IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double- Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

Machine Learning and Data Mining Techniques for Efficient IoT Data Transmission

Shwetha H L[1] and Dr. Deepak Kumar Sharma[2]
[1]Research Scholar, [2]Professor
Department of Computer Science
Sunrise University, Alwar (Rajasthan)

Abstract: *The rapid expansion of the Internet of Things (IoT) has resulted in the generation of enormous volumes of heterogeneous and continuously changing data from interconnected sensors, devices, and intelligent systems. Efficient transmission of such data is essential for reducing communication overhead, conserving energy, minimizing latency, improving network reliability, and supporting real-time IoT applications.*

Conventional data transmission techniques often face limitations because of restricted bandwidth, dynamic network conditions, redundant data generation, device heterogeneity, and limited computational resources at edge devices. Machine learning (ML) and data mining techniques provide promising solutions by enabling IoT networks to identify data patterns, predict network conditions, detect redundant information, optimize routing, and intelligently determine transmission priorities. This research paper examines the application of supervised, unsupervised, reinforcement, and deep learning approaches together with data mining techniques for improving IoT data transmission efficiency. The paper discusses how classification, clustering, association analysis, anomaly detection, feature selection, and predictive modelling can reduce unnecessary transmissions and improve the utilization of communication resources.

Keywords: Internet of Things, IoT Data Transmission, Data Aggregation, Intelligent Routing, Edge Computing, Energy Efficiency, Network Optimization, Anomaly Detection.

I. INTRODUCTION

The Internet of Things has emerged as an important technological paradigm in which physical objects, sensors, machines, vehicles, appliances, and infrastructure are connected through communication networks to collect and exchange information. The increasing deployment of IoT devices in smart cities, healthcare, agriculture, industrial automation, transportation, environmental monitoring, and smart homes has created a substantial demand for reliable and efficient data transmission. Unlike traditional computer networks, IoT environments generally contain large numbers of resource-constrained devices that operate with limited battery power, storage capacity, processing capability, and communication bandwidth. Consequently, transmitting every piece of generated information directly to centralized servers may result in congestion, high energy consumption, redundant traffic, and increased latency.

Machine learning offers the ability to make communication decisions based on historical and real-time data. Instead of relying exclusively on predefined rules, ML models can learn relationships between network conditions, device behaviour, traffic patterns, energy levels, and transmission requirements. For example, predictive models can estimate future network congestion, classification algorithms can identify important data, clustering algorithms can group similar sensor readings, and reinforcement learning can dynamically select communication routes. Data mining complements these techniques by extracting meaningful patterns and relationships from large IoT datasets.

Efficient IoT transmission is particularly important because communication is often one of the largest consumers of energy in battery-powered sensor networks. If redundant or insignificant data can be identified before transmission, the number of communication operations can be reduced considerably. Similarly, intelligent data aggregation can combine multiple sensor observations into a smaller representation without substantially reducing information quality. ML-based anomaly detection can also distinguish unusual observations from routine measurements and assign appropriate transmission priorities.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double- Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

The integration of ML, data mining, edge computing, and IoT therefore represents a promising direction for developing intelligent communication systems. Rather than treating transmission as a simple data-forwarding process, intelligent IoT networks can dynamically analyse information and determine what should be transmitted, when it should be transmitted, and through which route it should be transmitted. This paper examines these approaches and evaluates their potential contribution to efficient IoT data transmission.

II. IOT DATA TRANSMISSION CHALLENGES AND REQUIREMENTS

IoT data transmission involves several interconnected challenges. The first is the enormous volume of data generated by heterogeneous devices. Sensors may continuously produce temperature, pressure, motion, location, video, audio, health, or industrial measurements. Transmitting all generated data can create unnecessary traffic and consume considerable energy. The second challenge is data redundancy. Multiple nearby sensors can generate highly correlated measurements, meaning that transmitting every observation may provide little additional information.

Network conditions also change continuously. Wireless interference, device mobility, varying traffic loads, node failures, and fluctuating energy levels can affect transmission quality. Static routing mechanisms may therefore become inefficient when network conditions change. IoT applications additionally have different quality-of-service requirements. Healthcare monitoring may prioritize low latency and reliability, whereas environmental monitoring may emphasize energy conservation.

Security and privacy represent further concerns. IoT data can contain sensitive information, and intelligent transmission systems must prevent unauthorized access while maintaining communication efficiency. ML models themselves may also become targets of adversarial attacks or may produce inaccurate decisions when trained using poor-quality or biased data.

Challenge	Effect on IoT Transmission	ML/Data Mining Contribution
High data volume	Network congestion and bandwidth consumption	Data reduction and predictive transmission
Redundant sensor data	Unnecessary communication	Clustering and correlation analysis
Limited battery power	Reduced network lifetime	Energy-aware prediction and routing
Dynamic topology	Route instability	Reinforcement learning
Network congestion	Increased latency and packet loss	Traffic prediction
Noisy data	Incorrect transmission decisions	Data preprocessing and anomaly detection
Heterogeneous devices	Difficult resource management	Classification and adaptive models
Security threats	Data loss and unauthorized access	Anomaly and intrusion detection
Real-time requirements	Delayed decision-making	Edge-based intelligent processing
Large-scale deployment	High management complexity	Automated learning and optimization

An efficient transmission framework should consequently reduce communication overhead while maintaining data accuracy and application requirements. Important performance indicators include packet delivery ratio, end-to-end delay, throughput, energy consumption, network lifetime, bandwidth utilization, data reduction rate, and computational overhead.

III. MACHINE LEARNING TECHNIQUES FOR EFFICIENT IOT DATA TRANSMISSION

Machine learning can support IoT transmission at multiple stages, including data classification, route selection, traffic prediction, resource allocation, and transmission scheduling. Supervised learning techniques can be trained using labelled network and sensor data to predict specific outcomes. Classification algorithms such as decision trees, support vector machines, random forests, and artificial neural networks can determine whether sensor information is normal,

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double- Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 5, July 2026 Impact Factor: 8.2

redundant, critical, or anomalous. Regression models can estimate network traffic, energy consumption, or transmission delay.

Unsupervised learning is useful when labelled IoT datasets are unavailable. Clustering algorithms can group sensors or data points according to similarity. Sensors producing highly correlated observations can be represented through cluster-based aggregation, reducing the amount of information that needs to be transmitted. K-means, hierarchical clustering, and density-based approaches can therefore support efficient data aggregation and network organization.

Reinforcement learning is particularly relevant to dynamic routing and transmission scheduling. An agent can observe network states such as available energy, link quality, congestion, and queue length and then select actions that maximize a defined reward. Over time, the agent can learn communication policies that balance competing objectives such as energy consumption, latency, and reliability.

Deep learning can process complex and high-dimensional IoT data. Recurrent neural networks and related temporal models can learn patterns in time-series sensor data and predict future observations. If a future sensor value can be accurately predicted at the receiver, unnecessary transmission of predictable values may be avoided. Autoencoder-based models can also support dimensionality reduction and anomaly detection.

ML Technique	Major IoT Application	Transmission Benefit
Decision Tree	Data classification	Rapid transmission prioritization
Random Forest	Traffic and anomaly classification	Improved decision accuracy
Support Vector Machine	Sensor/event classification	Efficient identification of important data
K-means	Sensor/data clustering	Reduced redundant transmissions
Hierarchical Clustering	Network organization	Efficient cluster-based communication
Reinforcement Learning	Dynamic routing	Adaptive route selection
Deep Neural Networks	Traffic prediction	Improved resource planning
Recurrent Neural Networks	Time-series prediction	Reduced predictable data transmission
Autoencoders	Compression/anomaly detection	Data reduction and filtering
Regression Models	Delay/energy prediction	Better transmission scheduling

IV. DATA MINING TECHNIQUES FOR IOT DATA REDUCTION AND OPTIMIZATION

Data mining provides systematic methods for discovering useful patterns from large volumes of IoT-generated information. One important application is data preprocessing, where missing values, noise, inconsistencies, and duplicate observations are identified before transmission. Removing irrelevant information at the source or edge can substantially reduce network traffic.

Association analysis can identify relationships among sensor events. For example, if certain sensor conditions frequently occur together, the system can use these relationships to determine whether every individual observation needs to be transmitted. Correlation analysis can similarly identify highly related sensor readings. When two sensors provide almost identical information, transmitting one representative value may be sufficient for selected applications.

Data clustering is another important technique. Sensor nodes can be grouped according to geographic proximity, measurement similarity, energy status, or communication characteristics. A cluster-head can aggregate information received from cluster members and forward a compact representation to an edge server or gateway. This approach can reduce the number of long-distance transmissions.

Sequential pattern mining can discover recurring temporal patterns in IoT data. Such patterns can support predictive transmission, in which data is transmitted only when a new observation significantly differs from an expected pattern. Outlier mining and anomaly detection can identify unusual events that deserve immediate transmission, while routine observations can be compressed, aggregated, or transmitted less frequently.

Data Mining Technique	Function	Expected Transmission Improvement
Data preprocessing	Removes noise and inconsistencies	Lower unnecessary traffic

Correlation analysis	Identifies related sensor readings	Reduced duplicate data
Clustering	Groups similar devices/data	Efficient aggregation
Association mining	Finds relationships among events	Improved transmission decisions
Sequential pattern mining	Detects temporal patterns	Predictive transmission
Outlier detection	Identifies unusual observations	Priority-based transmission
Feature selection	Removes irrelevant attributes	Reduced data size
Data summarization	Creates compact representations	Lower bandwidth requirement
Classification mining	Categorizes observations	Selective transmission
Predictive mining	Forecasts future values	Reduced transmission frequency

V. INTEGRATED ML AND DATA MINING FRAMEWORK FOR EFFICIENT IOT TRANSMISSION

An integrated intelligent transmission framework can combine sensing, preprocessing, data mining, machine learning, edge computing, and communication optimization. At the sensing layer, IoT devices collect raw observations. Initial preprocessing can remove obvious noise and duplicate information. Data mining algorithms can then identify correlations, clusters, and relevant features. The resulting information can be supplied to ML models for prediction, classification, anomaly detection, or routing decisions.

Edge computing can play a significant role because processing data close to its source reduces the need to send raw data to distant cloud servers. An edge node can aggregate sensor readings, execute lightweight ML models, detect abnormal events, and transmit only relevant information. This architecture can decrease latency and bandwidth consumption while improving responsiveness for real-time applications.

The transmission decision can be formulated as an optimization problem involving several objectives. A system may attempt to minimize energy consumption and delay while maximizing reliability and data quality. The ML model can estimate the expected value of transmitting a particular observation, while data mining can determine whether the information is redundant or significant. Based on these results, the system can select among transmission, aggregation, compression, prediction, or local processing.

Framework Component	Primary Function	Contribution
IoT sensors	Data generation	Collect real-time observations
Preprocessing layer	Cleaning/filtering	Removes irrelevant information
Data mining layer	Pattern discovery	Identifies redundancy and relationships
ML layer	Prediction/classification	Supports intelligent decisions
Edge layer	Local computation	Reduces cloud communication
Routing layer	Path selection	Improves reliability and energy efficiency
Aggregation layer	Data summarization	Reduces transmitted volume
Security layer	Threat detection	Protects transmission
Cloud layer	Large-scale analysis	Supports long-term learning
Feedback mechanism	Model adaptation	Responds to changing conditions

The performance of such a framework should be evaluated using measurable indicators. Energy consumption indicates the cost of communication, while network lifetime measures how long the IoT infrastructure remains operational. Packet delivery ratio reflects reliability, and end-to-end delay indicates responsiveness. Data reduction ratio measures the percentage of information eliminated through filtering, aggregation, compression, or prediction.

VI. CONCLUSION

Machine learning and data mining provide powerful mechanisms for improving the efficiency of IoT data transmission. Traditional IoT communication approaches often transmit large quantities of raw data without considering redundancy, information importance, network conditions, or future data patterns. Intelligent approaches can overcome these limitations by analysing data before transmission and adapting communication decisions according to network and application requirements.

Supervised learning can support data classification and transmission prioritization, while unsupervised learning can identify clusters and redundant observations. Reinforcement learning can provide adaptive routing and scheduling in dynamic environments, and deep learning can support sophisticated prediction and feature extraction. Data mining techniques further strengthen these capabilities by discovering correlations, associations, temporal patterns, and anomalies within IoT data.

The integration of these approaches with edge computing is particularly promising. Processing information closer to IoT devices can reduce communication distance, bandwidth requirements, latency, and unnecessary cloud traffic. A future intelligent IoT network can therefore operate as a data-aware system in which only meaningful, non-redundant, and application-relevant information is transmitted.

Despite these advantages, several issues require further investigation. ML models can require substantial training data and computational resources, which may be difficult for highly constrained IoT devices. Model accuracy can decline when network environments change, and security threats can compromise both IoT data and learning processes. Future research should focus on lightweight learning models, federated and privacy-preserving learning, explainable AI, energy-aware model deployment, secure edge intelligence, and multi-objective optimization. Overall, the combination of machine learning and data mining represents an important pathway toward scalable, adaptive, reliable, and energy-efficient IoT data transmission systems.

REFERENCES

1. Begum, B. A., & Nandury, S. V. (2023). Data aggregation protocols for WSN and IoT applications – A comprehensive survey. *Journal of King Saud University – Computer and Information Sciences, 35*, 651–681.
2. Ding, Q., Zhu, R., Liu, H., & Ma, M. (2021). An overview of machine learning-based energy-efficient routing algorithms in wireless sensor networks. *Electronics, 10*(13), 1539.
3. Mutombo, J., & Hong, J. (2021). EER-RL: Energy-efficient routing based on reinforcement learning. *Mobile Information Systems, 2021*, 5589145.
4. Sanyal, S., & Zhang, P. (2018). Improving quality of data: IoT data aggregation using device to device communications. *IEEE Access, 6*, 67830–67840.
5. *Data aggregation mechanisms on the Internet of Things: A systematic literature review*. (2021). *Internet of Things, 15*, 100427.
6. *Performance comparison of machine learning algorithms for data aggregation in social Internet of Things*. (2021). *Global Transitions Proceedings, 2*(2), 212–219.
7. *An energy-efficient routing protocol for the Internet of Things networks based on geographical location and link quality*. (2021). *Computer Networks, 193*, 108116.
8. *Towards developing a machine learning-metaheuristic-enhanced energy-sensitive routing framework for the Internet of Things*. (2023). *Microprocessors and Microsystems, 96*, 104747.
9. *Offloading and transmission strategies for IoT edge devices and networks*. (2019). *Sensors, 19*(4), 835.
10. Oliveira, F., Costa, D. G., Assis, F., & Silva, I. (2024). Internet of Intelligent Things: A convergence of embedded systems, edge computing and machine learning. *Internet of Things, 26*, 101153.
11. *A survey of machine learning in edge computing: Techniques, frameworks, applications, issues, and research directions*. (2024). *Technologies, 12*(6), 81.
12. *IoT-Edge technology based cloud optimization using artificial neural networks*. (2024). *Microprocessors and Microsystems, 106*, 105049.
13. *Energy-efficient routing optimization for underwater Internet of Things using hybrid Q-learning and predictive learning approach*. (2024). *Procedia Computer Science, 235*, 45–55.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double- Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

14. Tan, N. D., & Nguyen, V.-H. (2025). Machine learning meets IoT: Developing an energy-efficient WSN routing protocol for enhanced network longevity. *Wireless Networks, 31*(4), 3127–3141.
15. Shekar, K., Reddy, N. R., Arvind, S., Kumar, T. V. S., Kodukula, S., et al. (2025). Implementation of novel learning based energy efficient routing protocols in wireless sensor networks for Internet of Things use cases. *Discover Computing, 28*, 190.
16. Tan, N. D., Quy, N. M., & Nguyen, V.-H. (2025). EE-AIRP: An AI-enhanced energy-efficient routing protocol for IoT-enabled WSNs. *Computer Networks*, 111770.
17. Thakur, S., Sarkar, N. I., & Yongchareon, S. (2025). AI-driven energy-efficient routing in IoT-based wireless sensor networks: A comprehensive review. *Sensors, 25*(24), 7408.