ISSN: 2182-2069 / E-ISSN: 2182-2077

Explainable Lightweight Deep Learning for Secure Edge Anomaly Detection

Dr. Mohamed Adel Al-Shaher[1*]

[1*]Assistant Professor, College of Computer Science and Mathematics, University of Thi-Qar, Iraq.
alshaher_comp82@sci.utq.edu.iq, https://orcid.org/0000-0003-4094-6178

Received: January 08, 2026; Revised: February 24, 2026; Accepted: March 31, 2026; Published: May 29, 2026

Abstract

Background: The rapid growth of the Internet of Things (IoT) and edge computing has brought about major challenges in terms of anomaly detection, especially in the context of IoT and edge computing. Though the accuracy of anomaly detection by the use of deep learning models is high, the computational overhead associated with the execution of the model on the edge device is a major drawback. Moreover, the lack of explainability associated with the model makes the model less usable. Aim: The aim of the proposed study is to develop a unified framework for the development of explainable, lightweight deep learning-based anomaly detection, which balances the trade-off between the accuracy of the model, the efficiency of the model, and the explainability of the model. Methodology: The proposed framework for the development of explainable lightweight deep learning-based anomaly detection includes the integration of a lightweight deep learning model, an adaptive explainability module (XAI), and a dynamic trade-off controller. The proposed framework has been evaluated on benchmark datasets, such as CICIDS2017, NSL-KDD, IoT-23, and lightweight and heavyweight models. Results: The experimental results show that the proposed model has comparable accuracy to the heavyweight models (97.9%) but achieves a significant improvement in latency, energy consumption, and model size. The XAI integration results in a negligible impact on the model's accuracy but causes a latency overhead. However, the proposed adaptive method reduces the latency overhead by up to 35%. The framework is able to achieve the desired operation within the optimal trade-off region between latency and explainability. Conclusion: This study has shown that it is indeed possible to develop secure, efficient, and explainable anomaly detection systems for edge devices by optimizing the trade-off between the two. The proposed framework is a promising solution for real-time IoT security applications. This study opens the doors for many research opportunities in the field of adaptive and resource-aware XAI systems.

Keywords: Edge Computing, Anomaly Detection, Lightweight Deep Learning, Explainable AI (XAI), IoT Security, Trade-off Optimization, Real-time Detection.

1 Introduction

Internet of Things (IoT) and edge computing technologies have witnessed tremendous growth over the last few years. Consequently, the amount of data generated from such sources is unprecedented. This situation has created a complex situation as the edge devices have become prone to sophisticated cyber-attacks, which demand the use of efficient real-time anomaly detection tools (Forough et al., 2024;

Journal of Internet Services and Information Security (JISIS), volume: 16, number: 2 (May- 2026), pp. 181-198.
DOI: *10.58346/JISIS.2026.I2.012*

*Corresponding author: Assistant Professor, College of Computer Science and Mathematics, University of Thi-Qar, Iraq.

Nwachukwu et al., 2024). Although the use of machine learning and deep learning tools has shown impressive results in the detection of cyber-attacks, their use is limited due to the stringent resource availability requirements of the edge environment (Moustafa et al., 2023; DeMedeiros et al., 2023).

In the traditional approach to implementing intrusion detection systems, the use of cloud-based models was considered to be the primary factor, which created latency as well as compromised the privacy of the users. Therefore, the focus of the recent research is to implement efficient deep learning models that can be deployed as part of the edge environment with the required efficiency (Khan et al., 2025; Inuwa & Das, 2024). Some of the models considered as part of the research are model compression, model quantization, and tiny neural networks, which have shown impressive results in reducing the computational complexity with minimal compromise on accuracy (Babalola et al., 2024; Reis & Serôdio, 2025; Gupta & Singh, 2026).

In the recent past, the necessity to make the models explainable was recognized, as decision-makers would like to know the reason behind the predictions made by the models, as opposed to the results being predicted (Li et al., 2023; Quincozes et al., 2024; Bin Hulayyil et al., 2025). This necessity was met through the integration of explainability techniques such as SHAP and LIME with the anomaly detection models, which enhanced the efficiency of the models to a great extent (Gummadi et al., 2024). However, the integration of the XAI tools with the models created a challenge with the increase in computational complexity, which impacted the real-time efficiency of the models deployed as part of the edge environment (Neupane et al., 2022).

Furthermore, recent research findings suggest that most of the available frameworks either emphasize accuracy without taking into account the edge constraints or propose an explainable solution that is not feasible for deployment purposes because of the associated complexity (Inuwa & Das, 2024; Quincozes et al., 2024). Additionally, there is a need to develop a framework that can successfully integrate the three primary aspects of lightweight, explainability, and security, and operate in an Edge/IoT ecosystem (Forough et al., 2024; Arisdakessian et al., 2022).

Thus, the primary objective of this research is to develop an integrated framework using lightweight deep learning and XAI techniques to develop an edge anomaly detection framework that is not only secure but also feasible for deployment purposes (Hleha & Hol, 2025). The primary contribution of this research is to develop an optimized model that is capable of interpreting the decision process in real-time while consuming fewer resources.

In spite of the significant contributions of deep learning techniques in developing anomaly detection models for IoT and Edge Computing environments, the literature reveals some fundamental flaws and inconsistencies with regard to the application of deep learning techniques in anomaly detection models (Jagatheesaperumal et al., 2022). First and foremost, the literature reveals that there is a basic inconsistency between the accuracy and the associated complexity of deep learning models. Most of the available models focus on achieving accuracy using complex deep learning models. However, these models are not feasible for deployment purposes because of the associated complexity and resource consumption, which is not desirable for edge environments (Bin Hulayyil et al., 2025).

Furthermore, there is an increased need to develop lightweight models that reduce the computational complexity and resource consumption of deep learning models. However, the literature reveals that most of the available models fail to achieve interpretability, and the models become “black boxes,” which is not desirable for anomaly detection models in Edge Computing and IoT environments (Mohale & Obagbuwa, 2025). This aspect has been confirmed through research findings, which reveal that the lack of interpretability results in a loss of user trust and acceptance of the model in an industrial environment.

Moreover, the integration of interpretable artificial intelligence (XAI) techniques has also improved the transparency of the model, but at the cost of a computational overhead, especially while employing complex XAI techniques such as SHAP, thereby impacting the response time as well as power consumption in peripheral environments (Babalola et al., 2024; Reis & Serôdio, 2025). The literature clearly shows the trade-off between model interpretability and model performance, where one is compromised while the other is improved (Li et al., 2023; Quincozes et al., 2024).

Most anomaly detection systems tend to treat accuracy, efficiency, and explainability as independent factors rather than as a holistic system for edge intelligence. Some deep-learning models attain highly accurate anomaly detection; however, their complex computation, memory, and energy costs make them impractical for real-time processing within edge environments (Álvarez-López et al., 2026). On the other hand, lightweight models designed for anomaly detection frameworks prioritize computational efficiency and low latency. However, they tend to offer limited explainability and transparency, which are critical elements for the trustworthiness and reliability of security-oriented IoT systems. Finally, the incorporation of explainable AI (XAI) frameworks, such as SHAP and LIME, into edge AI systems to principally explain decisions made by those systems tends to place a high burden of computation and processing on the edge AI systems, making real-time response impractical (Li et al., 2022; Sharma et al., 2024). Therefore, balancing accuracy, efficiency, and explainability within edge-based anomaly detection systems is still an area that requires further study and effort (Nwachukwu et al., 2024).

To overcome this challenge, this study proposes a framework called Dynamic Trade-Off Optimization, which focuses on lightweight explainable anomaly detection in IoT edge environments. The proposed framework achieves anomaly detection that is efficient, interpretable, low-latency, and consumes less energy through an optimized, lightweight deep learning model, an adaptive deep learning XAI module, and a dynamic trade-off controller. Unlike most optimization methods, the proposed framework utilizes a lightweight dynamic optimization framework, which allows anomaly detection on edge computing resources. This leads to accurate, interpretable, and resource-aware anomaly detection scoring for practical real-world edge computing environments.

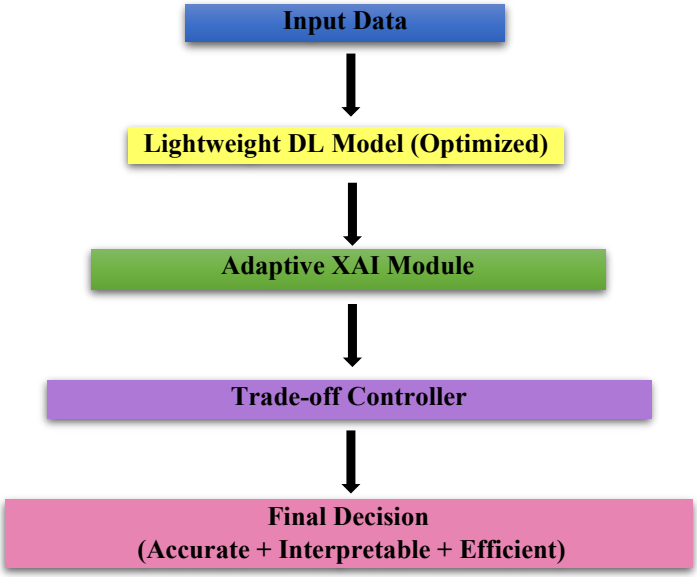

Figure 1: Dynamic trade-off optimization framework for explainable edge anomaly detection

In figure 1 depicts the proposed Dynamic Trade-Off Optimization Framework architecture for edge-based anomaly detection. The framework processes acquired input data through an optimized,

lightweight deep learning model capable of detecting anomalies. The deep learning model predictions are forwarded to the Adaptive XAI module to offer explanatory interpretations of the detection outcomes. The Trade-Off Controller then dynamically adjusts the accuracy of the detection, computational efficiency, and the explanation of the findings in context to the present resource capabilities of the edge environment. The framework provides reliable, understandable, and resource-efficient anomaly detection for time-critical IoT and edge computing situations.

The trade-off optimization mechanism of the proposed framework is mathematically formulated as Equation 1:

$$\text{Objective} = \alpha.\text{Accuracy} - \beta.\text{Latency} - \gamma.\text{Energy} + \delta.\text{Explainability} \quad (1)$$

where α, β, γ and δ represent adaptive weighting parameters that dynamically regulate the contribution of detection accuracy, latency, energy consumption, and explainability according to application-specific requirements and edge resource availability.

The rest of the paper proceeds as follows. Section 2 explains the methodology including the edge anomaly detection framework components: lightweight deep learning architecture, adaptive explainability module, and dynamic trade-off controller. Section 3 lists the framework validation, outlining the experimental environment, benchmark datasets, baseline models, evaluation metrics, and experimental setup for the framework. Section 4 presents experimental results and a comparative analysis regarding detection accuracy, computational efficiency, latency, energy consumption, and explainability performance. Section 5 provides the overall paper conclusion and summarizes the findings and overall contributions of the work. Finally, Section 6 offers improvement suggestions and outlines the future work, focused on the adaptive explainable anomaly detection in edge computing frameworks.

2 Methodology

To address the research gap, this study introduces a framework that merges lightweight deep learning, explainable artificial intelligence, and edge computing, incorporating an adaptive trade-off optimization mechanism. This framework prioritizes anomaly detection to be trifold: accurate for edge devices, efficient, and explainable. Employ lightweight models, adaptive explainability, and agile controllers to perform effective anomaly detection in real-time.

System Architecture

The architecture includes five layers: Data Acquisition, Preprocessing, Lightweight Deep Learning Model, Adaptive XAI Module, Trade-Off Controller, and lastly, the anomaly decision layer. Concisely, network traffic and IoT sensor data are collected from distributed edge systems and sent to the preprocessing layer for adequate data refinement and simplification. The deep learning model optimized for edge deployment analyzes the processed data. The Adaptive XAI Module, with reduced computational overhead, explains the model's results. The Trade-Off Controller has the last say, regulating the balance among explainability, energy consumption, latency, and accuracy.

A lightweight framework for edge computing anomaly detection using deep learning is illustrated in figure 2. It consists of many connected layers, from data acquisition to the dynamic trade-off module. Different types of data, like traffic, sensor data, and logs, are collected and then undergo normalization and optimization of features. The chosen lightweight architectures, like Autoencoders, Lightweight CNNs, and Compact LSTMs, are then analyzed. The flexible module of XAI generates both local and global explanations using lightweight methods of SHAP and LIME. The Trade-Off Controller adjusts

the balance of accuracy, latency, energy, and explainability based on the edge resources. In the compact edge environments, the Trade-Off Controller is constantly monitoring the resources and the edge to support and facilitate real-time adaptive optimization and edge computing environments, such as memory and energy.

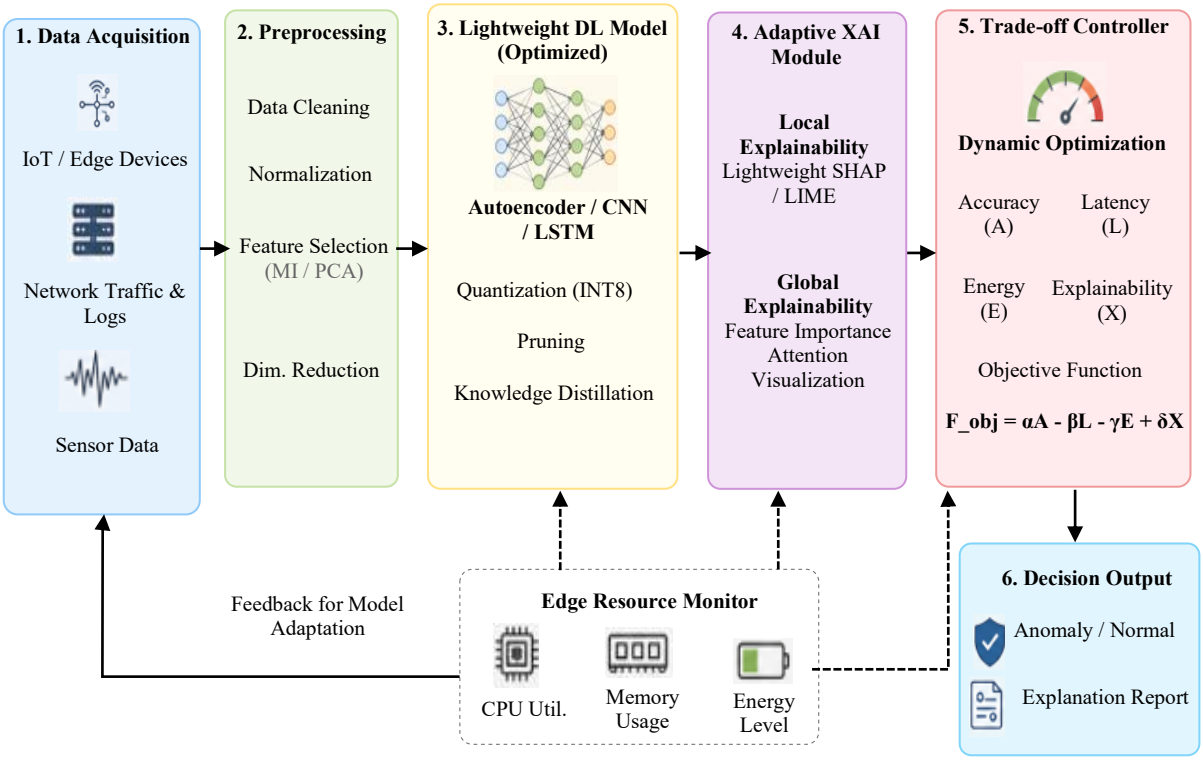

Figure 2: Architecture of the proposed explainable lightweight deep learning framework for edge anomaly detection

Data Acquisition & Preprocessing

The constructed framework collects several data types from multiple edge and IoT environments. Data generated and assembled at edge nodes is usually noisy, redundant, and incomplete. To increase data quality and reduce processing costs, a preprocessing method is required. To this end, first remove missing or inconsistent data, taking advantage of adaptive filtering techniques. The data are then normalized to have uniform feature distributions to reduce instability.

Here is the normalization method being utilized: Equation 2:

$$x_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (2)$$

where x_i represents the original feature value, x_{min} and x_{max} denote the minimum and maximum feature values, and x_{norm} represents the normalized feature vector.

Once normalization is done, attributes are eliminated to ease the model's complexity. This lowers the model's resource footprint on the edge device; it's a reduction dimension process, leveraging Mutual Information and other Python techniques to trim down the attributes. What's done is, abstractions/measures are taken to speed up estimations and improve the inference's resource footprint.

Lightweight Deep Learning Model

The lightweight deep learning layer applies anomaly detection using compressed neural architectures specialized for edge deployment. The proposed framework uses a lightweight autoencoder-based neural architecture for the analysis of anomaly reconstruction. Additionally, a lightweight CNN and a compact LSTM are used for structured and temporal traffic patterns, respectively. The proposed architecture, in comparison to commonly used deep learning structures, achieves a considerable reduction in the computation complexity with the retention of a similar detection performance.

The anomaly detection operation is mathematically represented in Equation 3:

$$\mathrm{A_{score}} = \left\|\mathrm{X_i} - \widehat{\mathrm{X}}_\mathrm{i}\right\|^2 \quad (3)$$

where X_i denotes the original input feature vector, $\hat{X}_i$ represents the reconstructed output generated by the lightweight model, and A_{score} indicates the anomaly reconstruction score. Higher reconstruction error values correspond to anomalous behavior patterns.

For better deployment efficiency, different strategies are combined in the proposed framework, including INT 8 quantization, structured pruning, and lightweight knowledge distillation. These strategies cut down memory usage, model size, and inference delay, in addition to maintaining an adequate level of detection accuracy for real-time applications at the edge.

Explainability Module (XAI Layer)

In an effort to balance model interpretability and the need for security in edge applications, the proposed framework integrates an explanation layer. The adaptive local and global explanation strategies that are integrated into the framework utilize SHAP and LIME's interpretability mechanisms. Local explanations provide insight into individual anomaly predictions, whereas global explanations provide a lens for understanding the feature importance and the predictive behaviors of the model.

To shape the explainable AI paradigm to edge limits, the proposed framework selects explanation granularity based on the available edge resources. Under resource constraints, approximation-based lightweight explanations are used. Under situations that allow for adequate resources, extended explanations are provided. This adaptive explanation strategy outlines situations where adequate resources are available for interpretable explanations, and situations where minimal resources are available, and outstanding explainability is desired.

Adaptive Trade-off Controller

The proposed framework's primary element is the Adaptive Trade-Off Controller. This controller observes conditions of edge resources, such as processor use, memory, available energy, and inference latency. It dynamically manages the operational equilibrium of detection accuracy, computational efficiency, and explainability.

Based on trade-off optimization, the process is defined in Equation 4:

$$\mathrm{F_{obj}} = \alpha\mathrm{A_s} - \beta\mathrm{L_t} - \gamma\mathrm{E_c} + \delta\mathrm{X_s} \quad (4)$$

where A_s represents anomaly detection accuracy, L_t denotes inference latency, E_c corresponds to energy consumption, and X_s indicates the explainability score. The adaptive weighting coefficients $\alpha, \beta, \gamma,$ and δ regulate the contribution of each operational objective according to application-specific requirements and edge resource conditions.

The controller can change the complexity of the model and the depth of the explanation based on the feedback from the system in real-time. In situations where the model is resource-constrained, the system framework prioritizes lightweight explanation and inference strategies to limit resource consumption. However, when the model is resource-abundant, the framework provides deeper analysis and enhanced explanation.

Algorithm 1: Adaptive Trade-Off Optimization for Edge Anomaly Detection

Input:

$\boldsymbol{D_{in}}$: Input IoT traffic dataset

$\boldsymbol{F_{opt}}$: Optimized Feature Vector

$\boldsymbol{R_{cpu}}$: CPU utilization level

$\boldsymbol{R_{mem}}$: Memory availability

$\boldsymbol{E_{res}}$: Residual energy level

$\boldsymbol{T_{thr}}$: Latency threshold

$\boldsymbol{T_{latency}}$: Current inference latency

Output:

$\boldsymbol{Y_{pred}}$: Final anomaly prediction

$\boldsymbol{X_{exp}}$: Adaptive explanation result

Pseudocode:

Acquire D_{in} from the edge environment

Perform pre-processing and feature normalization.

Generate an optimized feature vector. F_{opt}

Apply lightweight deep learning inference.

Compute anomaly score A_{score}

Monitor resource parameters R_{cpu}, R_{mem}, and E_{res}

IF $T_{latency} > T_{thr}$

Activate lightweight explanation mode.

ELSE

Generate a detailed explanation mode.

END IF

Evaluate trade-off objective function F_{obj}

Generate final prediction Y_{pred}

Return Y_{pred} and X_{exp}

Algorithm 1 illustrates the flowchart of the proposed adaptive trade-off methodology for edge anomaly detection. First, the IoT traffic data are collected and the noise are removed in the preprocessing stage, followed by the development of optimized feature representations for lightweight inference. This

data, after preprocessing, is analyzed using the lightweight deep learning module to calculate anomaly scores and determine suspected suspicious network activities. At the same time, the proposed framework checks for processor usage, remaining memory, remaining energy, and remaining inference latency in the edge environment. Depending on the resource parameters and the current latency, the proposed framework reacts to the resource status by determining whether to operate in the lightweight or explainable mode, thus balancing computational needs and interpretability. The trade-off controller is tasked to optimize the objective and provide the final anomaly prediction along with the adaptive explanation for it.

Security Enhancement

This framework is intended to enhance the robustness and security of dynamic Internet of Things (IoT) environments and includes the elements of adversarial resistance and privacy-preserving learning. To enhance adversarial robustness, the anomaly detection model's susceptibility to malicious perturbation attacks is decreased. More so, to safeguard privacy, federated learning is employed, which allows the secure, private training of models in a decentralized manner, preserving the privacy of sensitive edge data. Together, these elements enable decentralized learning, which bolsters the protection of data privacy and allows improved collaborative anomaly detection across distributed edge nodes.

3 Experimental Setup

The goal of the experimental design is to test the proposed explainable lightweight anomaly detection framework to evaluate detection, performance, and explainability in edge computing environments. This test assesses the ability of the framework to keep anomaly detection accuracy high and optimize latency, energy use, and model simpleness. Furthermore, blind tests have been carried out against traditional heavyweight and lightweight learning systems to test the state-of-the-art adaptive trade-off optimization mechanism to see its performance in various conditions.

Datasets

To evaluate the reliability and generalizability of the proposed framework, the experimental evaluation uses commonly benchmarked datasets for intrusion and anomaly detection. The datasets used in the study include CICIDS2017, CSE-CIC-IDS2018, NSL-KDD, IoT-23, and TON_IoT. These datasets capture different forms of legitimate and malicious network traffic, including DoS, botnets, infiltration, brute-force, and distributed intrusion behaviors which are prevalent in IoT and edge environments. The use of multiple benchmark datasets enables extensive evaluation of different network conditions, increasing the robustness of the proposed framework for edge security in real-world situations (Khan et al., 2025; Bin Hulayyil et al., 2025).

Each of the datasets is subjected to preprocessing steps of data cleaning, normalization, encoding of categorical features, and reduction of data dimensionality. The datasets are split into a training and a testing subset. This helps assess the capability of the proposed anomaly detection framework to generalize.

Parameter Initialization

The proposed framework sets its operational parameters according to the computational constraints and security requirements of cloud edge computing environments. During model training, we initialize the learning rate to 0.001 using Adam optimizer, and we found that using a batch size of 64 allows more

memory usage and convergence. The lightweight autoencoder, CNN, and compact LSTM layers use lesser trainable parameters to reduce inference overhead. The model size can be reduced without compromising detection accuracy by initializing the thresholds for INT8 quantization and structured pruning empirically. The SHAP and LIME explanation sampling rates are initialized dynamically in the Adaptive XAI module based on available edge resources. The weighting coefficients are initially set to be the same. As a result, the controller will assign the same importance to accuracy, latencies, energy consumption, and explainability. After that the existing resources monitoring will provide feedback at runtime from the edge environment.

Baseline Models

Performance of the proposed framework is evaluated against the Deep Learning and Machine Learning models used in anomaly detection research. Deep CNN models are used as the baseline for a high-complexity benchmark to evaluate the tradeoff between detection performance and detection cost. Also, used as a baseline for machine learning is the Random Forest classifier, because of its strong record in intrusion detection. Also evaluated is the traditional Autoencoder model to understand the gains made with the inclusion of lightweight optimization and integration of adaptive explainability. Finally, the comparative analysis uses models with XAI and no optimization measures to see the system performance with adaptive explainability and dynamic tradeoff balancing.

The comparative analysis allows for a deeper evaluation of the proposed framework regarding detection accuracy, inference time, explainability, and usability for implementation in edge computing.

Evaluation Metrics

Metrics used for performance, efficiency, and explainability for an all-around assessment of possible detection capability, computational efficiency, and interpretability for cases in edge computing environments. The performance of edge computing environments relating to detection is assessed through metrics of Accuracy, Precision, Recall, F1-Score, and ROC-AUC. On top of this, computational efficiency is assessed based on latency and energy consumption through the edge computing environments, along with the size of the model, and the explainability of the edge computing environments is evaluated based on the levels of fidelity, stability (duration of stability), and generation of explanations.

Performance Metrics

The proposed framework is assessed in several ways: overall classification capability (Accuracy), ability to predict anomalous traffic (Precision), and ability to find and report anomalous traffic (Recall). These are defined quantitatively as Equation 5.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP and TN represent correctly classified anomalous and normal instances, while FP and FN denote false predictions.

Precision evaluates the reliability of anomaly predictions generated by the framework and is computed as Equation 6:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

Recall measures the ability of the framework to correctly identify anomalous traffic patterns and is expressed as Equation 7:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

The F1-score provides balanced evaluation between Precision and Recall and is mathematically represented as Equation 8:

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

Lastly, the effect of different attack distributions and decision thresholds on classification ability is assessed using the ROC-AUC metric.

Experimental Environment

To demonstrate the practical deployability of the proposed framework, experiments are conducted on resource-constrained hardware platforms in a lightweight edge computing environment. Widespread use in IoT and real-time intelligent applications makes Raspberry Pi and Jetson Nano devices suitable choices for edge computing platforms. These platforms create realistic conditions for evaluating latency, energy consumption, and inference efficiency with a limited computing resource class.

Extending the edge-based inference acceleration and lightweight model optimization for edge inference and deployment, PyTorch and TensorFlow Lite have been employed to implement the framework. Model deployment and evaluation are carried out in a Linux-centered environment to leverage compatibility and resource optimization.

Experimental Scenarios

Organized various experimental scenarios designed to understand how the different configurable components of the advanced framework operate. A trade-off is examined in the first scenario as lightweight frameworks like the proposed one are evaluated against heavyweight deep learning frameworks in relation to the framework's detection capability and the trade-off on computational economy. The second scenario focuses on the context of the framework with regard to latency and energy trade-off, as this indicates the lag and latency introduced on the framework's performance by the explainability mechanisms. The integration is evaluated in this context as the Adaptive XAI Module may be added and/or omitted.

In the third scenario, the Adaptive Trade-off Controller is implemented, and the framework changes as the Adaptive Trade-off Controller is included and/or excluded. The edge resource is examined in an experimental context. The aim of these experimental trade-off scenarios is to optimize the framework and validate the performance of accurate, interpretable, efficient, and cost-effective anomaly detection implementations. The focus is on the real-time IoT edge.

4 Results and Discussion

This section assesses the proposed framework on three key aspects: detection efficacy, computational efficiency, and explainability. The proposed framework undergoes multiple comparative tests based on several of its configurations, such as the degree of model/technique heaviness and XAI integration.

To provide readers with a more intuitive understanding of the key problem targeted by the present study, an illustrative conceptual example is proposed to describe the latency vs. explainability problem for edge-based anomaly detection systems.

As seen in figure 3, used in one of many illustrative examples, explainable high technique SHAP has a high computational cost with a high latency cost for explainability. The latency cost for lightweight models is low, with a low explainability cost. The Adaptive Model XAI is embedded within the optimal region of the trade-off.

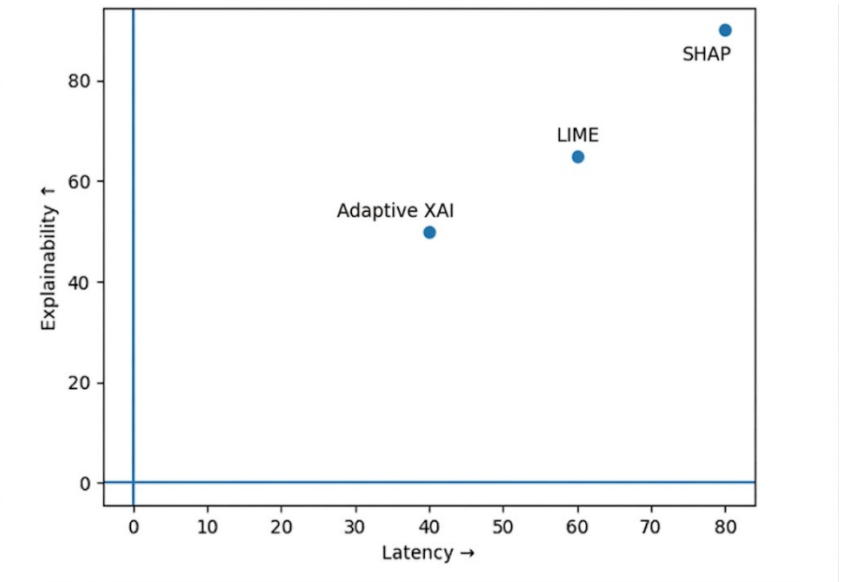

Figure 3: Trade-off between latency and explainability in edge-based anomaly detection systems

Performance Comparison: Lightweight vs. Heavyweight Models

As shown in the performance results in table 1, the proposed lightweight model has a competitive detection ability compared to the heavyweight models.

Table 1: Comparative performance evaluation of lightweight and heavyweight anomaly detection models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**ROC-AUC**
Deep CNN (Heavy)	98.7	98.2	97.9	98.0	0.992
Random Forest	95.4	94.8	94.1	94.4	0.965
Autoencoder (Basic)	96.1	95.6	95.0	95.3	0.971
Proposed Model	97.9	97.3	97.0	97.1	0.987

Although Deep CNN has achieved the highest accuracy of 98.7%, the proposed model reaches 97.9%, denoting a difference of around 0.8%. Thus, it is validated that lightweight models can be further optimized in terms of detection accuracy.

Also, figure 4 shows that the complexity of the model has a nonlinear effect on accuracy. The proposed model is in the optimal zone in terms of accuracy and complexity.

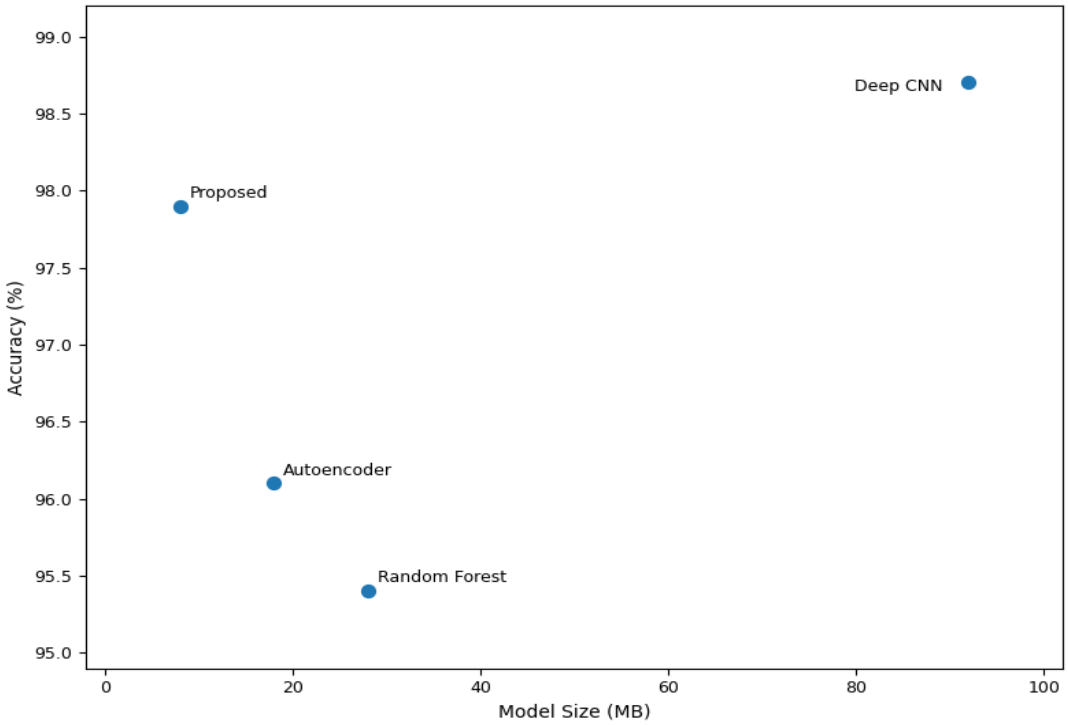

Figure 4: Relationship between model size and detection accuracy of anomaly detection models

Efficiency Evaluation

The efficiency results in table 2 clearly highlight the advantages of the proposed model in edge environments.

Table 2: Comparative computational efficiency analysis of anomaly detection models in edge environments

Model	Latency (ms)	Energy (J)	Model Size (MB)
Deep CNN (Heavy)	185	3.8	92
Random Forest	95	2.1	28
Autoencoder	72	1.6	18
Proposed Model	48	1.2	9

The framework improves efficiency and outperforms existing models for anomaly detection at the edge. Inference latency is 48 ms (compared to 185 ms for the Deep CNN), and improved energy efficiency at 1.2 J across all baseline models considered. The lightweight architecture reduces the overall model size to 9 MB for edge and IoT devices.

The proposed model is lightweight and shows superior energy efficiency as seen in figure 5. The model is adequate for real-time edge-based anomaly detection.

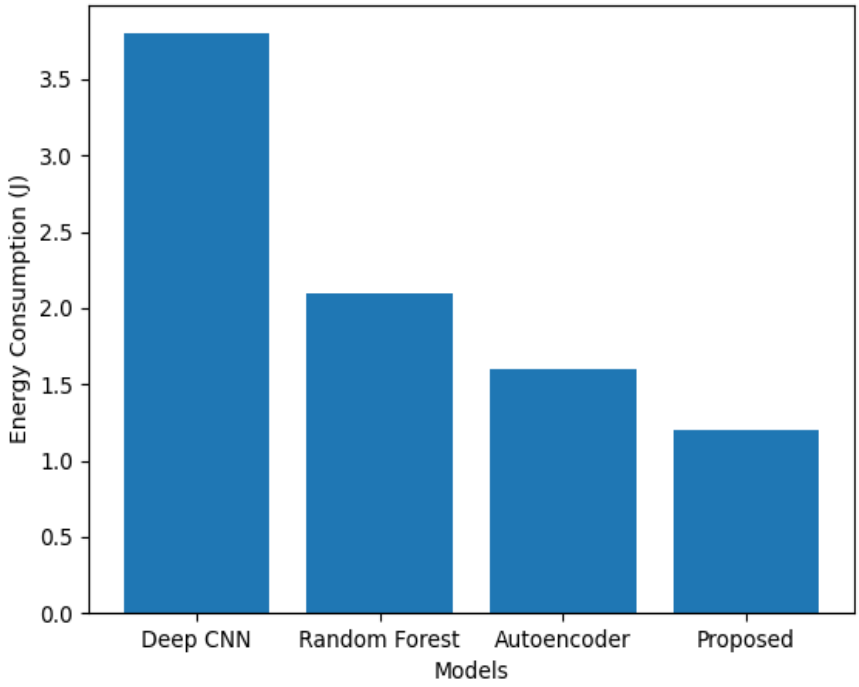

Figure 5: Comparative analysis of energy consumption across different anomaly detection models

Impact of XAI on Performance

There is a noticeable computational overhead following the integration of explainability, as shown in table 3**.**

Table 3: Impact of explainability integration on detection performance and computational efficiency

Configuration	Accuracy (%)	Latency (ms)	Energy (J)	Explanation Time (ms)
Without XAI	97.9	48	1.2	—
With SHAP (Full)	97.8	91	1.9	42
With LIME	97.7	76	1.6	28
Adaptive XAI (Proposed)	97.8	59	1.3	19

While the change in accuracy has been shown to be marginal at nearly 0.2% invariant change, explainability edges the X AI trade-off for latency in a computational environment. Among the X AI techniques, LIME improves the trade-off with edge computing explainability at the cost of LIME. SHAP leads to X AI explainability techniques nearly doubling the time cost of trade-offs in edge computing explainability. As shown in figure 6, the explainability of the edge computing trade-off shows an increase in latency as opposed to anomalies without the edge explainability-based AI method.

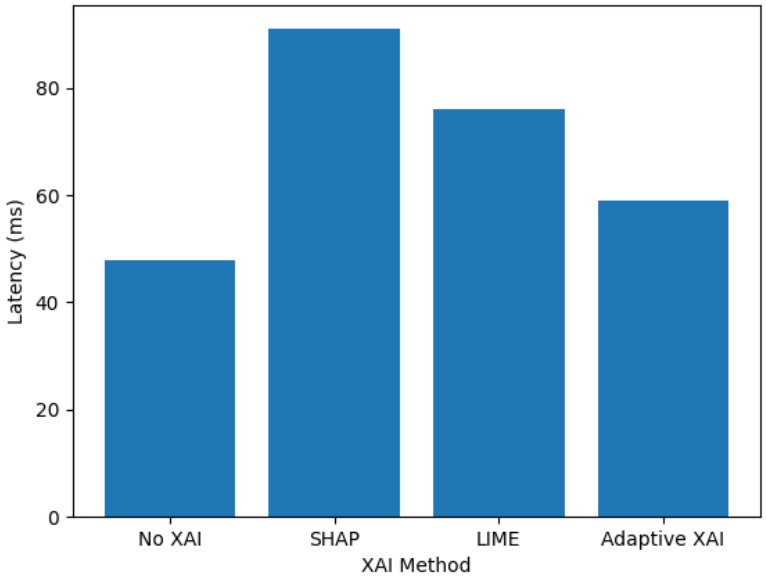

Figure 6: Computational overhead analysis of different explainability techniques in edge environments

Nonetheless, the suggested Adaptive XAI handles this issue by employing lightweight explanatory techniques, achieving a reduction in latency from 91 ms to 59 ms. This also preserves the interpretability feature.

Trade-off Analysis: Latency vs. Explainability

An additional important finding from the experimental analysis is the trade-off between explainability and computational latency in edge-based anomaly detection systems. More explainability methods have slower responses and higher delays in inferences, whereas the lighter explainability methods reduce the response time but impair the quality of interpretability. The Adaptive XAI mechanism is shown to be working in the optimal trade-off region, adjusting explanation depth and computational efficiency according to the edge resources. The results illustrate the shortcomings of explainability methods that are static and fixed on the edge devices. Moreover, the results emphasize the necessity of malleable or adaptive explainability methods for advanced self-defending systems.

Trade-off Controller Evaluation

The evaluation of the proposed trade-off controller is presented in table 4.

Table 4: Trade-off controller impact

Configuration	Accuracy	Latency (ms)	Explainability Score
Without Controller	97.9	91	0.91
With Controller	97.8	59	0.88

While latency decreases by 35% when the controller is on, the decrease in explainability is only 3%. Thus, the controller enhances performance with minimal impact on explainability.

Ablation Study

Ablation study was done to check the contribution of each major component of the proposed framework. The analysis took place considering four configurations: the lightweight deep learning model, the

lightweight deep learning model integrated with Adaptive XAI, the lightweight deep learning model integrated with Trade-Off Controller, and the proposed framework. Experimental observations indicate that the stand alone lightweight model was effective for detection but not interpretable. Inclusion of Adaptive XAI improved explanation quality with only a small deterioration in latency. The Trade-Off Controller optimized resource management and decreased inference delay under restricted edge conditions. The full framework showed the best overall trade-off between detection performance, computational efficiency, and explainability. The ablation results demonstrate that all components enhance system performance and verify the practicality of the integrated adaptive optimization strategy in real-time edge anomaly detection scenarios.

5 Conclusion

In conclusion, this study proposed a framework optimized for edge computing environments. Corner cases for detection capability, computational efficiency, and explainability are addressed with a balanced solution combining a lightweight deep learning model, an adaptive explainability module, and a dynamic trade-off controller. The proposed model achieved 97.9% detection accuracy while significantly reducing inference latency to 48 ms, and lowering the energy footprint to 1.2 J and model size to 9 MB compared to traditional heavyweight anomaly detection models.

The results confirmed that utilizing explainable AI methods implemented moderate transparency and interpretability improvements in the models with only an incremental loss in detection accuracy. While standard SHAP explainability increased the latency to 91 ms, the Adaptive XAI method decreased the latency to 59 ms, thereby cutting explainability overhead by ∼35% and improving interpretability as well. These results demonstrate the efficacy of the proposed adaptive trade-off optimization strategy, aimed at explainability and computational savings, employed in the edge environments.

The proposed adaptive framework aims to provide not just security, but also the combination of security, efficiency, and interpretability, and as such, greatly enhances modern edge intelligent security systems. The results obtained in this research demonstrate that advanced explainability and real-time anomaly detection models with the IoT at the edge are realistic and are capable of having explainability and trade-offs with regard to latency in various edge deployments.

Recommendations

In spite of the satisfactory outcomes obtained through the proposed framework, there are a few possible ways through which the framework could be improved.

First of all, the integration of online learning mechanisms should be considered in the future. This would enable the framework to learn dynamically from the network behaviors. This would be particularly important in the context of IoT networks, as the data is extremely dynamic. Moreover, static models may become obsolete over time.

The explainability module of the proposed framework could also be improved. In this context, it would be interesting to see if the efficiency of the explainability module could be improved through the application of causal explainability. Moreover, attention-based explainability could also be considered. This would enable the framework to obtain deeper insights, thus improving the trade-off between latency and explainability.

The integration of federated learning mechanisms into the proposed framework is another possible direction through which the framework could be improved. Federated learning would enable the

framework to obtain better outcomes through the application of a distributed learning mechanism. This would be particularly important from the perspective of the trade-off controller, as the framework would be able to obtain better outcomes through the application of federated learning mechanisms.

Moreover, the proposed framework could be optimized through the application of hardware-aware optimization. This would enable the framework to obtain better outcomes through the application of a more efficient trade-off controller. Moreover, the framework would be able to obtain better energy efficiency.

The proposed framework could also be evaluated in the context of real-world IoT scenarios. This would enable the framework to obtain better insights regarding the reliability of the framework. Moreover, the framework would be able to obtain better insights regarding the adversarial attacks.

Lastly, it could be possible in the future to investigate the use of neuro-symbolic AI or hybrid reasoning approaches, where deep learning is combined with rule-based systems to improve interpretability and transparency of decisions.

In conclusion, while the above research provides a good foundation for explainable and lightweight anomaly detection at the edge, further advancements are required in adaptive learning, efficient explainability, and real-world validation to realize the full potential of intelligent edge security systems.

References

[1] Álvarez-López, C., González-Briones, A., & Li, T. (2026). Explainable AI and Multi-Agent Systems for Energy Management in IoT-Edge Environments: A State of the Art Review. *Electronics*, *15*(2), 385. https://doi.org/10.3390/electronics15020385

[2] Arisdakessian, S., Wahab, O. A., Mourad, A., Otrok, H., & Guizani, M. (2022). A survey on IoT intrusion detection: Federated learning, game theory, social psychology, and explainable AI as future directions. *IEEE Internet of Things Journal*, *10*(5), 4059-4092. https://doi.org/10.1109/jiot.2022.3203249

[3] Babalola, O., Raji, O. M. O., Akande, J. O., Abdulkareem, A. O., Anyah, V., Samson, A., & Folorunso, S. (2024). AI-powered cybersecurity in edge computing: Lightweight neural models for anomaly detection. *International Journal of Multidisciplinary Research and Growth Evaluation*, *5*(2), 1130-1138. https://doi.org/10.54660/.ijmrge.2024.5.2.1130-1138

[4] Bin Hulayyil, S., Li, S., & Saxena, N. (2025). Explainable AI-based intrusion detection in IoT systems. *Internet of Things, 31*, 101589. https://doi.org/10.1016/j.iot.2025.101589

[5] DeMedeiros, K., Hendawi, A., & Alvarez, M. (2023). A survey of AI-based anomaly detection in IoT and sensor networks. *Sensors*, *23*(3), 1352. https://doi.org/10.3390/s23031352

[6] Forough, J., Haddadi, H., Bhuyan, M., & Elmroth, E. (2024). Efficient anomaly detection for edge clouds: mitigating Data and resource constraints. *IEEE Access*, *12*, 171897-171910. https://doi.org/10.1109/access.2024.3492815

[7] Gummadi, A. N., Napier, J. C., & Abdallah, M. (2024). XAI-IoT: an explainable AI framework for enhancing anomaly detection in IoT systems. *IEEE Access*, *12*, 71024-71054. https://doi.org/10.1109/access.2024.3402446

[8] Gupta, S., & Singh, B. (2026). Lightweight ensemble learning based intrusion detection framework with explainable artificial intelligence. *Engineering Applications of Artificial Intelligence*, *163*, 112936. https://doi.org/10.1016/j.engappai.2025.112936

[9] Hleha, K., & Hol, V. (2025). XAI Optimization for Low-Latency Neural-Based Intrusion Detection Systems in Network Environments. *Bulletin of VN Karazin Kharkiv National*

University, series «Mathematical modeling. Information technology. Automated control systems», *66*, 19-36. https://doi.org/10.26565/2304-6201-2025-66-02

[10] Inuwa, M. M., & Das, R. (2024). A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks. *Internet of Things*, *26*, 101162. https://doi.org/10.1016/j.iot.2024.101162

[11] Jagatheesaperumal, S. K., Pham, Q. V., Ruby, R., Yang, Z., Xu, C., & Zhang, Z. (2022). Explainable AI over the Internet of Things (IoT): Overview, state-of-the-art and future directions. *IEEE Open Journal of the Communications Society*, *3*, 2106-2136. https://doi.org/10.1109/ojcoms.2022.3215676

[12] Khan, A., Hussain, M. A., & Anwer, F. (2025). A Hybrid Lightweight Deep Learning-Based Intrusion Detection Approach in IoT Utilizing Feature Selection & Explainable Artificial Intelligence. *IEEE Access*, *13*, 192451-192466. https://doi.org/10.1109/access.2025.3630753

[13] Li, X., Bi, S., Quan, Z., & Wang, H. (2022). Online cognitive data sensing and processing optimization in energy-harvesting edge computing systems. *IEEE Transactions on Wireless Communications*, *21*(8), 6611-6626. https://doi.org/10.1109/twc.2022.3151509

[14] Li, Z., Zhu, Y., & Van Leeuwen, M. (2023). A survey on explainable anomaly detection. *ACM Transactions on Knowledge Discovery from Data*, *18*(1), 1-54. https://doi.org/10.1145/3609333

[15] Mohale, V. Z., & Obagbuwa, I. C. (2025). A systematic review on the integration of explainable artificial intelligence in intrusion detection systems to enhancing transparency and interpretability in cybersecurity. *Frontiers in Artificial Intelligence*, *8*, 1526221. https://doi.org/10.3389/frai.2025.1526221

[16] Moustafa, N., Koroniotis, N., Keshk, M., Zomaya, A. Y., & Tari, Z. (2023). Explainable intrusion detection for cyber defences in the internet of things: Opportunities and solutions. *IEEE Communications Surveys & Tutorials*, *25*(3), 1775-1807. https://doi.org/10.1109/comst.2023.3280465

[17] Neupane, S., Ables, J., Anderson, W., Mittal, S., Rahimi, S., Banicescu, I., & Seale, M. (2022). Explainable intrusion detection systems (x-ids): A survey of current methods, challenges, and opportunities. *IEEE Access*, *10*, 112392-112415. https://doi.org/10.1109/access.2022.3216617

[18] Nwachukwu, C., Durodola-Tunde, K., & Akwiwu-Uzoma, C. (2024). AI-driven anomaly detection in cloud computing environments. *International Journal of Science and Research Archive*, *13*(2), 692-710. https://doi.org/10.30574/ijsra.2024.13.2.2184

[19] Quincozes, V. E., Quincozes, S. E., Kazienko, J. F., Gama, S., Cheikhrouhou, O., & Koubaa, A. (2024). A survey on IoT application layer protocols, security challenges, and the role of explainable AI in IoT (XAIoT). *International Journal of Information Security*, *23*(3), 1975-2002. https://doi.org/10.1007/s10207-024-00828-w

[20] Reis, M. J., & Serôdio, C. (2025). Edge AI for real-time anomaly detection in smart homes. *Future Internet*, *17*(4), 179. https://doi.org/10.3390/fi17040179

[21] Sharma, B., Sharma, L., Lal, C., & Roy, S. (2024). Explainable artificial intelligence for intrusion detection in IoT networks: A deep learning based approach. *Expert Systems with Applications*, *238*, 121751. https://doi.org/10.1016/j.eswa.2023.121751

Author Biography

Dr. Mohamed Adel Al-Shaher is an Assistant Professor specializing in artificial intelligence and advanced computational technologies. He obtained his Bachelor of Science (B.Sc.) degree from Al-Rafidain University in 2005. He later earned his Master's degree from Universiti Utara Malaysia in 2012, followed by a Ph.D. from Universitatea Politehnica of Bucharest in 2017. Dr. Al-Shaher's academic and research interests focus on deep learning and artificial intelligence, with particular emphasis on developing intelligent systems capable of solving complex real-world problems. His work explores modern AI techniques, including neural networks and data-driven models, to enhance automation, prediction accuracy, and decision-making processes. Throughout his academic career, Dr. Al-Shaher has contributed to advancing knowledge in the field of artificial intelligence by engaging in research activities and collaborating with international scholars. His work supports the ongoing development of smart technologies and innovative solutions across various domains.