Journal of Information Systems Engineering and Management
2026, 11(2s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Edge AI for Autonomous Device Security Management

Naveen Kumar
Independent Researcher, USA

ARTICLE INFO

Received: 04 March 2026

Accepted: 08 March 2026

ABSTRACT

For industrial, healthcare, consumer and enterprise environments, the number of connected devices has fundamentally expanded the attack surface of the modern enterprise beyond the reach of centralized security architectures. Resource constrained endpoints are generally connected intermittently via diverse communication protocols. What is needed is not simply another security model: it need autonomous real-time enforcement solutions that do not rely on centralized infrastructure. Edge Artificial Intelligence enables solutions to the structural imbalance by incorporating clever threat detection and response directly into the distributed device ecosystem. The hierarchical deployment model supports on-device inference, gateway-level data aggregation, and federated learning, thus preserving data privacy while minimizing the communication burden over heterogeneous deployment environments. Flexible machine learning models, including decision tree classifiers, temporal sequence models, and compressed neural network architectures, support accurate low-latency threat detection on devices with limited resources. The system implementation comprises four layers: threat detection, access control, vulnerability management and automated incident response, forming a closed-loop enforcement system, which does not rely on a central controller. The challenges posed by adversarial machine learning, model drift, regulatory compliance with law, and model explainability are addressed using adversarial training, federated learning, and interpretable model explanation frameworks. The security architecture is a self-adaptive edge AI fabric, which is suitable for the perpetually-evolving distributed environment.

Keywords: Edge Artificial Intelligence, Federated Learning, Intrusion Detection Systems, Adversarial Machine Learning, Autonomous Security Management.

1. Introduction

The rapid emergence of ecosystems of connected devices across industrial, healthcare, consumer and enterprise use cases has fundamentally changed the modern attack surface. However, customary security architectures based on centralized processing and perimeter-based controls are, by design, insufficiently capable of protecting distributed endpoints that operate outside customary network controls. Enabling the Internet of Things (IoT) provides a set of challenges that cause it to be particularly vulnerable to cyber attack, including limited resources, its heterogeneous composition and lack of security standards. One high-profile incident was the use of Mirai to attack an Internet domain name system services provider with a distributed denial-of-service (DDoS) attack of 1.2 terabits per second, at the time one of the largest DDoS attacks reported in Frustaci et al. [1].

Device-level vulnerabilities are not the only issue. Centralized security models become practically cumbersome at scale. Many endpoints have resource, bandwidth, communication protocol, and connectivity constraints, making it inefficient to use signature-based detection systems and cloud-based threat intelligence for device security analysis in most operational scenarios. The three layers of IoT system architecture (i.e., Perception Layer, Transportation Layer, and Application Layer) discussed in Frustaci et al.'s [1] security framework are exposed to different types of attacks, with the

Journal of Information Systems Engineering and Management
2026, 11(2s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Perception Layer being the most vulnerable due to the physical accessibility of devices in open areas, limited resources, and the absence of conventional protective software on constrained nodes.

Edge Artificial intelligence (Edge AI) seeks to overcome these limitations by moving the autonomous security analytics close to the data source, or its approximation, removing the need for centralized architecture with its latency and bandwidth costs. The convergence of edge computing and smart sensing is now viewed as a foundational approach to future large scale distributed environments. Zhou et al. [2] report an important increase in the publication of edge computing applied to transportation and automotive applications from 2011 (21) to 2019 (199), indicating that the convergence of edge intelligence with a complex distributed sensing architecture is increasing. They list real-time requirement, large scale sensing and high intelligence among the three major goals of edge computing for advanced distributed systems operations, all of which can be applied to managing device security at scale.

Existing IT security models face limitations in heterogeneous device ecosystems. Customary IT systems mostly consist of resource-rich devices in closed networks; IoT environments on the other hand contain heterogenous technologies and components with a wide attack surface and limited available security solutions [1]. Edge AI addresses this issue by locally processing telemetry data, learning normal operational behavior, and autonomously deciding security rules, without needing constant connectivity to the cloud. This article surveys the enabling architectures, AI techniques, security service layers, deployment challenges, and future outlook for autonomous management of security in Edge AI-based IoT devices.

2. Related Work

Existing research on security for distributed devices and edge intelligence has laid the theoretical and technical foundations for autonomous security systems. The foundational taxonomy of IoT security vulnerabilities in the Perception, Transportation and Application layers presented by Frustaci et al. [1] highlighted the amplification of risks faced by resource-constrained and protocol-heterogeneous large-scale device networks. Theoretical work by Bonomi et al. [4] on the Fog Computing model, where computation can be tiered between endpoint devices and the cloud based on latency requirements, can also be applied to tiered edge security architectures.

Ferrag et al. [5] proposed RDTIDS, a detection framework that improved state-of-the-art intrusion detection by utilizing hierarchical classifier topologies composed of both rule-based and decision tree classifiers. This improved the detection rate on data sets collected from large IoT networks. HeteroFL [3] is a framework for addressing heterogeneity in FL and enabling clients of varying resource capabilities to collaboratively learn a global model without centralizing their telemetry data. Cheng et al. [6] summarized model compression techniques (pruning, quantization, knowledge distillation) making feasible the deployment of state-of-the-art neural networks to devices constrained by processing, power, memory, bandwidth, and latency. Papernot et al. [9] showed adversarial attacks against deep neural networks. Ribeiro et al. [10] proposed interpretable local explanation techniques for automated security system decisions to be trusted by analysts.

3. Architectural Foundations of Edge AI Security

Edge AI security systems are based on a layered deployment architecture with intelligence distributed across the layers, often following the capabilities of the edge devices regarding processing and latency

Journal of Information Systems Engineering and Management
2026, 11(2s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

requirements. The architecture was first described by Bonomi et al. [4] in the definition of the Fog Computing model, an extension of cloud computing to the edge of the network via adding compute, storage and networking capabilities between the end devices and the data clouds. Key attributes of fog computing are low-latency, location-awareness, geographic dispersion, mobility and heterogeneity, mirroring requirements for distributed device security architecture [4]. Low latency network transmission and wireless communication being the dominant mode of access to fog computing architecture indicates that security enforcement is more effective at the fog network edge than at central clouds.

First, AI on-device does inference at the endpoint and detects without having to communicate back over the network which is very important in Industrial and healthcare applications where millisecond response times do not allow for round trip latency to the cloud. Near-device AI placed at gateways or edge servers, enables aggregation and correlation of telemetry data from clustered endpoints on an intermediate scale. Bonomi et al. [4] presented a fog computing architecture for smart grid applications. Fog collectors at the edge of the grid ingest data from sensors and actuators, performing fast processing in milliseconds to subseconds timescales for protection and control loops, and progressively higher-level analytics in seconds to days timescales.

The federated learning layer is a privacy-preserving way to collaboratively update the local model without disclosing raw telemetry data from the device. Diao et al. [3] proposed HeteroFL to solve the problem of training over clients with diverse computation and communication capabilities, which fits a heterogeneous IoT and edge device environment. HeteroFL enables heterogeneous local models of varying complexity to be federated into a global inference model by partitioning the subnetworks of the local models proportional to the hidden channel shrinkage ratio of the clients. When tested on 100 clients, the weak learner was able to achieve a test accuracy of 77.09% when trained alone with the reduced-complexity model 'e' on the CIFAR10 dataset. When federated with clients training larger models under 'a-e' setting, global accuracy under HeteroFL reached 90.29%, which is close to 91.19% accuracy under all clients training model 'a' setting. HeteroFL demonstrated competitive performance with lower communication rounds (800 compared to 1800 communication rounds for the LG-FedAvg) and proved to be more communication-efficient and stable with heterogeneous client distributions [3].

This multi-level security logic was tuned to each potential source threat. The heterogeneous 'a-e' architecture reached an IID accuracy of 99.46% with only 782K parameters and 40.5M FLOPs, compared to the full 'a' architecture which had 1.6M parameters and 80.5M FLOPs. This reduces the computation and communication overhead by 50% while not considerably compromising accuracy [3]. These advantages also apply to resource-constrained edge security applications that have limited bandwidth and processor budgets. In addition to avoiding sending all telemetry data back to the central platform, hierarchical edge architecture allows the device ecosystem to be secured even if the network is disrupted and without relying on the cloud.

Journal of Information Systems Engineering and Management
2026, 11(2s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

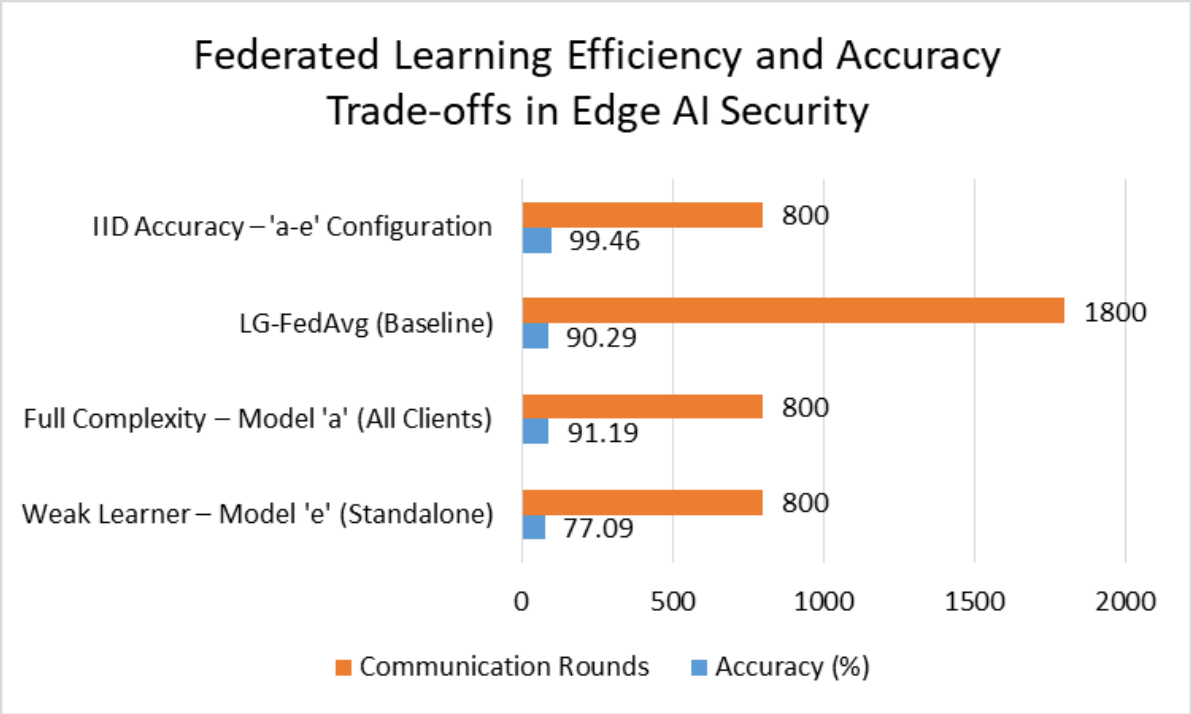

Figure 1: Federated Learning Efficiency and Accuracy Trade-offs in Edge AI Security [3]

4. Core Ai Techniques For Threat Detection

Smart security for Edge AI networks uses machine learning approaches that exploit the behavioral and temporal characteristics of the distributed activity of the devices. Decision tree and rules-based applied supervised classifiers have been used successfully in IoT network intrusion detection. Ferrag et al. [5] have developed a hierarchical intrusion detection system, RDTIDS, with REP Tree, JRip and Forest PA classifiers for implementation in a three-level fog computing network for the IoT. The proposed architecture was tested using the CICIDS2017 dataset that contains 2830743 records and has 79 features, as well as the BoT-IoT dataset that has more than 72000000 records and 46 features. The system has achieved an overall detection rate of 94.475% and 95.175%, accuracy of 96.665% and 96.995%, and false alarm rate of 1.145% and 1.120% for both datasets respectively [5]. The DDoS, PortScan, Heartbleed and Infiltration classifiers had a true positive rate of 99.879, 99.881, 100 and 100%, respectively. This shows the effectiveness of hierarchical classifier architecture in classifying multi-class attacks within fog and edge environments with limited resources [5]. The training time of 195.5s and testing time of 2.27s indicate that the models can be used in such environments where the computational and memory overhead should be limited [5].

Temporal sequence analysis applies to staged and persistent slow intrusions not visible in a single time snapshot. Long Short-Term Memory networks and temporal convolutional models track device activity over time and detect anomalies that only appear in a single snapshot when multiple snapshots are aggregated. They are especially relevant in cases such as industrial control systems and medical IoT devices, where threat patterns may evolve over time and not be detected by thresholding.

Model optimization is necessary in compute-constrained environments to make the model usable. Cheng et al. [6] summarized the major methods for deep neural networks compression and

Journal of Information Systems Engineering and Management
2026, 11(2s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

acceleration, which comprised parameter pruning and quantization, low-rank factorization, transferred convolutional filters, and knowledge distillation. As for quantization, it has been shown that quantizing a network to 8-bit precision can already give a meaningful speed-up in inference while losing negligible accuracy. The memory space and floating-point operations in ResNet-50 (over 95 MB of memory and 3.8 billion multiplications per image) can be compressed with more than 75% reduced model size and 50% computation speedup, while keeping functional equivalence, by setting all the redundant weights to zero [6]. Low-rank factorization such as CP decomposition in VGG-16 has achieved 2.05× speed-up at 2.75 compression rate and BN Low-rank decomposition has got 1.53× speed-up with 2.72× compression rate [6]. Knowledge distillation enables the transfer of learned representations from large teacher models to smaller student models deployed on edge-visible class processors, allowing for advanced threat detection in microcontrollers and other embedded systems with limited computational power and memory resources.

The successful combination of these approaches enables deployment of accurate, real-time threat detection at the edge, balancing the expressiveness of the machine learning model with the memory and power constraints of distributed edge-device networks.

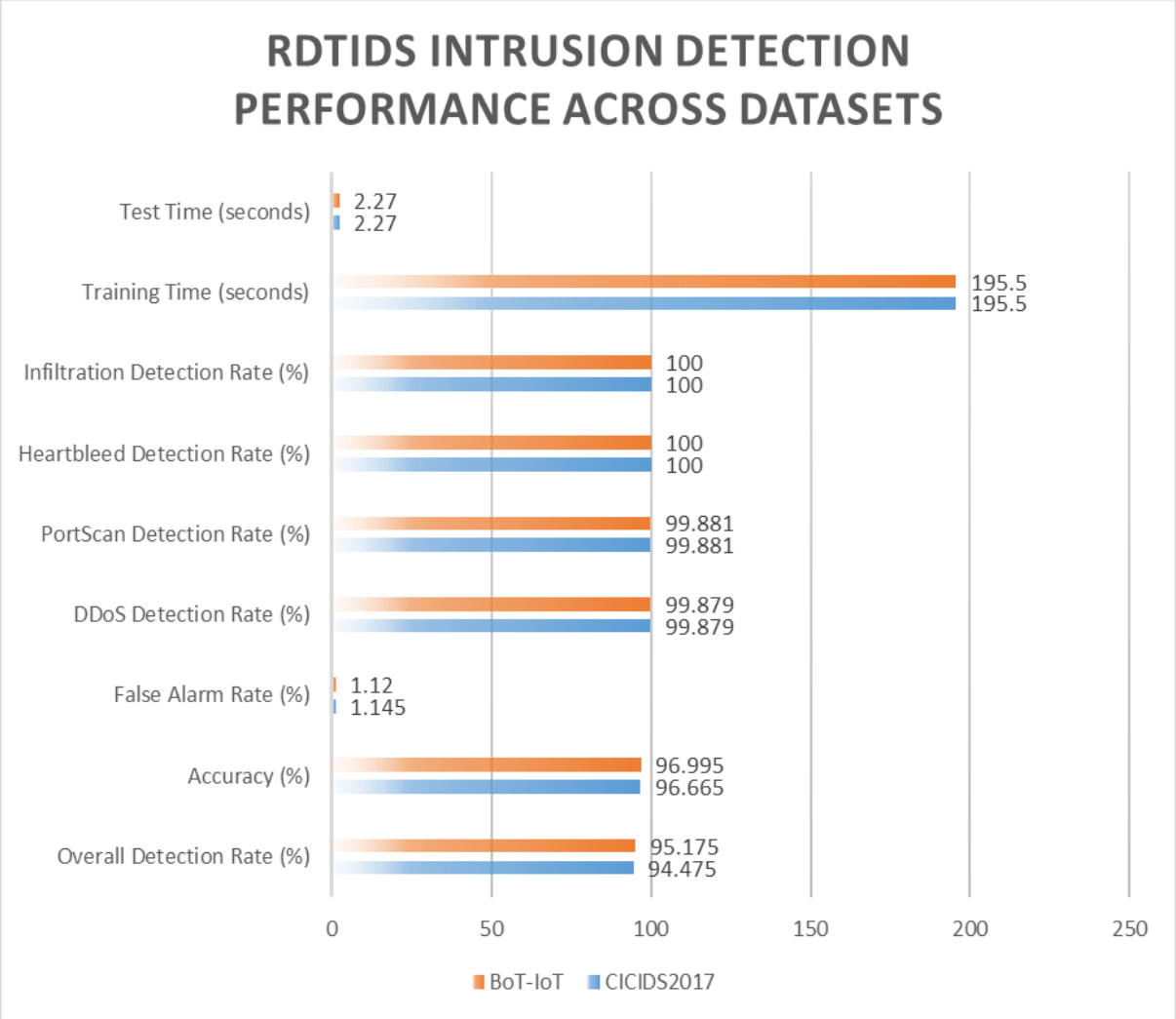

Figure 2: RDTIDS Intrusion Detection Performance Across Datasets [5]

Journal of Information Systems Engineering and Management
2026, 11(2s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5. Security Service Layers and Autonomous Operations

The closed-loop Edge AI security triangle has three phases: detect, decide and enforce. The detection aspect of Edge AI operates on device ecosystems and continuously monitors for malware, unauthorized process executions, ransomware activities and zero-day threats in real-time. The attack surface of a heterogeneous IoT ecosystem includes the perception, network and application layers of the IoT stack. Obaidat et al. [8] present an IoT attack taxonomy that is based on the IoT architecture layers. It consists of node tampering, malicious code injection and DoS attacks in the perception layer; MitM, sinkhole and routing information attacks in the network layer; and software vulnerability, code injection and data leakage attacks in the application layer. In addition, attacks on IoT devices were averaging out at around 5200 a month in 2017-2018, which means there is a need for systems that detect attacks at the source and in real-time rather than performing deferred analysis in a central location [8].

Access control services in the Edge AI security architecture can replace credential checking with continuous behavioral authentication. The risk score is based on context given by device signals. Protocol layer access control enforcement issues also exist in the cellular and wireless communications layers. The LTEInspector framework by Hussain et al. [7] used symbolic model checking and cryptographic protocol verification and identified 10 new attacks and 9 previously known attacks on the attach, detach and paging protocols of 4G LTE. 8 out of the 10 attacks were found to exist on a testbed, with a hardware cost of $3,900 [7]. Real attacks include the authentication synchronization failure attack, in which sequence numbers are desynchronized and a device is prevented from attaching, and the paging channel hijacking attack, due to the unobtrusive denial of incoming services. They are relevant when evaluating access control enforcement for edge-deployments made of mobile and IoT endpoints. For energy depletion attacks, a victim device can generate 1200 service request messages in one hour from paging messages that an opponent controls, versus an average of 156 paging responses to legitimate service requests over one hour, an approximately 8x increase in resources consumed [7].

Vulnerability management for the edge security stack includes risk-based patch sequencing and the ability to deploy virtual patches when firmware patching is not possible due to operational continuity constraints, which is common among legacy industrial and healthcare devices. Edge policy enforcement includes micro-segmentation and encryption compliance and protocol validation for heterogeneous device fleets. From the enforcement gap perspective, Obaidat et al. [8] highlight the lack of standardized authentication mechanisms for IoT devices, and argue that the wide range of memory constraints on IoT hardware (from tens of kilobytes of RAM on lowest end sensors to full OS-enabled devices) make it impossible to enforce cryptographic policies homogeneously without adaptive edge-based policy management.

Once a threat has been verified, automatic incident response initiates isolation of devices, telemetry forensics, and guided remediation workflows from the management console, without an operator having to take further action, which is useful in distributed environments, where lateral movement or continuing compromise within device clusters may acquire a foothold on the network.

Security Layer	Threat Category	Value
Network Layer—4G LTE	New Attacks Identified via LTEInspector	10
	Prior Attacks Confirmed via LTEInspector	9
	New Attacks Validated in Real Testbed	8

Journal of Information Systems Engineering and Management
2026, 11(2s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	Testbed Hardware Cost (USD)	3,900
	Service Requests Under Adversarial Condition (per hour)	1,200
	Service Requests Under Benign Baseline (per hour)	156
	Resource Consumption Amplification Factor	8×
IoT Device Layer	Average Monthly Attacks on IoT Devices (2017–2018)	5,200

Table 1: LTE Protocol Attack Identification and IoT Threat Volume Across Security Service Layers [7, 8]

6. Implementation: challenges and recommendations

The deployment of Edge AI security at scale presents a variety of challenges, including adversarial robustness, model reliability, compliance with regulation, and run-time performance. These challenges must be addressed under the often limited resource and latency requirements of distributed edge deployments.

As one of the more important challenges to the supervised detection models deployed on edge devices, Papernot et al. [9] show that one can generate adversarial samples consistently to force the victim deep neural networks to incorrectly classify certain input data into adversarial classes by combining forward derivative computation and adversarial saliency maps. Their created adversarial examples successfully tricked the model 97.10% of the time at 4.02% perturbation. Further qualitative experiments showed that the created perturbations were imperceptible to human observers, and that human classification accuracy remained above 90% for distortions up to 14.29%. These results show that edge AI threat detection models are structurally exposed to inputs as a result of under-generalization present in the finite training set used to train each model. Ensemble architectures and adversarial training methods reduce these vulnerabilities. For example, adding 18k adversarial examples to their training dataset resulted in a 7.2% decrease in adversarial algorithm success rate and a 37.5% increase in the distortion to obtain misclassification [9].

Model drift, caused by a change in device behavior or a new type of threat, requires continuous retraining pipelines and monitoring of detection performance throughout the lifespan of deployment to maintain the accuracy of detection. If no retraining is performed, the behavioral baselines learned during initial model training will deviate from reality, resulting in inaccurate detection.

Data residency regulations may also prohibit using a cloud or edge architecture in healthcare or industrial scenarios. Federated learning and on-device processing architectures, however, would allow for these scenarios because they do not require raw telemetry data to leave the device boundary (i.e., phones, edge devices). This allows for privacy while still enabling the group to study the problem together, as in distributed learning.

Explainability is also applied to security operations teams. Ribeiro et al. [10] proposed Local Interpretable Model-Agnostic Explanations (LIME) to explain the predictions of any classifier by approximating it locally with interpretable sparse linear models. For many classifiers and datasets, LIME achieves greater than 90% recall on true positive features. On average, LIME has an F1 score of 96.6% in trustworthiness tests involving logistic regression and support vector machine classifiers. LIME is considerably more accurate than other explanation approaches based on gradients or greedy methods when humans use it. In an experiment where LIME is used to pick a submodular set of

Journal of Information Systems Engineering and Management
2026, 11(2s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

examples, the better-generalizing classifier is chosen 89% of the time using LIME explanation, compared to 68% using greedy explanations. The use of these explanation frameworks supports decision-making in edge security operations by allowing analysts to examine, validate, and act upon automated detections without inspecting underlying model parameters.

All these operational requirements are achieved in real-time through hardware-accelerated inference and architectures designed specifically for edge-class processors, ensuring persistent protection without exceeding constrained deployments' computational budgets.

Metric	Baseline Value	Post-Mitigation Value
Adversarial Success Rate (%)	97.1	7.2 reduction after adversarial training
Average Input Features Modified per Sample (%)	4.02	A 37.5% increase in distortion required post-training
Human Imperceptibility Distortion Threshold (%)	14.29	Human accuracy sustained above 90 within threshold
LIME Feature Recall Across Classifiers %	Above 90	Greedy baseline below 65
LIME Trustworthiness F1 Score – LR and SVM (%)	96.6	Greedy baseline at 53.7
Correct Classifier Selection via SP-LIME (%)	89	Greedy with SP at 68

Table 2: Adversarial Sample Crafting Success Rates and LIME Explanation Fidelity Across Classifier Deployments [9, 10]

Conclusion

Edge computing necessities and AI security lead to an architectural shift from centralized monitoring after a security breach to self-automated execution at the origin point of the threat, embedded in and at the device level. In a world with billions of connected devices, perimeter-based and cloud-centric security architectures are both structurally inferior and introduce intolerable operational risk in industry, healthcare, or enterprise contexts. Hierarchical edge architectures, federated learning frameworks, and compressed neural networks can enable autonomous, continuous, and real-world security intelligence for heterogeneous device ecosystems, even when intermittent, and even with limited hardware resources. Behavioral anomaly detection, protocol-level access control, risk-based vulnerability management, and automated post-exploitation response now form a closed loop that can contain the attack on the device automatically without human intervention. Prioritization of adversarial robustness, model explainability, and regulatory compliance is no longer a luxury. These properties need to be architected within security solutions deployed at the edge. With an increasing number of connected devices deployed in infrastructure and hyper-sensitive enterprise operational environments, security that is enforced at the edge must be accurate, resilient, and transparent by default. Edge AI-driven security is an architectural model for creating resilient and self-adaptive security fabric that is tailored for the scale, heterogeneity, and operations of modern distributed environments.

Journal of Information Systems Engineering and Management
2026, 11(2s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

References

[1] Mario Frustaci et al., "Evaluating critical security issues of the IoT world: Present and Future challenges," IEEE Internet of Things Journal, vol. 5, no. 4, pp. 2483–2495, Aug. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8086136

[2] Xuan Zhou et al., "When Intelligent Transportation Systems Sensing Meets Edge Computing: Vision and Challenges," MDPI, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/20/9680

[3] Enmao Diao et al., "HETEROFL: COMPUTATION AND COMMUNICATION EFFICIENT FEDERATED LEARNING FOR HETEROGENEOUS CLIENTS," arXiv, 2021. [Online]. Available: https://arxiv.org/pdf/2010.01264

[4] Flavio Bonomi et al., "Fog Computing and Its Role in the Internet of Things," Proceedings of the first edition of the MCC workshop on Mobile cloud computing (August 2012) [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/2342509.2342513

[5] Mohamed Amine Ferrag et al., "RDTIDS: Rules and Decision Tree-Based Intrusion Detection System for Internet-of-Things Networks," MDPI, 2020. [Online]. Available: https://www.mdpi.com/1999-5903/12/3/44

[6] Yu Cheng et al., "A Survey of Model Compression and Acceleration for Deep Neural Networks," arXiv, 2020. [Online]. Available: http://arxiv.org/pdf/1710.09282

[7] Syed Rafiul Hussain et al., "LTEInspector: A Systematic Approach for Adversarial Testing of 4G LTE," Network and Distributed Systems Security (NDSS) Symposium 2018. [Online]. Available: https://par.nsf.gov/servlets/purl/10055689

[8] Muath A. Obaidat et al., "A Comprehensive and Systematic Survey on the Internet of Things: Security and Privacy Challenges, Security Frameworks, Enabling Technologies, Threats, Vulnerabilities and Countermeasures," MDPI, 2020. [Online]. Available: https://www.mdpi.com/2073-431X/9/2/44

[9] Nicolas Papernot, "The Limitations of Deep Learning in Adversarial Settings," 1st IEEE European Symposium on Security & Privacy, IEEE 2016. [Online]. Available: https://arxiv.org/pdf/1511.07528

[10] Marco Tulio Ribeiro et al., "Why Should I Trust You?" Explaining the Predictions of Any Classifier," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (August 2016). [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/2939672.2939778?.