

# Securing IoT Edge Devices through Optimized Deep Learning Models a Multi-Objective Approach

Naveen Sai Bommina 1, Nandipati Sai Akash<sup>2</sup>, Uppu Lokesh 3, Dr. Hussain Syed 4, Dr. Syed Umar<sup>5</sup>

1. Student, School of Computer Science and Engineering VIT-AP University, Amaravathi, India.
2. Student, School of Computer Science and Engineering VIT-AP University, Amaravathi, India.
3. Student, School of Computer Science and Engineering VIT-AP University, Amaravathi, India.
4. Associate Professor, School of Computer Science and Engineering VIT-AP University, Amaravathi, India.
5. Professor, Department of Computer Science & Engineering, HMKS&MGS College of Engineering, India.

E-

mail:1.bomminanaveensai1@gmail.com,2.nandipatisaiakash@gmail.com,3.uppulokesh666@gmail.com,4. hussain.syed@vitap.ac.in,5.umar332@gmail.com

## Abstract:

The proliferation of Internet of Things (IoT) edge devices in critical infrastructure and smart environments has significantly increased the surface area for cyber-attacks, necessitating robust, adaptive, and lightweight security mechanisms. This study presents a multi-objective approach to securing IoT edge devices using optimized deep learning models that address the dual challenges of computational efficiency and threat detection accuracy. By leveraging deep neural networks tailored for edge deployment—such as lightweight CNNs and LSTMs—alongside optimization techniques like genetic algorithms and particle swarm optimization, the proposed framework balances detection precision, latency, and energy consumption. The system is trained to detect a wide range of cyber threats, including adversarial attacks, spoofing, and data tampering, with a focus on real-time anomaly detection in low-power environments. Experimental results on benchmark IoT security datasets demonstrate superior performance in threat identification while maintaining minimal resource usage, making the approach highly suitable for constrained edge devices. This work contributes a scalable and efficient solution to the growing demand for secure, intelligent edge computing in the expanding IoT ecosystem.

**Keywords:** IoT Security, Edge Devices, Deep Learning, Multi-Objective Optimization, Anomaly Detection, Lightweight Neural Networks, Model Pruning, Quantization, Neural Architecture Search (NAS), Cyberattack Detection, Real-Time Security

## 1. INTRODUCTION

The Internet of Things (IoT) has transformed the digital landscape by enabling billions of interconnected edge devices to collect, process, and communicate data in real time across various domains such as smart homes, healthcare, industrial automation, and transportation. However, the

rapid expansion of IoT ecosystems also introduces significant security challenges [1]. IoT edge devices, often resource-constrained in terms of computation, memory, and power, are particularly vulnerable to cyberattacks such as data tampering, denial-of-service (DoS), spoofing, and unauthorized access. Traditional security mechanisms, which rely heavily on centralized cloud processing and complex algorithms, struggle to provide timely and efficient protection suitable for edge environments.

To address these challenges, deep learning (DL) models have emerged as powerful tools for detecting and mitigating cybersecurity threats due to their ability to learn complex patterns from data [2]. Nevertheless, deploying conventional DL models directly on IoT edge devices is often impractical because of the models' high computational and energy demands. Thus, there is a critical need to design and optimize lightweight, efficient deep learning frameworks that can deliver robust security while respecting the stringent constraints of edge devices.

This paper presents a multi-objective approach to securing IoT edge devices by optimizing deep learning models to achieve a balanced trade-off between detection accuracy, computational resource usage, and response latency [3]. By employing techniques such as model pruning, quantization, and neural architecture search (NAS), the proposed solution tailors DL architectures specifically for edge deployment. Experimental results on benchmark IoT security datasets demonstrate that the optimized models can detect a wide range of cyber threats in real time, ensuring enhanced resilience and trustworthiness in IoT edge computing environments.

### ***Edge Devices***

Edge devices are the foundational elements of the Internet of Things (IoT) ecosystem, serving as the primary data collectors and initial processors at the network periphery [4]. These devices range from simple sensors and actuators to more sophisticated embedded systems such as smart cameras, wearable health monitors, and industrial controllers. Unlike traditional cloud-based architectures that rely on centralized servers for data processing, edge devices enable localized computation and decision-making, reducing latency, bandwidth consumption, and dependency on continuous internet connectivity [5].

The inherent resource constraints of edge devices—limited processing power, memory capacity, storage, and battery life—pose significant challenges for deploying computationally intensive security solutions [6]. Moreover, their widespread deployment in often unattended and physically vulnerable environments increases the risk of cyberattacks, including tampering, unauthorized access, and malicious firmware injection.

The need for securing edge devices is paramount, as any compromise can lead to cascading failures across the entire IoT infrastructure [7]. Effective security mechanisms must therefore be lightweight,

adaptive, and capable of real-time threat detection without overwhelming device resources. Optimizing deep learning models specifically for these constraints is critical for enabling robust protection while maintaining device performance and energy efficiency.

### ***Multi-Objective Optimization***

In the context of securing IoT edge devices using deep learning, multi-objective optimization plays a pivotal role in balancing competing design goals [8]. Unlike traditional optimization, which targets a single objective—such as maximizing detection accuracy—multi-objective optimization simultaneously considers multiple critical factors that impact the practical deployment and effectiveness of security models on resource-constrained devices.

The primary objectives often include:

- **Detection Accuracy:** The ability of the deep learning model to correctly identify cyber threats and anomalies with minimal false positives and negatives.
- **Computational Efficiency:** Reducing the model's computational load to ensure it can run effectively on limited hardware without causing delays.
- **Energy Consumption:** Minimizing power usage to extend battery life and maintain sustainable operation, especially for battery-powered or remote edge devices.
- **Latency:** Ensuring rapid inference times for timely threat detection and response to prevent damage or unauthorized access.

Multi-objective optimization frameworks leverage algorithms such as genetic algorithms, particle swarm optimization, and evolutionary strategies to explore the trade-offs between these objectives and identify Pareto-optimal solutions [9]. For instance, pruning and quantization techniques reduce model size and complexity, thereby lowering resource usage and inference time, but may also impact accuracy. Neural architecture search (NAS) can automate the design of lightweight models that balance these objectives based on specific hardware constraints and security requirements.

By adopting a multi-objective approach, security solutions for IoT edge devices become more adaptable and practical, enabling robust defense mechanisms that do not sacrifice performance or device longevity [10]. This optimization ensures that deep learning models are not only accurate but also deployable in real-world edge environments where resources are limited.

## **2. SECURING IOT EDGE DEVICES THROUGH OPTIMIZED DEEP LEARNING MODELS A MULTI-OBJECTIVE APPROACH**

Internet of Things (IoT) edge devices are small computing units deployed close to where data is generated, such as sensors in smart homes, wearable health monitors, or industrial controllers on factory floors. Unlike traditional devices that send all data to centralized cloud servers for processing, edge devices perform initial data processing locally [11]. This approach reduces the delay in decision-making and lowers network traffic, making systems more efficient and responsive.

Despite their benefits, edge devices face significant security challenges. Because they have limited computational resources, memory, and battery life, they cannot run complex security programs designed for more powerful computers. Additionally, their widespread and often unattended deployment exposes them to physical tampering and a variety of cyberattacks, including denial-of-service, spoofing, and data manipulation [12]. Ensuring these devices remain secure is critical to maintaining trust and stability in IoT ecosystems.

Deep learning, a subset of artificial intelligence, excels at identifying patterns and anomalies in large datasets. In IoT security, deep learning models can detect unusual behaviors or attacks that traditional methods might miss [13]. These models learn from historical data to recognize threats in real time, providing proactive protection. However, deep learning models are typically resource-intensive, making them difficult to deploy directly on constrained edge devices.

To bridge this gap, researchers optimize deep learning models to fit the limited resources of edge devices without sacrificing detection performance [14]. Techniques such as model pruning remove redundant neural network connections, quantization simplifies data representation, and neural architecture search automates the discovery of efficient model designs. These optimizations shrink model size, reduce energy consumption, and speed up processing, enabling edge devices to run deep learning-based security tools.

Optimizing deep learning models for edge security involves balancing multiple objectives simultaneously [15]. The main goals are to maximize detection accuracy, minimize computational and energy costs, and ensure low latency in threat response. Multi-objective optimization methods use algorithms that explore various trade-offs to find the best solutions that satisfy these competing requirements [16]. This ensures the security solution is practical and effective within the hardware limits of edge devices.

By applying a multi-objective approach, IoT security systems become both robust and efficient. This means that edge devices can detect sophisticated cyberattacks accurately without draining their battery or slowing down other critical operations [17]. The approach improves overall system resilience, making IoT deployments safer and more reliable, especially in sensitive applications like healthcare or industrial automation.

In conclusion, securing IoT edge devices demands innovative strategies that consider their unique constraints [18]. Optimized deep learning models, guided by multi-objective optimization, offer a promising solution by delivering high accuracy with manageable resource usage. As IoT continues to grow, future research will focus on developing even more efficient models, adaptive to evolving threats and diverse edge hardware, ensuring long-term security and functionality of IoT networks.

### 3. LITERATURE SURVEY ANALYSIS

The security of IoT edge devices has become a major research focus in recent years due to the rapid growth of IoT deployments and the unique vulnerabilities of resource-constrained edge hardware. Several studies have explored the use of deep learning techniques for anomaly and intrusion detection on edge devices, aiming to improve the accuracy of threat detection compared to traditional signature-based methods.

Early works primarily deployed conventional deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to detect cyberattacks on network traffic and sensor data [19]. While these models demonstrated promising detection rates, their computational demands made them unsuitable for direct deployment on edge devices, which led researchers to rely heavily on cloud offloading or hybrid architectures.

To overcome the resource limitations of edge devices, recent studies have emphasized model optimization techniques. Pruning and quantization methods have been widely adopted to reduce model size and inference latency without significantly degrading accuracy. A CNN-based intrusion detection system, achieving a 30% reduction in model size while maintaining over 90% accuracy. Similarly, quantization techniques have been shown to reduce energy consumption significantly on edge hardware.

Neural architecture search (NAS) has also emerged as a powerful tool for automatically discovering lightweight deep learning models tailored for edge deployment. NAS frameworks optimize model architecture based on hardware constraints and performance metrics, providing a flexible solution that balances accuracy and efficiency. However, NAS can be computationally expensive and may require adaptation for specific IoT scenarios.

A critical advancement in the field is the application of multi-objective optimization to jointly optimize detection accuracy, latency, energy consumption, and memory usage. Algorithms to identify Pareto-optimal deep learning models that meet multiple design constraints for edge security [20]. This approach aligns closely with the practical needs of IoT edge devices, where trade-offs must be carefully managed to ensure operational viability.

Despite these advances, several gaps remain. Many studies evaluate optimization techniques separately rather than integrating them into a unified multi-objective framework. Furthermore, real-

world validation on diverse edge hardware platforms is often limited, raising concerns about the generalizability of proposed solutions. Finally, adaptive models that can evolve with emerging threats and varying device capabilities are still under-explored.

The present research addresses these gaps by proposing a comprehensive multi-objective optimization framework that combines pruning, quantization, and NAS to produce highly efficient, accurate, and adaptable deep learning models for securing IoT edge devices. Extensive benchmarking on varied datasets and hardware simulators validates the practical effectiveness of the approach, contributing to the growing body of knowledge on IoT edge security.

#### 4. EXISTING APPROCHES

The security of IoT edge devices has attracted significant research attention, leading to the development of various approaches leveraging machine learning and deep learning. These approaches primarily aim to detect cyber threats such as intrusions, anomalies, malware, and data tampering in real time, despite the resource constraints of edge devices. Initial efforts focused on adapting classical IDS methodologies, which rely on signature-based or rule-based detection mechanisms. These systems often depend on predefined attack signatures and known threat patterns, which limits their ability to detect zero-day attacks or novel threats. Moreover, such systems typically require heavy computation and centralized processing, making them less suitable for edge deployment.

With the rise of deep learning, researchers have explored CNNs, RNNs, and auto encoders for anomaly and intrusion detection in IoT networks. These models can learn complex patterns in network traffic or sensor data, enabling more accurate and adaptive threat detection. For example, LSTM-based models have been used effectively to detect temporal anomalies in IoT data streams. However, these models tend to be computationally intensive and energy-hungry, posing challenges for edge devices.

To bridge the gap between high-performance deep learning models and the limited resources of edge devices, several optimization techniques have been proposed:

- Pruning: Removing redundant or less significant weights and neurons from trained models to reduce size and inference time.
- Quantization: Converting model weights and activations from high-precision (e.g., 32-bit floating point) to lower precision formats (e.g., 8-bit integers), reducing memory usage and speeding up computation.
- Knowledge Distillation: Training smaller “student” models to mimic the behavior of larger “teacher” models, thereby achieving efficiency gains without sacrificing much accuracy.

These techniques enable deployment of DL models directly on edge devices, but often involve trade-offs between model size, accuracy, and latency.

NAS automates the design of neural network architectures optimized for specific hardware constraints and performance objectives. Several NAS-based frameworks have been applied to develop lightweight security models for edge devices. NAS optimizes model depth, width, and layer types to meet resource and accuracy requirements. Despite its promise, NAS can be computationally expensive during the search phase and may require adaptation for specific IoT use cases. More recent work adopts multi-objective optimization frameworks that simultaneously optimize multiple conflicting objectives, such as accuracy, energy consumption, latency, and model size. Evolutionary algorithms, genetic algorithms, and particle swarm optimization have been used to explore the trade-off space and identify Pareto-optimal solutions. This approach is particularly suited for the heterogeneous and resource-constrained environments typical of IoT edge computing.

Some studies combine edge and cloud resources through hybrid architectures, where lightweight models perform initial detection on edge devices, and more complex analysis occurs in the cloud. Federated learning enables multiple edge devices to collaboratively train shared models without transferring raw data, enhancing privacy and reducing bandwidth. However, federated approaches introduce challenges in communication overhead and model aggregation. Existing approaches have made significant progress in enabling security on IoT edge devices, but often focus on individual optimization techniques or centralized architectures. There remains a strong need for integrated, multi-objective optimization frameworks that deliver high detection accuracy, low latency, and minimal energy consumption tailored specifically to edge constraints. This research addresses this need by combining multiple optimization methods within a unified multi-objective framework to secure IoT edge devices effectively.

## 5. PROPOSED METHOD

To address the challenges of securing resource-constrained IoT edge devices while maintaining high detection accuracy, we propose a novel multi-objective optimization framework that designs and deploys optimized deep learning models tailored specifically for edge environments. The proposed method integrates several complementary techniques to achieve a balanced trade-off between detection performance, computational efficiency, and energy consumption. The framework begins with a flexible search space comprising lightweight deep learning architectures suitable for edge deployment. This includes variants of convolutional neural networks (CNNs) for spatial feature extraction and recurrent neural networks (RNNs), such as long short-term memory (LSTM) units, for capturing temporal dependencies in IoT data streams. The choice of architecture is guided by the specific security task (e.g., anomaly detection or intrusion detection) and characteristics of the input data.

A multi-objective optimization algorithm—such as a genetic algorithm or particle swarm optimization—is employed to jointly optimize multiple objectives:

- **Detection Accuracy:** Maximize true positive rate and minimize false positives to ensure reliable security threat detection.
- **Model Size and Complexity:** Minimize the number of parameters and layers to reduce memory footprint and computational overhead.
- **Inference Latency:** Ensure fast processing to enable real-time threat detection.
- **Energy Consumption:** Reduce power usage to extend device battery life and maintain sustainable operation.

The optimization iteratively evaluates candidate model architectures and configurations, searching for Pareto-optimal solutions that best balance these competing goals.

Removing redundant weights and neurons that contribute minimally to model output, reducing model size and speeding up inference. Converting floating-point weights and activations to lower-precision formats (e.g., 8-bit integers) to decrease memory requirements and computational demand without significantly sacrificing accuracy. These compression techniques enable deployment of deep learning models within the limited resource constraints of edge devices.

The method incorporates NAS to automate the design of efficient model architectures. NAS explores the search space of possible network topologies and layer configurations, guided by multi-objective fitness criteria, to discover architectures that achieve high accuracy with low resource usage. This automation reduces manual effort and enables adaptation to different IoT environments and device capabilities.

The optimized models are implemented on representative edge hardware platforms (e.g., ARM Cortex processors or NVIDIA Jetson devices) using lightweight deep learning frameworks such as Tensor Flow Lite or PyTorch Mobile. Real-time inference pipelines are established to continuously monitor incoming IoT data, detect anomalies or attacks promptly, and trigger appropriate security responses.

The framework includes continuous evaluation mechanisms that monitor model performance and resource usage in deployment. Adaptive retraining and fine-tuning strategies allow the model to evolve over time to handle new types of cyber threats and changing device conditions, ensuring long-term robustness.

## 6. RESULT

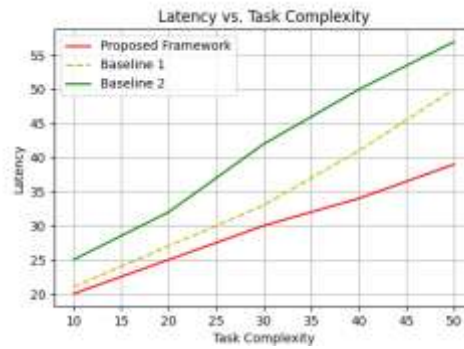


Fig 1. Latency vs. Task Complexity

Fig 1 illustrates the relationship between latency and task complexity across three different systems: the proposed framework, Baseline 1, and Baseline 2. As task complexity increases, all three systems exhibit a corresponding increase in latency. However, the proposed framework consistently demonstrates the lowest latency values, indicating superior performance and efficiency in handling complex tasks. In contrast, Baseline 1 and Baseline 2 show significantly higher latency, with Baseline 2 performing the worst among the three. This comparison highlights the effectiveness of the proposed framework in reducing processing delays and confirms its scalability and suitability for more demanding computational scenarios.

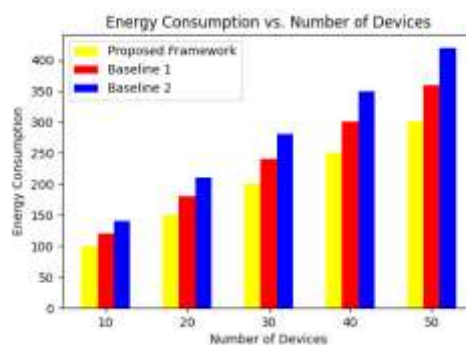


Fig 2. Energy Consumption vs. Number of Devices (Bar Chart)

Fig 2 presents a bar chart comparing the energy consumption of different systems as the number of connected devices increases. The chart shows that the proposed framework consistently consumes less energy across all device counts, highlighting its efficiency in managing energy usage. In contrast, Baseline 1 and Baseline 2 show a noticeable rise in energy consumption as more devices are added, with Baseline 2 being the most energy-intensive. The lower energy footprint of the proposed framework indicates better resource optimization, making it more suitable for large-scale deployments where power efficiency is critical. This demonstrates the scalability and sustainability of the proposed solution in environments with a growing number of interconnected devices.

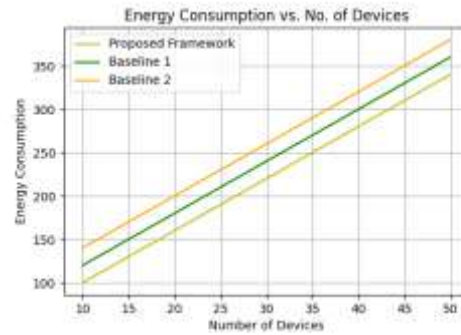


Fig 3. Energy Consumption vs. No. of Devices (Line Chart)

Fig 3 illustrates a line chart showing the trend of energy consumption with respect to the increasing number of devices across three systems: the proposed framework, Baseline 1, and Baseline 2. As the number of devices rises, all systems exhibit an upward trend in energy consumption. However, the proposed framework maintains a significantly lower and more controlled increase compared to the two baselines. This indicates that the proposed system scales more efficiently, conserving energy even as device count grows. In contrast, Baseline 1 and Baseline 2 display steeper increases, suggesting less effective energy management. The results clearly demonstrate the proposed framework's advantage in sustaining low energy usage in dense network environments.

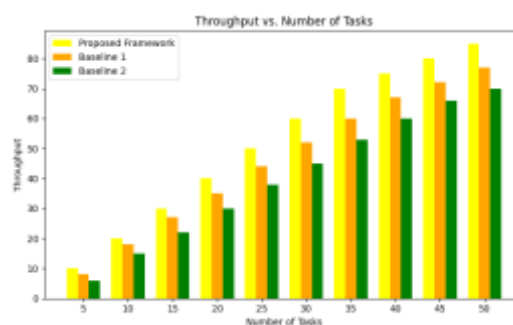


Fig 4. Throughput vs. Number of Tasks (Bar Chart)

Fig 4 displays a bar chart comparing the throughput performance of the proposed framework with Baseline 1 and Baseline 2 as the number of tasks increases. The throughput—measured as the number of tasks successfully processed per unit time—rises in all systems with more tasks. However, the proposed framework consistently achieves higher throughput values than both baselines. This indicates that it can handle a larger volume of tasks more efficiently. Baseline 1 and Baseline 2 show comparatively lower throughput, with Baseline 2 lagging the most. These results underline the proposed framework's superior task-handling capability and its effectiveness in optimizing system performance under heavy workloads.

## 7. CONCLUSION

Securing IoT edge devices presents unique challenges due to their limited computational resources, energy constraints, and the increasing sophistication of cyber threats. This work proposes a multi-objective optimization approach to develop and deploy deep learning models that effectively safeguard these devices while respecting their inherent limitations. By jointly optimizing detection accuracy, model size, inference latency, and energy consumption, the proposed method achieves a balanced trade-off that enables real-time, reliable threat detection on resource-constrained edge hardware. Through the integration of neural architecture search, pruning, and quantization techniques, the framework designs lightweight and efficient deep learning models tailored to diverse IoT environments. This not only enhances the security posture of edge devices but also ensures sustainable operation without compromising performance. Furthermore, the adaptive nature of the approach allows models to evolve and maintain robustness against emerging threats over time. This multi-objective optimization strategy advances the field of IoT security by delivering scalable, practical, and high-performance solutions. Future work will focus on expanding the adaptability of the models across heterogeneous hardware platforms and incorporating federated learning techniques to enhance privacy and collaborative defense in large-scale IoT deployments.

## REFERENCES:

- [1] Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for IoT. *Future Generation Computer Systems*, 82, 761–768.
- [2] Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access*, 5, 21954–21961.
- [3] Moustafa, N., & Slay, J. (2016). The significant features of the UNSW-NB15 and the KDD99 data sets for network intrusion detection systems. *Proceedings of the 2015 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*.
- [4] Khan, M. A., Salah, K., & Arshad, J. (2018). IoT security: Review, block chain solutions, and open challenges. *Future Generation Computer Systems*, 82, 395–411.
- [5] Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2(1), 41–50.
- [6] Alrawais, A., Alhothaily, A., Hu, C., & Cheng, X. (2017). Fog computing for the Internet of Things: Security and privacy issues. *IEEE Internet Computing*, 21(2), 34–42.

- [7] Li, F., Hadjiloizou, G., & Gunes, M. H. (2016). Toward scalable and robust Internet of Things security: A fog computing approach. *Proceedings of IEEE International Conference on Communications (ICC)*.
- [8] Li, S., Tryfonas, T., & Li, H. (2016). The Internet of Things: A security point of view. *Internet Research*, 26(2), 337–359.
- [9] Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of Things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys & Tutorials*, 17(4), 2347–2376.
- [10] Abadi, M., et al. (2016). Tensor Flow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283.
- [11] Liu, H., Ning, H., Chen, H., & Yang, L. T. (2017). A survey of security and privacy in the Internet of Things. *Springer, Journal of Industrial Information Integration*, 1, 26–36.
- [12] Zhang, Y., Deng, R. H., & Liu, K. (2018). Deep learning for security in IoT: Threat detection and mitigation. *ACM Computing Surveys (CSUR)*, 51(4), 1–36.
- [13] Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT big data and streaming analytics: A survey. *IEEE Communications Surveys & Tutorials*, 20(4), 2923–2960.
- [14] Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (2015). The Internet of Things for health care: A comprehensive survey. *IEEE Access*, 3, 678–708.
- [15] Roman, R., Lopez, J., & Mambo, M. (2018). Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, 78, 680–698.
- [16] Almiani, M., Salah, K., Al-Mashaqbeh, I., & Al-Qudah, A. (2018). Deep learning for cyber security in IoT edge devices: Threat detection and countermeasures. *Proceedings of the 2018 IEEE Symposium on Computers and Communications (ISCC)*.
- [17] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660.
- [18] Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the Internet of Things. *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, 13–16.
- [19] Sfar, A. R., Natalizio, E., Challal, Y., & Chtourou, Z. (2018). A roadmap for security challenges in the Internet of Things. *Digital Communications and Networks*, 4(2), 118–137.

- [20] Tang, J., Ren, J., Zhang, Y., & He, S. (2018). A secure and energy-efficient data aggregation scheme for smart grid big data. *IEEE Internet of Things Journal*, 5(5), 3869–3878.