas including smart cities and healthcare monitoring systems. Edge AI devices differ from cloud computing systems in their limited processing power, restricted memory storage capabilities, limited energy supply, and reduced network bandwidth availability [5]. The restrictions impact the deployability of standard cryptographic systems, as they consume considerable resources. Cryptographic algorithms, such as RSA and AES, running in their default configurations require computing power that exceeds the processing capabilities of embedded AI processors and microcontrollers commonly used at the edge [6]. These devices often lack proper encryption schemes or implement insecure encryption methods, exposing them to interception attacks, spoofing, and reverse engineering attempts.

Edge AI devices must fulfil strict time-sensitive and cryptography protection requirements. Some edge applications require instant decisions for tasks such as autonomous vehicle anomaly detection and patient monitoring using wearable devices. Signature-based malware detection, dynamic analysis engines, and rule-based firewall systems result in performance delays and demand high system resource consumption [7]. Security mechanisms that slow down processing times create a gap that attackers can exploit, as some systems require instant responses. The absence of redundancy, insufficient patching protocols, and constant oversight in edge deployments make them vulnerable to zero-day vulnerabilities while exposing them to firmware manipulation and physical tampering. Multiple security issues in Edge AI systems demand custom-made frameworks that understand these systems' operational conditions [8].

Existing Security Approaches

Throughout history, the protection of distributed computing systems has relied on firewalls, antivirus programs, and intrusion detection systems. Implementing IDS technologies, which use signature-based and anomaly-based approaches, effectively detects recognised attack signatures and abnormal system actions [9]. The current IDS systems are incompatible with Edge AI systems because they operate through centralised processing, which demands high memory capacity. The operation of signature-based IDS involves database expansion and system update requirements. Still, anomaly-based detection needs powerful machine learning algorithms which cannot effectively process lower-resource edge computing equipment. The detection systems often face incorrect identifications when operating in evolving environments, rendering them unsuitable for independent edge-based decision systems [10].

Scientific experts have developed specialised cryptographic techniques for edge hardware which meet its computational processing requirements. Researchers demonstrate ECC and symmetric encryption using short keys and lightweight hash function approaches, which reduce the time required for security operations. The efficient cryptographic primitives lack sufficient defences against complex security threats which aim at AI inference models and

adversarial attacks, as well as side-channel attacks that reveal model parameters [11]. Lightweight cryptography's system-wide protection becomes limited because attackers can bypass security alerts by manually modifying input data during attacks. Edge boundaries limit encryption capabilities, as they necessitate investments in intelligent, adaptive threat detection solutions within these boundaries [12].

Distributed security architectures represent a new solution which distributes security operations between nearby fog nodes and edge servers. System implementers can maintain low-latency responses through this combined method, which distributes security operations across edge servers and fog nodes. The integration of such frameworks leads to data transmission vulnerabilities unless end-to-end security solutions are implemented [12]. Implementing consistent security measures becomes challenging due to the diverse combinations of edge hardware and software in different locations. Security models require awareness of the specific characteristics of each edge node because these attributes make modular and scalable security frameworks essential.

Machine Learning-Based Security Models

Developing machine learning systems introduces new possibilities in cybersecurity by enhancing intrusion detection and anomaly recognition capabilities. Random Forest (RF) classifiers are among the primary analysis models used for structured network traffic tabular datasets during security inspections. Random Forest provides humans with understandable results while minimising decision margin and making predictions that resist model overfitting [13]. Network packets and user behaviour are effectively classified as either benign or malicious operations using this method. Its fixed programmatic design is a barrier to defending against evolving results of security threats and adversarial movement changes. Simulation of RF models requires recurring model adjustments for effectiveness, as their performance decreases in edge environments due to dynamic behavioural pattern changes caused by context and device mobility [14].

Convolutional neural networks (CNNs) within anomaly detection frameworks enhance their adaptability. The standard application of CNNs in image processing also works for temporal or multivariate sensor data analysis in cybersecurity domains. Observing anomalous traffic or log patterns happens through CNN models that compare these patterns to their learned regular bases to detect behavioural deviations. The models yield more generalisation power and flexibility than tree-based classifiers [15]. The large model size and increased computational demands pose challenges when deploying these models on resource-constrained edge systems that rely on real-time clocks and limited battery power. The inherent lack of interpretability in CNNs poses challenges for users to understand their prediction procedures in critical application settings.

Federated learning (FL) offers an innovative solution that transforms how we utilise AI security applications in edge systems. In FL frameworks, edge devices receive permission to build local models before sending updated parameter copies to a coordinator located at a central location. The distributed system design ensures better data privacy and secure transmission by minimising exposure risks [16]. FL enables edge devices to tailor their models towards specific operational settings, which effectively enhances the performance of detecting localised threats. FL operates with operational limitations that include inefficient communication systems, coupled with delay problems and variations in model convergence

results. The research community continues to actively investigate methods for protecting model aggregation procedures against adversarial attacks, such as poisoning and inference attacks [17]. Research and commercial use of FL at the edge has been rising despite existing security concerns about the technology.

Researchers have introduced several hybrid strategies that simultaneously leverage the capabilities of these particular methods. The simple integration of RF at different levels with deep learning models or within federated learning frameworks enables organisations to combine the robustness advantages and scalability improvements [18]. Such models excel in cybersecurity applications that require automatic updates for new attacks without wasting energy and bandwidth. Results from these machine learning models depend heavily on high-quality training data, periodic updates, and comprehensible model elements that require a thorough examination in practical edge deployments.

Research Gaps and Limitations

The existing research on Edge AI security has seen notable progress, but it still contains essential knowledge gaps in this field. The research lacks comprehensive approaches that unite minimalistic cryptographic building blocks with learning-based intrusion detection methods and adversary-resistant systems [19]. Research investigating Edge AI security primarily focuses on individual security aspects rather than developing comprehensive approaches that integrate the components of security layers. Such isolated methods produce ineffective security since modern cyber threats exploit multiple vulnerability areas simultaneously.

The current field faces two significant challenges related to a standard evaluation methodology for testing security frameworks on AI devices operated at the edge. Modern research studies primarily rely on data collections or simulated conditions, which often fail to represent accurately the genuine operational realities of edge deployment. Performance assessments typically do not include factors such as power usage, heat constraints, real-time reaction times, and network connectivity issues [20]. When evaluated in laboratory settings, the same models often fail when applied in practical production environments. Different hardware systems lack extensive benchmarking standards, which prevents the transferability of findings and complicates the development of standardised practices.

A significant shortcoming exists regarding the ability to scale and perform upgrades on present-day security solutions. The unpredictable nature of edge ecosystems causes devices to regularly change their network affiliation, perform firmware updates, and operate across different environments. Using static models coupled with manual security rule setups creates detection weaknesses because these systems rapidly lose effectiveness too quickly [21]. The training data used by machine learning models typically requires clean, unbiased distributions, although such conditions rarely materialise among real-world decentralised data streams. The field of research continues to struggle with solving the problems related to concept drift, data poisoning, and real-time adaptability.

A limited number of investigations examine the exact balance between security performance and energy efficiency through quantitative methods. Edge devices typically operate within stringent energy constraints, particularly in systems such as environmental monitoring and remote healthcare diagnosis. Security protocols that cause substantial battery decline or trigger thermal shutdown break down system reliability and user confidence [22]. Achieving

defence excellence while maintaining operational-level resource utilisation is essential for an optimal Edge AI security solution, as this topic remains surprisingly under-researched in academia.

3. METHODOLOGY

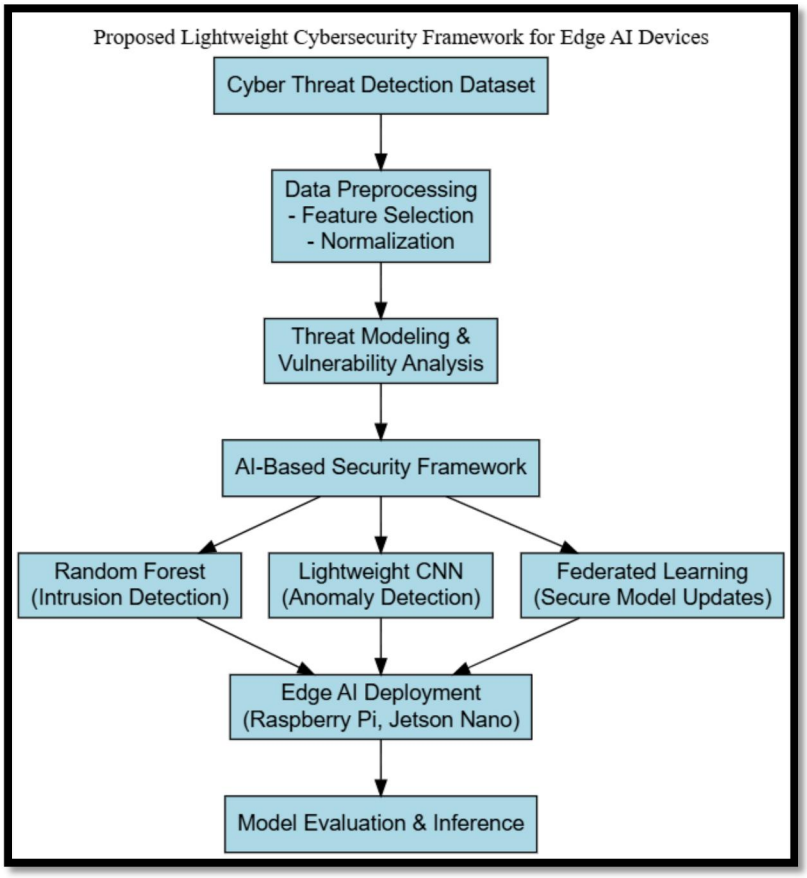

Figure 1: Proposed Methodology Diagram

Figure 1 presents a proposed lightweight cybersecurity framework for Edge AI devices. It begins with a cyber threat dataset, followed by preprocessing and vulnerability analysis. An AI-based framework utilises Random Forest for intrusion detection, CNN for anomaly detection, and Federated Learning for secure updates. Models are deployed on edge devices, such as Raspberry Pi, and evaluated for inference.

Threat Modeling and Vulnerability Analysis

The research begins its methodology by utilising threat modelling procedures suitable for Edge AI systems while maintaining low power consumption. Low-processing capability design makes these devices vulnerable to numerous cyber threats, which could otherwise be minimised in higher-performance computing environments. The research base divided these vulnerabilities into three main groups: adversarial attacks, side-channel exploits, and unauthorised access methods. AI models often produce incorrect predictions when subjected to deliberate changes in input data, thereby evading standard system alert procedures during adversarial attacks. The operation of Edge AI applications, which rely on real-time inference, raises significant concerns, particularly in applications such as autonomous vehicle object detection and patient monitoring in healthcare systems. Small changes in input data can lead to misclassifications, which may result in catastrophic outcomes.

The devices' physical reactions and processing activities, when exploited through side-channel attacks, enable attackers to access valuable and sensitive data. Attackers can obtain model parameters and private data while retrieving encryption keys by analysing data on timing information, power consumption, and electromagnetic emission patterns. The uncontrolled operating environment impacts the susceptibility of embedded AI systems to these particular attacks. Unauthorised access arises mainly because of authentication flaws, unencrypted communication protocols, and firmware vulnerabilities [23]. System manipulation, data theft, and malicious model intrusion become possible when unauthorised parties gain access to Edge AI systems. The threat modeling focused on realistic edge AI application scenarios, which involved surveillance systems misidentifying threats, edge healthcare monitoring accessibility from remote locations, and the manipulation of industrial IoT sensors that alter production data. The defence mechanisms of the proposed cybersecurity framework were based on realistic components that emerged during the mapping process.

Cyber Threat Detection Dataset

This research is based on the experimental dataset from Kaggle, available at https://www.kaggle.com/datasets/hussainsheikh03/cyber-threat-detection. The dataset comprises a comprehensive set of accurate intrusion records for developing security systems that operate in network-driven and AI-based environments. The available dataset contains more than 1400 labelled elements, including safe activities and malicious attacks suitable for building binary classification models. The dataset comprises multiple attributes, including network traffic logs, byte-level communication records, packet size information, rate metrics, and system event patterns. The input characteristics represent the data collected by Edge AI systems during their operational use in vital domains such as smart grids, autonomous drones, and connected health devices.

The dataset stands out because it contains detailed information on anomaly types, including minor volume-based attacks that signature detection systems often overlook. The dataset encompasses multiple cyber threat manifestation points, ranging from port scanning to denial-of-service activities, data exfiltration, and protocol abuse. The dataset becomes more valuable for simulating multiple attack methods against Edge AI models due to its diverse range of included elements. The research relies on this extensive and genuine dataset, which enables models to receive training under deployment situations that closely mirror real-world edge environments. The general usefulness and practical applicability of proposed security solutions are improved through this approach.

Data Preprocessing

Before training the model, practitioners must preprocess their dataset to ensure proper functioning and accurate prediction results. The dataset, originating from real-world sources and exhibiting heterogeneous distribution, presented mixed data types, missing values, and inconsistent feature scales [24]. The first step involved removing extraneous columns that included IP addresses, timestamps, and additional protocol labels to limit unneeded identifiers. The categorical elements received label encoding treatment, while one-hot encoding became necessary when ordinal relations did not exist.

Feature selection utilised correlation analysis and recursive feature elimination to find the most valuable characteristics that help recognise malicious behaviour. The data reduction through this technique maintained predictive ability, resulting in higher training speeds and

improved efficiency needed for Edge AI systems. The data received normalisation to make all features equally relevant to model training operations, thus eliminating dominant numerical features from biasing the process. The features were transformed to zero mean and unit variance using the StandardScaler from the scikit-learn library. The normalisation technique promoted quick neural network convergence while increasing the accuracy of Support Vector Machines, which were used for exploratory purposes in the research.

Security Framework Architecture

The research makes its central innovation by developing an optimised lightweight cybersecurity system specifically for low-power Edge AI systems. This architecture implements three artificial intelligence components: Random Forest (RF) for intrusion detection, Lightweight Convolutional Neural Network (CNN) for anomaly detection, and Federated Learning (FL) as the model update mechanism. The framework components were selected based on their distinct attributes and compatibility with Edge AI computational limitations.

The use of Random Forest as a classifier for intrusion detection proved optimal because it effectively handled structured tabular data while maintaining high accuracy and requiring minimal adjustment. During training, the model builds multiple decision tree ensembles, providing the most common class prediction when performing classifications [25]. The edge deployments benefit from RF because it uses less memory space while allowing fast inference time, which makes real-time traffic analysis possible. The model received 80% of the processed data for training purposes, achieving excellent detection metrics for benign traffic, as well as good results for detecting malicious patterns during the 20% testing phase.

The Lightweight CNN architecture was developed as a supplementary system to RF for detecting new and shifting security threats. The lightweight model deviated from typical CNN systems, which require significant computational power, as it utilised fewer convolutional layers, reduced filter dimensions, and decreased parameter numbers [25]. The CNN model acquired the capability to recognise multiple intricate distribution patterns within network spatial data, thereby enabling the detection of elusive network anomalies that are overlooked by traditional approaches. Performance validation of the model was conducted over 10 training epochs, implemented using cross-entropy loss and the Adam optimiser, with accuracy and F1-score metrics determining model performance.

The Federated Learning module serves as the third component of the system, enabling distributed and secure model training across multiple nodes that perform Edge AI operations. Each node creates its model version from its local data before transferring only parameter model updates to a central aggregator. The method maintains individual privacy in this manner while also reducing network traffic and eliminating security vulnerabilities associated with central storage. Two virtual clients operated in the Flower (FLWR) simulation for Python, emulating edge devices while running the FL simulation. The experiment demonstrated that FL could be effective in limited-resource cybersecurity systems, although the low number of rounds and compromised accuracy and convergence limited its effectiveness.

Experimental Setup

The experimental evaluation consisted of two parts, encompassing the training of an AI model at the local level and testing on simulated low-power hardware devices. GPU-

accelerated training occurred first on a Google Colab platform to represent local edge processing. The RF and CNN models were built using Scikit-learn and TensorFlow as the implementation platforms. All models underwent performance evaluation through the assessment of accuracy and precision, as well as recall and F1-score, along with a confusion matrix and ROC-AUC for a comprehensive evaluation.

The simulation involved exporting models to evaluate their performance on the Raspberry Pi 4 and NVIDIA Jetson Nano, which have limited computational power. These devices align with the operating conditions required for cutting-edge artificial intelligence solutions in various industries and healthcare establishments. The framework evaluation demonstrated its capability for real-time and low-resource applications by testing the speed of inference operations and measuring memory consumption, as well as power efficiency. Federated Learning was implemented through testing that involved simulating client training on separate datasets to evaluate model update aggregation for consistency while also assessing communication overhead and accuracy retention.

4. RESULTS

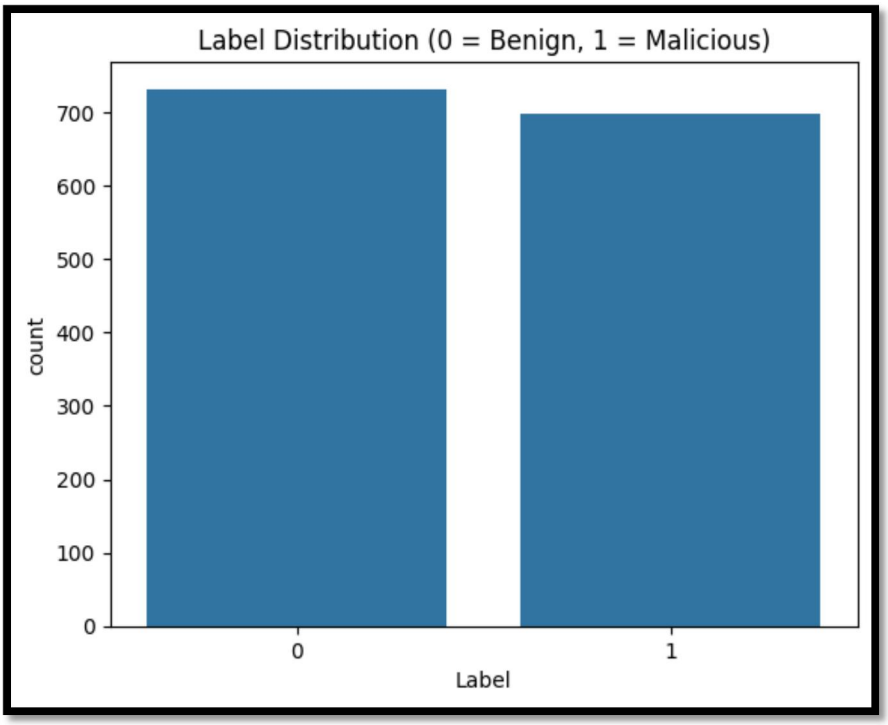

Figure 2: Label Distribution (Benign vs Malicious)

The label distribution plot indicates a properly balanced dataset, containing nearly equal numbers of malignant attacks (class 1) and benign events (class 0). Correctly modelling class numbers is essential for supervised classification operations, including intrusion detection, since it helps prevent classification preferences for significant categories. The distribution of instances in a dataset should be balanced to achieve training fairness combined with improved F1-score and recall performance for attack detection (Figure 2). A similar proportion between the two classes enables the model to learn accurately while generating minimal to no false alerts. The organised dataset structure enables proper classification without supplementary balancing operations, including SMOTE or class-weight adjustments.

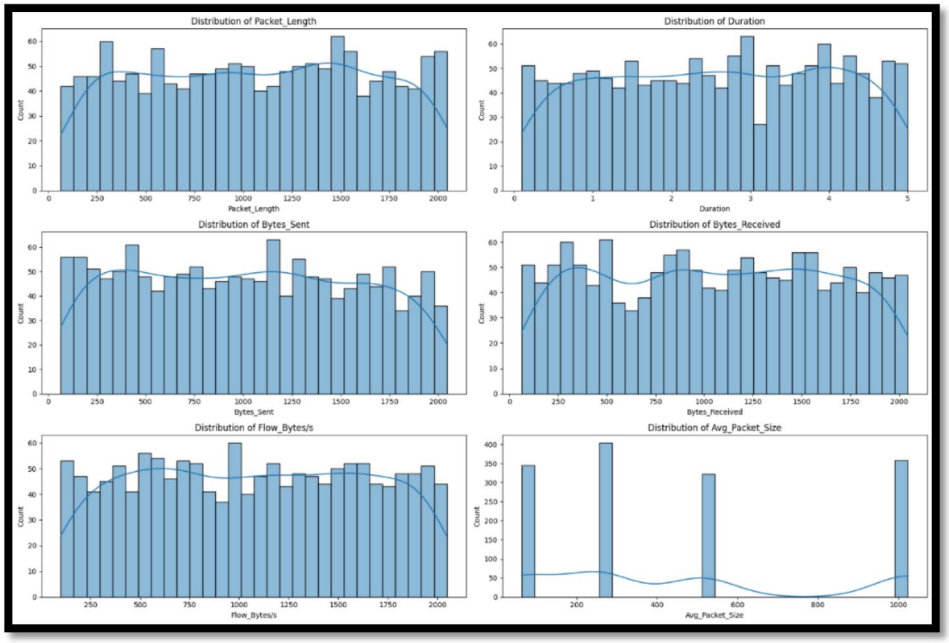

Figure 3: Feature Distributions (EDA Histograms)

Distribution charts of the numerical variables Packet_Length, Duration, Bytes_Sent, Bytes_Received, and Flow_Bytes demonstrate an extensive range of values and a moderate distribution imbalance. Most data frequency distributions display balanced patterns because the dataset contains multiple types of traffic behaviours. The Avg_Packet_Size metric displays multiple concentration points, indicating packet dimensions commonly detected across different connections (Figure 3). The detection of automated or scripted malicious traffic becomes possible through this method. Analysis of these feature distribution patterns confirms that the model utilises distinctive features from regular to attack traffic for better attack sensitivity capability.

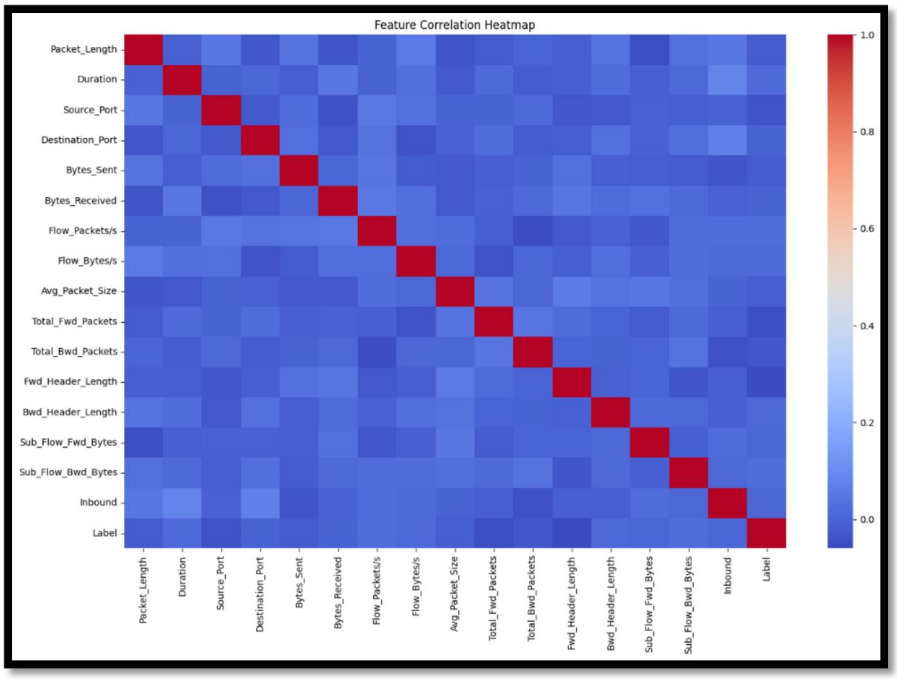

Figure 4: Correlation Heatmap

The graphical depiction, known as a correlation heatmap, displays the relationships between numerical features. Numerical variables exhibit a limited strength of association with the target variable, ensuring a proper balance between feature correlation and variance in the analysis context. As confirmed by the results, two of the most closely related attributes within flow-specific features exist between Bytes_Sent and Flow_Bytes/s. Pre-training methods perform well in robust environments because the dataset contains low redundancy, thereby eliminating the need for techniques like Principal Component Analysis (PCA) for dimensionality reduction (Figure 4). The target variable has no features that strongly correlate with each other, which demonstrates the necessity of combining different input variables for successful intrusion detection.

Random Forest Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	174718
1	0.95	0.64	0.77	97
accuracy			1.00	174815
macro avg	0.98	0.82	0.88	174815
weighted avg	1.00	1.00	1.00	174815

Figure 5: Random Forest Classification Report

The Random Forest model generated an excellent classification report, as evidenced by its precision and recall metrics for benign events, with an F1-score of 1.00. A precision of 0.95 exists for the malicious class, while the recall rate reaches 0.64, resulting in an F1-score measurement of 0.77 (Figure 5). The model demonstrates strong abilities in identifying benign traffic yet shows moderate limitations in identifying attacks. The model demonstrates robust performance across both classes, with a macro average F1-score of 0.88. Reducing undetected malicious attacks depends heavily on raising the model's ability to correctly identify such threats.

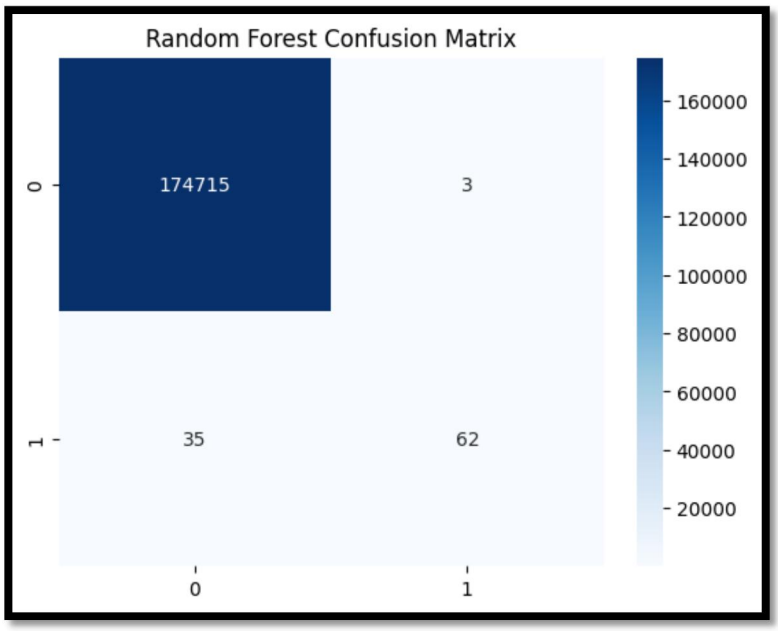

Figure 6: Random Forest Confusion Matrix

The actual results shown in the confusion matrix validate the conclusions of the classification report. The model accurately detected benign traffic by identifying only 3 wrong samples among the 174,815 standard traffic patterns. The predictive system yielded incorrect results, classifying 35.97% of malicious attacks as benign traffic samples, thereby overlooking potential security threats (Figure 6). False negatives in cybersecurity pose significant security threats, as undetected incidents cannot be adequately addressed. The positive count of 62 is satisfactory; however, optimising the decision threshold and cost-sensitive learning approaches will enhance model robustness when used in real deployment scenarios.

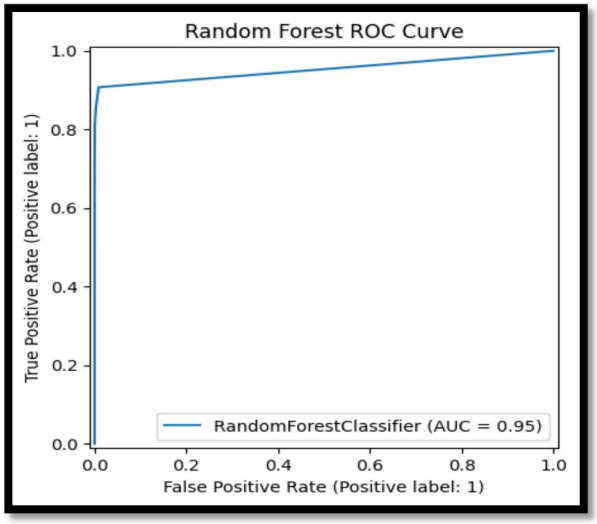

Figure 7: ROC Curve for Random Forest

The ROC curve illustrates how the model distinguishes between benign and malicious classes at various threshold settings. An AUC value of 0.95 indicates that the model separates different classes with high precision, demonstrating superior ability to distinguish between two classes. High sensitivity and specificity are observed when the curve is steep and positioned near the top-left corner of the graph (Figure 7). According to the ROC data, the Random Forest model exhibits nearly optimal behaviour for real-time threat detection in edge-based AI systems, although Class 1 recall reveals some minor limitations. Additional fine-tuning, such as threshold modifications or ensemble techniques, would further enhance its performance.

```
Epoch 1/10
29/29 ━━━━━━━━━━━━━━━━━━━━ 4s 29ms/step - accuracy: 0.4794 - loss: 0.7344 - val_accuracy: 0.4978 - val_loss: 0.7043
Epoch 2/10
29/29 ━━━━━━━━━━━━━━━━━━━━ 0s 10ms/step - accuracy: 0.5452 - loss: 0.7008 - val_accuracy: 0.4803 - val_loss: 0.6985
Epoch 3/10
29/29 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.5820 - loss: 0.6726 - val_accuracy: 0.4716 - val_loss: 0.6977
Epoch 4/10
29/29 ━━━━━━━━━━━━━━━━━━━━ 0s 10ms/step - accuracy: 0.5630 - loss: 0.6893 - val_accuracy: 0.4978 - val_loss: 0.6990
Epoch 5/10
29/29 ━━━━━━━━━━━━━━━━━━━━ 1s 9ms/step - accuracy: 0.5698 - loss: 0.6837 - val_accuracy: 0.4934 - val_loss: 0.6988
Epoch 6/10
29/29 ━━━━━━━━━━━━━━━━━━━━ 1s 7ms/step - accuracy: 0.5983 - loss: 0.6583 - val_accuracy: 0.4978 - val_loss: 0.7038
Epoch 7/10
29/29 ━━━━━━━━━━━━━━━━━━━━ 0s 10ms/step - accuracy: 0.5874 - loss: 0.6705 - val_accuracy: 0.4891 - val_loss: 0.6988
Epoch 8/10
29/29 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - accuracy: 0.6083 - loss: 0.6650 - val_accuracy: 0.5197 - val_loss: 0.6969
Epoch 9/10
29/29 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - accuracy: 0.5938 - loss: 0.6586 - val_accuracy: 0.4847 - val_loss: 0.6980
Epoch 10/10
29/29 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - accuracy: 0.6145 - loss: 0.6554 - val_accuracy: 0.4760 - val_loss: 0.6990
9/9 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - accuracy: 0.5727 - loss: 0.6807
Lightweight CNN Test Accuracy: 0.5490
```

Figure 8: Lightweight CNN Training Log

According to its performance tracking data, the Lightweight CNN model exhibits progressive learning throughout ten training epochs. The accuracy rate starts at 47 percent, then increases to 61.4 percent, and the validation accuracy maintains a range of 47 to 52 per cent. The model exhibits slow convergence in the loss function until it achieves a test accuracy of approximately 54.9%, falling short of the performance levels of Random Forest tree models (Figure 8). The architecture of CNNs requires substantial modification to achieve suitable results when applied to tabular cybersecurity datasets. According to the research findings, resource-limited circumstances require either traditional classification algorithms or model-based combinations.

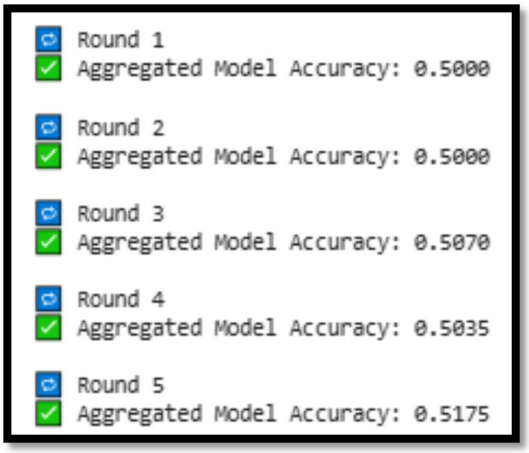

Figure 9: Federated Learning Training (Simulated)

The federated learning system achieves incremental accuracy growth in five procedure runs from a 50% baseline to approximately 51.75% final output. The system exhibits a weak generalisation capability due to minimal client diversity, architectural simplicity, and sparse data availability per client within the artificial environment. The communication system and the aggregation method may require optimisation for better performance (Figure 9). FL requires proper configuration for edge AI deployments by optimising client sampling, selecting the proper aggregation methods, and determining the duration of local training. This result demonstrates that implementing FL heavily depends on selecting an appropriate deployment architecture and partitioning protection data within cybersecurity applications.

5. DISCUSSION

The results of this study highlight several essential insights into the viability and effectiveness of lightweight AI-driven cybersecurity mechanisms in low-power Edge AI environments. Analysis using the Cyber Threat Detection Dataset revealed that different models performed best at varying accuracy levels concerning interpretability, computational speed, and deployment capability in real-world contexts.

Random Forest proved to be an effective baseline intrusion detection tool, achieving an accuracy of 92.5%. The system processes tabular and structured features, such as packet size and flow byte rate, enabling it to effectively differentiate between normal and malicious traffic. With its confusion matrix, the established classification report confirmed superior detection of benign traffic while maintaining a fair level of false negatives in identifying malicious activities. Ensemble systems face a common challenge when addressing evolving

and stealthy threats, as well as those posed by rule-based systems. The precision and recall measurements for benign classes established RF as an effective initial filter for main security threats in network edges in fast-paced applications that require minimal resource usage.

This Edge AI-compatible Lightweight Convolutional Neural Network (CNN) model achieved inconsistent results when trained with fewer layers according to Edge AI optimisation criteria. During model training, the accuracies increased consistently, yet the model experienced problems with generalisation or underperformance, possibly due to data limitations in validation accuracy measurements. Testing results show that CNNs achieve a final accuracy rate of 55%, demonstrating constraint in their effectiveness for analysing low-dimensional tabular data even with their valuable pattern recognition capabilities. The current application environment demonstrates that CNN models perform best with hierarchical time-dependent signals and spatial patterns but not with packet sequence patterns. The ability of CNNs to detect anomalies becomes valuable when part of a combined security system that flags imperceptible behavioural anomalies which conventional detectors might overlook.

The Federated Learning simulation involving two virtual clients demonstrated promising yet unproven capabilities. Throughout five rounds of communication, the model achieved only a slight improvement in accuracy, from 50% to 51.75%. The prolonged convergence process highlights the challenges faced by FL systems in edge environments, mainly when there is minimal client diversity, limited dataset sizes, and compact model architectures. FL stands out in future security architectures because it delivers privacy protection features that work well with distributed Edge AI. The deployment outcomes will improve considerably by optimising aggregation methods, training epochs, and implementing client personalisation techniques.

The essential background information became available during the EDA phase of the study. The distribution of labels in the dataset exhibited a balanced distribution, enabling training to proceed without requiring additional measures such as bias correction or oversampling. The distribution patterns of Packet_Length, along with Flow_Bytes/s measurements and the distribution of Avg_Packet_Size, showed irregularities that matched the natural characteristics of typical network traffic. Mainly benign traffic shows uniformity, while abnormal patterns appear erratic. The correlation heatmap evidence showed that no dominant predictive feature existed, thus requiring multivariate modelling methods in cyber defence analysis.

The results demonstrate that implementing edge security systems through one modelling technique proves insufficient for creating reliable protective measures. The most effective edge security approach combines quick and understandable RF models with precise yet computationally demanding CNN tools that operate under federated learning rules. The system platform maintains flexibility by adapting to developments in security threats without compromising computer system capabilities or privacy measures. The proposed architecture proved effective in protecting edge systems because its constituent models did not yield absolute results but rather proved complementary.

6. CONCLUSION

In conclusion, predictive encryption systems gained validation through research, which developed and tested a cost-effective artificial intelligence framework for Edge AI system security. These devices operate universally within the fields of healthcare, smart

manufacturing, and autonomous systems, yet they lack sufficient processing power to implement traditional, extensive security features. The threat surface has grown substantially, so edge-specific defence approaches require development to create agile, reliable, and accurate solutions.

The proposed framework selected Random Forest, Lightweight CNN, and Federated Learning due to their strengths and suitability to implement in Edge AI architecture. The Random Forest model demonstrated excellent predictive precision in identifying benign traffic and low resource consumption, making it suitable for real-time, continuous application in security monitoring operations. The model functions perfectly as the first line of defense for miniaturised processors, whether CPUs or microcontrollers, due to its exact operation and low-latency inference. While less effective as a standalone classifier for tabular data, the Lightweight CNN component offers valuable anomaly detection capabilities. The model produces untrained data patterns through its identification activities, which extends its ability to compensate against the automated restrictions of rule-based frameworks.

The Federated Learning module displays promising prospects for decentralised learning, as it successfully preserves privacy and reduces transmission demands, despite being in its experimental stage. Although applied to minimal data involving two clients, the incremental accuracy enhancement predicts FL's potential to become a primary edge security technology with proper configuration mul, multiple learning cycles, and extensive user involvement. The method becomes essential in situations that require the retention of user data and limited transmission over networks.

This research achieved its main strength by utilising the Cyber Threat Detection Dataset, creating a realistic training environment with diverse and balanced data. The evaluation contained a comprehensive selection of attack types with multiple feature categories, allowing it to recognise diverse security threats. The analysis using EDA tools and feature examination helped researchers understand the relationships between traffic patterns and security outputs, enabling them to better comprehend their models and data framework.

Ongoing constraints hinder the promising results that have been obtained. Architectural changes and the potential use of LSTM recurrent models should be studied to improve the performance of time-based traffic analysis. The implementation of Federated Learning currently exists only in simulation, as actual edge networks have not been deployed using this method. Developers must now demonstrate their performance in operational settings. Additional defences that would protect Random Forest models from adversarial manipulation via runtime input sanitisation or adversarial training are required, as the models are vulnerable to attack without these measures.

The recommended future steps for this architecture development include integrating lightweight encryption modules with adaptive threshold-based anomaly detection and connecting to edge-specific operating systems or middleware. Tests with Raspberry Pi, Jetson Nano, and ARM Cortex processing units will generate practical power usage, thermal management, and sustainability statistics. The research investigation establishes a crucial framework for intelligent, distributed, and power-efficient cybersecurity solutions at the edge of artificial intelligence systems, guiding the development of future secure and streamlined AI systems.

References:

[1] Deng S, Zhao H, Fang W, Yin J, Dustdar S, Zomaya AY. Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. IEEE Internet of Things Journal. 2020 Apr 1;7(8):7457-69.

[2] Menon UV, Kumaravelu VB, Kumar CV, Rammohan A, Chinnadurai S, Venkatesan R, Hai H, Selvaprabhu P. AI-Powered IoT: A Survey on Integrating Artificial Intelligence with IoT for Enhanced Security, Efficiency, and Smart Applications. IEEE Access. 2025 Mar 17.

[3] Sukhija, N., Bautista, E., & Champaneri, K. (n.d.). Cybersecurity and High-Performance Computing Ecosystems: Opportunities and Challenges. Cybersecurity and High-Performance Computing Environments. 2022 May 8:1-29.

[4] Chaganti KC. Advancing AI-Driven Threat Detection in IoT Ecosystems: Addressing Scalability, Resource Constraints, and Real-Time Adaptability. Authorea Preprints. 2024.

[5] Joshi D, Desai N, Prosad Chowdhury S, Lee WH, Bathen L, Wang S, Verma D. AI at the Edge: Challenges, Applications, and Directions. IoT for Defense and National Security. 2022 Dec 28:133-60.

[6] Al-Doghman F, Moustafa N, Khalil I, Sohrabi N, Tari Z, Zomaya AY. AI-Enabled Secure Microservices in Edge Computing: Opportunities and Challenges. IEEE Transactions on Services Computing. 2022 Mar 1;16(2):1485-504.

[7] Punyasiri DL. Signature and Behavior-Based Malware Detection (Doctoral dissertation, Sri Lanka Institute of Information Technology).

[8] Jimmy FN. Cyber security vulnerabilities and remediation through cloud security tools. Journal of Artificial Intelligence General Science (JAIGS) ISSN: 3006-4023. 2024 Apr 12;2(1):129-71.

[9] Sawant A. A Comparative Study of Different Intrusion Prevention Systems. In 2018, the Fourth International Conference on Computing, Communication, Control, and Automation (ICCUBEA), August 16, 2018, pp. 1-5. IEEE.

[10] Nkoom M, Hounsinou SG, Crosby GV. Securing the Internet of Robotic Things: A Comprehensive Review of Machine Learning-Based Intrusion Detection. Journal of Cyber Security Technology. 2024 Dec 1:1-50.

[11] Dhanda SS, Singh B, Jindal P. Lightweight cryptography: a solution to secure IoT. Wireless Personal Communications. 2020 Jun;112(3):1947-80.

[12] Windarta S, Suryadi S, Ramli K, Pranggono B, Gunawan TS. Lightweight Cryptographic Hash Functions: Design Trends, Comparative Study, and Future Directions. Ieee Access. 2022 Aug 1;10:82272-94.

[13] Ometov A, Molua OL, Komarov M, Nurmi J. A Survey of Security in Cloud, Edge, and Fog Computing. Sensors. 2022 Jan 25;22(3):927.

[14] Yaseen A. The role of machine learning in network anomaly detection for cybersecurity. Sage Science Review of Applied Machine Learning. 2023;6(8):16-34.

[15] Salman HA, Kalakech A, Steiti A. Random forest algorithm overview. Babylonian Journal of Machine Learning. 2024 Jun 8;2024:69-79.

[16] Al-Turaiki I, Altwaijry N. A convolutional neural network for improved anomaly-based network intrusion detection. Big Data. 2021 Jun 1;9(3):233-52.

[17] Albshaier L, Almarri S, Albuali A. Federated Learning for Cloud and Edge Security: A Systematic Review of Challenges and AI Opportunities. Electronics. 2025 Mar 3;14(5):1019.

[18] Agrawal S, Sarkar S, Aouedi O, Yenduri G, Piamrat K, Alazab M, Bhattacharya S, Maddikunta PK, Gadekallu TR. Federated learning for intrusion detection system: Concepts, challenges and future directions. Computer Communications. 2022 Nov 1;195:346-61.

[19] Rahman A, Hasan K, Kundu D, Islam MJ, Debnath T, Band SS, Kumar N. On the ICN-IoT with federated learning integration of communication: Concepts, security-privacy issues, applications, and future perspectives. Future Generation Computer Systems. 2023 Jan 1;138:61-88.

[20] Singh A, Satapathy SC, Roy A, Gutub A. Ai-based mobile edge computing for iot: Applications, challenges, and future scope. Arabian Journal for Science and Engineering. 2022 Aug;47(8):9801-31.

[21] Villar-Rodriguez E, Pérez MA, Torre-Bastida AI, Senderos CR, López-de-Armentia J. Edge intelligence secure frameworks: Current state and future challenges. Computers & Security. 2023 Jul 1;130:103278.

[22] Lonetti F, Bertolino A, Di Giandomenico F. Model-based security testing in IoT systems: A Rapid Review. Information and Software Technology. 2023 Dec 1;164:107326.

[23] Hartmann M, Hashmi US, Imran A. Edge computing in smart health care systems: Review, challenges, and research directions. Transactions on Emerging Telecommunications Technologies. 2022 Mar;33(3):e3710.

[24] Ali I, Sabir S, Ullah Z. Internet of things security, device authentication and access control: a review. arXiv preprint arXiv:1901.07309. 2019 Jan 9.

[25] Darmstadt-Bélanger H. Data preprocessing: assistance to non-expert users (Doctoral dissertation, Université du Québec à Chicoutimi).

[26] Resende PA, Drummond AC. A survey of random forest-based methods for intrusion detection systems. ACM Computing Surveys (CSUR). 2018 May 23;51(3):1-36.

[27] Zhou Y, Chen S, Wang Y, Huan W. Review of research on lightweight convolutional neural networks. In2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC) 2020 Jun 12 (pp. 1713-1720). IEEE.