Received: 8 May 2026 | Revised: 12 July 2026 | Accepted: 24 July 2026 | Published online: 18 August 2026

Medinformatics
2026, Vol. 00(00) 1–12
DOI: 10.47852/bonviewMEDIN620210326

RESEARCH ARTICLE

A Fog-Based AI Doctor Copilot for Real-Time Clinical Risk Triage in Telehealth IoT Systems

Nathan Guo[1], Bryan Guo[2], Yunyong Guo[1,*] and Alex Kuo[1]

[1] *Department of Health Information Science, University of Victoria, Canada*

[2] *Department of General Science, University of British Columbia, Canada*

Abstract: The increasing adoption of telehealth and remote patient monitoring has created new challenges for clinical decision support, particularly when timely responses are required for large numbers of geographically distributed patients. Rather than replacing clinical judgment, current artificial intelligence (AI) applications are increasingly designed to support healthcare professionals by identifying patients at elevated risk, prioritizing alerts, and assisting clinical workflows. Building on this direction, this study presents a fog-based AI Doctor Copilot for real-time clinical risk triage in Internet of Things-enabled telehealth environments. The proposed framework integrates continuous physiological monitoring with adaptive machine learning models deployed on fog computing nodes, enabling patient risk to be dynamically classified as low, medium, or high while preserving clinician oversight throughout the decision-making process. System performance was assessed through simulation using synthetic physiological data, including heart rate, blood pressure, and blood glucose measurements, and benchmarked against a conventional cloud-based clinical decision support system. Compared with cloud-only architecture, the proposed approach reduced alert latency by as much as 52%, decreased false alerts by approximately 31%, and lowered overall energy consumption by about 22%, while maintaining comparable risk stratification accuracy. These findings suggest that distributing AI inference to the fog layer can improve the responsiveness and efficiency of telehealth decision support without compromising clinical oversight, making the approach well suited to time-sensitive remote care environments. Because the evaluation was conducted using synthetic patient data in a simulation environment, the results should be interpreted as evidence of architectural feasibility rather than clinical effectiveness.

Keywords: AI doctor, clinical risk triage, energy-efficient healthcare

1. Introduction

Telehealth and remote patient monitoring (RPM) have evolved from optional innovations into key components of modern healthcare delivery. This shift is largely driven by aging populations, increasing rates of chronic disease, and persistent shortages of clinical resources. As these systems become more widely adopted, they continuously generate high-frequency physiological data—including heart rate, blood pressure, oxygen saturation, and glucose levels—which offers significant potential for early detection of patient deterioration but also introduces new operational challenges [1–3].

Incorporation of artificial intelligence (AI) into healthcare practices has been growing as a means of supporting data analytics, detecting clinical deterioration, and assisting in decision-making based on evidence. Even though there has been significant progress in model performance, their practical implementation has been difficult to achieve. It appears that despite the high predictive accuracy of the algorithms, other concerns such as lack of transparency, biases, and reliability as well as patients safety have remained obstacles to the adoption of these tools [4–6]. The focus has shifted from developing AI systems that can function autonomously to using them to augment, not to replace, clinical expertise. In this way, the role of an AI doctor should be considered as a copilot to a clinician who would assess risks and assist in decision-making but remain responsible for diagnosis and treatment decisions. This is consistent with the recent recommendations provided by international organizations regarding the implementation of AI in healthcare settings [7–9].

1.1. Clinical workflow integration

The success of AI adoption in the healthcare setting does not only depend on the accuracy of predictions; it also depends on how effectively innovations integrate into existing clinical workflows. Decision-supporting tools that involve the use of separate applications by clinicians or the production of a large volume of alerts of little value may lead to increased cognitive load and alert fatigue and reduce the usability of such innovations in

***Corresponding author:** Yunyong Guo, Department of Health Information Science, University of Victoria, Canada. Email: yunyong@uvic.ca

regular practice. To overcome these barriers, the suggested AI Doctor Copilot will complement the existing workflow, rather than add another one. Risk-stratified alerts created at the fog level can easily be added to the existing clinical dashboard or electronic health record (EHR) systems to allow clinicians to evaluate the results produced by AI within their regular workflow.

High-risk alerts are prioritized for immediate clinician review, while medium- and low-risk cases are queued for routine assessment or trend monitoring. Importantly, the system does not autonomously modify clinical orders, diagnoses, or treatment plans. Instead, it provides interpretable recommendations and longitudinal trend summaries that support clinician judgment. This design minimizes workflow disruption, reduces alarm fatigue, and aligns with prevailing governance policies that emphasize human-in-the-loop decision-making and limit AI autonomy in clinical settings [10–12].

1.2. Motivation for fog-based AI Doctor Copilot

Fog computing provides an alternative by shifting part of the intelligence closer to where the data are generated. This allows more of the processing to happen locally, which helps reduce communication overhead and improves overall system responsiveness. In the proposed design, the AI Doctor Copilot is deployed at the fog layer, where it can perform anomaly detection, track short-term trends, and estimate patient risk with minimal delay. Meanwhile, the cloud remains available for tasks that are less time-sensitive, such as long-term analysis and model updates [13–15].

1.3. Research objective and contribution

Compared with earlier fog-based healthcare systems, which tend to focus mainly on distributing computation or enabling general edge analytics, the proposed framework takes a slightly different direction. Here, AI is treated more explicitly as a clinician-oriented copilot. The system is designed not only to process data but also to prioritize risk in a way that is interpretable, aligned with clinical workflows, and consistent with current governance expectations. The main contribution of this work is therefore not a more complex predictive model, but a system design that is practical, deployable, and closer to real-world clinical requirements.

This research is focused on understanding whether a new AI Doctor fog copilot using the idea of a fog computing environment could be more effective for risk triage in the early stages compared to existing clinical decision support systems (CDSSs) operating exclusively within cloud computing environments. Unlike binary warnings used previously, this new solution utilizes risk and creates explanations that can be readily incorporated into the workflow. The proposed solution is tested in terms of several trade-offs, including alert time, false-alert rate, latency, and energy consumption.

2. Related Work

2.1. AI-assisted clinical decision support and risk stratification

CDSSs have traditionally been employed to support the interpretation of patient information and the detection of possible risks. Recent innovations in machine learning and AI have dramatically enhanced the scope of CDSS applications, especially in early warning systems and risk stratification. Several papers have proven that AI models can detect patient deterioration at an earlier stage than conventional rule-based systems, particularly in areas like sepsis detection, cardiac monitoring, and chronic disease management [16–18].

Still, despite having high predictive capabilities, these systems are currently employed for advisory purposes rather than as fully automated decision-makers due to the ongoing doubts about explainability, accountability, and trustworthiness. According to recent research [16, 17], the key factor in the successful implementation of such systems is not just the accuracy of a model but the compatibility with workflow as well.

2.2. Workflow integration and alert fatigue in AI-enabled systems

A critical challenge in deploying AI-enabled CDSS is alert fatigue, which arises when clinicians are exposed to excessive or poorly prioritized alerts. Studies have shown that high false-alert rates can reduce clinician responsiveness and undermine patient safety [19, 20]. To address this issue, recent research has focused on alert prioritization, tiered risk stratification, and explainable alerts that provide contextual information alongside risk scores [21, 22].

2.3. Fog and edge computing in healthcare IoT systems

Cloud-centric architectures have traditionally dominated healthcare AI deployments; however, their limitations in latency-sensitive and bandwidth-constrained environments have become increasingly apparent. To overcome these challenges, fog and edge computing paradigms have been introduced to bring computation closer to data sources [23–25].

2.4. Limitations of existing architectures and research gap

Despite substantial progress, existing AI-enabled healthcare systems exhibit several limitations. Many studies emphasize predictive accuracy without adequately considering clinical usability, explainability, or regulatory alignment [26, 27]. Moreover, fog-based healthcare architectures often treat intelligence as a generic analytics function, rather than explicitly framing AI as an assistive clinical copilot with clearly defined responsibility boundaries [28–30].

Recent work has begun to recognize the importance of human-in-the-loop designs and explainable AI in healthcare settings [31–33], yet few architectures integrate these principles with fog-based real-time risk triage. This gap highlights the need for system designs that simultaneously address latency constraints, energy efficiency, workflow integration, and ethical governance. The proposed fog-based AI Doctor Copilot architecture seeks to fill this gap by embedding assistive, explainable risk stratification at the fog layer while reserving the cloud for long-term learning and optimization. Collectively, these limitations motivate the proposed fog-based AI Doctor Copilot architecture. Rather than focusing solely on predictive performance, the proposed framework integrates workflow-aware risk stratification, fog-based real-time processing, explainable AI support, and clinician oversight into a unified system intended for practical telehealth deployment.

3. System Architecture

The architecture of the proposed fog-based AI Doctor Copilot is provided in Figure 1. The proposed architecture is built around four layers, namely, the data acquisition layer, fog computing layer, cloud analytics layer, and the clinical interface. As seen above, each of the layers serves its own purpose in the processing workflow, thus making it possible to process time-sensitive tasks locally, closer to the patient, whereas intensive computations are done in the cloud environment. It also helps deploy modules independently and facilitate maintenance of the system. Data about patients are obtained from various sources, namely, wearable sensors, home-based devices, bedside monitoring equipment, and EHR systems. Both continuous physiological data, such as heart rate, blood pressure, oxygen saturation, and blood glucose levels, and longitudinal clinical data to establish context for risk assessment are being gathered. Basic preprocessing, such as filtering and normalization of signals, is done to the incoming data before it gets to the fog computing layer. Clinical inference is not performed at this stage to save computational resources.

The fog layer is the main runtime environment of the AI Doctor Copilot. Being located near the devices used for monitoring, the physiological parameters of the patients can be analyzed with minimum delay in communications. Measurements coming from the sensors are being analyzed using machine learning models for the purposes of anomaly detection, trend analysis, and individual risk assessment. Unlike a binary warning system, the algorithm used provides for a triage system in which each patient is being assessed as either low, medium, or high risk. Besides the risk notification, the variable used to determine it is included in the notification and is provided to the clinicians to make their decision-making easier. Communications between fog nodes are carried out through the gateway, which is responsible for aggregating the results and delivering information to the relevant destination points. The gateway itself does not run machine learning models but only decides whether the patient data are going to be delivered to the cloud for further analysis or directly to the clinicians. High-risk events are sent instantly, while low-risk ones are sent to the cloud.

The long-term storage of information and computations that are resource-intensive take place at the cloud layer. The historical medical records of patients are stored for analysis, testing models, and population-based studies. Also, re-training and validation of models take place at the cloud level using the accumulated data from the clinical sector before re-distribution of the new models to the fog nodes. This is because these operations do not require any latency tolerance; hence, making sure that the process is done in the cloud prevents an increase in delay in the real-time monitoring system. The users in the clinical environment interact with the system through existing dashboards and EHR platforms as opposed to an application. Classification of risks, alert notifications, and their explanations take place together with other patient data through the already available interfaces.

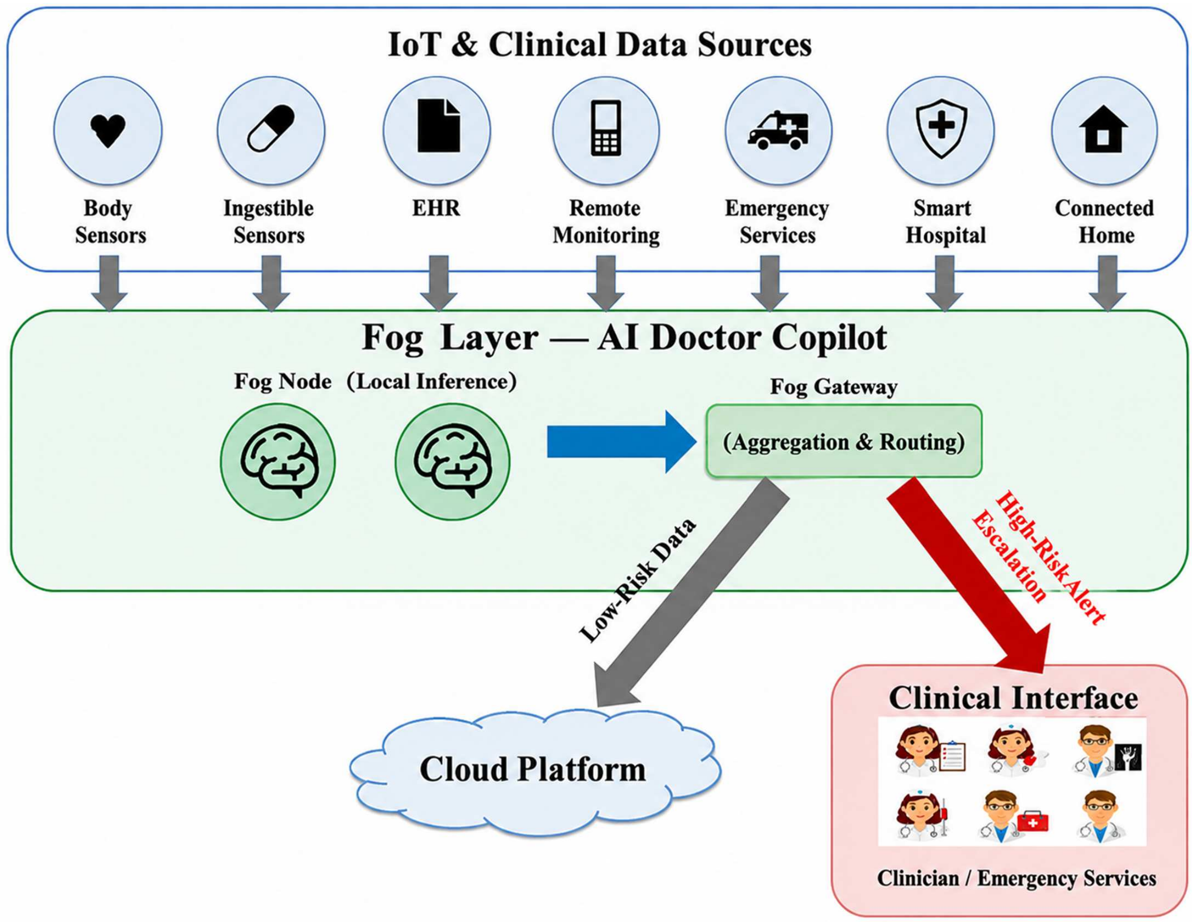

Figure 1. Fog-based AI Doctor Copilot system architecture

4. Methodology

The current research employs an experimental setup based on simulations to assess the effectiveness of a fog-based AI Doctor Copilot in early risk detection in telemonitoring clinical practice. Notably, while accuracy is one of the factors to consider when evaluating the system's performance, another important aspect is its behavior in real operating conditions. Thus, when designing experiments, special attention has been paid to making the comparison of different architectures consistent and reliable.

To have comparable results, it was decided to use the same data input, the same detection algorithm, and the same criteria for assessing the performance of the two types of system under study—the fog-based one and a CDSS running entirely in the cloud. The only thing that distinguishes the two systems from each other is the architecture—whether a model should run in the fog or in the cloud. It makes the results easier to interpret because any differences that have been observed should be related to architectural design rather than to different algorithms employed in each of the cases. The same setup has been used for all experiments.

4.1. Data generation and simulation environment

Continuous datasets with measurements obtained using telehealth applications are rare and are usually not available due to confidentiality restrictions. Thus, for evaluation purposes, a simulated dataset has been created in order to estimate the performance of the proposed architecture. For each individual simulation subject, physiological measurements were obtained in one-minute intervals during a 24-h monitoring period. The simulated parameters were the following: heart rate, systolic and diastolic blood pressure, and blood glucose as the most common physiological measures used in RPM.

The initial physiological parameter values were sampled from representative adult monitoring ranges: heart rate (60–100 bpm), systolic blood pressure (90–140 mmHg), diastolic blood pressure (60–90 mmHg), and blood glucose (70–180 mg/dL). The above ranges corresponded to the typical physiological measures reported in telehealth/clinical monitoring studies rather than some patient population. The interindividual variability was modeled with the help of Gaussian noise and circadian fluctuations.

To represent changes in patient condition, deterioration events were simulated by gradually shifting one or more physiological variables beyond predefined clinical thresholds. Additional disturbances, including random measurement noise, intermittent signal loss, and motion-related artifacts, were incorporated to approximate conditions commonly encountered during remote monitoring. Patients were subsequently assigned to predefined risk categories according to the simulated deterioration trajectories, providing consistent reference labels for evaluating the performance of the proposed AI Doctor Copilot across all experimental scenarios.

The generated dataset was made to mirror real-world conditions as much as possible. Variation in baseline data was created, as well as the addition of noise and temporary interruptions. This will help to make sure that the system is not evaluated on "too clean" data and tested under more realistic conditions.

To stress the system even further, additional disruptions were created, including noise related to movement and temporary interruptions of the signal. This will simulate the common problems that occur during telemedicine consultations. Although the data remain artificial, the entire simulation will aim to mirror practical implementation conditions.

4.2. Anomaly detection and feature extraction

Anomaly detection consists of two types of analyses. In the first instance, the data points will be compared with the baseline to detect the anomalies. Second, the trends will be identified using moving average and slope calculations. This allows detecting changes that happen slowly and would have been missed otherwise. It is important to use both types of analysis for a more effective model, as it will allow detecting the spikes very quickly and yet will not be too sensitive to the noise. The change is always slow and important.

4.3. Risk scoring and tier classification

The risk score for the AI Doctor Copilot is obtained by integrating anomaly signals with short-term trends. Rather than having an absolute set of rules, weighted scores are computed where various features have varying weightages. Such weightages have been derived through various calibration runs in order to balance between responsiveness and stability. As a result, transient changes do not have much influence on the score, but persistent changes matter more. According to this risk score, patients are classified into three groups—low risk, medium risk, and high risk. The thresholds are arrived at through iterative processes, where the objective has been sensitivity and reduction in false alarms.

The composite risk score was calculated as:

$$\text{Risk Score} = 0.4(A) + 0.3(T) + 0.3(V) \tag{1}$$

where:

A = anomaly severity score,
T = short-term trend score,
V = physiological variability score.

The weights were determined through calibration experiments designed to balance sensitivity and false-alert reduction. Greater weight was assigned to anomaly severity because sudden physiological deterioration often requires immediate clinical attention. Trend and variability measures were assigned slightly lower weights to improve stability and reduce transient false alerts. The weighting strategy was selected empirically through repeated simulation calibration rather than optimization on clinical datasets. Based on the resulting score, patients were classified as:

1) Low Risk: Risk Score < 0.30
2) Medium Risk: $0.30 \le$ Risk Score < 0.70
3) High Risk: Risk Score ≥ 0.70

Thresholds were selected iteratively using repeated simulation runs to maximize high-risk recall while minimizing false-alert generation.

4.4. Fog-level processing and alert prioritization

All real-time operations like anomaly detection, feature extraction, and scoring are performed by the fog layer. The fog layer decreases latency and reduces dependence on constant communication with the cloud. In the fog layer, all alarms are sorted according to priority. Urgent alarms are dealt with immediately, whereas less urgent alarms are processed slowly. This corresponds to the functioning of telehealth systems since there is restricted bandwidth and not every alarm needs an immediate response. It should also be noted that the AI Doctor Copilot does not make any clinical decisions. Instead, it produces prioritized and understandable alerts for the clinician to process.

4.5. Cloud-based model recalibration

Even if the actual decisions take place locally, there is an important role played by the clouds. Clouds are used for storing historical data and for recalibrating the model based on long-term trends. It happens asynchronously, meaning that there is no influence on the real-time process. After the recalibration process, the updated parameters are delivered to fog nodes.

4.6. Experimental protocol and evaluation methodology

Performance testing was done in a simulated environment with various sizes of patient cohorts being used for scalability analysis. In each case, the experiment was performed 30 times with different random seeds being used to minimize the impact of randomness on the results. The following indicators were taken into consideration as primary indicators of performance: alert latency time, energy usage, classification accuracy, false-alert ratio, and recall rate for risky patients. Performance testing was conducted for both the fog-based approach and cloud-only baseline under identical conditions to eliminate any bias. By keeping all data and algorithms the same but varying computation placement, the effect of the architecture can be measured.

5. Experimental Setup

The proposed fog-based AI Doctor Copilot was assessed using a simulation setting that aims to provide a comparison of the performance of the proposed fog-based AI Doctor Copilot and a cloud-based CDSS within controlled operating environments. The comparison involved not only the predictive performance but also the operating characteristics related to telemedicine implementation such as alert delay, energy consumption, false-alert production, and risk classification. Results of each experiment are given as the mean ± standard deviation (SD) and 95% confidence interval (CI).

To consider different levels of the load on the system, experiments were performed on 10, 50, and 100 patients. The fog nodes have been set up to mimic resource-limited edge devices that can be found in home monitoring gateways or hospital edge servers, while the cloud setting mimics the central computing architecture with much higher computational power. Network latency, limited bandwidth, and intermittent network congestion have been taken into account during the simulation to mimic real-life conditions for RPM.

Two processing architectures were evaluated using identical input data and machine learning models. In the proposed architecture, data processing and risk inference were performed at the fog layer, allowing alerts to be generated close to the point of data acquisition. The comparison architecture executed the same analytical workflow entirely within the cloud. Because both systems used identical datasets and prediction models, observed performance differences reflect the effect of computation placement rather than differences in the underlying AI algorithms.

The system performance was measured in terms of four metrics. Alert latency is defined as the amount of time that elapses between data collection and the generation of a clinical risk alert. The energy expenditure was estimated through a system-level model that considered the cost of sensing, wireless communication, and computational processes. Energy used for communication was assumed to be directly proportional to the amount of data that was transmitted, but processing energy was dependent on the computational load at the fog layer or the cloud layer. Even though this model cannot estimate the hardware-specific power consumption, it enables a consistent comparison of the alternative architectures under identical conditions. Classification accuracy is the ratio between the true risk classes and the simulated events, while the false alarm rate is the percentage of risk alerts generated for non-risk patients.

Each experiment configuration was run 30 times using varying random seeds in synthetic data generation. This helped reduce the effects of any stochastic variability to stabilize the results. This experimental methodology enables consistent comparison of the strengths of fog- and cloud-based CDSSs under telehealth operating conditions.

6. Results

An independent samples t-test was performed to analyze the difference between fog computing architecture and cloud-only computing baseline architecture in terms of latency, energy efficiency, and false alarm parameters. The observed differences were statistically significant ($p < 0.01$) in all patient load cases. Large effect sizes in terms of latency decrease and prioritization ability were confirmed using Cohen's d parameter. In this context, all the above-mentioned performance metrics were achieved in controlled simulations through the use of synthetic datasets. Therefore, the presented findings illustrate the architectural feasibility of the proposed AI Doctor Copilot based on fog computing but not its clinical efficacy. Clinical validation is needed for drawing any conclusion about clinical efficacy.

The paragraphs below contain the experimental results that have been gained in relation to fog-based AI Doctor Copilot and cloud-based CDSS solution. All numerical results have been given in the form of tables, whereas graphs have been used to show the changes in the results based on different patient loads. Mean, SD, and CI have been given in order to conduct statistical analysis of the results.

To test the scalability and stability of the system, the experiment was conducted for five different patient loads starting from 10 and ending up with 100 patients. Each experiment was conducted independently 30 times with different random seeds, and all the quantitative results are provided in terms of means and SDs and 95% CIs.

Table 1 summarizes results for the alert latencies of the suggested fog-based AI Doctor Copilot and cloud-based CDSS. The fog-based architecture was consistently faster than the cloud-based architecture when it came to generating alert signals in all the testing scenarios. The same trend has been depicted in Figure 2, where, even as latency rises with the number of patients, the fog-based architecture remains much faster.

Table 2 presents the energy consumption of both architectures across different patient loads. The results indicate that local processing at the fog layer substantially reduces overall system energy consumption compared with continuous cloud-based processing. As illustrated in Figure 3, energy consumption increases gradually as the monitoring workload grows, but the fog-based architecture consistently remains more energy efficient than the cloud-only baseline.

Table 3 summarizes the simulation-based clinical risk triage performance, including classification accuracy, false-alert rate, and high-risk recall. The proposed fog-based AI Doctor Copilot

Table 1. Alert latency results across patient loads

Patient load	Fog mean latency (ms)	Fog SD (ms)	Fog 95% CI (± ms)	Cloud mean latency (ms)	Cloud SD (ms)	Cloud 95% CI (± ms)
10	30.8	5.2	1.86	87.5	10.5	3.76
25	32.0	5.5	1.97	91.25	11.25	4.03
50	34.0	6.0	2.15	97.5	12.5	4.47
75	36.0	6.5	2.33	103.75	13.75	4.92
100	38.0	7.0	2.5	110.0	15.0	5.37

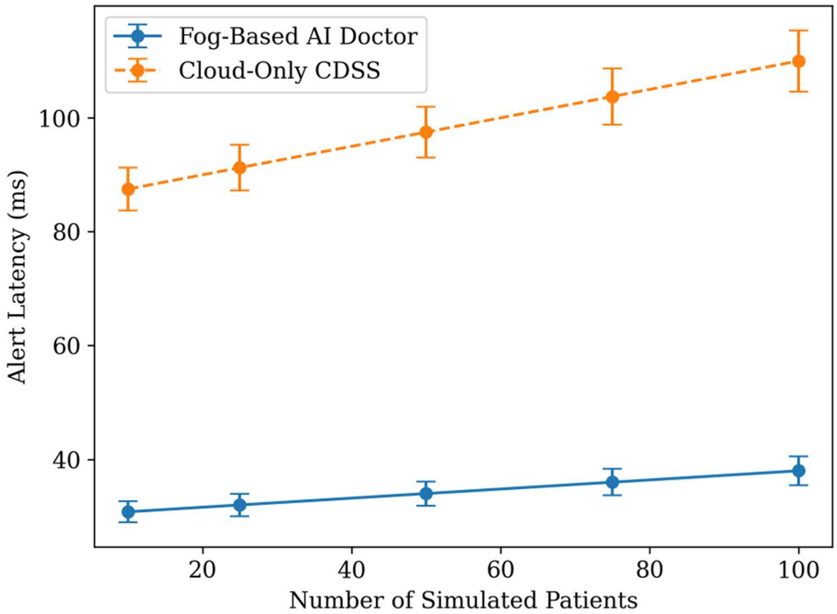

Figure 2. Latency trend visualization

Table 2. Energy consumption results across patient loads

Patient load	Fog mean energy (kWh)	Fog SD (kWh)	Fog 95% CI (± kWh)	Cloud mean energy (kWh)	Cloud SD (kWh)	Cloud 95% CI (± kWh)
10	116.0	11.0	3.94	171.0	19.5	6.98
25	125.0	12.5	4.47	187.5	21.75	7.78
50	140.0	15.0	5.37	215.0	25.5	9.13
75	155.0	17.5	6.26	242.5	29.25	10.47
100	170.0	20.0	7.16	270.0	33.0	11.81

achieves consistently higher classification accuracy and high-risk recall while simultaneously reducing false alerts across all patient loads. These overall performance trends are further visualized in Figure 4, demonstrating the clinical advantages of the proposed architecture as monitoring demand increases.

7. Discussion

First of all, the results demonstrate that the fog-based AI Doctor Copilot outperforms the cloud-only CDSS in practical terms. This distinction becomes obvious in terms of latency and energy consumption as well as the usefulness of the solution in the clinical environment. One of the key reasons for the improved performance is the local processing of the input data at the fog layer level, which leads to reduced network latency and data transfer overhead.

The choice of relatively simple anomaly detection and trend analysis algorithms was not made by chance. In this case, the idea was not to create the most sophisticated algorithm. Much more important was to make sure that the system would be able to respond quickly and work stably with limited computing resources. That is why the main goal of this experiment was to make the practical implementation and not to test model performance. As for the latency issue, the distinction between the two architectures is quite evident. Detection and scoring at the fog layer allow one to react to the changes in the patient's condition faster. It is especially important in telehealth applications where any delay may influence the reaction speed of the clinician.

Though the decrease in the processing time is given in milliseconds, the benefit will become increasingly evident when the number of observed patients increases. With lower latency of alerts, the system can work more efficiently, decreasing the number of queues that may be formed during the period of high load

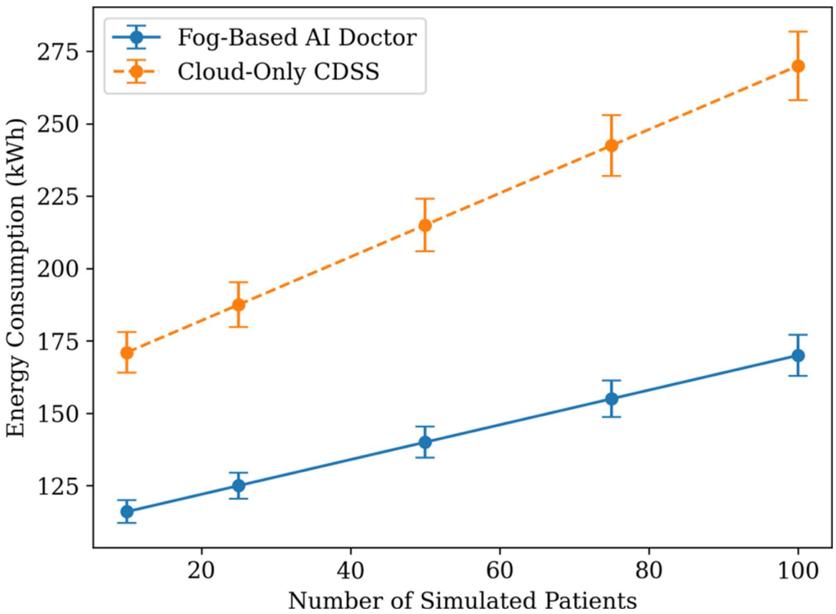

Figure 3. Energy trend visualization

Table 3. Simulation-based risk triage performance metrics across patient loads

Patient load	Fog accuracy (%)	Cloud accuracy (%)	Fog false-alert rate (%)	Cloud false-alert rate (%)	Fog high-risk recall (%)	Cloud high-risk recall (%)
10	96.4	90.2	8.2	12.4	94.85	88.2
25	96.25	89.75	8.5	13.0	94.62	87.75
50	96.0	89.0	9.0	14.0	94.25	87.0
75	95.75	88.25	9.5	15.0	93.88	86.25
100	95.5	87.5	10.0	16.0	93.5	85.5

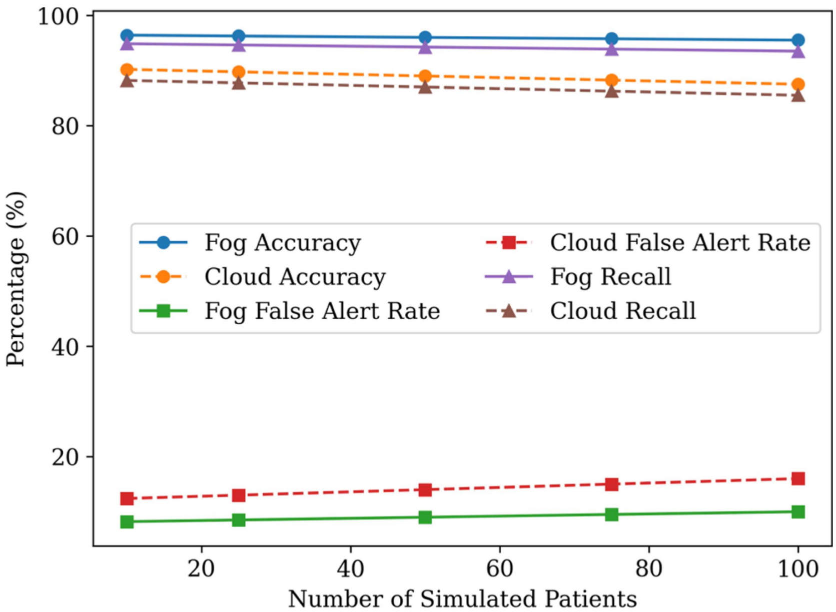

Figure 4. Clinical utility trend visualization

and increasing the efficiency of notifying the clinicians. The benefits mentioned above can be very useful for the telehealth scenario because delay may influence the priority of those patients who should be examined urgently. The same tendency could be noticed in the case of energy consumption. The traditional cloud-based architecture presupposes continuous transmission of the physiological data to the servers; thus, communication overheads are higher. In our case, when part of the workload is processed at the fog level, there is less amount of transmitted data, and thus, there is lower energy consumption.

The proposed architecture also affects the creation of clinical alerts. Instead of binary alerts, the system sorts patients into predefined risk levels, making it possible for clinicians to prioritize patients depending on their level of clinical urgency. A risk-based approach helps reduce the number of unnecessary alerts and can be used as a solution to alert fatigue problems that are faced by CDSSs.

Fog-based architecture is distributed, which makes it have other implementation issues. Management of multiple fog nodes, software versions, and distributed resources is much more difficult compared to using a centralized cloud architecture. It means that when implementing a fog-based CDSS, one should consider the trade-off between performance improvements and added implementation complexity.

8. Clinical Implications

There are several applications of the findings from the above study for the use of AI in telehealth and RPM. The AI Doctor Copilot does not aim to replace any clinical decision-making; rather, its application is meant to complement that of a clinical doctor. The most important part of this application is in how it helps a clinician to perceive risks of the patients and in detecting any early warning signs and organizing the alarms efficiently. In most cases of monitoring, knowing about the risk of the patient may be much more important than getting automated diagnostic results.

Another interesting outcome of this study is the ability to decrease the cognitive load of a clinician. Telehealth solutions provide huge amounts of data and many alerts that are usually not clinically significant. With the help of grouping the patients into low-, medium-, and high-risk categories, the system allows the doctor to focus on the cases that require immediate attention. However, low-risk changes can also be checked without interrupting any critical processes.

Furthermore, the use of fog-level processing has an added benefit for the operation of the system. As risk assessment is done in proximity to data collection, the process becomes faster and independent from any network inconsistencies, which will be beneficial in cases where the number of monitored patients grows. This could come in handy for community or home monitoring, where network conditions and staffing could be problematic.

The last fact to point out is how the solution integrates with the workflow that is already being used. There is no need for additional interfaces or another parallel system; on the contrary, the copilot integrates into the software that is already being used by clinicians, for example, dashboards and EHRs. It is time-saving and makes clinicians less reluctant to implement the solution while still making it clinicians who must interpret information and make decisions.

All these considerations can lead us to the conclusion that it is possible to develop an effective fog-based AI copilot system to solve the existing problem of risk detection. If developed with clear automation boundaries and taking into account all the restrictions of real-life use, such a system will detect risks in advance while reducing alert fatigue.

The principal clinical implications of the proposed architecture are summarized in Table 4. The table highlights how fog-based AI-assisted risk triage supports earlier clinical awareness, improves workflow integration, reduces alert fatigue, enhances scalability, and maintains clinician oversight through a human-in-the-loop decision-support model.

9. Ethical and Regulatory Considerations

Utilization of AI through telehealth and RPM brings up several ethical and regulatory concerns that include but are not limited to accountability issues, transparency, and safety concerns. The findings of the current study give clear indications that utilization of AI can be done in such a way that makes it play an assisting role that enables early identification of risks without impacting the clinical decision-making process. The current

Table 4. Clinical implications of the fog-based AI Doctor Copilot

Clinical dimension	System capability	Implication for clinical practice
Early risk awareness	Real-time risk-tiered alerts generated at the fog layer	Supports earlier identification of patient deterioration without replacing clinician judgment
Workflow integration	Alerts delivered through existing dashboards and EHR interfaces	Minimizes workflow disruption and reduces training burden for clinicians
Alert prioritization	Low/medium/high risk stratification with prioritization logic	Enables clinicians to focus attention on high-risk cases and mitigates alert fatigue
Scalability of monitoring	Fog-based processing supports increasing patient loads	Facilitates scalable telehealth programs without proportional increases in clinician workload
Operational resilience	Localized inference reduces dependence on continuous cloud connectivity	Improves reliability in home monitoring and community care environments
Clinical safety	AI functions as an assistive copilot with human oversight	Maintains clinician accountability and aligns with patient safety principles

research utilizes AI Doctor as a copilot and not an autonomous system, which is consistent with regulatory expectations.

In this case, the term “AI Doctor” is not used to describe the role of an independent decision-maker. On the contrary, it refers to the decision-making support solution that is designed for facilitating early risk triaging and prioritizing. The solution described does not produce any recommendations for treatment, modify any existing clinical orders, or conduct any activities independently from the clinician. It produces tiered alerts and short descriptions of the situation for clinical judgment.

From an ethical perspective, the hierarchical alerting system does not only take care of pragmatic issues such as alert fatigue and cognitive overload that have long been an issue in the clinical world. This is because there may be very serious consequences from such issues both for the clinicians and the patients. The use of the hierarchical system ensures that the attention of the clinicians is focused better on more important alerts.

The ethical and regulatory considerations associated with the proposed fog-based AI Doctor Copilot are summarized in Table 5. The table outlines how the proposed architecture addresses clinical autonomy, regulatory compliance, explainability, data governance, system resilience, and patient safety while maintaining clinician responsibility for all clinical decisions.

Additionally, using fog computing brings some issues from ethical and regulatory viewpoints. As fog computing makes some processing occur locally, near the data collection points, it decreases the dependence of the system on a stable Internet connection to the cloud. The benefits here include not only improved system availability but also a decrease in the amount of sensitive data that should be transferred through the network. All these features are especially significant when talking about in-home or community-based health monitoring.

Nevertheless, as we are talking about telehealth Internet of Things (IoT), it should be clear that the system operates with highly sensitive physiological data, so the issue of data governance should not be overlooked. Any implementation of the described architecture would require standard security measures, including encrypted communication channels, fog node access control, and regular model updates. Logging system operation history is also necessary for auditing purposes. One of the key advantages of fog computing in this context is that it allows reducing unnecessary data transfer to the cloud. This may be helpful in compliance with standards like Health Insurance Portability and Accountability Act (HIPAA) and Personal Information Protection and Electronic Documents Act (PIPEDA).

Therefore, these points suggest that fog-based AI Doctor Copilot systems can be deployed in a way that is both ethically responsible and compliant with existing regulations. However, this depends on how the system is implemented. Clear limits on autonomy, transparent communication of risk, and continuous human oversight are all necessary. When these elements are built directly into the architecture, the system is more likely to be accepted in clinical practice.

10. Limitations and Future Work

10.1. Limitations

Several limitations can be identified in this study that need to be considered to interpret the results properly. First, this is the fact that this assessment is based on simulation rather than on actual implementation. While simulation is a convenient way to experiment, it does not give a proper representation of the real-world environment in all its complexity and diversity. Datasets like MIMIC and eICU are suitable for retrospective analysis, but it is not very appropriate to test a whole system that incorporates aspects like latency, energy consumption, and distributed computation on those datasets. For this reason, a simulation-based approach was chosen in this research. In future studies, the actual implementation of this system should be done. One possibility is to validate this system using real clinical datasets like MIMIC-IV or eICU. It will be interesting to check how the system performs with real-life patient trajectories and deterioration cases.

Yet another constraint is the utilization of artificial patients' data. It is useful for scalability testing purposes and control, but it cannot represent the variety present in a real-life situation. All

Table 5. Ethical and regulatory considerations of the fog-based AI Doctor Copilot

Dimension	Design approach in this study	Ethical/regulatory implication
Clinical autonomy	AI Doctor operates as an assistive copilot providing risk-tiered alerts only; no diagnosis or treatment decisions are generated	Maintains human-in-the-loop decision-making and aligns with regulatory guidance restricting autonomous clinical actions
Regulatory alignment	System explicitly avoids autonomous decision-making and supports clinician review through existing workflows	Consistent with current governance frameworks emphasizing clinician accountability and oversight
Alert fatigue	Tiered alerting mechanism (low/medium/high risk) prioritizes clinically significant events	Reduces unnecessary interruptions, mitigates alert fatigue, and supports patient safety
Transparency and explainability	Risk scores are derived from interpretable anomaly and trend indicators	Facilitates clinician trust, auditability, and ethical use of AI-assisted decision support.
Data governance	Physiological data are processed at the fog layer with limited reliance on continuous cloud transmission	Supports data minimization principles and reduces exposure risks associated with centralized data processing
System resilience	Fog-based processing enables continued operation during network disruptions	Improves availability and reliability in home and community telehealth settings

efforts have been made in order to introduce variety, noise, and transient effects, but more validation is required.

The model utilized for this research project is also deliberately kept simple. The selection of techniques for anomaly detection and scoring has been made in a way that is simple and easy to understand. This is necessary for deployment within the fog layer. Using more complex machine learning algorithms can lead not only to increased performance but also to higher costs in terms of resources and lower levels of transparency.

In addition to this, it should be noted that the current simulation does not account for all real-life conditions of operation. For example, hardware failures, security issues, or any changes in the behavior of clinicians are not considered. Rather, the emphasis is placed on performance criteria of the system, which are essential for clinical usability, such as latency and accuracy of alerts. Inclusion of these factors will make the experiment more complete. Several ways can be pursued. First of all, the system can be improved in terms of inclusion of additional types of physiological information or risk models tailored to each patient. Second, integration with EHRs is very important. Moreover, clinicians' behavior should be considered over an extended period of time.

10.2. Threats to validity

The following are some of the possible threats to validity that need to be considered. The internal validity can be undermined due to various assumptions made in the process of designing a simulation model and calibrating parameters as well as generating stochastic data. The construct validity will depend on how well the simulated deterioration patterns reflect real-life ones. The external validity of the research is constrained since it is based only on the artificially generated patient data. Lastly, the operational validity is limited by the fact that the behavior of clinicians, adaptations of their work processes, and potential network and hardware issues have not been included in the simulation.

11. Conclusion

In the context of this research, a fog-based AI Doctor Copilot architecture was presented as an approach to performing clinical risk triage in a real-time manner in telemedicine IoT environments. The aim of the proposed combination of inference at the fog level and model management in the cloud is to improve speed while enabling efficient use of resources and integration into the existing clinical process. In comparison with the cloud-based solution, fog implementation showed better results in terms of alert latency, energy consumption, and false alerts, while maintaining similar results in risk classification depending on the increase in patient load.

One of the important aspects of this architecture is the focus on decision support rather than clinical decision-making by the algorithm. The purpose of risk assessment performed by the AI Doctor Copilot is to help clinicians by providing information about patients who need their attention, while the actual diagnosis and treatment decisions are made by the medical specialists themselves. Such an approach is currently expected in terms of ethical and trustworthy AI.

The findings indicate some of the potential benefits of pushing AI inference to the edge in telehealth scenarios. It is, however, crucial to acknowledge that there are several limitations to the work. All the evaluations performed were purely based on simulations using artificial physiological signals, and thus, the reported findings should only be seen as an assessment of architectural feasibility but not clinical efficacy. Validation using real patient data and clinical trials would be required for understanding the usefulness, robustness, and generalizability of the approach. Further research will also examine the application of the developed system within real telehealth facilities and integration into EHRs, as well as the expansion of AI models to cover more medical conditions.

As telehealth and RPM continue to expand, fog-based AI copilot architectures represent a practical and scalable pathway for integrating AI into clinical care. By explicitly aligning system design with operational constraints and clinical practice, this work contributes toward the development of safe, efficient, and trustworthy AI-assisted telehealth systems.

Acknowledgments

This work was supported by the National Natural Science Foundation of Canada and the Department of Health Information Science, University of Victoria.

Ethical Statement

This study did not involve human participants or animal subjects. The research was conducted entirely using synthetically generated physiological data in a simulation environment. Therefore, institutional ethics approval and informed consent were not required. The reported findings are based on simulation and should be interpreted as evidence of architectural feasibility rather than clinical validation.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support this work are available upon reasonable request to the corresponding author.

Author Contribution Statement

Nathan Guo: Methodology, Data curation, Writing – review & editing, Visualization. **Bryan Guo:** Software, Writing – review & editing, Visualization. **Yunyong Guo:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Project administration. **Alex Kuo:** Conceptualization, Validation, Resources, Data curation, Writing – review & editing, Supervision.

References

[1] Bates, D. W., Levine, D., Syrowatka, A., Kuznetsova, M., Craig, K. J. T., Rui, A., . . . , & Rhee, K. (2021). The potential of artificial intelligence to improve patient safety: A scoping review. *npj Digital Medicine*, *4*(1), 54. https://doi.org/10.1038/s41746-021-00423-6

[2] Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

[3] Mittermaier, M., Venkatesh, K. P., & Kvedar, J. C. (2023). Digital health technology in clinical trials. *npj Digital Medicine*, *6*(1), 88. https://doi.org/10.1038/s41746-023-00841-8

[4] Wornow, M., Xu, Y., Thapa, R., Patel, B., Steinberg, E., Fleming, S., . . . , & Shah, N. H. (2023). The shaky foundations of large language models and foundation models for electronic health records. *npj Digital Medicine*, *6*(1), 135. https://doi.org/10.1038/s41746-023-00879-8

[5] Preti, L. M., Ardito, V., Compagni, A., Petracca, F., & Cappellaro, G. (2024). Implementation of machine learning applications in health care organizations: Systematic review of empirical studies. *Journal of Medical Internet Research*, *26*, e55897. https://doi.org/10.2196/55897

[6] European Parliament, & Council of the European Union. (2024). *Artificial Intelligence Act*. https://eur-lex.europa.eu/eli/reg/2024/1689

[7] World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. https://www.who.int/publications/i/item/9789240029200

[8] Warraich, H. J., Tazbaz, T., & Califf, R. M. (2025). FDA perspective on the regulation of artificial intelligence in health care and biomedicine. *JAMA*, *333*(3), 241–247. https://doi.org/10.1001/jama.2024.21451

[9] You, J. G., Hernandez-Boussard, T., Pfeffer, M. A., Landman, A., & Mishuris, R. G. (2025). Clinical trials informed framework for real world clinical implementation and deployment of artificial intelligence applications. *npj Digital Medicine*, *8*(1), 107. https://doi.org/10.1038/s41746-025-01506-4

[10] Muralidharan, V., Adewale, B. A., Huang, C. J., Nta, M. T., Ademiju, P. O., Pathmarajah, P., . . . , & Olatunji, T. (2024). A scoping review of reporting gaps in FDA-approved AI medical devices. *npj Digital Medicine*, *7*(1), 273. https://doi.org/10.1038/s41746-024-01270-x

[11] Damoiseaux-Volman, B. A., van der Velde, N., Ruige, S. G., Romijn, J. A., Abu-Hanna, A., & Medlock, S. (2021). Effect of interventions with a clinical decision support system for hospitalized older patients: Systematic review mapping implementation and design factors. *JMIR Medical Informatics*, *9*(7), e28023. https://doi.org/10.2196/28023

[12] Newton, N., Bamgboje-Ayodele, A., Forsyth, R., Tariq, A., & Baysari, M. T. (2024). How are clinicians' acceptance and use of clinical decision support systems evaluated over time? A systematic review. In J. Bichel-Findlay, P. Otero, P. Scott, & E. Huesing (Eds.), *Studies in health technology and informatics* (pp. 259–263). IOS Press. https://doi.org/10.3233/SHTI230967

[13] Iftikhar, S., Gill, S. S., Song, C., Xu, M., Aslanpour, M. S., Toosi, A. N., . . . , & Uhlig, S. (2023). AI-based fog and edge computing: A systematic review, taxonomy and future directions.. *Internet of Things*, *21*, 100674. https://doi.org/10.1016/j.iot.2022.100674

[14] hatia, J., Italiya, K., Jadeja, K., Kumhar, M., Chauhan, U., Tanwar, S., ..., & Raboaca, M. S. (2023). An overview of fog data analytics for IoT applications. *Sensors*, *23*(1), 199. https://doi.org/10.3390/s23010199

[15] Dahmke, H., Fiumefreddo, R., Schuetz, P., de Iaco, R., & Zaugg, C. (2023). Tackling alert fatigue with a semi-automated clinical decision support system: Quantitative evaluation and end-user survey. *Swiss Medical Weekly*, *153(7)*, 40082. https://doi.org/10.57187/smw.2023.40082

[16] Tomašev, N., Glorot, X., Rae, J. W., Zielinski, M., Askham, H., Saraiva, A., . . . , & Mohamed, S. (2019). A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*, *572(7767)*, 116–119. https://doi.org/10.1038/s41586-019-1390-1

[17] van der Vegt, A. H., Scott, I. A., Dermawan, K., Schnetler, R. J., Kalke, V. R., & Lane, P. J. (2023). Deployment of machine learning algorithms to predict sepsis: Systematic review and application of the SALIENT clinical AI implementation framework. *Journal of the American Medical Informatics Association*, *30*(7), 1349–1361. https://doi.org/10.1093/jamia/ocad075

[18] Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *npj Digital Medicine*, *3*(1), 17. https://doi.org/10.1038/s41746-020-0221-y

[19] London, A. J. (2019). Artificial intelligence and black-box medical decisions: *Accuracy versus explainability*. *Hastings Center Report*, *49*(1), 15–21. https://doi.org/10.1002/hast.973

[20] Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, *3*(11), e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9

[21] Beaulieu-Jones, B. K., Yuan, W., Brat, G. A., Beam, A. L., Weber, G., Ruffin, M., & Kohane, I. S. (2021). Machine learning for patient risk stratification: Standing on, or looking over, the shoulders of clinicians? *npj Digital Medicine*, *4*(1), 62. https://doi.org/10.1038/s41746-021-00426-3

[22] Kairouz, P., & McMahan, H. B. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, *14*(1-2), 1–210. https://doi.org/10.1561/2200000083

[23] Ukoba, K., Olatunji, K. O., Adeoye, E., Jen, T.-C., & Madyira, D. M. (2024). Optimizing renewable energy systems through artificial intelligence: Review and future prospects. *Energy & Environment*, *35*(7), 3833–3879. https://doi.org/10.1177/0958305X241256293

[24] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., . . . , & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[25] Ali, S., Akhlaq, F., Imran, A. S., Kastrati, Z., Daudpota, S. M., & Moosa, M. (2023). The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. *Computers in Biology and Medicine*, *166*, 107555. https://doi.org/10.1016/j.compbiomed.2023.107555

[26] Jung, J., Lee, H., Jung, H., & Kim, H. (2023). Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. *Heliyon*, *9*(5), e16110. https://doi.org/10.1016/j.heliyon.2023.e16110

[27] How to support the transition to AI-powered healthcare. (2024). *Nature Medicine*, *30*, 609–610. https://doi.org/10.1038/s41591-024-02897-9

[28] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17(1)*, 195. https://doi.org/10.1186/s12916-019-1426-2

[29] Phillips, P. J., Hahn, C. A., Fontana, P. C., Yates, A. N., Greene, K., Broniatowski, D. A., & Przybocki, M. A. (2021). *Four principles of explainable artificial intelligence* (Report No. NISTIR 8312). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8312

[30] Hassan, M., Borycki, E. M., & Kushniruk, A. W. (2025). Artificial intelligence governance framework for healthcare.

Healthcare Management Forum, *38*(2), 125–130. https://doi.org/10.1177/08404704241291226

[31] Sadeghi, Z., Alizadehsani, R., Cifci, M. A., Kausar, S., Rehman, R., Mahanta, P., . . . , & Pardalos, P. M. (2024). A review of explainable artificial intelligence in healthcare. *Computers & Electrical Engineering*, *118*, 109370. https://doi.org/10.1016/j.compeleceng.2024.109370

[32] Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(11), 4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314

[33] van de Sande, D., Chung, E. F. F., Oosterhoff, J., van Bommel, J., Gommers, D., & van Genderen, M. E. (2024). To warrant clinical adoption AI models require a multi-faceted implementation evaluation. *npj Digital Medicine*, *7*(1), 58. https://doi.org/10.1038/s41746-024-01064-1

How to Cite: Guo, N., Guo, B., Guo, Y., & Kuo, A. (2026). A Fog-Based AI Doctor Copilot for Real-Time Clinical Risk Triage in Telehealth IoT Systems. *Medinformatics*. https://doi.org/10.47852/bonviewMEDIN620210326