A Lightweight Autonomous and Adaptive Scheduler for Next-Generation Fog Computing Systems

Abhijeet Mahapatra [1], Ganesh Khekare [2]

[1]Postdoctoral Researcher, Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia; [2]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India. Email IDs: pdf.abhijeet@lincoln.edu.my, khekare.123@gmail.com.

Abstract: Advanced healthcare apps create latency-sensitive and privacy-sensitive IoMT workload which cannot be effectively managed by faraway Cloud computing alone. LAAS is a lightweight contextual bandit-based scheduler proposed for self-managed adaptive Fog-Cloud computing. LAAS works at the healthcare gateway and hospital Fog level where task criticality classification, workload and unreliability prediction, action filtering based on hard QoS, privacy, energy, and queue constraints, and the execution resource selection based on LinUCB-based contextual bandit policies are executed. Critical tasks are guaranteed to be handled through a fallback mechanism that gives priority to local gateway, hospital Fog, nearby Fog, Fog cluster, and finally to Cloud when needed. Experimental results conducted in an IoMT gateway to Fog computing hardware setup indicate that LAAS minimizes emergency delay time, SLA violations, Cloud fallback, and energy consumption as well as enhancing critical task reliability, privacy-friendly local computation and balanced Fog-node utilization compared to recent IoMT offloading and Fog scheduling methods.

Keywords: Contextual bandit, Emergency response, Fog computing, Healthcare task scheduling, IoMT, Privacy-aware offloading, SLA-aware resource allocation.

Experimental Setup

Instead of simply using Cloud simulation, the experiment setup replicates a real-life IoMT to gateway to Fog execution flow. Raspberry Pi-class devices are used as constrained healthcare gateways which run LAAS. On the other hand, Jetson Nano and mini-PC-class devices serve as Fog computing execution nodes for hospitals and cooperative Fogs. The interaction between ESP32 and Raspberry Pi reflects light-weight medical sensors events submission.

Figure 2(a) depicts the edge computing cluster that works for cooperative Fog execution. Figure 2(b) depicts the edge inference gateway that is based on Raspberry Pi. Figure 2(c) depicts the devices used as gateways and Fog devices. Figure 2(d) depicts the communication between ESP32 and Raspberry Pi-based sensor emulator. Figure 2(e) depicts the edge computing cluster for executing tasks at the Fog level.

LAAS is evaluated against seven state-of-the-art approaches, including Energy Optimized IoMT Cloud Offloading [2], which deals with energy optimization for offloading tasks in healthcare Cloud applications. Adaptive Edge-Fog Emergency Healthcare Framework [5] is about managing emergencies in healthcare Fog computing networks. Fog-assisted Resource-efficient IoMT Scheduling [6] aims at resource-efficient scheduling for Fog-assisted IoMT systems. Hybrid Edge-Cloud IoMT Offloading [7] is about offloading and optimizing IoMT healthcare applications. DRL-based Healthcare Fog Scheduling [8] makes use of deep reinforcement learning in healthcare Fog computing task scheduling. Latency and Energy-aware DQN

Offloading [9] makes use of DQN in IoT-Fog-Cloud system task offloading. Joint Energy and Privacy-aware DRL Offloading [14] is related to energy and privacy considerations in healthcare monitoring systems.

The MIT-BIH Arrhythmia Database [15] is selected as the healthcare workload benchmark since it offers annotated ambulatory electrocardiogram records that can be used in modelling arrhythmia and other heart abnormalities. The performance measures include emergency reaction time, critical task reliability, healthcare-sensitive SLA violation ratio, privacy-preserving local task execution ratio, energy expenditure, Cloud fallback ratio, and hospital Fog node usage.

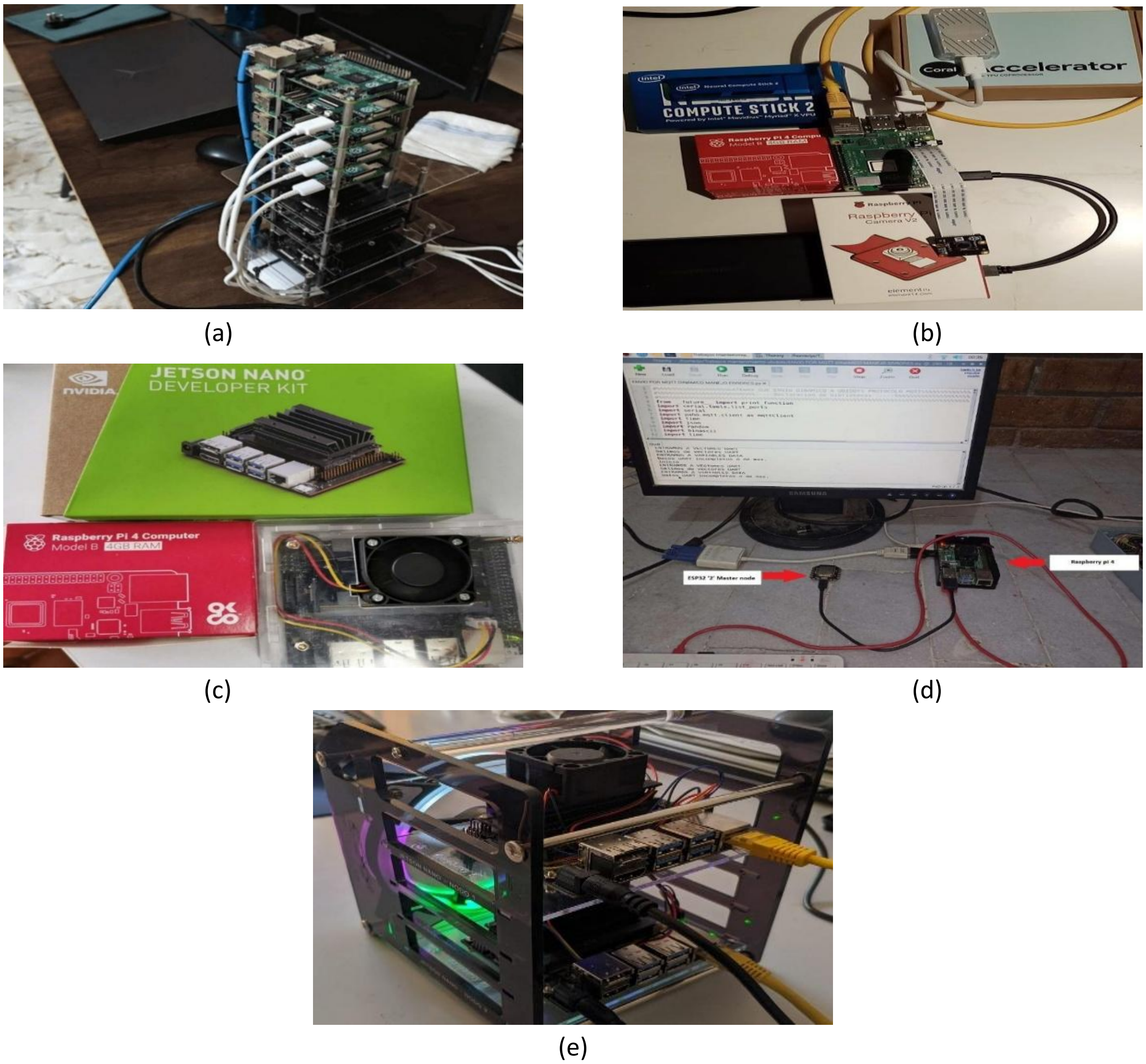

Figure 2. Experimental Setup: (a) edge computing cluster for cooperative Fog execution, (b) Raspberry Pi based edge inference gateway, (c) Raspberry Pi and Jetson Nano hardware used as gateway and Fog devices, (d) ESP32 to Raspberry Pi sensor emulator communication setup, (e) Jetson Nano based edge cluster for Fog level task execution

Result Analysis

Figure 3 shows the variation in response time with the increase in the number of healthcare tasks. The response times are observed to increase in all cases due to increased load in the queue and competition for the resources. LAAS obtains the least emergency response time among all tested workloads. It has an average emergency latency of 164.8ms compared to 205.2ms of the DRL based scheduling approach [8], showing a reduction by 19.7%. At 500 tasks, LAAS has 220ms latency, while [8],[9] and [14] have 280ms, 298ms and 305ms respectively.

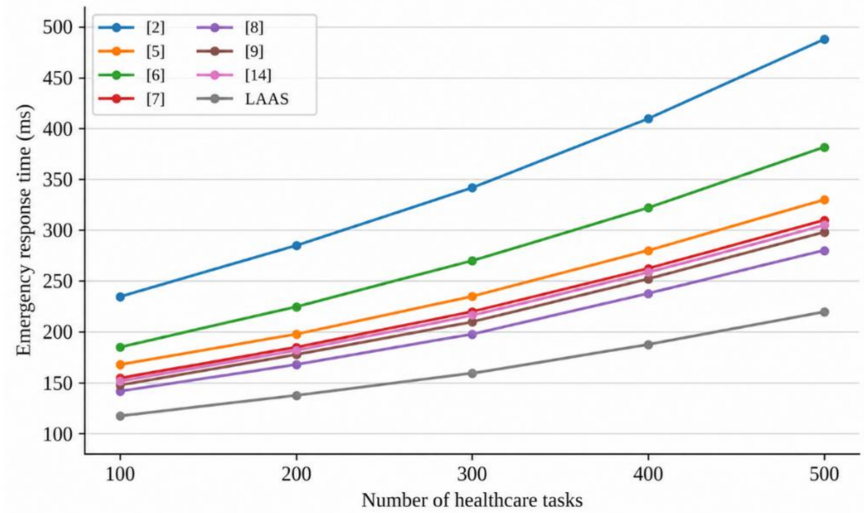

Figure 3. Emergency Response Time under Increasing Task Load

Figure 4 shows the fluctuations in critical task reliability with an increase in the failure rate of the Fog node. The reliability of all the schemes drops owing to the adverse effect caused by unreliable Fog nodes on the probability of success in the task execution process. LAAS shows the maximum average value of critical task reliability, which is equal to 0.967 against 0.947 achieved by the adaptive edge-Fog scheme [5]. This performance gain is attributed to the filtering process with respect to reliability threshold and reliability risk estimation before decision making.

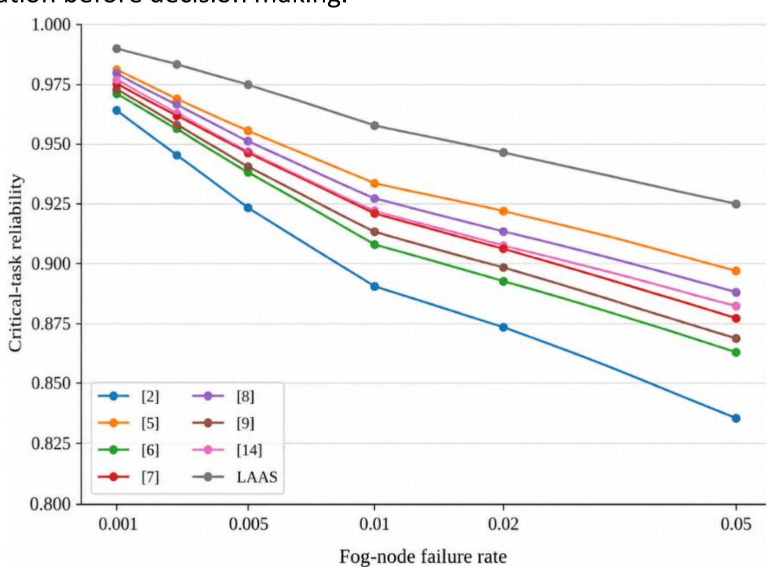

Figure 4. Critical task reliability under increasing Fog node failure rates

SLA violation is illustrated in Figure 5 for normal and burst workloads. SLA violations include deadline misses, reliability violations, and inappropriate task placement with regard to privacy considerations. LAAS achieves 5.2% and 10.8% SLA violations for normal and burst workloads, respectively. These numbers are less compared to those obtained from baseline algorithms due to the safe action filtering employed by the proposed algorithm before the contextual bandits scores are computed. This result is important since in the case of health care SLA violations, apart from delay, privacy and reliability are considered.

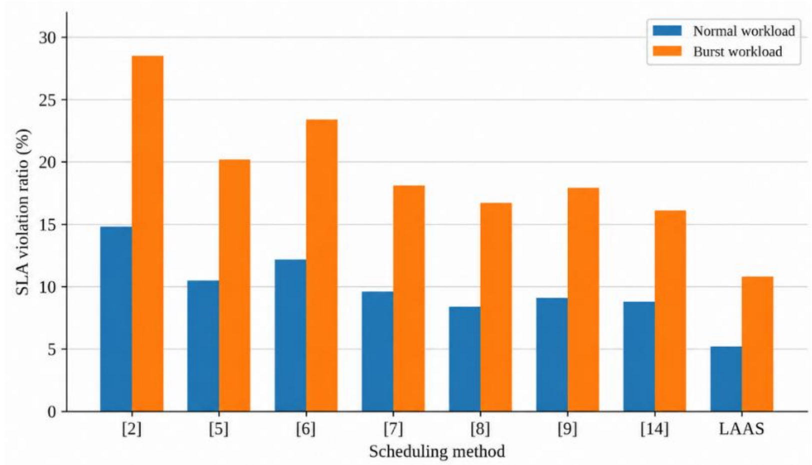

Figure 5. Healthcare-aware SLA violation ratio under normal and burst workloads

In Figure 6, we illustrate healthcare tasks that can be locally carried out using trusted Fog in hospitals resources in the areas of emergencies, urgencies, diagnostics, monitoring, and routine. The LAAS approach places more sensitive tasks near trusted local or hospital Fog resources. The average percentage of local execution that preserves privacy is 78.6%, showing a 9% increase from the privacy-aware baseline [14]. Tasks classified as emergencies and urgencies are more prone to execute locally compared to routine tasks which can also run in Clouds if privacy constraints allow.

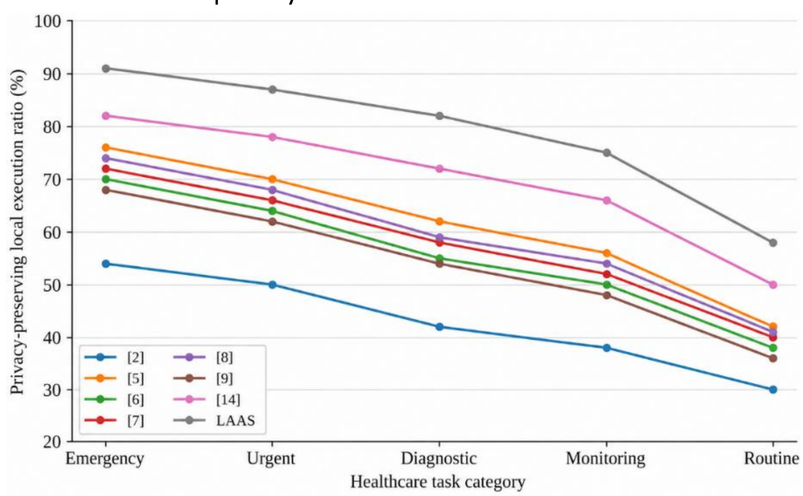

Figure 6. Privacy preserving local execution ratio across healthcare task categories

Figure 7 shows the comparison of the energy consumption that is needed to complete the healthcare tasks by using total energy consumption in terms of communications and computations. The LAAS method uses the minimum amount of energy that is 44.5 J. It is a decrease of 15.7% compared with the energy usage by privacy-aware DRL method [14] and 43.3% compared with Cloud-based offloading method [2]. The main reason for energy saving lies in avoiding unnecessary remote communications and Cloud escalations. Energy efficiency does not affect emergency latency because the scheduler finds the critical tasks near the data source.

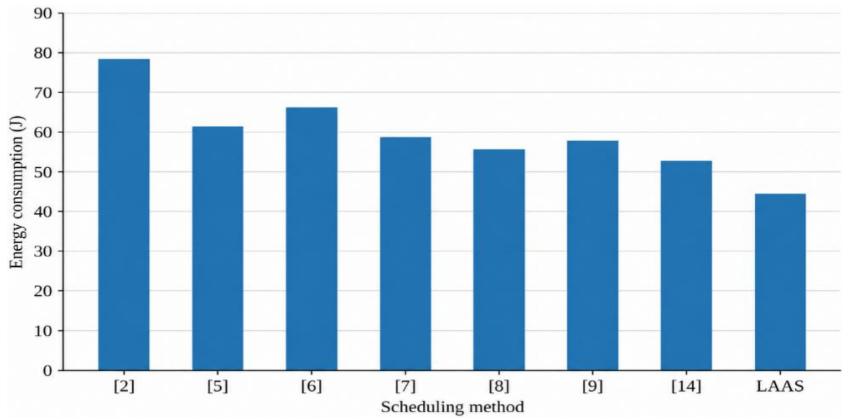

Figure 7. Energy consumption comparison for healthcare workloads

The Cloud fallback frequency of each scheduler in case of a fault within the Fog node and workloads spikes is shown in Figure 8. Cloud fallback ensures that services remain uninterrupted but when used excessively, the process becomes inefficient due to higher latency, bandwidth usage, and exposure of data privacy. On average, LAAS shows Cloud fallback frequency of 18.4%, which is less compared to the frequencies shown in [5], [8], and [14]. The fallback manager checks local gateway, hospital Fog, adjacent Fogs, and Fog cluster before Cloud fallback is initiated.

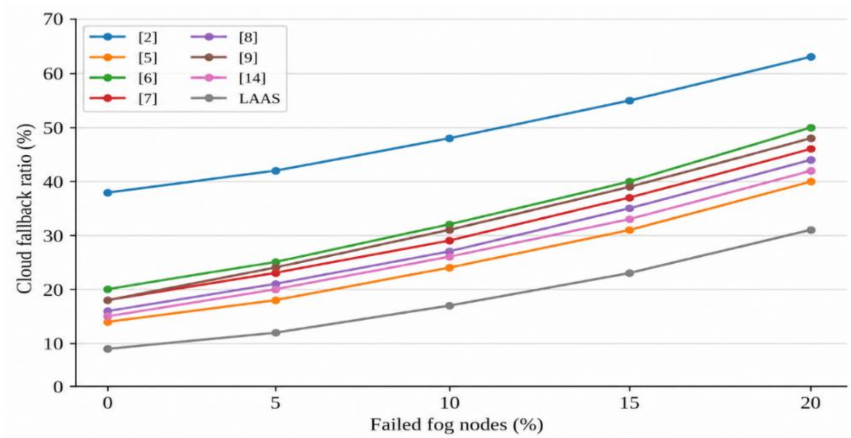

Figure 8. Cloud fallback ratio under Fog node failure and workload burst

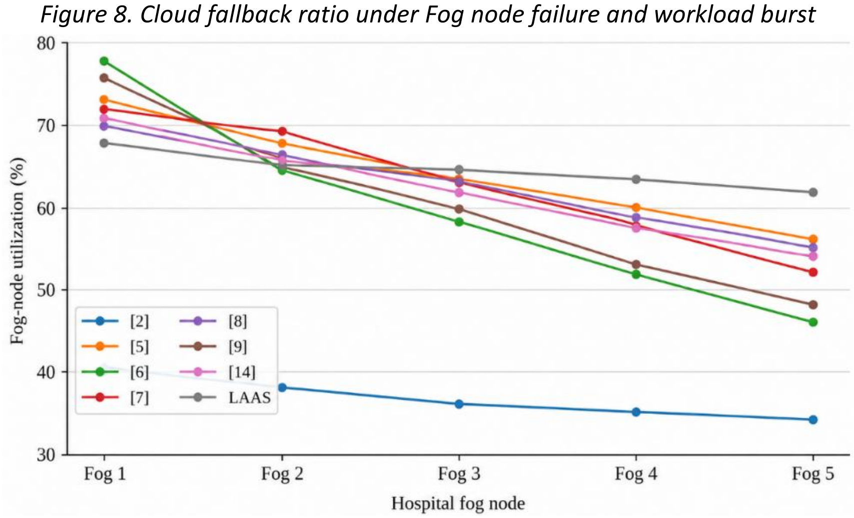

Figure 9. Hospital Fog node utilization across scheduling methods

Figure 9 shows the comparison of utilization in hospital Fog nodes. It can be seen from the results that balanced utilization is better than overloading the resources on one Fog node, since there should always be some buffer capacity for handling unexpected emergency situations in healthcare environments. With the use of the LAAS framework, the average utilization rate for Fog resources is 65%, with a difference of just 2% between 5 Fog nodes. This proves that resource utilization is balanced but also reserve capacity is available.

Conclusions and Future Works

LAAS is proposed in this work to represent an adaptive and autonomous scheduler for smart healthcare based on Fog–Cloud computing. The scheduler is defined by a combination of task criticality classification, forecasting of workload and reliability risks, constraint-based action filtering, action selection, emergency fallback methods, and adaptation through feedback. The formulation of the scheduler is performed by taking into account the trade-off among latency, energy cost, service-level agreement violation, load balance, scheduling overhead, reliability loss, and privacy risk under a set of constraints.

These experimental findings show that the LAAS model outperforms current state-of-the-art approaches for IoMT offloading and scheduling in terms of response time, reliability of critical tasks, SLA adherence, privacy-preserving local processing, energy usage, fallback to the Cloud, and load balancing of Fog nodes. The proposed model can be used as a gateway or a hospital Fog scheduling layer to place tasks. However, it is not intended to replace clinical diagnosis and physician decisions.

The future work of LAAS will be on the implementation of multi-hospital federation of policies, uncertainty aware prediction, improvement in security audits for cross-hospital Fog cooperation, and real-time validation in large-scale health care systems. Further evaluations using longer ECG sequences and more diverse patient events will help in generalization.

References

[1] Damera, V. K., Cheripelli, R., Putta, N., Sirisha, G., & Kalavala, D. (2025). Enhancing remote patient monitoring with AI-driven IoMT and Cloud computing technologies. *Scientific Reports*, *15*(1), 24088.

[2] Bhasker, B., Kaliraj, S., Gobinath, C., & Sivakumar, V. (2025). Optimizing energy task offloading technique using IoMT Cloud in healthcare applications. *Journal of Cloud Computing*, *14*(1), 9.

[3] Hussain, A. A., & Dawood, B. A. (2025). A survey on IoT-Cloud task in healthcare system. *The Journal of Supercomputing*, *81*(1), 141.

[4] Islam, U., Alatawi, M. N., Alqazzaz, A., Alamro, S., Shah, B., & Moreira, F. (2025). A hybrid Fog-edge computing architecture for real-time health monitoring in IoMT systems with optimized latency and threat resilience. *scientific reports*, *15*(1), 25655.

[5] Alwakeel, A. M. (2025). Adaptive edge-Fog healthcare networks: a novel framework for emergency response management. *Journal of Cloud Computing*, *14*(1), 48.

[6] Mohammadi, R. (2026). Resource efficient task scheduling in Fog assisted IoMT using low rate wireless personal area networks. *Discover Internet of Things*.

[7] Khan, S., Liu, S., Pan, L., & Mei, G. (2025). Optimization-based hybrid offloading framework for IoMT in edge-Cloud healthcare systems. *Future Generation Computer Systems*, 108163.

[8] Choppara, P., & Lokesh, B. (2025). Efficient task scheduling and load balancing in Fog computing for crucial healthcare through deep reinforcement learning. *IEEE Access*, *13*, 26542-26563.

[9] Benaboura, A., Bechar, R., Kadri, W., Ho, T. D., Pan, Z., & Sahmoud, S. (2025). Latency-aware and energy-efficient task offloading in IoT and Cloud systems with DQN learning. *Electronics*, *14*(15), 3090.

[10] Murthy, C. V. B., & Shri, M. L. (2026). An Adaptive Blockchain Framework for Federated IoMT with Reinforcement Learning-Based Consensus and Resource Forecasting. *Scientific Reports*.

[11] Alamri, M., Haseeb, K., Humayun, M., & Alshammeri, M. (2025). AI-embedded IoT healthcare optimization with trust-aware mobile edge computing: M. Alamri et al. *Scientific Reports*.

[12] Xie, Q., & Ding, Z. (2025). Provably secure and lightweight blockchain based cross hospital authentication scheme for IoMT-based healthcare. *Scientific Reports*, *15*(1), 6461.

[13] Al-Alshaikh, H. A. (2026). A federated learning–enabled energy-aware anomaly detection algorithm for secure big data analytics in IoT-based smart healthcare systems. *Scientific Reports*.

[14] Zhang, J., Hu, Y., Shao, M., & Li, X. (2025). Joint energy-aware task offloading and privacy protection in healthcare monitoring systems via deep reinforcement learning. *Scientific Reports*.

[15] Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE engineering in medicine and biology magazine*, *20*(3), 45-50.