Jana Nexus: Journal of Computer Science

ISSN(O): 3108-1916
Journal Homepage: jananexuscs.com
JN:JHST Volume 02 Issue 07 ***37-42***

| **RESEARCH ARTICLE**

Article DOI: 10.21474/JNCS01/145
DOI URL: http://dx.doi.org/10.21474/JNCS01/145

EDGE ARTIFICIAL INTELLIGENCE FOR SECURE AND INTELLIGENT INTERNET OF THINGS SYSTEMS: A COMPREHENSIVE REVIEW

Aarav Sharma, Mehak Khan, Rohan Verma and Sana Malik

Corresponding Author: Aarav Sharma

| **ABSTRACT**

The rapid expansion of the Internet of Things (IoT) has resulted in the deployment of billions of connected devices across healthcare, transportation, agriculture, manufacturing, smart cities, and domestic environments. These devices continuously generate large volumes of data that traditionally require transmission to centralized cloud platforms for processing and analysis. Although cloud computing provides substantial computational resources, dependence on remote servers can introduce latency, bandwidth consumption, privacy concerns, and security risks. Edge Artificial Intelligence (Edge AI) addresses these challenges by moving machine learning and intelligent decision-making closer to the location where data are generated. This paper presents a comprehensive review of Edge AI for secure and intelligent IoT systems, focusing on architectural principles, machine learning techniques, security mechanisms, privacy preservation, resource optimization, and practical applications. The paper discusses how lightweight models, model compression, federated learning, and distributed intelligence can improve the efficiency of resource-constrained IoT devices. It also examines challenges related to computational limitations, energy consumption, adversarial attacks, data privacy, model management, and interoperability. Finally, future research directions are discussed, including trustworthy Edge AI, collaborative edge-cloud architectures, efficient learning techniques, explainable models, and sustainable intelligent computing. The study concludes that the integration of Edge AI and IoT can provide responsive, privacy-aware, and scalable intelligent systems when security and resource constraints are considered throughout the system lifecycle.

| **KEYWORDS**

Edge Artificial Intelligence, Internet of Things, Machine Learning, Edge Computing, Cybersecurity, Federated Learning, Privacy, Intelligent Systems

|**ARTICLE INFORMATION**

RECEIVED: 23 May 2026 **ACCEPTED:** 29 June 2026 **PUBLISHED:** July 2026

Introduction:-

The Internet of Things has become one of the most influential developments in modern computer science. IoT enables physical objects to communicate through networks and exchange information using sensors, embedded processors, communication technologies, and software platforms. Applications range from smart homes and industrial automation to healthcare monitoring, intelligent transportation, environmental monitoring, and smart agriculture. The continuous growth of these applications has produced an enormous amount of data that must be collected, processed, stored, and analyzed.

Traditional IoT architectures frequently rely on cloud computing for data processing. In a cloud-centered architecture, information generated by sensors is transferred through a network to centralized data centers. Machine learning models can then analyze the information and return decisions or recommendations to the connected devices. While this approach provides significant computational power, it is not always appropriate for applications that require immediate responses. Network delays, bandwidth limitations, intermittent connectivity, and privacy concerns can negatively influence system performance.

Edge computing offers an alternative architecture in which computing resources are positioned closer to data-generating devices. Edge AI extends this concept by executing artificial intelligence and machine learning operations on edge devices or nearby edge servers. Instead of transferring every piece of information to the cloud, selected data can be processed locally. This approach can reduce communication requirements and support faster decision-making.

Recent research describes Edge AI as a significant computing paradigm for real-time analytics, privacy preservation, and bandwidth optimization. Research has also emphasized the importance of lightweight machine learning, TinyML, federated learning, and specialized hardware for resource-constrained environments.Security is particularly important because IoT environments contain large numbers of heterogeneous devices. A compromised sensor, gateway, or edge server can potentially affect an entire distributed system. Conventional security mechanisms may be insufficient when devices have limited memory, processing capacity, and energy. Consequently, intelligent security mechanisms that can operate efficiently at the edge have become an important research direction.This paper reviews the relationship between Edge AI and IoT, examines major technical approaches, discusses security and privacy challenges, and identifies future research opportunities.

Background of Edge AI and IoT:-

Internet of Things:-

The Internet of Things refers to a network of physical objects equipped with sensors, communication interfaces, computing capabilities, and software. These devices can collect information from their surroundings and exchange data with other devices or computing platforms.

A typical IoT ecosystem consists of four major components:-

1. **Sensing layer:** Sensors collect information such as temperature, location, motion, pressure, images, or physiological measurements.
2. **Communication layer:** Networks transfer information between devices, gateways, edge nodes, and cloud platforms.
3. **Processing layer:** Computing systems analyze collected data and generate useful information.
4. **Application layer:** Processed information is used by applications, organizations, or end users.

The number and diversity of IoT devices make centralized management increasingly difficult. Devices can differ significantly in hardware architecture, operating systems, communication protocols, energy availability, and computational resources.

Edge Computing:-

Edge computing moves computation and storage resources closer to end users and data sources. Rather than transmitting all data to distant cloud data centers, edge nodes perform selected processing operations locally.

The major advantages include:-

- Reduced network latency
- Lower bandwidth requirements
- Faster response times
- Improved availability
- Greater opportunities for local data processing
- Potential improvements in privacy

Edge computing is particularly valuable for applications where decisions must be made rapidly. Autonomous systems, industrial monitoring, intelligent transportation, and real-time healthcare monitoring are examples where excessive communication delay may reduce system effectiveness.

Edge Artificial Intelligence:-

Edge AI combines artificial intelligence with edge computing. Machine learning models are deployed directly on edge devices, gateways, mobile systems, or nearby edge servers.

A basic Edge AI architecture can be represented as:-

IoT Devices → Edge Node → Cloud Platform

IoT devices collect data. The edge node performs preliminary processing, inference, filtering, or anomaly detection. The cloud can remain responsible for computationally intensive tasks, long-term storage, global model training, and system-wide analytics.This hybrid architecture allows organizations to balance computational requirements between local and centralized infrastructure.

Literature Review:-

Recent research demonstrates growing interest in combining artificial intelligence, IoT, and edge computing. A 2026 structured review of AI for cybersecurity in IoT-edge systems emphasized the importance of considering not only model performance but also datasets, evaluation practices, and deployment constraints. Another recent review of Edge AI analyzed the evolution from distributed computing toward modern approaches such as TinyML and federated learning. It identified resource limitations,

security vulnerabilities, and model-management complexity as important challenges for practical deployment.Research in cybersecurity also increasingly examines AI as a tool for detecting vulnerabilities, identifying abnormal behavior, and supporting security testing. However, AI-based approaches do not eliminate the need for human expertise because different classes of vulnerabilities require different analytical techniques.Current Computer Science research topics also emphasize trustworthy AI, AI-enabled cybersecurity, IoT security, privacy-preserving machine learning, and edge intelligence. These trends indicate that future intelligent systems are likely to integrate several of these technologies rather than treating them as isolated fields.

Proposed Edge AI Architecture:-
A secure Edge AI system for IoT can be organized into five functional layers.

Data Acquisition Layer:-
The first layer contains sensors and IoT devices. These devices collect information from physical environments. Depending on the application, data may include environmental measurements, images, audio signals, machine conditions, network traffic, or device status.Because IoT devices are often resource constrained, the data acquisition layer should minimize unnecessary computation and communication.

Edge Processing Layer:-
The edge processing layer contains gateways, embedded processors, microcomputers, or specialized AI accelerators. This layer performs data preprocessing, feature extraction, classification, anomaly detection, and other inference tasks.Local processing can reduce the quantity of information transmitted to cloud servers. For example, instead of continuously transmitting raw sensor data, an edge device can identify an abnormal event and send only the relevant information.

Security Layer:-
Security mechanisms should operate throughout the architecture. Authentication, authorization, encryption, secure boot, intrusion detection, and anomaly detection can help protect devices and communication channels.AI-based intrusion detection systems can analyze network or device behavior and identify patterns that differ from normal operation. Such systems can complement traditional security mechanisms.

Cloud Coordination Layer:-
The cloud remains useful even when intelligence is distributed across the edge. It can provide large-scale model training, long-term data storage, centralized management, and global analytics.A practical system therefore does not necessarily require choosing between cloud and edge computing. Instead, cloud and edge resources can cooperate according to computational requirements, latency requirements, privacy policies, and network conditions.

Application Layer:-
The application layer provides services to end users. Examples include smart-city management, industrial monitoring, healthcare support, agricultural monitoring, and intelligent transportation.

Machine Learning Techniques for Edge Systems:-
Lightweight Deep Learning:-
Large deep learning models may require substantial memory and computational power. Such requirements can make direct deployment on small IoT devices difficult.Lightweight neural network architectures can reduce computational requirements while maintaining useful predictive performance. Efficient architectures are therefore important for Edge AI applications.

Model Compression:-
Model compression reduces the size and computational cost of machine learning models.

Common approaches include:-
- Quantization
- Pruning
- Knowledge distillation
- Weight sharing
- Architecture optimization

Quantization converts model parameters from higher-precision representations into lower-precision formats. This can reduce memory requirements and computational cost.Pruning removes parameters that contribute relatively little to model performance. Knowledge distillation trains a smaller model to reproduce useful behavior learned by a larger model.

Federated Learning:-
Federated learning allows multiple devices or edge nodes to participate in model training without directly sharing their raw datasets.

In a simplified federated learning process:-
1. A central model is distributed to participating devices.
2. Each device trains the model using local data.
3. Devices send model updates rather than raw data.
4. A coordinating server aggregates the updates.
5. The improved model is redistributed.

This approach can reduce direct data sharing and is particularly relevant when information is sensitive or expensive to transfer. However, federated learning introduces additional challenges, including communication overhead, malicious model updates, heterogeneous devices, and statistical differences between local datasets.

Security Challenges in Edge AI:-
Device-Level Attacks:-
IoT devices can be attractive targets because they are often deployed in large numbers and may have limited security capabilities. Attackers who gain control of an individual device may attempt to use it to access other parts of the system.
Secure authentication and device management are therefore essential.

Adversarial Machine Learning:-
AI models can be manipulated through specially designed inputs. An attacker may attempt to cause an AI system to classify an input incorrectly or behave unexpectedly.Edge AI systems require defensive techniques that account for adversarial manipulation while maintaining acceptable computational efficiency.

Model Theft:-
Machine learning models deployed at the edge may be physically accessible to unauthorized individuals. If model parameters are exposed, an attacker may attempt to copy the model or analyze its behavior.Model encryption, secure hardware, access controls, and protected execution environments can reduce these risks.

Data Privacy:-
IoT devices can collect sensitive information. Continuous transmission of raw information to centralized servers may increase privacy concerns.Local processing can reduce unnecessary data transmission. Federated learning can further reduce direct sharing of raw datasets, although it does not automatically eliminate all privacy risks.

Network Security:-
Communication between devices, edge nodes, and cloud servers must be protected against interception, manipulation, and unauthorized access.Encryption and authentication should be combined with continuous monitoring and anomaly detection.

Applications of Edge AI:-
Smart Healthcare:-
Edge AI can support healthcare monitoring systems by processing information close to sensors. Wearable devices can locally analyze signals and identify potentially important patterns before sending selected information to healthcare platforms.
Local processing can reduce latency and limit unnecessary transmission of personal data.

Smart Cities:-
Smart-city environments contain cameras, traffic sensors, environmental sensors, and connected infrastructure. Edge AI can process information locally to support traffic management, energy optimization, environmental monitoring, and public infrastructure management.Instead of sending every sensor reading to a centralized system, edge devices can perform preliminary analysis and transmit relevant events.

Smart Agriculture:-
Agricultural IoT systems can use sensors to monitor soil conditions, temperature, humidity, irrigation requirements, and crop environments. Edge AI can analyze this information locally and support timely decisions.For rural environments with limited connectivity, local intelligence can be particularly useful because systems may continue operating even when cloud connectivity is temporarily unavailable.

Industrial Automation:-
Industrial environments generate continuous streams of machine and production data. Edge AI can support predictive maintenance by identifying unusual machine behavior.
Early identification of abnormal patterns can help organizations schedule maintenance and reduce unexpected equipment failures.

Intelligent Transportation:-
Connected vehicles and transportation infrastructure require rapid processing. Edge AI can support traffic monitoring, vehicle detection, road-condition analysis, and other intelligent transportation applications.The ability to process information close to its source can reduce communication delays.

Performance Considerations:-
The performance of an Edge AI system depends on several factors.

Latency:-
Latency represents the time required to collect data, process it, and generate a response. Edge computing can reduce latency by avoiding unnecessary communication with distant cloud servers.

Energy Consumption:-
Many IoT devices operate using batteries or limited power sources. AI algorithms therefore need to be computationally efficient.Hardware acceleration, model compression, adaptive inference, and efficient scheduling can contribute to lower energy consumption.

Computational Capacity:-
Edge devices vary significantly in processing capability. A small microcontroller cannot execute the same models as a powerful edge server.Consequently, model selection should consider device memory, processor capability, available accelerators, and application requirements.

Bandwidth:-
Continuous transmission of raw IoT data can consume significant network bandwidth. Local processing can reduce the volume of information transmitted to centralized systems.

Reliability:-
Edge systems should remain functional when network connectivity is unstable. Local inference can support operational continuity by allowing devices to make certain decisions without continuous cloud access.

Challenges and Limitations:-
Despite its advantages, Edge AI faces several limitations.First, edge devices generally have fewer computational resources than cloud servers. This creates difficulties when deploying large and computationally expensive AI models.Second, distributed deployment increases system-management complexity. Organizations may need to update and monitor large numbers of models across heterogeneous devices.Third, security risks can increase because edge devices are physically distributed and may operate in environments that are difficult to control.Fourth, AI models may behave differently when deployed on devices with different hardware and software configurations. Reproducibility and standardization are therefore important.Fifth, privacy-preserving approaches can introduce computational or communication overhead. Federated learning, for example, reduces direct data sharing but requires repeated communication between participating devices and coordinating servers.Finally, AI models may produce incorrect predictions. In applications involving critical decisions, human oversight, appropriate validation, and explainability remain important.

Future Research Directions:-
Trustworthy Edge AI:-
Future research should focus on developing Edge AI systems that are reliable, transparent, secure, and explainable. Trustworthiness will become increasingly important as AI systems participate in real-world decision-making.

Edge-Cloud Collaboration:-
Rather than completely replacing cloud computing, future systems are likely to combine edge and cloud resources dynamically. Research can explore intelligent workload allocation in which computational tasks are assigned according to latency, energy, privacy, and resource requirements.

TinyML:-
TinyML focuses on deploying machine learning models on highly resource-constrained microcontrollers and embedded devices. Efficient TinyML systems can expand intelligent computing to devices that previously lacked sufficient resources for machine learning.

Privacy-Preserving AI
Future systems should investigate stronger privacy mechanisms that can operate efficiently on edge hardware. Combining federated learning with secure aggregation, differential privacy, and other privacy-enhancing technologies represents an important research direction.

Sustainable AI:-
AI computation can consume significant energy, particularly during large-scale training. Future Edge AI research should investigate energy-aware model architectures, efficient hardware, adaptive inference, and workload scheduling.

Explainable AI:-
When Edge AI systems make important decisions, users may need to understand why a particular prediction was produced. Explainable AI can improve transparency and support debugging, auditing, and user trust.

Conclusion:-

The convergence of Edge AI and the Internet of Things represents an important direction in modern Computer Science. IoT systems continuously generate large volumes of data, while conventional cloud-centered architectures may face challenges related to latency, bandwidth, privacy, and connectivity. Edge AI addresses these issues by bringing intelligent processing closer to the source of data.

The integration of lightweight machine learning, model compression, federated learning, and edge-cloud collaboration can enable intelligent applications across healthcare, smart cities, agriculture, manufacturing, and transportation. At the same time, security, privacy, resource limitations, model management, and reliability remain important challenges.

Future research should therefore move beyond simply improving AI model accuracy. Researchers should consider the complete system, including hardware limitations, energy consumption, communication requirements, privacy, cybersecurity, explainability, and long-term maintainability. A trustworthy Edge AI ecosystem will require cooperation among machine learning researchers, computer scientists, cybersecurity specialists, hardware engineers, and application-domain experts.
Overall, Edge AI has significant potential to transform IoT from a predominantly data-collection technology into a distributed intelligent computing infrastructure. Continued research into secure, efficient, privacy-preserving, and sustainable Edge AI will be essential for realizing this potential.

References:-

1. Recent research on Edge Artificial Intelligence, TinyML, federated learning, and edge-cloud architectures. Engineering Applications of Artificial Intelligence, 2026.
2. Recent structured review of artificial intelligence for cybersecurity in IoT-edge systems, including AI methods, datasets, evaluation, and deployment challenges. Electronics, 2026.
3. Recent review of AI-based cybersecurity paradigms, applications, deployment considerations, and emerging trends. Algorithms, 2026.
4. Recent research on edge-based artificial intelligence, including hardware, software, lightweight architectures, and deployment frameworks. Engineering Applications of Artificial Intelligence, 2026.
5. Frontiers in Computer Science. Research topics covering computer security, trustworthy AI, IoT security, edge computing, and AI-enabled cybersecurity. 2026.
6. Recent research topics in Computer Science covering AI, cybersecurity, edge computing, IoT security, federated learning, and green computing. 2026.