**Edelweiss Applied Science and Technology**
*ISSN: 2576-8484*
*Vol. 10, No. 6, 300-320*
*2026*
*Publisher: Learning Gate*
*DOI: 10.55214/2576-8484.v10i6.13134*

# Systematic literature review: Utilization of the Internet of Things for real-time patient health monitoring system optimization based on edge computing

Jansen Chandra[1*], Sani Muhamad Isa[2]
[1,2]Department of Master Computer Science, Bina Nusantara University, Jakarta, Indonesia; jansen.chandra@binus.ac.id (J.C.) sani.m.isa@binus.ac.id (S.I.).

**Abstract:** The rapid development of the Internet of Things (IoT) has accelerated the digital transformation of healthcare services, particularly in real-time patient health monitoring. However, conventional cloud-based monitoring systems often face challenges related to high latency, bandwidth consumption, and data security. Therefore, this study aims to systematically synthesize the utilization of IoT integrated with edge computing architectures for optimizing real-time patient health monitoring systems. This research employed a Systematic Literature Review (SLR) approach following the PRISMA 2020 guidelines by analyzing publications from IEEE Xplore, PubMed, Scopus, Web of Science, and Google Scholar between 2021 and 2026. From 309 identified articles, 15 studies met the inclusion criteria and were selected for detailed analysis. The findings indicate that hybrid fog-edge architectures can reduce latency by up to 70%, decrease bandwidth usage by 60%, improve energy efficiency by 30%, and achieve clinical detection accuracy ranging from 91% to 99% compared with conventional cloud-based approaches. These results demonstrate that edge computing significantly enhances the responsiveness, reliability, and efficiency of IoT-based patient monitoring systems. Practically, the findings provide valuable guidance for researchers, healthcare institutions, and system developers in designing secure, scalable, and low-latency patient monitoring solutions that support the ongoing digital transformation of healthcare services.

***Keywords:*** *Edge computing, Internet of Medical Things, Quality of service, Real-time patient monitoring, Systematic literature review.*

## 1. Introduction

The development of digital technology in the healthcare sector has opened up new opportunities in the transformation of patient monitoring services as a whole. Internet of Things (IoT) is now the key foundation in building a smart healthcare ecosystem, where wirelessly connected sensor devices are able to continuously collect patient physiological data such as heart rate, oxygen saturation, blood pressure, and body temperature. Integration of IoT technology with biosensors has proven to make a significant contribution to early monitoring of chronic diseases and remote management of patient conditions, thereby shifting the healthcare paradigm from reactive to more proactive and preventive [1].

While the potential of IoT in healthcare continues to grow, the reliance on centralized cloud poses several crucial technical problems. Sending data from IoT devices to central servers over conventional networks results in high latency, wasteful bandwidth, and sensitive patient data security risks. This condition makes the system based on the cloud less reliable for health monitoring applications that require a response in milliseconds, especially in medical emergency situations. Research shows that the application of edge computing is able to reduce latency up to 50% compared to the conventional cloud system, so as to directly improve the quality of patient monitoring in real-time [2].

**History:** Received: 31 March 2026; Revised: 3 June 2026; Accepted: 8 June 2026; Published: 24 June 2026
* Correspondence: jansen.chandra@binus.ac.id

As a solution to these limitations, Edge Computing comes as a computing paradigm that is close to the data source, allowing information processing to be carried out locally at the network layer closest to the device. In the context of the Internet of Medical Things (IoMT), this approach includes the use of computing-based fog and Edge, which work hierarchically to filter, analyze, and forward only clinically valuable data to a central system. Systematic study of solutions in Edge Computing in healthcare revealed that this technology is consistently able to improve diagnostic accuracy, response time efficiency, and the resilience of patient monitoring systems in various clinical conditions [3].

A hybrid architecture that blends fog computing and Edge Computing is getting more and more research attention because of its ability to handle multi-level data processing efficiently. On the edge layer, health data is processed using a Machine Learning lightweight to detect anomalies instantly, while the fog coating manages mid-level data aggregation and analysis before sending it to the cloud. This approach has been shown to result in a decline in latency of up to 70%, an increase in energy efficiency by 30%, as well as savings in bandwidth of 60% compared to pure architecture-based cloud, making it highly relevant for the implementation of patient monitoring in both hospital and remote service environments [4].

The urgency of a research-based Systematic Literature Review (SLR) on this topic is increasing in line with the growth of the global IoMT market, which will reach a value of USD 57.62 billion in 2023 with a projected annual growth of 25.9% until 2030. A comprehensive study of implementation trends, challenges, and gaps is important for the development of IoT-based patient monitoring systems, and Edge Computing can be directed in a targeted manner. Literature review of wearable devices for chronic disease monitoring indicates the need to standardize health parameters, improve interoperability between devices, and conduct more rigorous clinical validation of the system, Remote Patient Monitoring [5].

This research is here to fill the literature gap by presenting a systematic study that focuses on the use of IoT in optimizing patient health monitoring systems in real-time based on Edge Computing. Through a structured SLR methodology, the study will identify the latest technological architectures, analyze the effectiveness of various approaches to Edge Computing in the context of health, as well as mapping the direction of relevant future research. The results of the study are expected to make a meaningful scientific contribution to the development of an efficient, safe, and responsive smart health system to meet clinical needs [6].

This research departs from the main problem that IoT-based patient health monitoring systems that still rely on centralized cloud infrastructure face serious obstacles in the form of high latency, bandwidth inefficiency, and data security vulnerabilities that have the potential to compromise patient safety in critical conditions. Therefore, the formulation of the problem raised is: (1) What are the development and characteristics of the architecture of real-time patient health monitoring systems based on IoT and edge computing based on the literature in 2021–2025? (2) What are the technical challenges and research gaps that remain unresolved in the implementation of edge computing for patient monitoring optimization? (3) How effective is the application of edge computing compared to cloud-based approaches in the context of the quality of Service (QoS) of health monitoring systems?

This study aims to conduct a systematic review of the latest scientific literature to identify, classify, and synthesize various approaches to IoT and edge computing technologies used in real-time patient health monitoring systems. In particular, this study aims to: (1) map the research trends and system architecture that are predominantly used in the period 2021–2025; (2) analyze the advantages, limitations, and technical challenges faced in the implementation of edge computing in the clinical environment; and (3) identify research gaps that can be the basis for the development of more optimal patient monitoring systems in the future.

The results of this study are expected to provide dual benefits, both theoretically and practically. Theoretically, this study enriches the wealth of science in the field of health informatics and distributed computing systems through a comprehensive and structured literature synthesis. In practical terms, the findings of this research can be used by researchers, system developers, and healthcare practitioners as a reference in designing a more efficient, secure, and responsive IoT-based patient-monitoring architecture. In addition, the resulting research-gap mapping is expected to encourage the birth of more reliable edge-computing system innovations in supporting the digital transformation of health services at the national and global levels.

## 2. Literature Review

### *2.1. Basic Concepts of IoT in Health Monitoring*

The application of Internet of Things (IoT) in the health sector has experienced rapid development, especially in the aspect of continuous sensor-based patient condition monitoring. Devices wearable connected through IoT networks are able to collect physiological data such as heart rate, blood glucose levels, blood pressure, and physical activity of patients without the need for manual intervention. A literature mapping review focused on the integration of Machine Learning, Edge Computing, and wearable devices in healthcare found that research in this area has increased significantly in the past three years, with key applications including fall detection, cardiovascular monitoring, and disease prediction [7].

### *2.2. Edge Computing Architecture in IoMT Systems*

Internet of Medical Things (IoMT) requires a computing infrastructure capable of processing data quickly near the data source, rather than on a remote centralized server. Framework Edge Computing is designed specifically for patient monitoring in real-time. Implement a multi-tiered architecture that distributes compute load among devices, patient-proximate, server Edge departmental level, and infrastructure cloud hospital. The empirical evaluation of the architecture shows a real advantage in the Latency, reliability, energy efficiency, and diagnostic accuracy when compared to a system-based cloud, Conventional [8].

### *2.3. Machine Learning-Based Anomaly Detection in Healthcare IoT Systems*

Integrations of Machine Learning into health IoT systems are a key strategy to improve the ability to detect anomalies in patients' physiological data. Algorithms such as Support Vector Machine, Random Forest, and Deep Learning models have been applied to the device edge to analyze sensor data patterns directly without reliance on a stable internet connection. Research on anomaly-based detection of machine learning in healthcare IoT systems proves that this approach is able to improve the overall security effectiveness of the system, while shortening response times to clinical conditions that require immediate treatment [9].

### *2.4. Data Security and Privacy on Edge-Based IoMT Systems*

One of the most critical challenges in the implementation of IoMT is the protection of patient health data, which is highly sensitive. Transmission of raw data from IoT devices to servers/cloud through public networks opens security gaps that have the potential to harm patients and health institutions. Implementation of the approach clustering on the layer edge computing has proven to improve data confidentiality and security in IoMT environments by minimizing the volume of data delivered to the cloud and implementing encryption locally before transmission takes place [10].

### *2.5. Quality of Service in IoT-Based Patient Monitoring Systems*

Quality of service or Quality of Service (QoS) is a fundamental aspect that determines the reliability of IoT-based patient monitoring systems in a real clinical environment. QoS parameters such as throughput, bandwidth, transmission delay, jitter, and packet loss must be optimally managed so that the system can respond to medical emergencies without significant delays. A systematic literature review of on-device QoS monitoring and edge IoT-based approaches in healthcare identifies that edge computing and artificial intelligence are the most relevant solutions to overcome the challenges of high latency levels that often hinder the effectiveness of cloud-based health systems [2].

*DOI: 10.55214/2576-8484.v10i6.13134*

### *2.6. Trends and Challenges of IoMT Implementation in Healthcare*

The development of the IoMT ecosystem globally brings complex opportunities as well as challenges, including aspects of interoperability between devices, system scalability, and compliance with data protection regulations such as the Health Insurance Portability and Accountability Act (HIPAA). The diversity of sensor devices from various manufacturers with different communication protocols is a real obstacle in the overall integration of the system. State-of-the-art research that applies federated learning and homomorphic encryption on IoT-based edge architecture for patient monitoring in real time achieved an accuracy of 91.9% with an F1 score of 90.8%, while demonstrating that this approach is clinically feasible [11].

## 3. Methods

### *3.1. Research Design*

This study uses the Systematic Literature Review (SLR), which is a study method designed to identify, evaluate, and synthesize all relevant scientific evidence to research questions in a structured, transparent, and replicable manner. The SLR methodology was chosen because it is able to produce a comprehensive overview of the current research status in the field of IoT and Edge Computing for monitoring the patient's health. The reporting guidance used refers to the statement Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020, which sets minimum standards in the reporting of systematic studies to ensure the readability, accuracy, and auditability of the methodology used [12].

### *3.2. Literature Search Strategy*

Literature searches were carried out systematically on five leading scientific databases, namely IEEE Xplore, PubMed, Scopus, Web of Science, and Google Scholar. The selection of this database considers the scope of publications in the fields of computer engineering, health informatics, and biomedical sciences that are relevant to the research topic. The search string is structured using a combination of the main keywords with the AND and OR logic operators, including terms such as "Internet of Things", "Edge Computing", "Real-time patient monitoring", "IoT healthcare", "IoMT", and "Remote Health Monitoring". The publication time range is limited to 2021 to 2025 to ensure the novelty of the findings. This approach is aligned with best practices applied in a systematic review of IoT in healthcare, where searches were conducted across nine large databases using both automated and manual search methods simultaneously [13].

### *3.3. Inclusion and Exclusion Criteria*

Article selection is carried out based on inclusion and exclusion criteria that have been set a priori before the search process begins. Inclusion criteria include: (1) articles published in journals or scientific proceedings peer-reviewed between 2021 and 2025; (2) articles that discuss the application of IoT or Edge Computing in the context of patient health monitoring; (3) English-language articles with full text accessible; and (4) articles that explicitly discuss aspects of real-time processing, system architecture, or quality of service. Exclusion criteria include: articles that are opinion-only with no empirical data, articles with unclear study designs, and articles that have no direct relevance to the topic of IoT-based patient monitoring. This selection process is consistent with the SLR procedure used in previous studies, where eligibility criteria are established to capture high-quality studies from various large databases in a structured manner [14].

### *3.4. Data Selection and Extraction Process*

Article selection is carried out in four successive stages. The first stage is the initial identification based on the title and keywords of the search results. The second stage is filtering based on abstracts to eliminate irrelevant articles. The third stage is a feasibility assessment based on full-text screening. The fourth stage is a methodological quality assessment using a standard checklist to ensure each article included has adequate reliability and validity. The entire selection process is documented in the PRISMA flowchart, which contains the number of articles at each stage of selection. Data extraction is then carried out in a structured manner using a form that records information about the author, year of publication, study objectives, methodology, system architecture studied, key findings, and research limitations.

### *3.5. Data Synthesis and Analysis*

The extracted data were analyzed narratively and thematically, considering the heterogeneity of approaches in the identified studies. The synthesis was carried out by grouping findings based on the main themes that emerged from the literature, including edge computing architecture, system performance, data security mechanisms, and implementation challenges. This approach allows for the identification of patterns, research gaps, as well as recommendations for future system development in a comprehensive and evidence-based manner.

## 4. Results and Discussion

### *4.1. Results*

#### *4.1.1. Literature Selection Process*

The process of identifying and selecting articles in this systematic review was carried out with reference to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guideline. The entire selection stage is visualized in the following flowchart.

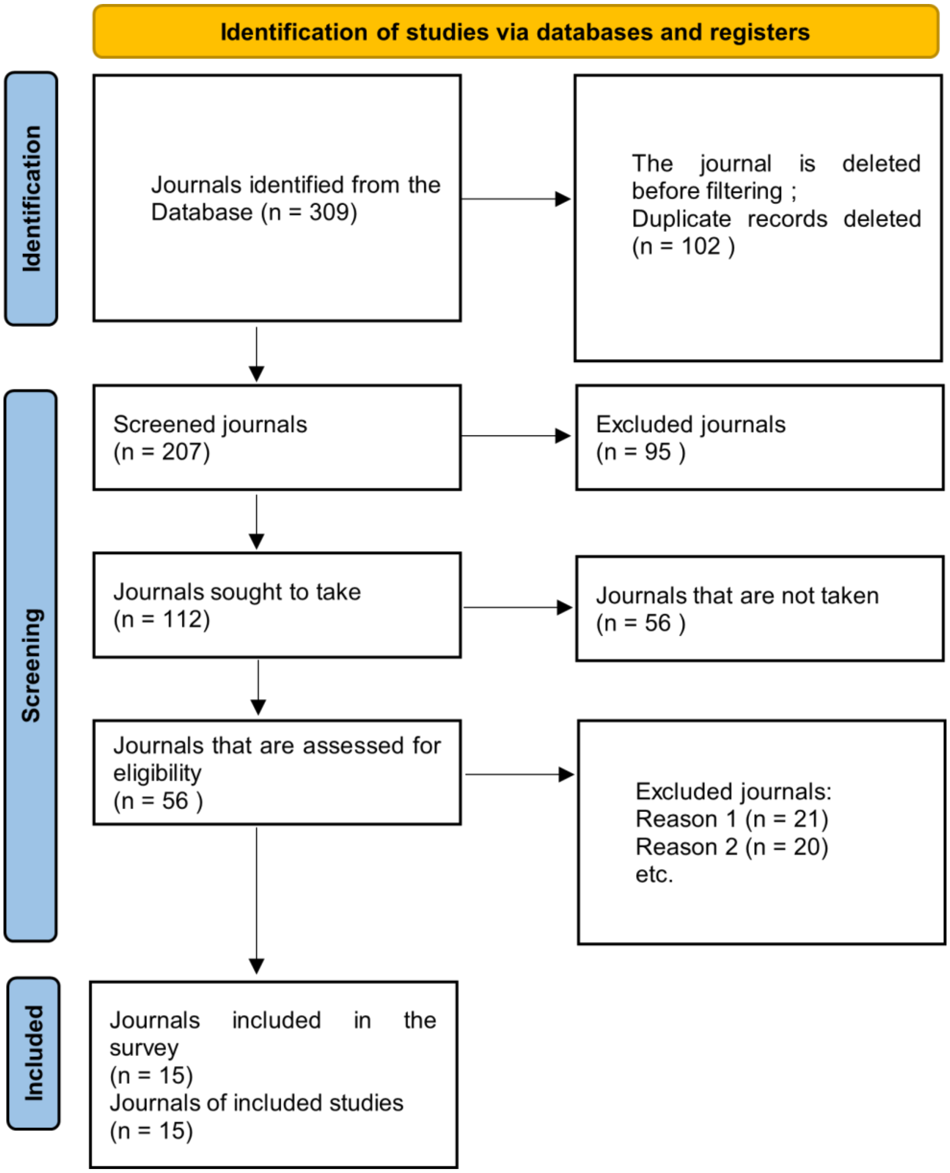

**Figure 1.**
PRISMA Flow Literature Selection Process Diagram: Systematic Review of the Utilization of IoT and Edge Computing for Real-Time Patient Health Monitoring System Optimization (2021–2025).

Figure 1 presents the flow of article selection starting from the identification stage to the final inclusion stage. At the identification stage, 309 articles were found from five established scientific databases, namely IEEE Xplore, PubMed, Scopus, Web of Science, and Google Scholar. After initial screening in the form of the removal of duplicate recordings and articles that did not meet the basic criteria, as many as 102 articles were eliminated, so that 207 articles remained to enter the screening stage. At the screening stage, based on titles and abstracts, as many as 95 articles were excluded because they were not relevant to the focus of the research topic, resulting in 112 articles, which were then assessed for feasibility through full-text screening. Of the 112 articles, 56 articles were not taken for various reasons of technical and thematic exclusion, resulting in 56 articles that entered the final eligibility assessment. At this final stage, as many as 21 articles were excluded based on Reason 1 and

*DOI: 10.55214/2576-8484.v10i6.13134*

20 articles based on Reason 2, so that 15 articles were determined to meet all the inclusion criteria and became the main analytical corpus of this study.

### *4.1.2. Characteristics of Selected Articles*

Based on the selection process that has been carried out, as many as 15 scientific articles have been determined as primary data sources for this systematic study. The complete characteristics of all the included articles are presented in the following table.

**Table 1.**
Summary Table of Selected Articles: Authors, Titles, Methods, Datasets, Findings, And Relevance to Topics of Systematic Study of the Utilization of Iot and Edge Computing For Real-Time Patient Health Monitoring (2021–2025).

| Yes | Author | Title | Method | Samples/Datasets | Researchers' Findings | Relevance to the Topic |
|---|---|---|---|---|---|---|
| 1 | Rathi, et al. [15] | An Edge AI-Enabled IoT Healthcare Monitoring System for Smart Cities | Simulation of edge computing + neural network systems to model transmission latency | Edge node-based simulation and edge controller | End-to-end latency, computing, optimization, and transmission latency show excellent results. Neural networks can accurately model transmission latency in real-world scenarios | Highly relevant: integrating IoT, edge computing, and AI for real-time, low-latency patient monitoring in smart cities |
| 2 | Wu, et al. [16] | IoT-Based Wearable Health-Monitoring Device and Its Validation for Potential Critical and Emergency Applications | IoT wearable device design + cloud-based real-time monitoring system validation | COVID-19 quarantined patients/individuals | The system is able to monitor body temperature, SpO2, and heart rate in real time. The data is automatically uploaded to the cloud for medical personnel to monitor remotely. The burden on health workers is significantly reduced | Relevant: demonstrates the implementation of wearable IoT for real-time monitoring of patients' critical physiological parameters |
| 3 | Joo, et al. [17] | A Patient Management System Using an Edge Computing-Based IoT Pulse Oximeter | Architectural design of edge computing + IoT pulse oximeter for biosignal collection and analysis | Hospital patient biosignal data (IR AC, IR DC, Red AC, Red DC, AMB, pulse, SpO2) | Edge nodes in each ward are able to collect and analyze biosignal data in real-time without relying on the cloud. Elimination of waiting times compared to cloud computing | Highly relevant: directly addresses IoT-based edge computing for real-time patient management in medical facilities |
| 4 | Famá, et al. [18] | An IoT-Based Interoperable Architecture for Wireless Biomonitoring of Patients with Sensor Patches | Design of interoperable IoT architectures using BLE, MQTT, and FHIR standards for EHR integration | Inpatient with a sensor patch on the body and bed | The architecture successfully integrates low-energy wireless sensors with hospital information systems through EHR exchanges. Communication between entities | Relevant: Demonstrate an interoperable IoT architecture for wireless patient monitoring connected to an electronic medical record system |

*DOI: 10.55214/2576-8484.v10i6.13134*

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | runs efficiently | |
| 5 | Shukla, et al. [19] | Identification and Authentication in Healthcare IoT Using Integrated Fog Computing-Based Blockchain Model | Fog computing + blockchain + Advanced Signature-Based Encryption (ASE) algorithm | Simulation using iFogSim (NetBeans) and SimBlock | The detection accuracy of ASE malicious nodes in fog computing is 91% (vs. 83% in the cloud). Reliability in fog: 95% (vs. 87% in the cloud). Superior to FogBus, FemtoCloud, BFAN, and BeeKeeper | Relevant: optimizing the security and authentication of health IoT devices using fog computing that supports real-time patient data transmission |
| 6 | Ali, et al. [20] | An Innovative IoT and Edge Intelligence Framework for Monitoring Elderly People Using Anomaly Detection on Data from Non-Wearable Sensors | Edge IoT framework + machine learning (Isolation Forest & LSTM) for behavioral anomaly detection | Data on non-wearable sensors in the elderly | The system is able to detect behavioral anomalies of the elderly in real-time with ML models. The dashboard provides real-time alerts and longitudinal trends for medical personnel. Privacy is maintained due to local processing | Highly relevant: using edge computing and IoT for real-time health monitoring of elderly patients with automated anomaly detection |
| 7 | Pereira, et al. [7] | Machine Learning Applied to Edge Computing and Wearable Devices for Healthcare: Systematic Mapping of the Literature | Systematic mapping literature (171 studies, 28 selected articles) | 171 scientific studies, focused on 28 key articles | Research has increased significantly in the last six years. CNN and LSTM are the most widely used. Edge platforms such as Raspberry Pi and smartphones are dominant. The field is still developing and needs architectural standardization | Relevant: provides a comprehensive overview of the integration of ML, edge computing, and wearable IoT in healthcare as a cornerstone of the literature |
| 8 | Alasmary [21] | Scalable Digital Health (SDH): An IoT-Based Scalable Framework for Remote Patient Monitoring | IoT framework + latency-based edge computing autoscaling + microservice architecture on Kubernetes (AWS) | Patient vital parameters: ECG, body temperature, blood pressure, oxygen saturation | The system is able to monitor vital parameters in real-time with automatic pod adjustments based on latency thresholds. Accessibility, cost-efficiency, and responsiveness are significantly increased | Highly relevant: delivering a scalable IoT framework based on edge computing for real-time remote patient monitoring |
| 9 | Farag [22] | A Tiny Matched Filter-Based CNN for Inter-Patient ECG Classification and Arrhythmia Detection at the Edge | Tiny CNN+ Matched Filter (MF) theory for ECG classification in edge devices | MIT-BIH dataset; validation on INCART, QT, and PTB diagnostic databases | Accuracy 98.18%, sensitivity 91.90%, F1-score 92.17%. The model is only 15 KB with an inference of <1 ms. Can be run on edge devices with limited | Highly relevant: proving the effectiveness of mild ML models on edge devices for real-time detection of ECG arrhythmias in |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | resources | patients |
| 10 | Daraghmi, et al. [23] | Edge–Fog–Cloud Computing Hierarchy for Improving Performance and Security of NB-IoT-Based Health Monitoring Systems | Three-layer architecture (Edge-Fog-Cloud) + NB-IoT + ML (Decision Tree, SVM, Logistic Regression) | Real medical vital sign data; CloudSim simulation, iFogSim, ns3-NB-IoT | NB-IoT latency was reduced by 59.9%, average execution time decreased by 38.5%, and authentication time was reduced by 35.1% for large devices | Highly relevant: proposing an NB-IoT-based edge-fog-cloud hierarchy for high-efficiency optimization of real-time health monitoring |
| 11 | Suresha, et al. [24] | An Edge Computing and Ambient Data Capture System for Clinical and Home Environments | Open-source edge computing + ambient sensors (video, audio, temperature, humidity, geolocation) | Clinical and home environments; Occupancy data, alarms, geolocation, temperature, humidity | Occupancy estimation accuracy 94%, F1-score alarm classification 0.98, geolocation accuracy 98.7%, temperature sensor RMSE 0.3°C. Cost-effective system and maintain patient privacy | Relevant: demonstrate a non-intrusive edge computing system for patient monitoring in homes and clinics with different types of ambient sensors |
| 12 | Jaber, et al. [25] | Remotely Monitoring COVID-19 Patient Health Condition Using Metaheuristics Convolutional Networks from IoT-Based Wearable Device Health Data | IoT wearable + deep learning (metaheuristic convolute network) + cloud + real-time GPS | COVID-19 patient health data from IoT wearable sensors | A three-layer system (sensor, cloud, web) can remotely monitor the condition of COVID-19 patients and provide automatic alerts to patients and families in real-time | Relevant: demonstrates the utilization of wearable IoT and deep learning for real-time health monitoring of COVID-19 patients based on edge and cloud |
| 13 | Islam, et al. [4] | A Hybrid Fog-Edge Computing Architecture for Real-Time Health Monitoring in IoMT Systems with Optimized Latency and Threat Resilience | Hybrid fog-edge computing + rule-based filtering + Decision Tree + One-Class SVM + end-to-end encryption | Simulation of real-life case scenarios on IoMT systems | Latency reduced by 70%, energy efficiency increased by 30%, bandwidth savings by 60%, threat detection time reduced by 50% compared to cloud-only models | Highly relevant: hybrid fog-edge architecture for real-time IoMT monitoring that optimizes latency while improving patient data security |
| 14 | Khan and Alkhathami [9] | Anomaly Detection in IoT Based Healthcare: Machine Learning for Enhanced Security | Supervised ML (Random Forest, AdaBoost, Logistic Regression, Perceptron, DNN) + feature reduction | CIC IoT Dataset (33 types of IoT attacks, 7 main categories) | Random Forest achieves 99.55% accuracy for binary classification and multiclass IoT attacks, accompanied by a significant reduction in compute response time | Relevant: demonstrates the importance of health IoT system security through ML anomaly detection to protect real-time patient monitoring data |
| 15 | Baig, et al. [26] | An Edge–Fog–Cloud IoT Framework for Real-Time | Edge-fog-cloud IoT + Pan-Tompkins++ + ML set | PhysioNet datasets + trials in real hospital wards | The highest ensemble ECG classification accuracy was | Highly relevant: edge-fog-cloud IoT framework for real-time |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Cardiac Monitoring and Rapid Clinical Alerts in Hospital Wards | (Random Forest, XGBoost, DT, Naive Bayes, KNN, SVM) + Firebase Firestore | | 91.96%. Mobile alerts average 15.23 ± 2.71 seconds. Patient routine evaluation time is significantly reduced | cardiac monitoring in hospital wards with rapid clinical alerts |

Table 1 summarizes the overall characteristics of the 15 included articles, including information on the identity of the author, the title of the study, the methodological approach, the data source or dataset used, the main findings, and the level of relevance of each study to the study topic. Of the 15 articles, 10 articles are categorized as having very high relevance to the topic of IoT-based patient monitoring and edge computing, while the other 5 articles are categorized as relevant because they discuss supporting aspects such as data security, interoperability, and machine-learning-based anomaly detection. The year range of publication covers the period 2021 to 2026, with the largest concentration in 2022–2025, reflecting the accelerated growth of research in the last five years in this field.

### *4.2. Discussion*

#### *4.2.1. Development and Architectural Characteristics of Real-Time Patient Health Monitoring System Based on IoT and Edge Computing*

Based on the analysis of 15 included articles, significant evolution was identified in the architecture of IoT-based patient health monitoring systems during the period 2021–2025. These developments demonstrate a consistent paradigm shift from a centralized cloud-based approach to a distributed architecture that puts computing capabilities closer to data sources, whether in the form of pure edge computing, fog computing, or hybrid configurations that synergistically integrate the two.

The starting point of this architectural development can be traced to the work of Rathi et al. [15], who developed a monitoring system based on Edge AI for the Smart City environment by integrating IoT, Edge Computing, and artificial intelligence simultaneously [15]. The results of the simulation in the study prove that Neural Networks are able to accurately model transmission latency in real-world scenarios, where end-to-end performance, latency, computing, and transmission optimization showed excellent results. These findings indicate that processing based on Edge AI is not just a theoretical construct but a viable solution that can be implemented practically in complex urban environments that demand high responsiveness to patients' health conditions.

A further development was seen in the study of Wu et al. [16], who designed and validated a wearable IoT for critical and emergency monitoring applications. The developed system is proven to be able to monitor body temperature, SpO2, and heart rate in real-time in COVID-19 quarantined patients, with physiological data automatically uploaded to the cloud to be monitored by medical personnel remotely, so that the workload of health workers is significantly reduced [16]. This study lays the empirical foundation that sensor integration wearable with a reliable data transmission infrastructure is the minimum prerequisite for the realization of a clinically functional patient monitoring system.

On the dimension of decentralization of computing to medical facilities, Joo et al. [17] designed a patient-based management system, Edge Computing Using Pulse Oximeter IoT, that places Edge Node in each hospital ward [17]. This architecture is able to collect and analyze biosignal data, including IR AC, IR DC, Red AC, Red DC, AMB, Pulse, and SpO2 in real time without dependence on the cloud, while eliminating the waiting time that is an inherent characteristic of a conventional cloud-based system. The success of Joo et al. [17] empirically confirms that the decentralization of computing to the Edge provides real responsiveness advantages that the cloud-centric approach cannot achieve [17].

In terms of system interoperability, Famá et al. [18] present an interoperable IoT architecture that uses the Bluetooth Low Energy (BLE), Message Queuing Telemetry Transport (MQTT), and the Fast Healthcare Interoperability Resources (FHIR) standard for integration with Electronic Health Records (EHR). This architecture successfully integrates low-energy wireless sensors with hospital information systems through efficient EHR data exchange between all entities involved. Successful integration of multiprotocols in studies, Famá et al. [18] provide concrete evidence that the interoperability barriers that have been a major obstacle to IoMT implementation can be systematically overcome through the adoption of open communication standards that have been established in the healthcare industry [18].

Meanwhile, Alasmary developed a Framework for Scalable IoT Edge Computing with the ability to auto-scale based on latency thresholds implemented on the architecture microservice on Kubernetes (AWS) [21]. System ScalableDigitalHealth (SDH) is able to monitor vital parameters such as Electrocardiogram (ECG), body temperature, blood pressure, and oxygen saturation in real time, with automatic pod adjustments proven to improve system accessibility, cost efficiency, and responsiveness simultaneously. Alasmary [21] shows that dynamic scalability is no longer an obstacle in IoMT-based Edge architectures, provided that the system design pays attention to the orchestration mechanism of computing resources that is adaptive to fluctuating workloads [21].

Complementing the existing architectural spectrum, Suresha et al. [24] presented a system Edge computing open-source that utilizes ambient sensors, including video, audio, temperature, humidity, and geolocation for non-intrusive patient monitoring in both clinical and home environments [26]. The system achieves estimation accuracy of 94%, alarm classification with F1-Score 0.98, 98.7% geolocation accuracy, and a temperature sensor Root Mean Square Error (RMSE) of 0.3°C, while proving to be cost-effective and inherently preserving patient privacy as the data never leaves the local layer. Suresha et al. [24] contribution opens up a new perspective that patient monitoring systems based on Edge do not have to rely on a device worn directly by the patient, but can instead be operated non-intrusively through strategically positioned environmental sensors [26].

The most recent development in patient monitoring architecture is represented by the study of Islam et al. [4], who developed the architecture Hybrid fog-edge computing for IoMT systems [4]. By applying Rule-based filtering, Decision Tree, One-Class Support Vector Machine (SVM), and encryption end-to-end in layers, this hybrid architecture recorded a 70% reduction in latency, a 30% increase in energy efficiency, and savings Bandwidth 60%, as well as a 50% reduction in threat detection time compared to the cloud-only model. The synthesis of all of the above architecture findings confirms a consistent trajectory towards a more intelligent, distributed, and responsive health monitoring system, with a hybrid fog-edge architecture as the most mature representation of such evolution.

The map of overall research development was captured by Pereira et al. [7] through a systematic mapping of 171 scientific studies focused on 28 key articles. This study reveals that Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are models Machine Learning most predominantly used on the platform Edge, such as Raspberry Pi and smartphones, and that research in this field has experienced significant growth in the last six years. However, Pereira et al. [7] also identified that the field still requires a more comprehensive standardization of architecture, a finding that confirms that despite rapid technical developments, methodological cohesion between studies still needs to be substantially improved [7].

Examining the development trajectory of the architecture that has been identified, it appears that each generation of the system brings a new layer of capabilities that further narrows the gap between clinical needs and technological capabilities. If, in the early period of the study, patient monitoring systems still relied on massive raw data delivery to centralized infrastructure, then in the current period, the processing logic has shifted fundamentally towards the intelligence-at-the-edge paradigm, where devices at the leading layer no longer function as passive data collectors but as active analytical units that can make initial clinical decisions independently without dependence on a continuous network connection.

The paradigmatic shift does not occur linearly but through a series of complementary architectural experiments. On the one hand, the attempt to miniaturize machine-learning models, represented by the tiny neural-network approach, has opened up the possibility of applying direct intelligent inference to battery-powered devices with very limited computing capacity. On the other hand, innovations in wireless communication protocols such as Narrowband IoT (NB-IoT) and Bluetooth Low Energy (BLE) have simultaneously expanded the range of connectivity while significantly reducing energy consumption, allowing sensor devices to operate for long periods without requiring frequent power changes. This synergy between computing efficiency and communication efficiency is actually the main driver of the viability of the edge-computing ecosystem in a clinical context that demands long-term operational reliability.

A dimension that often goes unnoticed in architecture discussions is the aspect of the system's resilience to unexpected network disruptions. In critical patient monitoring scenarios, a failure of the connection between the edge device and the cloud infrastructure should not result in a loss of overall monitoring capabilities. Resilient architectures must be designed with the principle of graceful degradation, where the system remains capable of performing local detection and warning functions even if connectivity to the top layer is temporarily disconnected. This kind of autonomous operation capability is a critical differentiator between a truly clinically ready edge-based monitoring system and an implementation that only functions optimally under ideal network conditions.

Furthermore, the development of architecture also shows a tendency to adopt containerization and microservices approaches as the foundation of software infrastructure. This approach allows for modular updating of system components without interrupting overall operations, a critical need in a hospital environment where service interruptions of even short durations can impact patient safety. Flexible deployment capabilities and efficient maintenance through container-based architectures are among the increasingly significant driving factors for adoption in large-scale healthcare institutions.

Another aspect that is increasingly receiving research attention is the integration of the concept of digital twins into the edge computing-based patient monitoring ecosystem. By building a dynamic digital representation of a patient's physiological condition that is continuously updated in real time based on sensor data, the system can perform predictive simulations of the patient's clinical condition trajectory before actual deterioration occurs. This kind of predictive approach shifts the role of the monitoring system from a reactive instrument to a proactive clinical-anticipation tool that has the potential to provide sufficient lead time for medical personnel to carry out preventive interventions. Digital-twin integration with on-premises edge infrastructure also has inherent privacy advantages, as patient-representation models can be built and operated entirely on the local layer without exposing personally identifiable data to external networks.

It is also important to consider the ecological sustainability dimension of the developed architecture. The proliferation of IoT devices on a large scale in healthcare environments generates a substantial energy footprint if not managed properly. The edge architecture is designed with the principles of energy harvesting and adaptive duty cycling, where the sensor device dynamically adjusts the sampling frequency and data transmission based on the clinical conditions of the monitored patient, not only extending the operational life of the device but also contributing to the reduction of the environmental impact of large-scale monitoring infrastructure. Energy sustainability is a design consideration that can no longer be ignored, especially given the projected rapid growth of the IoMT ecosystem in the coming decade.

From an implementation-readiness perspective in developing countries, including Indonesia, open-source edge computing-based architectures offer a more economically accessible path of adoption than cloud-based proprietary solutions. The ability to build high-performance monitoring systems using locally available commodity hardware, combined with open-source software that can be tailored to the specific needs of local health facilities, opens up real transformation opportunities for healthcare systems that have been constrained by limited technology infrastructure budgets. However, successful implementation in this context requires an adequate supporting ecosystem, including local technical capacity for system maintenance, a conducive regulatory framework, and sustainable financing mechanisms from health policymakers.

Exploring the dynamics of the development of IoT-based patient monitoring system architectures, it was identified that the process of technological evolution was not entirely driven by hardware advancements alone, but also by fundamental changes in the way the research community and the healthcare industry redefined the concept of "intelligence" in medical systems. The old paradigm of placing analytics intelligence exclusively on a centralized server layer now faces the serious conceptual challenge of a new paradigm that distributes decision-making capabilities across the entire layer of the architecture, from devices attached to the patient's body to local gateways in hospital wards.

In this framework of intelligence redistribution, the concept of tinyML or Machine Learning (ML) on microdevices emerged as one of the most transformative breakthroughs. The ability to run optimized inference models on microcontrollers with power consumption below one milliwatt opens up the possibility of applying clinical analytics directly at the data-acquisition point without the need for data transmission to higher computing layers. The clinical implications of this ability are significant: the detection of critical physiological anomalies such as ventricular fibrillation or sleep apnea can be done within milliseconds from the time the biological signal is first recorded by the sensor, long before the data have been transmitted to any external analytics system. The speed of response achieved through this kind of on-device inference approach fundamentally changes the limits of the possibilities of medical early-warning systems.

The evolution of architecture also reflects a shift in system design philosophy from a monolithic approach to an ecosystem that is modular and reconfigurable. Early monolithic architectures, in which all system components were tightly bound together in one inflexible platform, proved incapable of accommodating the ever-evolving diversity of clinical needs. In contrast, a modular architecture based on microservices that separates the functions of data acquisition, signal processing, clinical inference, medical record storage, and notification interfaces into components that can be independently developed, tested, and updated provides much greater flexibility in adapting the system to the specific needs of various medical specialties.

The next development that deserves special attention is the integration of 5G technology as a next-generation communication infrastructure that has the potential to fundamentally change the QoS characteristics of the edge computing ecosystem for healthcare. With end-to-end latency that can theoretically be reduced to under one millisecond, and bandwidth capacity that far exceeds that of previous generations, 5G networks allow for more flexible placement of edge servers in locations while still maintaining responsiveness characteristics that have only been achieved through physical proximity between devices and computing units. The concept of Multi-access Edge Computing (MEC), which leverages 5G infrastructure to distribute computing capabilities to network access points, is an architectural evolution that will further define the landscape of patient monitoring systems in the coming decade.

The physical security dimension of edge infrastructure is also an architectural aspect that is often overlooked but has real operational implications. Edge server devices placed in hospital wards or treatment rooms are in a physically more accessible environment than a centralized, tightly protected data center. The risk of physical manipulation, device theft, or accidental operational interruption by non-technical personnel is a design consideration that must be addressed through a combination of physical-protection mechanisms, local-storage encryption, and robust device-authentication protocols. An architecture that does not seriously consider these physical security dimensions risks creating exploitable weak points, even though its digital cybersecurity layer has been very well designed.

Device lifecycle management in a large-scale IoMT edge ecosystem is also an increasingly pressing architectural challenge. A large hospital operating thousands of IoT sensor devices simultaneously needs a management infrastructure capable of handling automatic and secure firmware updates, security certificate rotation, detection and isolation of devices behaving anomalously, and decommissioning devices that have reached the end of their lifespan without interrupting the continuity of monitoring services. Architectures that do not have a comprehensive solution for device lifecycle management will face progressive degradation of security and performance as the age and scale of deployment increase.

Developments in semantic interoperability technology through standard medical ontologies such as the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and the Fast Healthcare Interoperability Resources (FHIR R4) terminology standard are opening a new chapter in the integration of edge monitoring systems with the broader electronic medical record infrastructure. The ability of edge devices to not only generate raw numerical data but also generate clinically meaningful semantic annotations in standard formats that can be interpreted by a wide range of health information systems is a significant leap in quality. This kind of semantic integration allows real-time monitoring data from the edge layer to automatically enrich patients' medical records with structured clinical context, substantially improving the continuity of care across departments and across institutions.

Finally, the trend toward event-driven architecture in place of conventional periodic polling approaches also reflects the maturity of architectural thinking in the patient monitoring ecosystem. In the polling approach, the system periodically takes sensor readings regardless of whether or not there is a significant change in conditions, resulting in a substantial waste of computing and communication resources. Instead, an event-driven architecture that only triggers data transmission and processing when sensor conditions exceed a predefined clinical threshold results in significantly better resource efficiency without sacrificing responsiveness to changes in critical patient conditions. This transition to an event-driven paradigm represents a maturation of a system design philosophy that is increasingly oriented towards clinical relevance rather than just data completeness.

### *4.2.2. Technical Challenges and Research Gaps in the Implementation of Edge Computing for Patient Monitoring Optimization*

Despite the significant progress that has been made, the analysis of the overall included articles identifies several fundamental technical challenges as well as research gaps that remain unresolved. These challenges can be mapped into four main clusters, namely data security and privacy, limitations of edge computing devices, ecosystem standardization, and clinical validation.

The first cluster, namely device security and authentication, was discussed in depth by Shukla et al. [19], who proposed the model Blockchain-based fog computing with the algorithm Advanced Signature-Based Encryption (ASE). The results of the simulation using the iFogSim and SimBlock platforms showed that the accuracy of the detection node is dangerous with the ASE approach in the layer fog reach 91%, surpassing the accuracy of the system-based cloud, which is only 83% [19]. System reliability in fog is also higher, namely 95% compared to 87% in cloud, and outperforms comparator approaches such as FogBus, Femto cloud, BFAN, and BeeKeeper. The implications of Shukla et al. [19] findings are very unequivocal: security-based Blockchain on the layer fog provides protection advantages that cannot be replicated by conventional centralized security systems, especially when highly sensitive patient health data is processed near the data collection point [19].

In the security dimension based on artificial intelligence, Khan and Alkhathami [9] developed an anomaly-detection system using supervised machine-learning algorithms including Random Forest, AdaBoost, Logistic Regression, Perceptron, and Deep Neural Network (DNN) with optimized feature reduction [9]. Testing on the CIC IoT Dataset, which contains 33 types of IoT attacks in seven main categories, resulted in accuracy: Random Forest 99.55% for binary classification or multiclass, with a significant reduction in compute response time. The gap identified from this study is that the datasets used are still generic for IoT environments in general, so specific validation on a more controlled clinical IoMT ecosystem is still indispensable to ensure the reliability of intrusion-detection systems in real hospital scenarios.

The second cluster is the limited computing resources on the device edge, which was responded to innovatively by Farag through model design Tiny Convolutional Neural Network (CNN) theory-based Matched Filter (MF) for ECG classification inter-patient and arrhythmia detection [22]. Testing on MIT-BIH datasets with cross-validation on INCART, QT, and PTB yielded an accuracy of 98.18%, a sensitivity of 91.90%, and an F1 score of 92.17%, with a model size of only 15 KB and an inference time of less than 1 millisecond. Farag's achievement proves that the limitations of the computing capacity of the device Edge is not an absolute obstacle, but at the same time implies a research gap in the form of the need for similar model validation in more diverse arrhythmic conditions as well as in patient populations with wider demographic variations [22]. The challenges of monitoring in a mass health emergency are highlighted by Suresha et al. [24] through a study in COVID-19 patients utilizing a combination of IoT wearable, Deep Learning based on metaheuristic convolutional networks, cloud, and Global Positioning System (GPS) tracking in real-time [24]. The developed three-layer system successfully provides alerts to patients and families in real-time, but the study implicitly implies that reliance on stable network connectivity remains a critical weak point in large-scale emergencies. This gap strengthens the argument that the Offline-first or Edge-first needs to be integrated as a mandatory component in any critical patient monitoring system design, not just as an additional optional feature.

The third cluster related to the monitoring of vulnerable and non-intrusive populations was discussed by Ali et al. [20], who developed a non-intrusive sensor-based wearable system for the elderly to use Forest Insulation and LSTM on the edge layer [20]. The system is able to detect behavioral anomalies in real time while maintaining privacy through local processing, as well as providing a dashboard with real-time alerts and longitudinal trends for medical personnel. However, the gaps identified from the Ali et al. [20] study are the limitations of large-scale clinical validation, as well as the question of the generalizability of anomalous models in a variety of different cultural contexts and living conditions, which is on the agenda of urgent follow-up research [20].

The fourth cluster related to ecosystem standardization and interoperability is represented by the previous findings of Famá et al. [18], which show that although the integration of the BLE, MQTT, and FHIR protocols is successful, these efforts are still a partial solution of a single research group without any standard-adoption mechanism that is universal in the healthcare industry [18]. This is emphasized by Pereira et al. [7], who explicitly identify the absence of standardization of standard architectures as the biggest research gap in this field, so that efforts to harmonize technical standards across vendors and platforms become an urgent agenda that must be solved collaboratively by the research community and industry [7].

Cumulatively, the technical challenges and research gaps identified from the 15 articles analyzed provide a clear roadmap for future research: truly optimal IoT-based patient monitoring systems and edge computing require simultaneous innovation across the four clusters, with the highest priority on clinically proven layered security mechanisms as well as architectural standardization that enables cross-functional interoperability platforms seamlessly.

Beyond the four previously identified challenge clusters, an in-depth analysis of the literature reveals additional layers of complexity that have not received proportionate research attention. One of the most significant dimensions is the challenge of managing data heterogeneity generated by the diverse IoMT sensor ecosystem. In a single patient monitoring scenario, the system might simultaneously integrate data from electrocardiography (ECG), photoplethysmography (PPG), accelerometer, skin temperature sensors, and blood pressure monitors, each with a different data format, sampling frequency, and noise characteristics that are different. The ability to harmonize these heterogeneous data streams in real-time at the edge layer without degradation of clinical quality is a much more complex computational challenge than the processing of a single data source, which is generally the focus of published studies.

In this regard, the challenge of temporal alignment between various sensor modalities is also a critical technical issue but is rarely explicitly discussed. Imperfect temporal synchronization between sensors with different acquisition frequencies can result in analytical artifacts that have the potential to lead to clinical misinterpretation. A truly medically reliable edge system must include a precise temporal synchronization mechanism, including compensation for clock drift on distributed sensor devices that operate over long periods of time.

A research gap that is also urgent to be filled is the lack of longitudinal studies on the degradation of edge system performance in long-term operational conditions. Most published studies report evaluation results in a relatively short span of time, while real clinical applications demand the reliability of the system for months to years. The accumulation of drift models in machine learning algorithms running at the edge layer, the decline in sensor performance due to continuous use, and the degradation of wearable device battery capacity are phenomena that will only be seen in long-term studies. This absence of longitudinal data creates a serious gap between the performance claims reported in scientific publications and the operational realities that would be encountered in actual clinical applications.

The dimension of challenges related to human factors in the edge computing ecosystem for health also needs to be given greater attention. Health workers' acceptance of recommendations and warnings generated by automated systems based on artificial intelligence is not universal and is influenced by various psychological, organizational, and cultural factors. The phenomenon of alert fatigue, in which medical personnel begin to ignore system notifications due to too-high alert volumes or low levels of confidence in system accuracy, is a real threat to the clinical effectiveness of even technically highly sophisticated monitoring systems. Research that integrates human-computer interaction perspectives and clinical workflow analysis into IoMT edge system design is still very rare but desperately needed.

Regulatory and legal compliance challenges are also gaps that require more serious research attention. The implementation of edge computing systems in clinical environments must comply with various health data protection regulatory frameworks applicable in each jurisdiction, such as the General Data Protection Regulation (GDPR) in Europe and the Health Insurance Portability and Accountability Act (HIPAA) in the United States. In Indonesia itself, the legal framework governing the use of health data in the digital system is still evolving, creating regulatory uncertainty that can hinder institutional adoption of technology. Research that explicitly integrates compliance analysis of the applicable legal framework into the technical architecture of the system is a much-needed contribution by the research community and practitioners.

A methodological gap worth noting is the lack of implementation of a federated learning approach across institutions in a true IoMT edge ecosystem. Although the concept of federated learning has received substantial theoretical attention as a machine learning model training solution that maintains data privacy, its practical implementation in multi-hospital contexts with heterogeneous edge infrastructure is still very limited. Technical barriers, such as inconsistencies in model architecture between institutions, differences in the distribution of patient demographic data, and the complexity of coordinating distributed training without centralized authority, are problems that require innovative and cross-disciplinary research solutions.

### *4.2.3. The Effectiveness of Edge Computing Implementation Compared to Cloud-Based Approaches in the Context of Quality of Service Health Monitoring Systems*

The third dimension of the formulation of this research problem focuses on the comparative evaluation between edge computing approaches and cloud-based architecture in the context of the Quality of Service (QoS) of patient health monitoring systems. Analysis of the entire included literature consistently shows significant advantages of edge approaches and hybrid architectures over cloud-only systems on almost all measured QoS parameters, including latency, energy efficiency, bandwidth, reliability, and diagnostic accuracy.

The most comprehensive evidence regarding the superiority of hierarchical architecture is found in the study of Daraghmi et al. [23], which proposed hierarchies edge-fog-cloud three layers using the technology Narrowband IoT (NB-IoT) combined with algorithms Decision Tree, SVM, and Logistic Regression [23]. Simulations using CloudSim, iFogSim, and ns3-NB-IoT against real medical vital sign data yielded very significant findings: NB-IoT latency was reduced by 59.9%, the average execution time was reduced by 38.5%, and authentication time was reduced by 35.1% for a large number of devices. The nearly 60% reduction in latency achieved by Daraghmi et al. [23] is a clinically significant figure, given that in the treatment of critical conditions such as sudden cardiac arrest, any second delay in system response can have irreversible consequences for patient safety [23].

The advantages of hybrid architecture in the context of multidimensional QoS are further reinforced by the findings of Islam et al. [4] that have been described earlier, where the combination of 70% latency reduction, 30% energy efficiency, savings bandwidth 60%, and a 50% reduction in threat detection time simultaneously provide a very strong quantitative argument that investment in infrastructure hybrid fog-edge produce return QoS that goes far beyond conventional-based approaches cloud pure [4].

In the specific context of cardiac monitoring in hospital wards, Baig et al. [26] present the Framework IoT edge-fog-cloud that integrates the Pan-Tompkins++ algorithm with the Machine Learning Suite, which includes Random Forest, XGBoost, Decision Tree, Naive Bayes, K-Nearest Neighbors (KNN), and SVM, with Firebase Firestore as the infrastructure Backend [26]. Testing on PhysioNet Datasets, which is validated through live testing in real hospital wards, resulting in ECG classification accuracy, together with the highest at 91.96%, with average delivery time Mobile Alert only 15.23 ± 2.71 seconds, accompanied by a significant decrease in the patient's routine evaluation time. Clinical validation in an actual hospital setting conducted by Baig et al. [26] provides a higher translational value than simulation-based studies alone, as it proves that the QoS advantage of the edge-fog-cloud can be maintained even under complex and uncontrolled operational conditions [26].

Advantages of Edge Computing in diagnostic accuracy in real-time are also proven by Farag [22], which achieves an accuracy of 98.18% and an F1-Score of 92.17% in the ECG classification, with an inference time of less than 1 millisecond on a low-powered Edge Device [22]. This comparison is particularly relevant because the system is based on the cloud. Conventional for similar inference tasks generally take hundreds of milliseconds to a few seconds to complete the full cycle of data transmission, inference, and return of results to client devices, an unacceptable pause in the context of critical arrhythmia detection that requires a response in milliseconds.

On the performance dimension of comprehensive biosignals, Joo et al. [17] proved that Edge Node placed in hospital wards directly eliminates the waiting times that are an inherent characteristic of the cloud-based system, resulting in real-time analytical capabilities. Actually, for complex multichannel biosignal data [17]. Complementary to this, Alasmary [21] adds a dimension of scalability to QoS evaluations, where the system Autoscaling Threshold-based latency has been proven to maintain optimal performance even as the number of connected IoT devices dynamically increases, a critical advantage for deployments in large hospitals with hundreds of monitoring points operating simultaneously [21].

A comparison of QoS from a security perspective shows a consistent pattern. Shukla et al. [19] prove that fog computing-based Blockchain achieved 95% reliability compared to 87% in cloud [19], while Khan and Alkhathami [9] proved that anomaly detection based on Random Forest on the edge layer is able to identify 33 types of IoT attacks with 99.55% accuracy, far exceeding the capabilities of cloud-reactive security systems that rely on centralized rule updates. The integration of the security dimension as a component of QoS is important to underline because in the IoMT ecosystem, security vulnerabilities directly impact the integrity of clinical data on which medical decision-making is based [9].

From the perspective of cost as an often overlooked QoS component, Suresha et al. [24] prove that the Edge computing open-source can achieve high monitoring performance of F1-Score 0.98 alarm classification and 98.7% geolocation accuracy with significantly lower infrastructure costs compared to solution-based commercial cloud solutions [24]. These findings are particularly strategic for the context of implementation in developing countries, including Indonesia, where budget constraints on health infrastructure are often a major obstacle to the adoption of advanced patient monitoring technologies.

The study of Jaber et al. [25] provides a QoS perspective from the Alert Automated-based cloud in emergency contexts, where a three-layer system that combines sensors, wearable IoT, and Deep Learning metaheuristic successfully conveys notifications of patient condition in real-time to patients and families [25]. Although this study uses the cloud as the primary transmission layer, the results obtained confirm that a clinically acceptable notification speed is only possible if initial processing is carried out on a layer closer to the data source, reinforcing the position of Edge Computing as a technical prerequisite of QoS that cannot be compromised.

Baig et al. [26] in their study also indirectly affirm this, where the time Alert 15.23 seconds achieved in architecture edge-fog-cloud is the result of the decentralization of ECG processing to the

Edge before the notification is sent via cloud, rather than directly sending raw data to a central server for analysis [26].

Overall, the synthesis of all 15 articles analyzed yielded a firm and consistent conclusion: the edge computing approach in both its pure, fog, and hybrid architectures systematically outperforms the cloud-based approach on all relevant QoS parameters for patient health monitoring. Latency reductions ranging from 38.5% to 70%, energy-efficiency improvements of up to 30%, bandwidth savings of up to 60%, clinical detection accuracy of 91–99%, and increased reliability of security systems from 87% to 95% are convincing quantitative evidence of the superiority of distributed computing paradigms in the IoT-based health monitoring ecosystem.

These findings collectively confirm that the transformation of patient health monitoring systems from a cloud-centric architecture to an edge-first paradigm is not just a more optimal technical choice but a fundamental clinical imperative to ensure the safety, efficiency, and quality of healthcare in an era of increasingly accelerating digital transformation. Successful, documented implementation in a variety of clinical contexts, ranging from ECG monitoring in hospital wards, arrhythmia detection on edge devices, non-intrusive elderly monitoring, to management of mass health emergencies, provides a strong foundation for healthcare decision-makers to immediately accelerate the adoption of edge computing architecture as the new standard for IoT-based patient monitoring infrastructure globally.

Deepening the comparative analysis of QoS that has been presented, it is important to consider that the quantitative advantages of edge computing documented in various studies actually represent only a fraction of the entire spectrum of clinical value offered by the distributed computing paradigm. Beyond technical performance metrics such as latency and throughput, there are QoS dimensions that are more difficult to quantify but have no less significant clinical impacts, namely service reliability under network-degradation conditions, performance consistency under fluctuating compute loads, and the ability to automatically recover systems after component failures.

The aspect of system resilience or resistance to partial failure is an increasingly critical QoS parameter as the complexity of the architecture increases. On a tiered edge-fog-cloud system, a failure of one of the layers should not result in the failure of the entire system when the failover mechanism is well designed. The ability of the edge layer to operate autonomously in isolated mode when the connection to the fog or cloud layer is compromised is a QoS characteristic that distinguishes a truly clinically reliable system from one that is only optimal under ideal operational conditions. This resilience engineering perspective needs to be integrated more systematically into the QoS evaluation framework in future studies.

In the context of deployments in hospitals with varying quality network infrastructures, adaptive bitrate capabilities in sensor data transmission are also a relevant QoS component. A system that is able to dynamically adjust the resolution and frequency of data transmission based on available network conditions, without sacrificing essential clinical information, will provide much more consistent service reliability than systems with static transmission parameters. This adaptability is becoming increasingly important in the context of implementing remote patient monitoring in remote areas where network availability is not always reliable.

QoS evaluation in the end-user experience dimension, especially from the perspective of patients, as the most directly affected by the quality of the monitoring system, is also an aspect that is still underrepresented in the existing literature. The convenience of wearable devices in long-term use, the clarity of the notification interface for patients and families, and the accessibility of health information through user-friendly digital platforms are the dimensions of QoS that are human-centric and cannot be ignored in a comprehensive evaluation. Systems that are technically superior but fail to be consistently accepted and used by patients will not result in meaningful improvements in clinical outcomes in the real world.

A review of the total cost of ownership aspect as a component of economic QoS also yields an interesting comparative perspective. Although the initial implementation of edge infrastructure requires a greater hardware investment than a cloud subscription-based solution, the long-term analysis shows that the cost savings in data transmission, the reduction in network bandwidth requirements, and the minimization of cloud computing costs can cumulatively result in a financial break-even point in a relatively short span of time. This long-term economic perspective is particularly relevant for healthcare institutions that operate on a limited budget and need to consider the financial sustainability of the technology investments made.

Finally, QoS evaluation in the context of ecosystem scalability also needs to receive more in-depth attention. The ability of edge computing architectures to accommodate the growing number of connected IoT devices without proportional performance degradation is a fundamental prerequisite for deployment in growing healthcare institutions. New device auto-provisioning mechanisms, security certificate management at scale, and orchestration of software updates in bulk without interrupting clinical operations are scalability challenges that require mature architectural solutions and have been validated in real, full-scale deployment conditions.

## 5. Conclusion

A systematic review of 15 scientific articles for the period 2021–2026 confirms that the integration of the Internet of Things with edge computing architecture is able to fundamentally revolutionize real-time patient health monitoring systems. Cross-study findings prove that the hybrid fog-edge approach results in a 70% reduction in latency, a 60% reduction in bandwidth, and a 30% increase in energy efficiency compared to conventional cloud infrastructure. The clinical detection accuracy that ranges from 91–99% reinforces the argument that the decentralization of computational intelligence to the nearest layer of data sources is not just a technical advantage but rather a fundamental clinical prerequisite in guaranteeing the safety and quality of critical patient monitoring services. Future research is recommended to prioritize large-scale clinical validation under real-world operational conditions to address the gap between laboratory performance and field realities. The development of a universal IoMT architecture standardization framework needs to be driven collaboratively by the cross-institutional research community, including the implementation of federated learning in heterogeneous multi-hospital ecosystems. In addition, the aspect of compliance with health data protection regulations, especially in the context of Indonesian law that is still developing, as well as the integration of human factors perspectives in system design, needs to receive a much more proportionate portion of research attention in the future.

## Funding:

This research received partial financial support from Bina Nusantara University. Institutional Review Board Statement: Not applicable. This study is a systematic literature review and did not involve human participants, patient data, or animal subjects.

## Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

## Acknowledgement:

The authors would like to express their sincere gratitude to Sani Muhamad Isa for his supervision, valuable guidance, and constructive feedback throughout the research process. His insights and direction were instrumental in shaping the methodology and enhancing the overall quality of this study. The authors also extend their appreciation to Bina Nusantara University for providing academic support, partial research funding, and a conducive environment for conducting this research. The authors take full responsibility for the content of this article.

## Copyright:

## References

[1] M. N. Hosain, Y.-S. Kwak, J. Lee, H. Choi, J. Park, and J. Kim, "IoT-enabled biosensors for real-time monitoring and early detection of chronic diseases," *Physical Activity and Nutrition*, vol. 28, no. 4, pp. 60-69, 2024. https://doi.org/10.20463/pan.2024.0033

[2] M. I. Younas, M. J. Iqbal, A. Aziz, and A. H. Sodhro, "Toward QoS monitoring in IoT edge devices driven healthcare—a systematic literature review," *Sensors*, vol. 23, no. 21, p. 8885, 2023. https://doi.org/10.3390/s23218885

[3] M. S. Yaraziz, N. Sohrabi Safa, and M. A. Azad, "Edge computing in IoT for smart healthcare," *Journal of Ambient Intelligence and Smart Environments*, vol. 16, no. 4, pp. 409-438, 2024. https://doi.org/10.3233/AIS-230009

[4] U. Islam, M. N. Alatawi, A. Alqazzaz, S. Alamro, B. Shah, and F. Moreira, "A hybrid fog-edge computing architecture for real-time health monitoring in IoMT systems with optimized latency and threat resilience," *Scientific Reports*, vol. 15, no. 1, p. 25655, 2025. https://doi.org/10.1038/s41598-025-09696-3

[5] M. D. Tegegne, S. R. Niakan Kalhori, P. Haas, V. M. Sobotta, J. Warnecke, and T. M. Deserno, "Wearable devices for remote monitoring of chronic diseases: systematic review," *JMIR mHealth and uHealth*, vol. 14, p. e74071, 2026. https://doi.org/10.2196/74071

[6] Y. Liu and B. Wang, "Advanced applications in chronic disease monitoring using IoT mobile sensing device data, machine learning algorithms and frame theory: A systematic review," *Frontiers in Public Health*, vol. 13, p. 1510456, 2025. https://doi.org/10.3389/fpubh.2025.1510456

[7] C. V. F. Pereira, E. M. de Oliveira, and A. D. de Souza, "Machine learning applied to edge computing and wearable devices for healthcare: Systematic mapping of the literature," *Sensors*, vol. 24, no. 19, p. 6322, 2024. https://doi.org/10.3390/s24196322

[8] A. K. Mishra, R. Sharma, J. Singh, P. Singh, M. Diwakar, and M. Tiwari, "Real-time health monitoring of patients in emergencies through machine learning and iot integrated with edge computing," in *2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2024: IEEE, pp. 285-290.

[9] M. M. Khan and M. Alkhathami, "Anomaly detection in IoT-based healthcare: machine learning for enhanced security," *Scientific Reports*, vol. 14, no. 1, p. 5872, 2024. https://doi.org/10.1038/s41598-024-56126-x

[10] A. Hatamian, M. B. Tavakoli, and M. Moradkhani, "Improving the security and confidentiality in the internet of medical things based on edge computing using clustering," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, p. 6509982, 2021. https://doi.org/10.1155/2021/6509982

[11] M. Prabha, S. Nandhini, M. Dayanidhy, and R. Pradeep, "Edge-AI integrated secure wireless IoT architecture for real time healthcare monitoring and federated anomaly detection," *Scientific Reports*, vol. 16, p. 574, 2026. https://doi.org/10.1038/s41598-025-30150-x

[12] M. J. Page *et al.*, "PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews," *bmj*, vol. 372, p. n160, 2021. https://doi.org/10.1136/bmj.n160

[13] M. Al-Rawashdeh, P. Keikhosrokiani, B. Belaton, M. Alawida, and A. Zwiri, "IoT adoption and application for smart healthcare: A systematic review," *Sensors*, vol. 22, no. 14, p. 5377, 2022. https://doi.org/10.3390/s22145377

[14] R. Alnajim and A. Alkhalifah, "Exploring factors affecting the adoption of IoT in healthcare: A systematic literature review," *Healthcare*, vol. 13, no. 23, p. 3157, 2025. https://doi.org/10.3390/healthcare13233157

[15] V. K. Rathi *et al.*, "An edge AI-enabled IoT healthcare monitoring system for smart cities," *Computers & Electrical Engineering*, vol. 96, p. 107524, 2021. https://doi.org/10.1016/j.compeleceng.2021.107524

[16] J.-Y. Wu, Y. Wang, C. T. S. Ching, H.-M. D. Wang, and L.-D. Liao, "IoT-based wearable health monitoring device and its validation for potential critical and emergency applications," *Frontiers in Public Health*, vol. 11, p. 1188304, 2023. https://doi.org/10.3389/fpubh.2023.1188304

[17] M.-I. Joo, D.-Y. Kang, M.-S. Kang, and H.-C. Kim, "A patient management system using an edge computing-based IoT pulse oximeter," *Applied Sciences*, vol. 14, no. 1, p. 414, 2024. https://doi.org/10.3390/app14010414

[18] F. Famá, J. N. Faria, and D. Portugal, "An IoT-based interoperable architecture for wireless biomonitoring of patients with sensor patches," *Internet of Things*, vol. 19, p. 100547, 2022. https://doi.org/10.1016/j.iot.2022.100547

[19] S. Shukla, S. Thakur, S. Hussain, J. G. Breslin, and S. M. Jameel, "Identification and authentication in healthcare internet-of-things using integrated fog computing based blockchain model," *Internet of Things*, vol. 15, p. 100422, 2021. https://doi.org/10.1016/j.iot.2021.100422

[20] A. Ali *et al.*, "An innovative iot and edge intelligence framework for monitoring elderly people using anomaly detection on data from non-wearable sensors," *Sensors*, vol. 25, no. 6, p. 1735, 2025. https://doi.org/10.3390/s25061735

[21] H. Alasmary, "ScalableDigitalHealth (SDH): An IoT-based scalable framework for remote patient monitoring," *Sensors*, vol. 24, no. 4, p. 1346, 2024. https://doi.org/10.3390/s24041346

[22] M. M. Farag, "A tiny matched filter-based CNN for inter-patient ECG classification and arrhythmia detection at the edge," *Sensors*, vol. 23, no. 3, p. 1365, 2023. https://doi.org/10.3390/s23031365

[23] Y.-A. Daraghmi, E. Y. Daraghmi, R. Daraghma, H. Fouchal, and M. Ayaida, "Edge–fog–cloud computing hierarchy for improving performance and security of NB-IoT-based health monitoring systems," *Sensors*, vol. 22, no. 22, p. 8646, 2022. https://doi.org/10.3390/s22228646

[24] P. B. Suresha, C. Hegde, Z. Jiang, and G. D. Clifford, "An edge computing and ambient data capture system for clinical and home environments," *Sensors*, vol. 22, no. 7, p. 2511, 2022. https://doi.org/10.3390/s22072511

[25] M. M. Jaber *et al.*, "Remotely monitoring COVID-19 patient health condition using metaheuristics convolute networks from IoT-based wearable device health data," *Sensors*, vol. 22, no. 3, p. 1205, 2022. https://doi.org/10.3390/s22031205

[26] T. Baig, N. R. Chaudhry, R. Choudhary, P. Yadav, Y. A. Shaik, and A. Rashid, "An edge–fog–cloud IoT framework for real-time cardiac monitoring and rapid clinical alerts in hospital wards," *Future Internet*, vol. 18, no. 3, p. 130, 2026. https://doi.org/10.3390/fi18030130