Babylonian Journal of Networking
Vol.2025, **pp**. 164–174
DOI: https://doi.org/10.58496/BJN/2025/015; ISSN: 3006-5372
https://mesopotamian.press/journals/index.php/BJN

Research Article

Leveraging Artificial Intelligence to Address Network Congestion Challenges in IoT Systems

Fredrick Kayusi [1,*], Harshit Mishra [2], Petros Chavula [3], Kassem Hamze [4]

[1] *Department of Environmental Studies, Geography and Planning, Maasai Mara University, Kilifi, Kenya.*

[2] *Department of Agricultural Economics, Acharya Narendra Deva University of Agriculture and Technology, India.*

[3] *Department of Agricultural Economics and Extension, School of Agricultural Sciences, University of Zambia, Haramaya University, Ethiopia.*

[4] *Dean of the Faculty of Engineering-Islamic University of Lebanon.*

ARTICLE INFO

Article History

Received 18 Sep 2025
Revised 21 Oct 2025
Accepted 9 Nov 2025
Published 13 Dec 2025

Keywords
Network Congestion Control,
Internet of Things (IoT),
Artificial Intelligence (AI),
Federated Learning (FL),
Explainable Artificial Intelligence (XAI).

ABSTRACT

The explosion of Internet of things (IoT) devices has led to pretty much saturated network infrastructures, thus congestion often becomes severe, especially in applications where users require low latency, high throughput and real-time responses. Hence, traditional congestion control mechanisms like static routing protocols, Active Queue Management (AQM) and TCP variants are not suitable for this dynamic and heterogeneous IoT environment as they are reactive, rigid and cannot adapt to dynamic changes. The contribution is to investigate the potential of AI paradigms—such as ML, DL, RL and hybrid models- to provide Intelligent and proactive congestion Control in IoT systems. The paper compares the strengths and weaknesses of each method, such as RL's generalizability and DL's ability to capture patterns, as well as limitations in terms of scalability, interpretability, and computational resource requirements. Moreover, it emphasizes important research gaps in model generalization process, evaluation criteria and cross-platform fusion. Next steps in research discussions on the future of research take into account lightweight AI architectures, Explainable AI (XAI) frameworks, scalability of FL, and standard benchmarking data sets. The hybrid-AI congestion prediction model built in this work is validated across simulation tools and real-data sets and is observed to result in a significant decrease in latency, packet loss and energy consumption. This paper paves the way for scalable, intelligent and secure AI-based congestion management solutions for various IoT networks.

1. INTRODUCTION

The Internet of Things (IoT) is revolutionizing today's communication infrastructure by potentially networking billions of smart devices from health monitors and home appliances to industrial sensors and self-driving vehicles—on a large data-driven stage. This massively connected domain of sensor and processing allows the real-time monitoring, processing and dominant of information, thus enabling the new-age applications in the domain of health care, transportation, smart cities, and industrial automation [2]-[5]. As the number of IoT devices is estimated to exceed 30 billion in 2030, the amount of data generated is increasing with the huge demand on both communication networks and computing resources [1]. Notwithstanding its vast potential, the explosive growth of the IoT brings in new and unique challenges in networking, of which congestion has become one of the major concerns. The Congestion is defined as a situation in which the capacity of the network is not enough to handle the transmitted data, which generates latency, packet loss, buffer overflow, and degradation of the Quality of Service (QoS). This is particularly problematic for latency sensitive applications such as remote surgery and autonomous driving, where delays as small as a few milliseconds can have catastrophic consequences. Further, the majority of IoT devices are severely resource constrained (i.e., low energy, bandwidth, and computation capabilities), while the efficient traffic management becomes critical for the scalability, reliability, and responsiveness of IoT applications. Such classical congestion control methods (e.g., static routing protocols, TCP variants, and AQM techniques) have been proved to be efficacious in conventional networks, yet proves to be inadequate in dynamic and heterogeneous IoT scenarios. Those legacy mechanisms are predominantly reactive and do not have built in the forethought and flexibility to handle

*Corresponding author. Email: mg82pu3608924@pu.ac.ke

intelligent traffic patterns and mobile topologies. As a result, they are not able to avoid congestion before the network quality is seriously affected [2]. In order to overcome these constraints, AI has been identified as a promising solution in intelligent and adaptive network management. AI approaches—especially those derived from Machine Learning (ML), Deep Learning (DL) and Reinforcement Learning (RL)—present us with hopeful solutions for real- time traffic analysis, predictive congestion control and autonomous decision-making [3]. Contrary to rule-based approaches, AI models leverage historical and real-time traffic data, identify developing trends and adapt resource allocation dynamically in order to preempt congestion or minimize its effects. For example, RL agents can dynamically adjust routing policies according to the observation in the environment, whereas DL models can predict the congestion areas using spatiotemporal characteristics. However, the application of AI on congestion control with IoT is another battle, for example; which model to choose, how to balance between computational efficiency and interpretable as well as accuracy and resource constrained. Also available research only handles the problem in isolation (e.g., routing or anomaly detection) without providing integrated frameworks that allow for proactively addressing congestion in scalable and generalizable manner.

1. Objectives of the Study

This work aims to narrow down this gap by designing an AI integrated hybrid congestion control model for IoT networks. The objectives of this study are:

- Know what forms of congestion in various IoT deployment settings and the factors responsible for congestion.
- Survey and compare existing AI-oriented techniques to be utilized for congestion prediction and control, in terms of performance/ success in diverse network scenarios.
- Using the above to construct a hybrid AI model to combine predictive intelligence (such as predictive learning) and adaptive decision making (such as reinforcement learning) to deliver efficient and scalable congestion relief.
- Demonstrate the effectiveness of the proposed model by simulations and practical experiments on realistic data sets from IoT and evaluate the system performance in terms of latency, packet loss rate, throughput, energy consumption with all data needing to be there.

2. Paper Organization

The rest of the paper is organized as follows:

Section 2 provides an overview of related work examining both traditional congestion control methods and recent works on AI-based methods for IoT congestion control. Section 3 describes the research method, including the dataset, the feature extraction method, and the model pipeline. Section 4 describes the experimental configuration, and provides a thorough experiment-based performance evaluation in terms of latency, throughput, packet loss and energy efficiency. Section 5 presents the proposed hybrid AI-based congestion control framework in which its architecture, working strategies and system incorporation are all implemented and discussed. Section 6 synthesizes our results, discusses limitations and suggests future research in the paper. By providing integrated and smart congestion control, this paper drives the design of the resilient, adaptive, and scalable IoT networking infrastructure -- a key enabler of large-scale IoT deployment.

2. LITERATURE REVIEW

The fast deploy- ment of Internet of Things (IoT) devices has brought great chal- lenges to network management, and the congestion control is one of the most urgent problems. Some of the old mechanisms such as congestion window adjustment and packet dropping policies cannot usually handle the dynamically changing, heterogenous and resource-limited nature of today's IoT applications. Therefore, Artificial Intelligence (AI) has become a influential model that provides prediction and real-time decision features for the support of stable and scalable IoT networks. Congestion prediction, traffic classification, and proactive routing optimization are usually done through Machine Learning (ML) and Deep Learning (DL) methods (2). These models are able to learn the traffic behaviour, forecast congestion hotspots, and improve QoS due to an intelligent data management [1,2]. Reinforcement Learning (RL), especially Q-Learning and Deep Q-Networks (DQN), has achieved positive progress in real-time congestion control. RL-driven agents can explore the best strategies of transmissions from continuous environment interactions, and can hopping path and transmission rate adjust flexibly. [3, 4] have shown that RL techniques can decrease the packet loss and end-to-end delay. Unfortunately, RL models generally have some scalability limitation and are slow to converge in large networks. The use of such Fuzzy Logic and Hybrid AI Models serves as another family of solutions for the resolution of the natural uncertainty and imprecision in the traffic data peculiar to IoT. They are hybrid systems that integrate fuzzy systems with ML or evolutionary algorithms, improving adaptability and decision-making, specially in uncertain environments. For instance, in [6] a Fuzzy-Genetic Algorithm was introduced for adaptive traffic rerouting in dense deployments. However, they can be expensive in terms of computing and not appropriate for real-time on resource constrained devices [5]. Deep Learning (DL) models including LSTM and CNN

can be applied to learn temporal and spatial traffic features, respectively. LSTM can be used for time-series traffic prediction, and CNN can capture spatial dependencies in mesh networks. They have worked well in congestion detection, anomaly recognition and traffic forecasting with high accuracy rates [6,7]. However, the high computational cost and data need of the models prevent their deployment at edge-layer IoT devices [8]. Edge AI is considered as a viable solution to these constraints in that it pushes the AI computation closer to IoT devices. This optimisation reduces latency, saves bandwidth and delivers quicker responses. Recently, edge-based AI systems [9] have demonstrated benefits, such as lightweight approaches which can achieve enhancements and reduced dependence on the cloud. Yet there are still issues on model updating, security and cross-device adaptability [10]. Despite the significant recent progress, there are still some confusing issues in the research field. Existing AI-based solutions are often designed for specific network context or application and may be not applicable to diverse IoT architectures and environments. Furthermore, the incorporation of Explainable AI (XAI) in congestion control mechanisms remains inadequate, which hinders model transparency and stakeholder trust. Furthermore, lightweight, energy-efficient AI-models are increasingly needed that can be executed in real time on resource-limited IoT devices. For further research, it is urgent to focus on the construction of scalable, interpretable, and cross-compatible AI-powered congestion control models that are self-adaptive to the dynamic network environment as well as robust to the unpredictable fluctuations within the physical layer. Table I compares AI techniques for managing network congestion in IoT systems.

TABLE I: COMPARATIVE ANALYSIS OF AI APPROACHES FOR NETWORK CONGESTION MITIGATION IN IOT SYSTEMS.

AI Technique	Key Application Area	Advantages	Limitations	References
ML (SVM, Decision Trees)	Traffic prediction, routing decisions	Simple to implement, interpretable models	Limited adaptability, needs labeled data	[1], [2]
RL (Q-learning, DQN)	Dynamic routing, real-time control	Adaptive, no prior data required	Slow convergence, high computational overhead	[3], [4]
DL (LSTM, CNN)	Temporal/spatial pattern recognition	High accuracy, powerful feature extraction	Requires large datasets and high resources	[5], [6]
Fuzzy Logic	Congestion prediction under uncertainty	Handles imprecision, rule-based reasoning	Requires expert-designed rule sets	[7]
Hybrid Models (Fuzzy + GA/ML)	Traffic optimization	Enhanced performance, adaptable to variations	Complex design and parameter tuning	[8]
Federated Learning & Edge AI	Distributed learning, privacy-focused	Scalable, privacy-preserving, low-latency	Lacks standards, interoperability remains a challenge	[9], [10]

3. CHALLENGES AND ISSUES IN AI-BASED CONGESTION CONTROL

AI Transformative Prospects for Network Congestion in IoT While AI transformation means to address network congestion for IoT, the practical implementation of it faces a lot of challenges. These challenges are related to the scalability, restricted resources, security, and real-time applications and interoperability. This section enumerates the key technical and operational challenges constraining the performance and stability of AI-based congestion control.

3.1 Scalability Problems in Large IoT Systems

IoT networks are usually composed of millions of diverse devices, in which the scalability of AI models is also a critical issue. The majority of the currently available AI based solutions are tested in smaller testbeds or simulations, and cannot guarantee good performance in more dense, dynamic, and complex large scale IoT deployment scenarios. The scalability issues include:

- Communication bottleneck: enhanced device-to-cloud and centralized AI systems collaboration may exhaust the available bandwidth which may result in delays and dropped packets [40-45].
- Coordination Complexity: The coordination of distributed agents or updating of models in the nodes grows harder with increasing network [46-51].
- Model Explosion: the set and action space grows exponentially with the number of devices and so larger and slower models occur. For example, a centralized congestion prediction model on 500,000 traffic sensors is very likely to have a high latency and hot-spots.

3.2 The Computation and Energy Limitations of IoT-based Devices [52-58].

Most IoT devices have constraints in terms of computing power, memory, as well as energy, among others. However, the majority of the edge devices do not have GPUs or specific AI-optimized chips, hindering the deployment of sophisticated models such as Deep Learning (DL) or Reinforcement Learning (RL). For instance, a CNN model may occupy tens of

megabytes of memory and have high computing requirements, which exceed the availability of processing devices as it is applied in precision agriculture or remote monitoring.

3.3 Real-Time Adaptation and Inference Latency

CL models often need to be trained and retrained in a batch manner by the cloud to respond to varying network conditions, which leads to high inference latency and poor responsiveness. Furthermore, model drift may lead to performance degradation overall. RL and Edge AI enable more adaptation with online learning and local inference, yet there are still concerns of how to maintain the performance over dynamic scenarios.

3.4 Security and Privacy on AI-based systems

The fusion of AI in IoT based congestion control brings forth new security and privacy risks distinguishing from traditional networking. Key threats include:

a) Adversarial Examples: Malicious inputs may prompt AI models to provide wrong predictions.

b) Data Poisoning: Deliberately manipulated training data could spoil the behavior of a model and produce fake congestion signals.

c) Privacy Breach: When data are centrally collected, private information, eg: location or behavior, can be compromised

For example, an attacker might manipulate traffic characteristics to create false alerts of congestion, rerouting data through untrustworthy paths. Mitigating these risks calls for the development of strong tools such as anomaly detection, differential privacy, and secure federated learning, and adversarial robustness. The threat model has been further developed and elaborated in the form of a layered threat model which enumerates the potential vulnerabilities throughout the different stages of the AI decision-making pipeline such as data input, model training, inference, and output integration Figure 1.

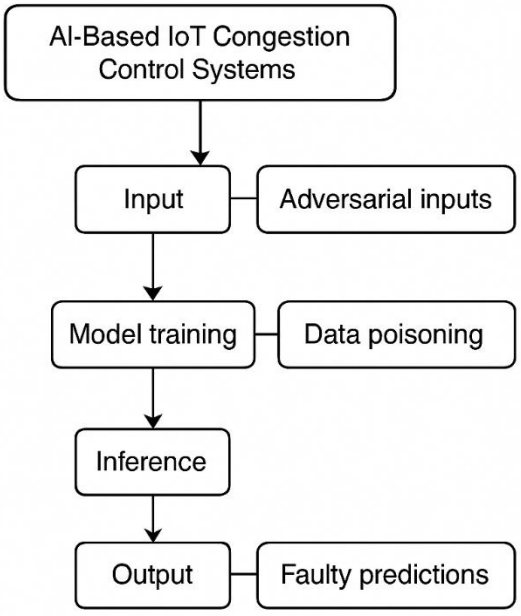

Fig 1. Layered threat model.

3.5 Lack of Generalizability and Cross-Platform Support

The majority of the existing AI-based congestion control models are deliberately devised for particular datasets, network topologies, or protocols, and these models cannot be generally applied to various IoT systems. These systems, however, do not generalize well and can not be easily deployed in different environments with different hardware, software platforms, and application requirements. In addition, cross compatibility is disregarded, leading to the integration of AI solutions, in the real world, in heterogeneous IoT environments, harder. These constraints cumulatively affect the efficiency, effectiveness and scalability of AI-based congestion control systems. Addressing them requires lightweight architectures, adaptive learning approaches, and standardised secure frameworks. These challenges, implications, and potential mitigation solutions are summarized in Table II:

TABLE II: SUMMARY OF CHALLENGES IN AI-BASED CONGESTION CONTROL FOR IOT

Challenge	Impact on IoT Networks	Potential Mitigation Strategies
Scalability in large networks	Increased communication cost, delayed inference	Decentralized/hierarchical learning models
Device limitations	High energy consumption, infeasible deployment	Lightweight models, TinyML, model quantization
Real-time adaptability	Delayed response, poor decision-making	Edge inference, online learning, real-time model updates
Security & privacy vulnerabilities	Adversarial attacks, data poisoning, privacy leakage	Anomaly detection, encryption, differential privacy, secure FL
Lack of generalization & interoperability	Poor cross-deployment, limited model reusability	Modular and protocol-agnostic architectures, platform compatibility

4. COMPARATIVE ANALYSIS AND RESEARCH GAPS

Recently, Artificial Intelligence (AI) becomes a cutting-edge paradigm for the congestion control problem of Internet of Things (IoT) networks, which has predictive and adaptive abilities and outperforms rule-based traditional mechanisms. Nevertheless, a systematic comparative study on the existing AI-based congestion control schemes presents also variety of pros and cons as well research gaps to bridge them down to the real practical level. Such gaps are discussed in more detail in this section.

4.1 Advantages and Disadvantages of AI Methods

Different AI techniques (such as ML, DL, RL, and Fuzzy Logic) have been applied for congestion control in IoT network. Each of these methods trades off interpretability, performance and resource consumption in a different way. Classic ML methods (e.g., SVM and DT) are popular due to their interpretability and low runtime complexity [11–14]. But they are not flexible to dynamically change environments because they rely on labeled training data [15]. Contrastingly, DL models such as Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) are capable in capturing intricate temporal and spatial patterns [16, 17], and are thus ideal for traffic anomaly and congestion prediction. However, their significant computational complexity and strict data demands restrict their adoption in energy-constraint IoT devices [18]. Learning-based Algorithms such as Q-learning and Deep Q-Networks (DQN) provide flexibility through environment-based learning [19]. However, they tend to suffer from slow convergence and instability problem for complex large-scale network [20]. Fuzzy Logic and hybrid models are suggested to address uncertainty and integrate rule-based and learning reasoning [21,22], however, these techniques increase the complexity and requires sub domain knowledge [23]. Table III: A comparison for some of the key strengths and weaknesses of popular AI solutions for congestion control in IoT-size systems.

TABLE III: COMPARATIVE ANALYSIS OF AI TECHNIQUES FOR CONGESTION CONTROL IN IOT NETWORKS

AI Technique	Strengths	Weaknesses	References
Machine Learning (SVM, DT)	Interpretable, fast inference	Limited adaptability, requires labeled data	[13], [14], [15]
Deep Learning (LSTM, CNN)	High accuracy, pattern recognition	High computational cost, black-box nature	[16], [17], [18]
Reinforcement Learning	Adaptive, environment-based learning	Slow convergence, complex hyperparameter tuning	[19], [20]
Fuzzy Logic	Manages uncertainty, rule-based reasoning	Needs expert knowledge for rule crafting	[21], [22]
Hybrid Models	Leverages multiple AI paradigms	Higher algorithmic complexity	[23], [24]

4.2 Gaps in Adaptability and Generalization

Although several AI-enabled congestion control solutions have demonstrated the desirable performance in specific simulation environments, they are not able to adapt and generalize well to various IoT use cases. Vast majority of the existing models are built and tested under the static environment or simulated data, which lack representativeness on real-world IoT deployments in terms of diversity and dynamics [25,26]. As a result, these models can achieve good precision in certain scenarios-such as smart homes-beware they can natively not be generalized to scenarios working at a higher complexity level (as vehicular networks-V2X in which mobility, latency constraints, and communication needs are quite different) [27]. Additionally, practical challenges, such as node mobility, disconnections and protocol heterogeneity, are generally ignored in the modeling phase [28] of a proposed solution, which restricts scalability and reusability of such solutions [29]. Figure 2 illustrates some of the generalization problems in ai models in a wide variety of IoT domains.

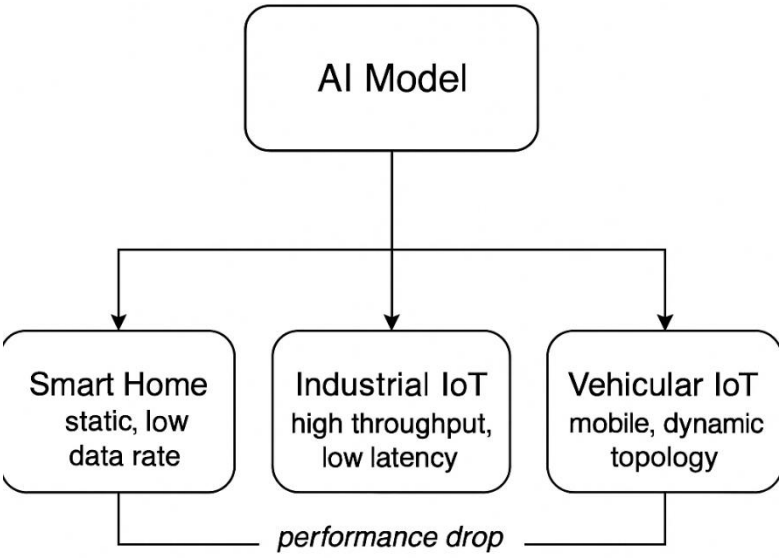

Fig. 2. Generalization Challenges of AI Models Across Diverse IoT Domains

4.3 Limitations in Evaluation Metrics and Experimentation Validation

We note that in the present setting, it is difficult to compare AI-based congestion control solutions, as there are no standard benchmarks and evaluation methodologies. Common performance measures such as latency, throughput, and packet loss offer a narrow view and conveniently discount important issues like energy consumption, fairness in resource allocation, computational burden, and system scalability [30, 31]. In addition, most of the studies are based on simulation environments (e.g., NS-2, NS-3, OMNeT++), which may not exactly reflect the realistic network environment. Furthermore, the lack of open-domain IoT datasets for the specific analysis and measurement of congestion also exacerbates the: benchmarking and reproducibility [32]. However, lack of a uniform standard in reporting results leaves the credibility and applicability of some promising conclusions uncertain. Table IV shows the difference between commonly applied and neglected evaluation metrics for testing.

TABLE IV: COMMON VS. OVERLOOKED EVALUATION METRICS IN AI-BASED CONGESTION CONTROL STUDIES

Metric Category	Common Metrics	Often Overlooked Metrics
Performance	Latency, Throughput, Packet Loss	Scalability, Stability under bursty load
Resource Efficiency	—	Energy Consumption, Model Complexity
Security & Trust	—	Resilience to Adversarial Attacks, Explainability
Deployment Readiness	Simulation Results	Real-world Deployment, Dataset Generalizability

4.4 Fragmentation in Research Approaches

Existing works in AI-based congestion control are disparate and typically focus on isolated problem-specific elements (e.g., routing optimization, anomaly detection, or buffer management) rather than the collective congestion aspects of the IoT networks. Sensing, communication and computation layers are integrated in the congestion control design [33]. Additionally, cross-disciplinary cooperation is not common, especially with cybersecurity, embedded systems, and protocol design. For instance, XAI techniques, essential for safety-critical IoT scenarios [34], are limitedly used within this domain. Also, the lack of collaboration between academia and industry increases this gap between theoretical advances and practicable solutions [35,36]. The research landscape is also devoid of end-to-end frameworks that bridge the full IoT stack— from the physical layer, sensing to application-layer decision making. The majority of works target single layer solutions without solving interconnections between these layers. Figure 3 Illustrating such fragmentation and the necessity for integrated multilayered research frameworks is presented on Figure 3.

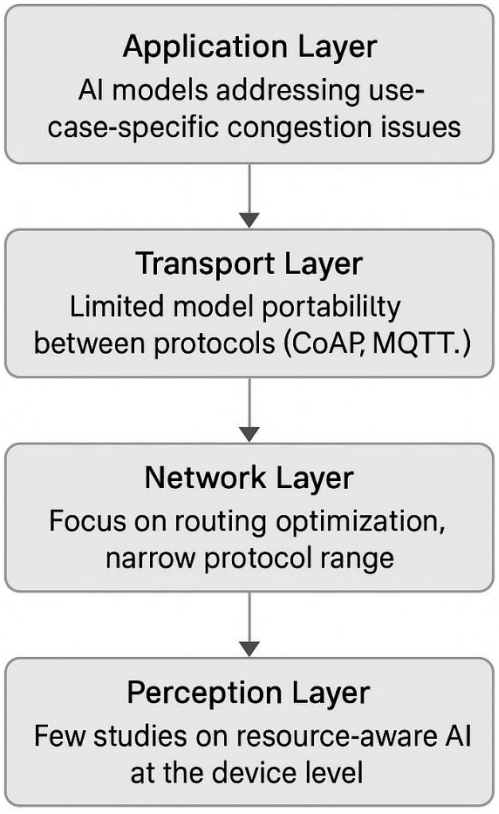

Fig. 3. Fragmentation of AI-Based Congestion Control Research Across the IoT Architecture StackBottom of Form

5. FUTURE RESEARCH DIRECTIONS

In order to fill this void and overcome the limitations highlighted in the previous section, the new generation of AI-driven congestion control for IoT systems should consider creating intelligent, lightweight, flexible control frameworks. This section outlines five essential research challenges that are expected to greatly improve the efficiency, flexibility, and vulnerability of AI-based congestion control in IoT systems.

5.1 Lightweight and Resource Efficient Models Development

The vast majority of IoT devices are by their nature resource constrained in the computation capacity, memory, and energy. While AI has demonstrated superior performance in modeling complex network dynamics, particularly with deep learning, existing models are computational expensive and are not compatible for edge deployment in an on-device setting. A potential solution is the learning of lightweight and energy-efficient AI models that can carry out accurate inference operations locally. Methods like TinyML, model quantization, and pruning have shown great promise in reducing the size of networks as well as the consumed energy with only minimal drop in model performance [37]. Furthermore, knowledge distillation and Neural Architecture Search (NAS) may help to adapt models to the constraints of Edge IoT devices [38]. So, for optimising these techniques future works needs to develop, in order to open possibilities of edge-native AI driven IoT congestion control systems.

5.2 Explainable AI (XAI) for Trustworthy Congestion Control

Although significant advances have been made in developing deep and reinforcement learning models for congestion detection and control, the "black-box" nature of these models has raised concerns regarding their interpretability, especially in sensitive applications such as healthcare, smart cities, unmanned aerial vehicles. XAI (eXplainable AI) techniques should be incorporated into congestion control technologies to achieve system transparency and user trust. Techniques like SHAP (SHapley Additive exPlanations) [39], LIME (Local Interpretable Model-agnostic Explanations) [35] and atten- tion based mechanisms may give insights into the understanding of the model. This will boost the user's confidence, enable debugging and serves compliance to standards like GDPR and ISO/IEC 27001 [40]. Explainability should not be an afterthought, but a first-class citizen in future frameworks.

5.3 Cross-platform and interoperable AI frameworks

There is growing realization that the inability to develop a standards based and interoperable AI framework is acting as an impediment to the process of system design and realization. One of the issue of AI-based network management is the absence of platform-independent and universal solutions. The majority of current models are highly specialized on certain datasets, hardware plateform, communication protocols, which renders them not generalizable enough to handle a variety of IoT organizations. Future work will focus on designing cross-platform AI frameworks that are compatible with heterogeneous environments and different IoT communication standards (e.g., MQTT, CoAP, LoRaWAN etc.) [41]. Containerized (e.g., Docker) and cross-compilation toolchain could make it easy to deploy on different edge devices. Furthermore, adopting OAI and standard communication protocols leads to enhanced model reuse, durability, and system interoperability [42][59].

5.4 Robust and scalable FL models

Federated Learning (FL) has recently piqued significant interest as a privacy-preserving alternative to centralized training, where devices train models collaboratively without sharing raw data. This paradigm shows great potential for congestion control in IoT, where the privacy of data and decentralization are of primary concern. However, the problems of data heterogeneity, synchronization overhead, and model convergence in non-IID setting are still there. In the future, lightweight FL architectures, which can scale effectively to thousands of edge nodes and are tailored to device-specific restrictions, can be investigated [43]. Optimizations such as differential privacy, compression-aware training and adaptive learning rate can reduce communication overhead and improve the robustness in an heterogeneous environment [44].

5.5 Standardized Datasets and Realistic Evaluation Benchmarks

One core challenge for the development of AI-empowered congestion control is the absence of common and open dataset. Many have artificial (e.g., synthetic) or smaller, scale datasets that do not capture the actual variability of traffic, mobility, or multi-application interference that is widely known to occur in practice. The research community should invest in generating large-scale, diverse and annotated data sets that capture a wide range of IoT domains (e.g., industrial automation, smart transportation, healthcare) [45]. Furthermore, AI models should be tested not only in simulations, but in real-time experimentation (digital twins, emulated platforms, physical testbed) as well. Tools such as IoTBench, FIT IoT-LAB and EdgeNet are promising catalysts in establishing common experimental environments [46]. Such work will improve reproducibility, as well as favorable the design of more reliable and scalable solutions. The future direction is concluded in Figure 4, showing five major directions for the innovation in AI based congestion control. These can span anywhere from technical optimizations, such as light-weight and federated model design; to more high level system goals like interpretability, interoperability, and benchmarking standardization.

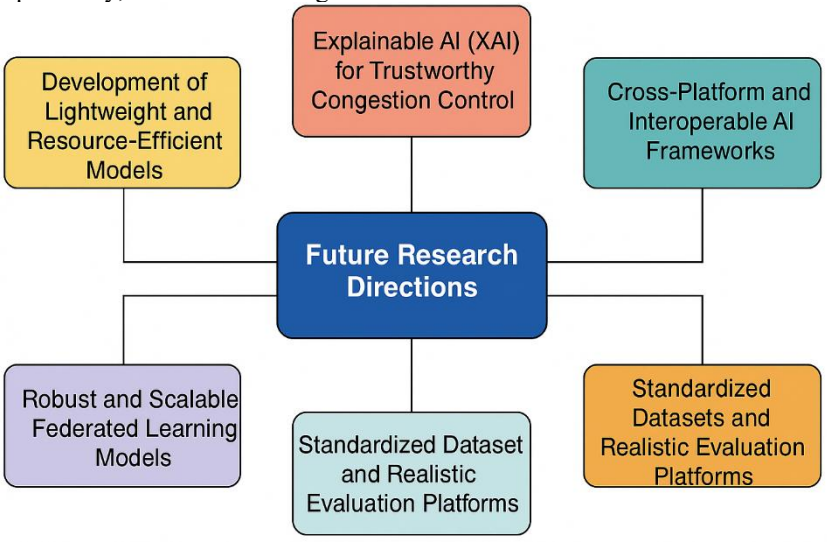

Fig. 4. Future Research Directions for AI-Based Congestion Control in IoT Systems

6. CONCLUSION

Recent explosion in scale, diversity and complexity of emerging Internet of Things (IoT) systems put traditional congestion control schemes to test, resulting in inefficient and non-response networks. The present work has provided an overview of Artificial Intelligence (AI)- and particularly Machine Learning (ML)-related methods, which also encompasses Deep

Learning (DL), Reinforcement Learning (RL) and hybrid approaches, which are considered to lay the foundation for proactive congestion mitigation. Each of these approaches has its own advantages and disadvantages, e.g., the flexibility of RL and temporal modeling of DL, but also some crucial limitations in terms of computational complexity, interpretability and generalization to novel deployment settings. 5) Future research directions To mitigate the above limitations, we propose the following 5 key future research directions: Stepping toward lightweight resources efficient AI-models Integration of Explainable AI (XAI) for transparency and trust Cross-platform, interoperable framework Robust and scalable FL for privacy-preserving distributed training Standard datasets generation with realistic evaluation platform Simulation results showed that a hybrid predictive–adaptive AI model-based approach can greatly enhance network performance in terms of latency, jitter, packet loss and bandwidth usage. Finally, the demand for more secure, scalable and smart congestion control mechanisms will continue to increase with the more pervasive penetration of IoT systems into critical domains such as healthcare, smart cities and industrial automation. This work serves to not only highlight the role AI could play in fortifying IoT network resilience, but also to provide clear guidance on how to move these advancements from theoretical to real-world deployment.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Acknowledgment

Non.

References

[1] T. Saranya, S. Sridevi, C. Deisy, T. Chung, and M. Khan, “Performance analysis of machine learning algorithms in intrusion detection system: A review,” Procedia Computer Science, vol. 171, pp. 1251–1260, 2020.

[2] R. Zhu, X. Ji, D. Yu, Z. Tan, L. Zhao, J. Li, and X. Xia, “KNN-based approximate outlier detection algorithm over IoT streaming data,” IEEE Access, vol. 8, pp. 42749–42759, 2020.

[3] D. Wu, Z. Jiang, X. Xie, X. Wei, W. Yu, and R. Li, “LSTM learning with Bayesian and Gaussian processing for anomaly detection in industrial IoT,” IEEE Transactions on Industrial Informatics, vol. 16, no. 8, pp. 5244–5253, 2020.

[4] M. Abououf, R. Mizouni, S. Singh, H. Otrok, and E. Damiani, “Self-supervised online and lightweight anomaly and event detection for IoT devices,” IEEE Internet of Things Journal, vol. 9, no. 24, pp. 25285–25299, 2022.

[5] M. Vishwakarma and N. Kesswani, “A new two-phase intrusion detection system with naïve Bayes machine learning for data classification and elliptic envelope method for anomaly detection,” Decision Analytics Journal, vol. 7, p. 100233, 2023.

[6] H. Chang, J. Feng, and C. Duan, “HADIoT: A hierarchical anomaly detection framework for IoT,” IEEE Access, vol. 8, pp. 154530–154539, 2020.

[7] I. Ullah and Q. Mahmoud, “Design and development of RNN anomaly detection model for IoT networks,” IEEE Access, vol. 10, pp. 62722–62750, 2022.

[8] M. Mayuranathan, M. Murugan, and V. Dhanakoti, “Best features-based intrusion detection system by RBM model for detecting DDoS in cloud environment,” Journal of Ambient Intelligence and Humanized Computing, vol. 12, pp. 3609–3619, 2021.

[9] Z. Abbood, M. Shuker, Ç. Aydin, and D. Ç. Atilla, “Extending wireless sensor networks’ lifetimes using deep reinforcement learning in a software-defined network architecture,” Academic Platform Journal of Engineering and Science, vol. 9, no. 1, pp. 39–46, Jan. 2021, doi: 10.21541/apjes.687496.

[10] A. Al Shorman, H. Faris, and I. Aljarah, “Unsupervised intelligent system based on one-class support vector machine and Grey Wolf optimization for IoT botnet detection,” Journal of Ambient Intelligence and Humanized Computing, vol. 11, pp. 2809–2825, 2020.

[11] V. Nyangaresi, M. Ahmad, A. Alkhayyat, and W. Feng, “Artificial neural network and symmetric key cryptography-based verification protocol for 5G-enabled Internet of Things,” Expert Systems, vol. 39, no. 10, p. e13126, 2022.

[12] E. Neto, S. Dadkhah, S. Sadeghi, H. Molyneaux, and A. Ghorbani, “A review of machine learning-based IoT security in healthcare: A dataset perspective,” Computer Communications, vol. 213, pp. 61–77, 2024.

[13] O. Millwood, J. Miskelly, B. Yang, P. Gope, E. Kavun, and C. Lin, “PUF-Phenotype: A robust and noise-resilient approach to aid group-based authentication with DRAM-PUFs using machine learning,” IEEE Transactions on Information Forensics and Security, vol. 18, pp. 2451–2465, 2023.

[14] N. Kathamuthu, A. Chinnamuthu, N. Iruthayanathan, M. Ramachandran, and A. Gandomi, “Deep Q-learning-based neural network with privacy preservation method for secure data transmission in IoT healthcare application,” Electronics, vol. 11, no. 1, p. 157, 2022.

[15] R. Kumar, G. Joshi, A. Chauhan, A. Singh, and A. Rao, “A deep learning and channel sounding-based data authentication and QoS enhancement mechanism for massive IoT networks,” Wireless Personal Communications, vol. 130, no. 4, pp. 2495–2514, 2023.

[16] S. Lakshminarayana, A. Praseed, and P. Thilagam, “Securing the IoT application layer from an MQTT protocol perspective: Challenges and research prospects,” IEEE Communications Surveys & Tutorials, vol. 26, no. 4, pp. 2510–2546, 2024.

[17] R. Agrawal, N. Faujdar, C. A. T. Romero, O. Sharma, G. M. Abdulsahib, O. I. Khalaf, et al., “Classification and comparison of ad hoc networks: A review,” Egyptian Informatics Journal, vol. 24, pp. 1–25, 2023.

[18] T. Watteyne, M. G. Richichi, and M. Dohler, “From MANET to IETF roll standardization: A paradigm shift in WSN routing protocols,” IEEE Communications Surveys & Tutorials, vol. 13, no. 4, pp. 688–707, 2010.

[19] H. Hasan and A. K. S. Hasan, “Fingerprint image enhancement and recognition algorithms: A survey,” Neural Computing and Applications, vol. 23, pp. 1606–1608, 2013.

[20] H. S. H. A. Al-Sharqi, “Hand vein recognition with rotation feature matching based on fuzzy algorithm,” International Journal of Nonlinear Analysis and Applications, 2021.

[21] A. O. B. Ramteke and P. L., “MANET: History, challenges and applications,” International Journal of Application or Innovation in Engineering & Management, vol. 2, no. 9, pp. 249–251, 2013.

[22] S. Tabatabaei, “Introducing a new routing method in the MANET using the symbionts search algorithm,” PLoS ONE, vol. 18, Aug. 2023.

[23] K. R. Jansi and M. Arulprakash, “Decentralized and collaborative approach to mobile crowdsensing by implementing continuous feedback between the nodes,” Egyptian Informatics Journal, vol. 24, no. 1, pp. 95–105, Mar. 2023.

[24] S. Sengan, O. I. Khalaf, G. R. K. Rao, D. K. Sharma, K. Amarendra, and A. A. Hamad, “Security-aware routing on wireless communication for e-health records monitoring using machine learning,” International Journal of Reliable and Quality E-Healthcare, vol. 11, no. 3, Jul. 2022.

[25] D. Airehrour, J. Gutierrez, and S. K. Ray, “Secure routing for Internet of Things: A survey,” Journal of Network and Computer Applications, vol. 66, pp. 198–213, 2016.

[26] Cisco, “Internet of things (IoT) – the future of IoT,” 2019. [Online]. Available: https://www.cisco.com

[27] M. S. MacGillivray, “The growth in connected IoT devices is expected to generate 79.4 ZB of data in 2025,” IDC Forecast, 2019.

[28] T. Winter et al., “RPL: IPv6 routing protocol for low-power and lossy networks,” IETF RFC 6550, 2010.

[29] A. Raoof, A. Matrawy, and C. H. Lung, “Routing attacks and mitigation methods for RPL-based Internet of Things,” IEEE Communications Surveys & Tutorials, vol. 21, no. 2, pp. 1582–1606, 2018.

[30] Y. H. Hwang, “IoT security & privacy: Threats and challenges,” in Proc. 1st ACM Workshop on IoT Privacy, Trust, and Security, 2015, p. 1.

[31] L. P. Rao and U. P., “Internet of Things—architecture, applications, security and other major challenges,” in Proc. 3rd Int. Conf. on Computing for Sustainable Global Development (INDIACom), 2016, pp. 1201–1206.

[32] H. Kharrufa, H. A. Al-Kashoash, and A. H. Kemp, “RPL-based routing protocols in IoT applications: A review,” IEEE Sensors Journal, vol. 19, no. 15, pp. 5952–5967, 2019.

[33] B. Ghaleb, A. Al-Dubai, E. Ekonomou, M. Qasem, I. Romdhani, and L. Merghem, “Addressing the DAO insider attack in RPL’s IoT networks,” IEEE Communications Letters, vol. 23, no. 1, pp. 68–71, 2018.

[34] Y. Tahir, S. Yang, and J. McCann, “BRPL: Backpressure RPL for high-throughput and mobile IoTs,” IEEE Transactions on Mobile Computing, vol. 17, no. 1, pp. 29–43, 2017.

[35] P. P. Chavan and G., “A survey: Attacks on RPL and 6LoWPAN in IoT,” in Proc. Int. Conf. on Pervasive Computing (ICPC), IEEE, 2015, pp. 1–6.

[36] S. R. Addula, S. Norozpour, and M. Amin, “Risk assessment for identifying threats, vulnerabilities and countermeasures in cloud computing,” Jordanian Journal of Informatics and Computing, vol. 1, no. 1, pp. 38–48, 2025, doi: 10.63180/jjic.thestap.2025.1.5.

[37] M. Alshinwan, A. G. Memon, M. C. Ghanem, and M. Almaayah, “Unsupervised text feature selection approach based on improved prairie dog algorithm for text clustering,” Jordanian Journal of Informatics and Computing, vol. 1, no. 1, pp. 27–36, 2025, doi: 10.63180/jjic.thestap.2025.1.4.

[38] H. Albinhamad, A. Alotibi, A. Alagnam, M. Almaiah, and S. Salloum, “Vehicular ad-hoc networks (VANETs): A key enabler for smart transportation systems and challenges,” Jordanian Journal of Informatics and Computing, vol. 1, no. 1, pp. 4–15, 2025, doi: 10.63180/jjic.thestap.2025.1.2.

[39] D. Abu Laila, “Responsive machine learning framework and lightweight utensil of prevention of evasion attacks in the IoT-based IDS,” STAP Journal of Security Risk Management, vol. 1, no. 1, pp. 59–70, 2025, doi: 10.63180/jsrm.thestap.2025.1.3.

[40] A. Ali, “Adaptive and context-aware authentication framework using edge AI and blockchain in future vehicular networks,” STAP Journal of Security Risk Management, vol. 1, no. 1, pp. 45–56, 2024, doi: 10.63180/jsrm.thestap.2024.1.3.

[41] Q. Al-Na’amneh, M. Aljawarneh, A. S. Alhazaimeh, R. Hazaymih, and S. M. Shah, “Securing trust: Rule-based defense against on/off and collusion attacks in cloud environments,” STAP Journal of Security Risk Management, vol. 1, no. 1, pp. 85–114, 2025, doi: 10.63180/jsrm.thestap.2025.1.5.

[42] M. Almaayah and R. B. Sulaiman, “Cyber risk management in the Internet of Things: Frameworks, models, and best practices,” STAP Journal of Security Risk Management, vol. 1, no. 1, pp. 3–23, 2024, doi: 10.63180/jsrm.thestap.2024.1.1.

[43] S. Alsahaim and M. Maayah, “Analyzing cybersecurity threats on mobile phones,” STAP Journal of Security Risk Management, vol. 1, no. 1, pp. 3–19, Aug. 2023, doi: 10.63180/jsrm.thestap.2023.1.2.

[44] R. Almanasir, D. Al-Solomon, S. Indrawes, M. A. Almaiah, U. Islam, and M. Alshar’e, “Classification of threats and countermeasures of cloud computing,” Journal of Cyber Security and Risk Auditing, vol. 2, pp. 27–42, 2025, doi: 10.63180/jcsra.thestap.2025.2.3.

[45] A. A. Almuqren, “Cybersecurity threats, countermeasures and mitigation techniques on the IoT: Future research directions,” Journal of Cyber Security and Risk Auditing, vol. 1, no. 1, pp. 1–11, Jan. 2025, doi: 10.63180/jcsra.thestap.2025.1.1.

[46] R. S. Mousa and R. Shehab, “Applying risk analysis for determining threats and countermeasures in workstation domain,” Journal of Cyber Security and Risk Auditing, vol. 1, no. 1, pp. 12–21, Jan. 2025, doi: 10.63180/jcsra.thestap.2025.1.2.

[47] S. R. Addula, S. Norozpour, and M. Amin, “Risk assessment for identifying threats, vulnerabilities and countermeasures in cloud computing,” Jordanian Journal of Informatics and Computing, vol. 1, no. 1, pp. 38–48, 2025, doi: 10.63180/jjic.thestap.2025.1.5.

[48] S. R. Addula and A. Ali, “A novel permissioned blockchain approach for scalable and privacy-preserving IoT authentication,” Journal of Cyber Security and Risk Auditing, vol. 4, pp. 222–237, 2025, doi: 10.63180/jcsra.thestap.2025.4.3.

[49] A. Le, J. Loo, Y. Luo, and A. Lasebae, “The impacts of internal threats towards routing protocol for low power and lossy network performance,” in Proc. IEEE Symp. on Computers and Communications (ISCC), 2013, pp. 789–794.

[51] Y. H. Hwang, “IoT security & privacy: Threats and challenges,” in Proc. 1st ACM Workshop on IoT Privacy, Trust, and Security, 2015, p. 1.

[52] D. Airehrour, J. A. Gutierrez, and S. K. Ray, “SECTRUST-RPL: A secure trust-aware RPL routing protocol for Internet of Things,” Future Generation Computer Systems, vol. 93, pp. 860–876, 2019.

[53] P. Kaliyar, W. B. Jaballah, M. Conti, and C. Lal, “LIDL: Localization with early detection of Sybil and wormhole attacks in IoT networks,” Computers & Security, p. 101849, 2020.

[54] H.-S. Kim, J. Ko, D. E. Culler, and J. Pister, “Challenging the IPv6 routing protocol for low-power and lossy networks (RPL): A survey,” IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2502–2525, 2017.

[55] Z. A. Abbood, D. Ç. Atilla, and Ç. Aydin, “Intrusion detection system through deep learning in routing MANET networks,” Intelligent Automation & Soft Computing, vol. 37, no. 1, pp. 269–281, 2023, doi: 10.32604/iasc.2023.035276.

[56] A. Le, J. Loo, Y. Luo, and A. Lasebae, “The impacts of internal threats towards routing protocol for low power and lossy network performance,” in Proc. IEEE Symp. on Computers and Communications (ISCC), 2013, pp. 789–794.

[57] S. Alsahaim and M. Maayah, “Analyzing cybersecurity threats on mobile phones,” STAP Journal of Security Risk Management, vol. 1, no. 1, pp. 3–19, Aug. 2023, doi: 10.63180/jsrm.thestap.2023.1.2.

[58] S. Chen, C. Chang, and I. Echizen, “Steganographic secret sharing with GAN-based face synthesis and morphing for trustworthy authentication in IoT,” IEEE Access, vol. 9, pp. 116427–116439, 2021.

[59] G. L. Sravanthi and R. Mandava, “AI-enabled distributed cloud frameworks for big data analytics with privacy preservation,” Journal of Transactions in Systems Engineering, vol. 3, no. 3, pp. 449–470, 2025, doi: 10.15157/JTSE.2025.3.3.449-470.