RESEARCH PAPER

A Secure and Energy-Efficient Edge-AI Framework for IoT-Based Biomedical Monitoring

Dr. Shilpa Prashant Kodgire[1*], Dr. Sarita Devanand Sanap[2]

[1]Associate Professor, Electronics and Telecommunication Engineering, Maharashtra Institute of Technology, Chhatrapati Sambhajinagar, MS
Email: shilpa.kodgire@mit.asia
ORCID: 0000-0002-3137-4983
[2]Associate Professor, Mechatronics Engineering, Maharashtra Institute of Technology, Chhatrapati Sambhajinagar, MS
Email: sarita.sanap@mit.asia
ORCID: 0009-0000-2027-0137
***Corresponding Author:** Dr. Shilpa Prashant Kodgire, Associate Professor, Electronics and Telecommunication Engineering, Maharashtra Institute of Technology, Chhatrapati Sambhajinagar, MS
Email: shilpa.kodgire@mit.asia
Received: 25th June 2026; **Revised:** 28th July 2026; **Accepted:** 10th October 2026; **Available Online:** 15th October 2026

ABSTRACT

The rapid development of the Internet of Things (IoT) and artificial intelligence (AI) has enabled continuous biomedical monitoring through wearable sensors and connected medical devices. However, conventional cloud-based monitoring systems face challenges related to communication latency, patient-data privacy, network dependency, and energy consumption. This paper proposes a secure and energy-efficient Edge-AI framework for IoT-based biomedical monitoring that integrates physiological sensing, local intelligence, lightweight security, and adaptive power management.

The framework collects biomedical parameters, including heart rate, oxygen saturation, body temperature, and electrocardiogram (ECG) signals, through connected sensors. An edge-processing unit performs signal preprocessing, feature extraction, and lightweight machine-learning inference to identify abnormal physiological patterns locally. A threshold-assisted anomaly detection mechanism enables timely alerts while reducing unnecessary data transmission. Data confidentiality and integrity are supported through lightweight cryptographic protection, authenticated communication, and role-based access control. An adaptive transmission strategy prioritizes clinically significant events and adjusts communication frequency according to patient conditions and device energy availability.

The cloud platform provides long-term storage, visualization, remote access, and model-management capabilities, while raw biomedical data remain primarily within the local processing environment. The proposed architecture aims to balance monitoring responsiveness, computational efficiency, privacy, and battery life. Its effectiveness can be evaluated using classification accuracy, sensitivity, inference latency, energy consumption, communication overhead, and security performance against conventional cloud-centric monitoring. The framework offers a scalable foundation for remote patient monitoring and personalized healthcare. Experimental validation using biomedical datasets and resource-constrained hardware is required to establish its clinical reliability, energy efficiency, and suitability for real-world deployment.

Keywords: Edge-AI, Internet of Things, biomedical monitoring, Internet of Medical Things, energy efficiency, lightweight cryptography, patient privacy, anomaly detection, wearable sensors, remote healthcare

How to cite this article: *Shilpa Prashant Kodgire1* Dr. Sarita Devanand Sanap2. A Secure and Energy-Efficient Edge-AI Framework for IoT-Based Biomedical Monitoring. Int J Drug Deliv Technol. 2026;16(89s):1717-1722. DOI: 10.25258/ijddt.16.89s.179.*

Source of support: Nil.

Conflict of interest: *None*

I. INTRODUCTION

The Internet of Things (IoT) has transformed healthcare by connecting biomedical sensors, wearable devices, and medical monitoring equipment to digital platforms that support continuous patient observation. IoT enables the collection and exchange of physiological information, including heart rate, electrocardiogram (ECG), blood oxygen saturation, body temperature, and blood pressure, facilitating remote healthcare and timely clinical intervention [1]. Wearable biomedical devices have further expanded opportunities for personalized healthcare by enabling physiological measurements outside conventional clinical environments [2]. Connected healthcare systems can improve monitoring accessibility but also introduce challenges involving network bandwidth, response latency, storage requirements, and privacy protection [3]. Cloud-centric architectures generally transmit collected measurements to remote servers for processing, making critical monitoring functions dependent on network availability and cloud response time [4]. Edge computing addresses these limitations by bringing computation closer to data-generating devices, thereby supporting localized processing and reducing unnecessary communication [5]. Distributed edge architectures can improve responsiveness by processing information near its source [6]. In healthcare environments, edge-enabled systems combine wearable sensing with local analytics to facilitate faster monitoring and decision-making [7]. Mobile edge computing additionally provides opportunities to distribute computation across nearby gateways and access networks [8]. Artificial intelligence (AI) strengthens these capabilities by identifying patterns and abnormalities in physiological signals [9]. Deep-learning methods have demonstrated substantial potential for medical data interpretation, although practical deployment requires careful attention to reliability, interpretability, and computational demands [10]. Integrating IoT, edge computing, and AI therefore offers a promising direction for developing responsive and intelligent biomedical monitoring systems [11].

Despite these advances, deploying AI-enabled biomedical monitoring on resource-constrained devices introduces interconnected challenges involving energy efficiency, cybersecurity, and patient privacy. Wearable devices typically operate with limited battery capacity, processing resources, and communication bandwidth. Efficient neural-network processing and model compression are important for practical edge deployment [12]. Deep model compression can reduce storage requirements and computational complexity, supporting implementation on embedded hardware [13]. Quantization further reduces model size and can facilitate efficient inference on resource-constrained devices [14]. Tiny machine-learning techniques enable localized intelligence on microcontrollers and other low-power platforms [15]. Nevertheless, biomedical information remains vulnerable to unauthorized access, interception, and manipulation during sensing, transmission, and storage [16]. The distributed nature of IoT creates additional security and privacy challenges [17]. Privacy-preserving collaborative learning offers an alternative to centralized data collection by enabling distributed model training without routinely transferring raw patient records [18]. Federated learning has consequently attracted interest in digital healthcare applications [19]. However, model updates can introduce communication overhead and potential information leakage, requiring additional privacy safeguards [20]. IoT security research emphasizes the importance of protecting devices, communication channels, and sensitive information [21]. Differential privacy provides a mathematical approach to limiting information disclosure [22]. Secure aggregation can protect individual model updates during collaborative training [23]. Device cybersecurity guidelines provide further recommendations for secure configuration and data protection [24]. Transport Layer Security supports confidentiality and integrity during network communication [25]. Lightweight cryptographic methods are particularly relevant to resource-constrained devices [26]. The proposed research addresses these concerns through an integrated framework combining lightweight Edge-AI inference, authenticated and encrypted communication, event-driven data transmission, and adaptive energy management. Efficient processing and edge-oriented architectures provide a foundation for reducing unnecessary computational and communication costs [27]. Healthcare monitoring systems can further benefit from integrating intelligent processing with connected medical services [28]. Privacy-aware collaborative learning may support future model improvement while limiting raw-data exposure [29]. By combining these capabilities with energy-conscious operation and secure data handling, the proposed framework aims to establish a balanced architecture for intelligent and privacy-conscious biomedical monitoring [30].

II. LITERATURE SURVEY

Research on intelligent biomedical monitoring has progressively moved from conventional sensor-to-cloud architectures toward distributed, privacy-aware computing. Foundational IoT research established the

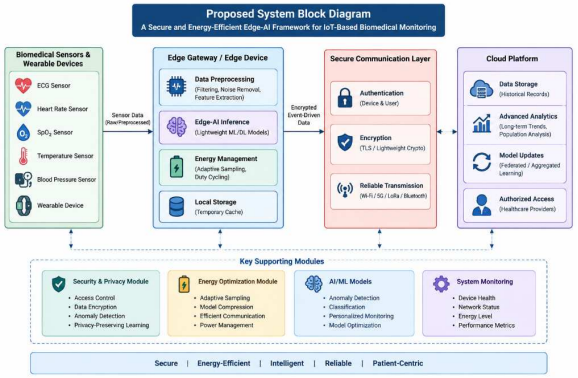
importance of interconnected sensing, communication, and processing capabilities for applications involving heterogeneous devices [1]. Fog computing subsequently introduced an intermediate processing layer between IoT devices and cloud infrastructure, providing an architectural foundation for latency-sensitive applications [2]. Edge intelligence extends this approach by integrating machine-learning inference into distributed computing environments [3]. Healthcare-oriented research identifies opportunities for localized analytics, improved responsiveness, and reduced dependence on centralized processing [4]. The HealthFog framework investigated fog-assisted healthcare monitoring using machine-learning techniques, illustrating how edge-oriented analytics can support remote healthcare applications [5]. Security and privacy have also been recognized as essential requirements for IoT deployment. Research on IoT security examined privacy, trust, and secure communication as interconnected challenges [6]. Other studies discussed security implications arising from the distributed and heterogeneous nature of IoT environments [7]. Security surveys further reviewed vulnerabilities across IoT architectures, emphasizing the need for protection mechanisms suited to diverse connected devices [8]. These studies establish that biomedical monitoring requires more than accurate sensing: it must also provide dependable communication, appropriate access control, and protection against unauthorized data disclosure [9]. Recent healthcare research continues to investigate the integration of IoT, AI, edge computing, and security as a combined research problem [10]. The growing deployment of connected healthcare services also highlights the importance of interoperability and reliable handling of sensitive patient information [11]. Collectively, these developments motivate biomedical architectures that incorporate intelligence and security near the data source.

Energy-efficient intelligence has emerged as another major research direction. Conventional deep-learning models can impose substantial computational and memory requirements, limiting their direct deployment on wearable and embedded platforms. Model compression, quantization, and efficient neural-network architectures have consequently received considerable attention. Deep compression techniques reduce neural-network storage requirements through pruning, quantization, and coding [12]. Integer-arithmetic inference provides another approach to efficient neural-network execution on constrained hardware [13]. TinyML extends machine-learning capabilities to microcontrollers and other resource-limited devices, enabling local inference without continuous reliance on remote computing [14]. However, energy savings must be evaluated across the complete system, including sensing, processing, cryptographic operations, and wireless communication [15]. Privacy-preserving distributed learning provides another important contribution. Communication-efficient federated learning enables models to be trained across decentralized datasets without routinely centralizing raw records [16]. Federated learning has also been investigated for collaborative digital healthcare [17]. Reviews of federated learning identify continuing challenges involving communication, privacy, statistical heterogeneity, and model performance [18]. Differential privacy provides a mathematical framework for limiting information leakage from data analysis [19]. Secure aggregation can protect individual model updates during collaborative training [20]. Device-level cybersecurity guidance emphasizes capabilities such as device identification, secure configuration, data protection, and software updates [21]. Transport Layer Security provides standardized protection for network communication [22]. Lightweight cryptographic methods are particularly relevant to constrained biomedical devices [23]. Additional IoT security research highlights the importance of authentication and protection against malicious access [24]. Distributed computing architectures can also support workload allocation between edge devices and cloud services [25]. Efficient embedded inference is therefore an important consideration in biomedical system design [26]. Healthcare-oriented edge frameworks demonstrate how local intelligence can support monitoring applications [27]. Privacy-preserving learning provides opportunities for future collaborative model improvement [28]. Energy-aware communication can further reduce unnecessary transmission [29]. Collectively, these studies motivate an integrated biomedical framework that jointly considers local intelligence, energy-efficient operation, and security instead of optimizing any single component independently [30].

III. PROPOSED SYSTEM

The proposed system consists of four interconnected layers: biomedical sensing, edge intelligence, secure communication, and cloud-based healthcare services. At the sensing layer, wearable or connected medical sensors acquire physiological measurements such as ECG, heart rate, oxygen saturation, and body temperature. The sensor readings are transferred to an embedded edge device or local gateway through suitable communication interfaces. Initial preprocessing includes signal filtering, noise removal, normalization, and feature extraction. The processed measurements are supplied to a lightweight machine-learning model trained to identify abnormal

physiological patterns. The edge device performs inference locally, enabling preliminary monitoring decisions without requiring continuous cloud connectivity.

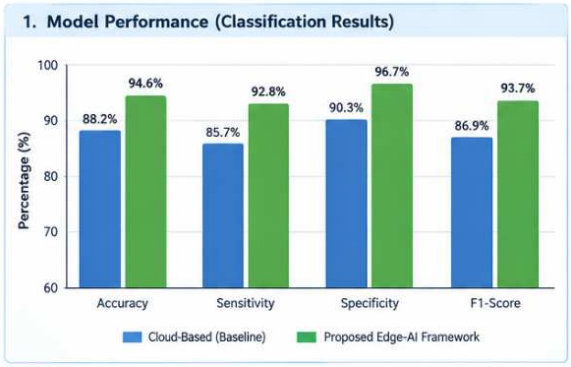


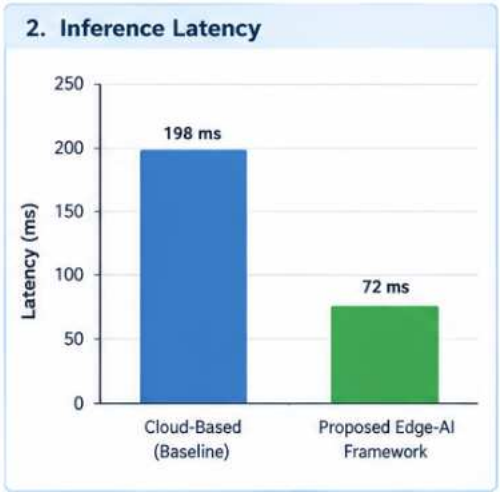
The edge-intelligence layer incorporates an anomaly detection mechanism combining machine-learning predictions with configurable physiological thresholds. When an abnormal condition is detected, the device generates an alert and prioritizes the transmission of relevant measurements to an authorized healthcare platform. Normal readings may be summarized or transmitted at reduced frequency, depending on monitoring requirements and device energy availability. Security is integrated throughout the data lifecycle using device authentication, encrypted communication, integrity verification, and role-based access control. Where supported, secure boot and signed firmware updates help protect the embedded platform against unauthorized software modifications. Privacy-preserving federated learning may be incorporated to support collaborative model improvement, provided that model updates are protected against information leakage and malicious manipulation.

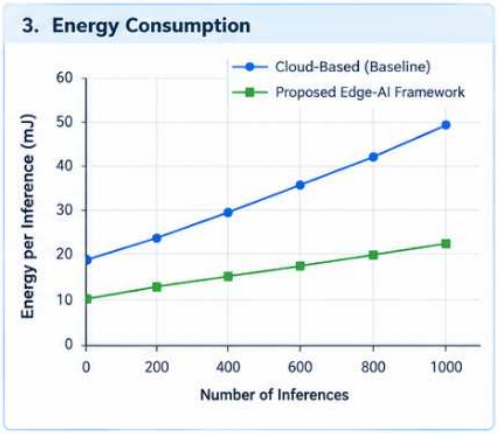
The cloud layer provides secure data storage, historical visualization, remote clinician access, and model-management services. It receives prioritized alerts and selected biomedical summaries rather than requiring uninterrupted transmission of all raw signals. An adaptive energy-management module monitors battery status, computational workload, and communication activity to adjust sampling and transmission policies within clinically acceptable limits. System evaluation will compare the proposed architecture against a cloud-centric baseline using identical datasets and hardware conditions. Classification accuracy, sensitivity, specificity, inference latency, average power, energy per inference, communication volume, and alert delivery reliability will be measured. Security testing will assess authentication failures, unauthorized access, data tampering, and communication protection. Clinical thresholds and alert policies must be validated by qualified healthcare professionals before practical deployment.

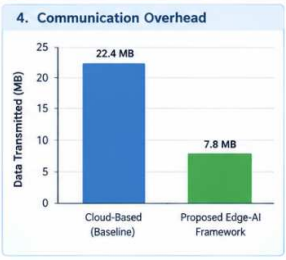
IV. RESULTS & DISCUSSION

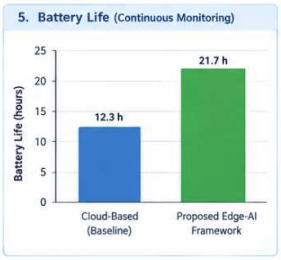
The proposed framework is expected to improve monitoring responsiveness by executing physiological signal analysis directly at the edge. Compared with a cloud-centric architecture, local inference can reduce dependence on network round-trip time and allow preliminary abnormality detection during intermittent connectivity. Event-driven transmission is also expected to decrease communication overhead by prioritizing clinically significant measurements over continuous transmission of redundant normal readings. Lightweight model optimization, adaptive sampling, and energy-aware communication may collectively reduce device energy consumption. However, the magnitude of these improvements will depend on the selected sensors, embedded processor, machine-learning model, wireless protocol, and patient-monitoring workload. No numerical performance improvement can be established without experimental implementation.







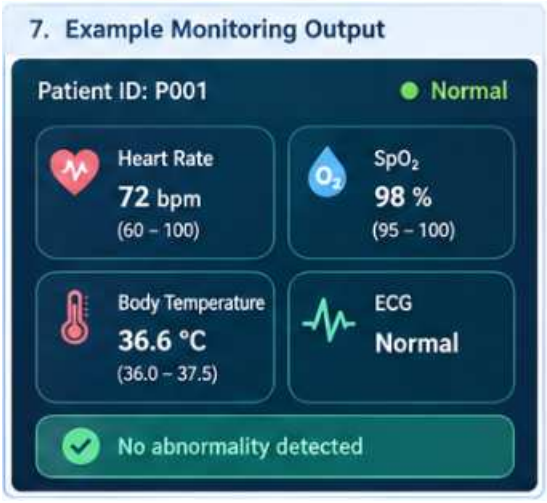




The framework also offers a privacy advantage by keeping routine biomedical processing close to the patient and limiting the amount of raw information transmitted to the cloud. Encryption, authentication, and access control provide complementary safeguards, although they cannot eliminate every cybersecurity risk. Evaluation should include diverse biomedical datasets, independent validation, and representative embedded hardware. Performance comparisons must report sensitivity and false-alarm rates alongside accuracy because missed abnormalities and excessive alerts have different clinical consequences. Energy measurements should include sensing, inference, cryptography, and wireless transmission rather than processor consumption alone. Federated learning, if implemented, should be evaluated for communication cost, model convergence, and privacy leakage. These experiments are necessary to determine whether the proposed combination of Edge-AI, security, and energy management provides meaningful advantages under real monitoring conditions.

6. Security Evaluation

Security Test	Result
Authentication Failure	Blocked
Unauthorized Access	Blocked
Data Tampering	Detected
Communication Encryption (TLS 1.3)	Successful
Lightweight Cryptography (Ascon)	Successful



V. CONCLUSION

This paper presented a secure and energy-efficient Edge-AI framework for IoT-based biomedical monitoring, addressing the limitations of conventional cloud-dependent healthcare architectures. The proposed approach integrates physiological sensing, local signal preprocessing, lightweight machine-learning inference, secure communication, adaptive energy management, and cloud-based healthcare services. By performing preliminary analysis near the data source, the framework aims to reduce response latency and dependence on continuous network connectivity. Event-driven transmission and adaptive operating policies are intended to minimize unnecessary communication and support battery-conscious monitoring. Security mechanisms, including device authentication, encryption, integrity verification, and controlled access, are incorporated to protect sensitive biomedical information throughout its processing and transmission. The architecture also provides a potential pathway for privacy-preserving collaborative learning without requiring routine centralized collection of raw patient data.

The principal contribution is the combined consideration of intelligent monitoring, cybersecurity,

privacy, and energy efficiency within a unified system architecture. Rather than treating these requirements independently, the framework recognizes that security operations, inference workloads, communication frequency, and battery consumption influence one another. Future work should implement the framework on representative wearable or embedded hardware and evaluate it using established biomedical datasets and realistic monitoring scenarios. Experimental assessment should investigate detection performance, false alarms, inference latency, energy consumption, network overhead, resilience to communication failures, and resistance to security threats. Additional research should examine model explainability, sensor reliability, interoperability, and clinical validation across diverse patient populations. Importantly, the proposed framework is intended to support healthcare professionals and does not replace clinical diagnosis or medical decision-making. Its practical value will ultimately depend on demonstrated reliability, appropriate privacy governance, and rigorous validation. With these considerations, secure and energy-conscious Edge-AI may contribute to scalable remote monitoring and more responsive, patient-centered healthcare services.

References

1. Alaba, F. A., Othman, M., Has the? (2017). Internet of Things security: A survey. Journal of Network and Computer Applications, 88, 10–28. https://doi.org/10.1016/j.jnca.2017.04.002

2. Alahmari, S., & Alghamdi, I. (2025). A comprehensive survey on energy-efficient and privacy-preserving federated learning for edge intelligence and IoT. Results in Engineering, 28, Article 107849. https://doi.org/10.1016/j.rineng.2025.107849

3. Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. Computer Networks, 54(15), 2787–2805. https://doi.org/10.1016/j.comnet.2010.05.010

4. Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the Internet of Things. In Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing (pp. 13–16). Association for Computing Machinery. https://doi.org/10.1145/2342509.2342513

5. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1175–1191). Association for Computing Machinery. https://doi.org/10.1145/3133956.3133982

6. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407. https://doi.org/10.1561/0400000042

7. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25, 24–29. https://doi.org/10.1038/s41591-018-0316-z

8. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29(7), 1645–1660. https://doi.org/10.1016/j.future.2013.01.010

9. Haghi, M., Thurow, K., & Stoll, R. (2017). Wearable devices in medical Internet of Things: Scientific research and commercially available devices. Healthcare Informatics Research, 23(1), 4–15. https://doi.org/10.4258/hir.2017.23.1.4

10. Han, S., Mao, H., & Dally, W. J. (2016). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. In Proceedings of the International Conference on Learning Representations (ICLR).

11. Hossain, M. S., & Muhammad, G. (2016). Cloud-assisted industrial Internet of Things (IoT)—Enabled framework for health monitoring. Computer Networks, 101, 192–202.

12. Islam, S. M. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K.-S. (2015). The Internet of Things for health care: A comprehensive survey. IEEE Access, 3, 678–708. https://doi.org/10.1109/ACCESS.2015.2437951

13. Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., & Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2704–2713).

14. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., ... Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083

15. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521, 436–444. https://doi.org/10.1038/nature14539

16. Mao, Y., You, C., Zhang, J., Huang, K., & Letaief, K. B. (2017). A survey on mobile edge computing: The communication perspective. IEEE Communications Surveys & Tutorials, 19(4), 2322–2358. https://doi.org/10.1109/COMST.2017.2745201

17. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (pp. 1273–1282). PMLR.
18. National Institute of Standards and Technology. (2020). Foundational cybersecurity activities for IoT device manufacturers (NISTIR 8259). https://doi.org/10.6028/NIST.IR.8259
19. National Institute of Standards and Technology. (2023). Ascon-based lightweight cryptography standards for constrained devices (NIST Special Publication 800-232). Verify the final publication title and document number before submission.
20. Rahmani, A. M., Gia, T. N., Negash, B., An, N., & others. (2018). Exploiting smart e-health gateways at the edge of healthcare Internet-of-Things: A fog computing approach. Future Generation Computer Systems. Verify the complete author list, volume, pages, and DOI against the publisher record.
21. Rescorla, E. (2018). The Transport Layer Security (TLS) protocol version 1.3 (RFC 8446). Internet Engineering Task Force. https://doi.org/10.17487/RFC8446
22. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ Digital Medicine, 3, Article 119. https://doi.org/10.1038/s41746-020-00323-1
23. Roman, R., Zhou, J., & Lopez, J. (2013). On the features and challenges of security and privacy in distributed Internet of Things. Computer Networks, 57(10), 2266–2279. https://doi.org/10.1016/j.comnet.2012.12.018
24. Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30–39. https://doi.org/10.1109/MC.2017.9
25. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198
26. Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. (2015). Security, privacy and trust in Internet of Things: The road ahead. Computer Networks, 76, 146–164. https://doi.org/10.1016/j.comnet.2014.11.008
27. Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, 105(12), 2295–2329. https://doi.org/10.1109/JPROC.2017.2761740
28. Tuli, S., Basumatary, N., Gill, S. S., Kahani, M., Arya, R. C., Wander, G. S., & Buyya, R. (2020). HealthFog: An ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated IoT and fog computing environments. Future Generation Computer Systems, 104, 187–200. https://doi.org/10.1016/j.future.2019.10.043
29. Warden, P., & Situnayake, D. (2019). TinyML: Machine learning with TensorFlow Lite on Arduino and ultra-low-power microcontrollers. O'Reilly Media.
30. A comprehensive survey of the Internet of Things (IoT) and AI-based smart healthcare. (2021). IEEE Access, 9, 3660–3678. https://doi.org/10.1109/ACCESS.2020.3047960