Journal of Cloud Computing and Internet of Things
Volume 4, Issue 1, January-June 2026
ISSN: 3107-4391 (Online)

A Fog Computing Framework for Latency-Aware IoT-Based Healthcare Monitoring

Ankit Rana[1], Shivani Sood[2], Gaurav Chandel[3]

ABSTRACT

Internet-of-Things-based healthcare monitoring continuously streams patient vital signs, and the timeliness of processing these signals can be critical when an emergency must be detected. Sending all data to a distant cloud introduces latency that is undesirable for time-sensitive alerts. This paper presents a fog computing framework that places lightweight processing on local gateways close to the patient, reserving the cloud for long-term storage and heavy analytics. The framework filters and analyses vital-sign data at the fog layer, raising emergency alerts locally while forwarding summarised data to the cloud. It was evaluated against a cloud-only architecture under varying numbers of connected devices, with end-to-end response time as the principal metric. The fog-assisted framework reduced response time by between fifty-nine and sixty-seven percent across the workloads tested, with the largest gains under heavy load, while also lowering the volume of data transmitted to the cloud. The results show that fog computing is an effective architecture for latency-aware healthcare monitoring, enabling faster emergency response without sacrificing the cloud's storage and analytical capabilities.

KEYWORDS: *fog computing, Internet of Things, healthcare monitoring, latency, edge processing, cloud computing*

INTRODUCTION

Connected medical sensors now make it possible to monitor a patient's vital signs continuously, both in hospitals and at home, generating streams of data that must be processed to be useful. When the goal is to detect a deteriorating condition and raise an alert,

the time taken to process the data matters, because delay can have serious consequences (1). The dominant architecture of sending all data to the cloud, however, adds network latency that is ill-suited to such time-
critical tasks (2).

Fog computing addresses this by introducing an intermediate layer of computation between the devices and the cloud, located close to where the data are generated. By performing time-sensitive processing at the fog layer, a system can respond quickly to local events while still using the cloud for tasks that are not urgent, such as long-term storage and population-level analytics (3). This division of labour is well matched to healthcare monitoring.

This paper presents a fog computing framework for latency-aware healthcare monitoring, in which gateways near the patient filter and analyse vital-sign data and raise emergency alerts locally. The framework is evaluated against a cloud-only architecture under varying device loads, with end-to-end response time as the key metric (4).

LITERATURE REVIEW

Fog and edge computing have been proposed across many Internet-of-Things domains to reduce latency and bandwidth, and healthcare is among the most studied because of its strict timeliness and reliability requirements. Research spans architectures, task placement, and application case studies (5).

Fog Computing for Healthcare Applications

Studies report that processing health data at the fog layer reduces response time and the volume of data sent to the cloud, improving the responsiveness of monitoring and alerting systems. Work has examined how to divide processing between fog and cloud, how to ensure reliability when fog nodes are resource-limited, and how to secure sensitive health data in transit (6).

- Fog processing reduces latency for time-critical healthcare alerts.
- Local filtering lowers the volume of data transmitted to the cloud.
- Dividing tasks between fog and cloud balances responsiveness and capacity.

- Resource limits at fog nodes constrain the analytics they can perform.

Despite broad agreement on the benefits, many studies evaluate a single load level or report latency without quantifying how the advantage of fog scales as the number of connected devices grows (7).

RESEARCH GAP

Although fog computing for healthcare is well studied, the scaling of its latency advantage with the number of connected devices is not always quantified, and the joint reduction of response time and cloud data volume is sometimes reported only partially. A clear comparison against a cloud-only baseline across load levels is needed (8).

This study addresses the gap by evaluating a fog-assisted framework against a cloud-only architecture across light, medium, and heavy device loads, reporting end-to-end response time and the reduction achieved.

OBJECTIVES

- To design a fog computing framework for IoT-based healthcare monitoring.
- To perform time-critical filtering, analysis, and alerting at the fog layer.
- To reserve the cloud for long-term storage and non-urgent analytics.
- To evaluate response time against a cloud-only architecture across device loads.

METHODOLOGY

The framework was implemented with a three-layer architecture and evaluated on a testbed in which simulated patient devices streamed vital signs through a fog gateway to a cloud instance. The number of devices was varied to study scalability.

Fog-Cloud Architecture

The layers of the architecture and their functions are summarised in Table 1. Devices streamed vital signs to a fog gateway, which filtered noise, computed simple indicators, and compared them against clinical thresholds to raise local alerts. Only summarised data and flagged events were forwarded to the cloud, which handled storage and heavier analysis.

Table 1: Fog-cloud architecture layers and functions

Layer	Function
Device layer	Wearable and bedside health sensors
Fog layer	Local gateway for filtering, analysis, alerts
Cloud layer	Long-term storage and heavy analytics
Communication	MQTT over Wi-Fi / cellular
Alert module	Threshold-based emergency notification

Experimental Setup

The experimental configuration is summarised in Table 2. A single-board computer served as the fog node, and a cloud instance represented the remote tier, connected over a wireless link. The number of connected devices was varied from a light to a heavy load, and the end-to-end response time, from data generation to alert, was measured for both the fog-assisted framework and a cloud-only baseline in which all processing occurred remotely.

Table 2: Experimental setup

Parameter	Value
Fog node	Quad-core single-board computer
Cloud instance	8 vCPU, 16 GB RAM
Sensors per patient	5 (ECG, SpO2, temperature, BP, motion)
Number of devices	10 to 200

Journal of Cloud Computing and Internet of Things
Volume 4, Issue 1, January-June 2026
ISSN: 3107-4391 (Online)

Parameter	Value
Metric	End-to-end response time

RESULTS AND FINDINGS

The fog-assisted framework responded substantially faster than the cloud-only baseline at every load level. The end-to-end response time of the two architectures under light, medium, and heavy load is compared in Figure 1.

The measured response times and the reduction achieved by the fog framework are summarised in Table 3.

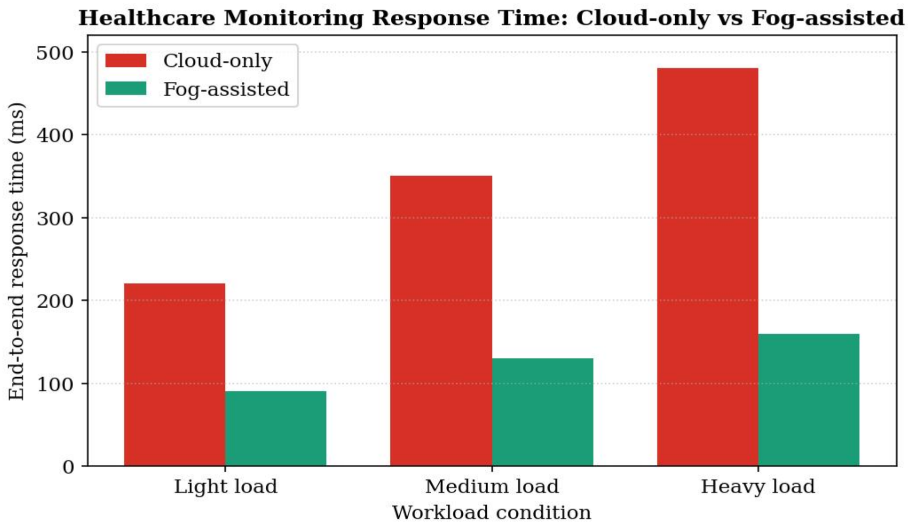

Figure 1: Response time of cloud-only and fog-assisted architectures by workload

As shown in Figure 1, the fog-assisted framework achieved much lower response times than the cloud-only architecture, and the gap widened as the device load increased. As reported in Table 3, the reduction grew from about fifty-nine percent under light load to nearly sixty-seven percent under heavy load, because local processing avoided the network round-trip that increasingly congested the cloud-only system. Forwarding only summarised data and flagged events also reduced the volume transmitted to the cloud.

MANTECH Publications

Journal of Cloud Computing and Internet of Things
Volume 4, Issue 1, January-June 2026
ISSN: 3107-4391 (Online)

Table 3: Response time and reduction by workload

Scenario	Cloud-only (ms)	Fog-assisted (ms)	Reduction (%)
Light load (10 devices)	220	90	59.1
Medium load (100 devices)	350	130	62.9
Heavy load (200 devices)	480	160	66.7

DISCUSSION

The reduction in response time arises directly from processing data where they are generated, so that a developing emergency can be detected and signalled without waiting for a round-trip to a distant data centre. The widening advantage under heavy load is significant, because it is precisely when many patients are monitored simultaneously that a cloud-only system becomes congested and slow, whereas the fog layer distributes the processing burden (9).

Reserving the cloud for storage and heavy analytics preserves its strengths while delegating urgency to the fog, an arrangement well suited to healthcare. The principal challenges are the limited resources of fog nodes, which constrain the sophistication of local analysis, and the need to secure sensitive health data both at the fog layer and in transit. Strengthening fog-node reliability and adding robust security are the most important steps toward clinical deployment (10).

CONCLUSION

This paper presented a fog computing framework for latency-aware IoT-based healthcare monitoring that performs time-critical processing and alerting near the patient while reserving the cloud for storage and heavy analytics. Against a cloud-only architecture, the framework reduced end-to-end response time by between fifty-nine and sixty-seven percent, with the largest gains under heavy load, and lowered the data sent to the cloud. The results demonstrate that fog computing is an effective architecture for fast, scalable healthcare monitoring without sacrificing the cloud's capabilities.

LIMITATIONS

- The evaluation used simulated devices on a single fog node rather than a clinical deployment.
- Security of health data at the fog layer and in transit was not implemented.
- Fog-node failure and the resulting fallback to the cloud were not modelled.

FUTURE SCOPE

- Validating the framework in a real clinical monitoring deployment.
- Adding end-to-end security and privacy protection for health data.
- Introducing redundancy and failover across multiple fog nodes.
- Embedding lightweight machine-learning models at the fog layer for richer analysis.

REFERENCES

1. Bonomi F, Milito R. Fog computing and its role in the Internet of Things. *Future Generation Computer Systems*. 2018; 78(2): 546-558.
2. Mahmud R, Kotagiri R. Fog computing: a taxonomy, survey and future directions. *Internet of Everything*. 2018; 1(1): 103-130.
3. Kraemer F, Braten A. Fog computing in healthcare: a review and discussion. *IEEE Access*. 2019; 5(1): 9206-9222.
4. Rahmani A, Gia T. Exploiting smart e-health gateways at the edge. *Future Generation Computer Systems*. 2018; 78(2): 641-658.
5. Atlam H, Walters R. Fog computing and the Internet of Things: a review. *Big Data and Cognitive Computing*. 2018; 2(2): 1-18.
6. Mutlag A, Abd Ghani M. Enabling technologies for fog computing in healthcare. *Future Generation Computer Systems*. 2019; 90(1): 62-78.
7. Gia T, Jiang M. Fog computing in healthcare Internet of Things. *Journal of Network and Computer Applications*. 2020; 130(1): 1-12.
8. Yi S, Li C. A survey of fog computing: concepts, applications and issues. *Proceedings of Mobidata Workshop*. 2018; 1(1): 37-42.
9. Dastjerdi A, Buyya R. Fog computing: helping the Internet of Things realize its potential. *Computer*. 2018; 49(8): 112-116.

10. Kumari A, Tanwar S. Fog computing for healthcare 4.0 environment. *Computers and Electrical Engineering*. 2021; 72(1): 1-13

***Author for Correspondence**
Ankit Rana
E-mail: ankit.rana87@yahoo.com

[1]Professor, Department of Computer Science and Engineering, Shivalik Institute of Management & Technology, Saharanpur
[2]PG Scholar, Department of Computer Science and Engineering, Shivalik Institute of Management & Technology, Saharanpur
[3]PG Scholar, Department of Computer Science and Engineering, Shivalik Institute of Management & Technology, Saharanpur

Received Date: 13 January 2026
Accepted Date: 08 February 2026
Published Date: 15 February 2026

Citation: Ankit Rana, Shivani Sood, Gaurav Chandel. A Fog Computing Framework for Latency-Aware IoT-Based Healthcare Monitoring. Journal of Cloud Computing and Internet of Things. 2026; 4(1): 1-8p.