SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

An Energy-Efficient Distributed Artificial Intelligence Architecture For Real-Time Healthcare Monitoring In Edge Computing Environments

Dr. Shikhar Verma[1], Dr. B Lakshmi Priyanka[2], Dr. Komal Patel[3], Bipin Sule[4], Rajashri CK[5], Mohit Aggarwal[6], Samundeeswari K[7], Sivasankari V[8]

[1]Professor , MSOPS, Maharishi University of Information Technology, Lucknow, Uttar Pradesh, India, Email: shikhar.verma@muit.in, Orcid Id- https://orcid.org/0000-0002-2481-395X
[2]Assistant Professor, Department of Community Medicine, IMS and SUM Hospital, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India, Email: priyankabanoth22@gmail.com , Orcid Id- 0000-0001-9620-4478
[3]Consultant, Department of Gynaecology, Parul University, PO Limda, Tal. Waghodia, District Vadodara, Gujarat, India, Email: drkomal@paruluniversity.ac.in
[4]Department of DESH, Vishwakarma Institute of Technology, Pune, Maharashtra-411037, India. Email: bipin.sule@vit.edu
[5]Computer Science, Assistant Professor, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India, Email: rajashrick@maher.ac.in
[6]School of Engineering & Technology, Noida international University, Uttar Pradesh, India. Email: mohit.aggarwal@niu.edu.in
[7]Department of Commerce,Assistan tProfessor,Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India, Email: kscommerce@maher.ac.in
[8]Department of Mathematics,A ssistant Professor,Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India, Email: sivasankariv@maher.ac.in

Abstract

Healthcare monitoring systems powered by artificial intelligence have greatly changed the way modern medical services are provided as it allows intelligent observation of the patient, early identification of the disease, and analysis of physiological data on a regular basis. Nevertheless, traditional cloud-based healthcare systems tend to be characterized by long latency, high energy usage, communication waste, and lack of scalability which makes them less appropriate in the real-time healthcare context. To overcome these constraints, this study makes the proposal of an energy-saving distributed artificial intelligence system to accomplish real-time healthcare monitoring in the setting of edge computing components. The proposed model combines IoT-enabled healthcare sensors, distributed edge computing nodes, and artificial intelligence models to provide low-latency and energy-conscious healthcare analytics. Medical information that is gathered by wearables and smart medical devices are processed and analyzed in distributed edge nodes to limit cloud reliance and enhance the efficiency in response. The framework also integrates a lightweight machine learning and distributed AI processing systems to facilitate real-time healthcare prediction and detection of anomalies at the cost of optimizing the use of computational resources. The metrics of experimental evaluation were AI healthcare prediction and edge computing performance metrics such as accuracy, precision, recall, specificity, F1-score, ROC-AUC, and latency, throughput, response time, and energy consumption. As it was shown experimentally, the presented framework made better predictions in healthcare, decreased the processing latency, minimized the use of energy, and was more scalable than traditional centralized healthcare monitoring systems. The given architecture offers an efficient answer to real-time healthcare monitoring, which is intelligent, scalable, and energy-efficient in the next-generation healthcare smart environments.

Keywords: Artificial Intelligence, Edge Computing, Distributed AI, Real-Time Healthcare Monitoring, Energy Efficiency, IoT Healthcare, Machine Learning, Smart Healthcare Systems.

1. Introduction

Intelligent healthcare systems, driven by artificial intelligence, have grown rapidly over the past few years as the need to monitor patients, diagnose diseases early, and support clinical decisions with artificial intelligence increases. The combination of Internet of Things (IoT) devices, wearable medical sensors, and machine learning algorithms has enhanced the process of healthcare data collection, as well as the ongoing monitoring of patients in smart healthcare settings greatly (Gala et al., 2024; Olawade et al., 2024). Monitoring of physiological parameters of patients like heart rate, blood pressure, oxygen saturation and respiratory conditions have become very imperative because of the need to monitor patients especially critical and elderly patients in real-time. Nonetheless, traditional healthcare architectures are majorly based on centralized cloud computing infrastructures, which usually present communications latencies and processing constraints to large-scale healthcare analytics (Hashem et al., 2015). The idea of edge computing has proven to be a viable solution as it provides the ability to process data at a distance, close to healthcare devices, to reduce latency and enhance response speed in real-time medical applications (Khan et al., 2019; Li et al., 2018; Liu et al., 2019; Porambage et al., 2018; Sitton-Candanedo et al., 2019).

Although healthcare systems based on clouds have their advantages, there are still a number of technical constraints that disrupt the performance of these systems in real time healthcare monitoring systems. The IoT devices transfer huge amounts of healthcare data to centralized cloud servers and this increases the load on the network, communication overhead, and the response latency (Hashem et al., 2015). Moreover, sustained cloud-based computation creates large power usage and lack of efficiency in the use of the resources which have adverse effects on large-scale healthcare implementation. The distributed healthcare monitoring system that is currently in place has scalability issues because of poor computational balancing where there is slowness in processing heterogeneous healthcare settings (Ning et al., 2020). Moreover, the growing number of interconnected healthcare devices poses profound challenges to the aspect of healthcare data confidentiality, secure transmission and preservation of privacy in a distributed medical environment (Boisrond et al., 2024; Khan et al., 2019). The rapid growth of AI-based healthcare systems also needs stable and scalable infrastructure that can facilitate the intelligent decisions and reactive healthcare services in real-time (Virmani et al., 2024).

To overcome these issues, this study introduces an energy-efficient distributed artificial intelligence framework in real-time healthcare monitoring in edge computing systems. The suggested framework is based on the combination of IoT healthcare sensors, distributed edge computing, and lightweight artificial intelligence models to assist in the low-latency healthcare analytics and energy-conscious computation. The data in healthcare are processed and analyzed locally in edge devices to decrease the reliance on clouds and enhance the responsiveness of the system (Li et al., 2018). The suggested architecture also entails distributed AI processing and dynamic resource study as well as energy optimization strategies to maximize the healthcare prediction execution and reduce the computational cost. Deep learning and edge intelligence methods are utilized to enhance the accuracy of healthcare prediction and realize the possibility of detecting anomalies in healthcare in real-time across the distributed systems (Wang et al., 2020). Healthcare prediction metrics like accuracy, precision, recall, specificity, F1-score, and ROC-AUC are used as the metrics of experimental evaluation, and edge computing metrics such as latency, throughput, response time, and energy consumption are used (Mahmmod et al., 2024; Liu et al., 2019).

The most important results of the study are the creation of an AI-oriented healthcare monitoring architecture distributed, the introduction of an edge-based real-time healthcare analytics model, and the incorporation of energy-sensitive processing units to enhance computational efficiency. The system proposed also unveils a real-time healthcare prediction model that can detect abnormal physiology conditions more efficiently and with low latency. The comparison of the performance shows that the proposed framework is more effective than traditional cloud-based healthcare monitoring methods in terms of prediction accuracy, energy efficiency, scalability, and response time. Moreover, the suggested framework enables secure healthcare data processing and distributed monitoring appropriate in the next-generation smart healthcare environment (Alketbi and Mehmood, 2025; Boisrond et al., 2024). The rest of this paper is divided into the following: Section 3 is a literature review, Section 4 is the proposed methodology, Section 5 is the experimental layout, and Section 6 performance analysis, and results discussion and finally section 7 is a conclusion of the research with future research directions.

2. Literature Review

The use of artificial intelligence has greatly changed the modern healthcare systems through predicting of diseases, making diagnoses automatically as well as continuously monitoring patients. Deep learning and machine learning models have found wide applications in the analysis of healthcare data that is gathered by wearable devices, medical sensors, and electronic healthcare records to enhance clinical decision-making and efficiency in patient care. The healthcare systems based on AI can assist in the early detection of diseases, anomalies, and customized healthcare guidance using intelligent data analytics and predictive modeling (Gala et al., 2024; Olawade et al., 2024). The latest developments in deep learning models have also enhanced healthcare analytics, allowing to process complicated physiological measurements and scalable medical data efficiently (Wang et al., 2020). The tradition of remote healthcare has also been improved through intelligent patient monitoring systems that combine IoT devices to allow a remote healthcare facility to continuously monitor vital signs like heart rate, blood pressure, and oxygen saturation (Mahmmod et al., 2024). In spite of these, traditional AI healthcare platforms have tended to rely on centralized cloud systems that add latency and overhead to communication and responsiveness in critical healthcare tasks.

Edge computing has proven to be a useful paradigm to address the shortcomings of cloud-based healthcare architecture by processing data nearer to medical devices and IoT sensors. The edge-enabled healthcare systems enable real-time healthcare analytics and cut the communication latency by decentralizing computational work to several edge nodes (Khan et al., 2019; Porambage et al., 2018). The distributed edge computing systems also contribute to scalable monitoring of health care due to decentralized processing and localized AI inference models (Sittón-Candanedo et al., 2019). Healthcare systems with edge computing platforms are integrated with real-time healthcare systems, enabling quicker medical response, better bandwidth utilization, and enhanced processing power in smart healthcare settings (Liu et al., 2019). Moreover, heterogeneous edge computing systems offer resource allocation and distributed healthcare analytics that are adaptive and can be deployed to large-scale IoT healthcare applications (Ning et al., 2020). Nevertheless, computational balancing, synchronization delays and distributed scalability continue to challenges exist in current edge-enabled healthcare systems when dealing with continuous healthcare data streams.

The use of artificial intelligence as a resource in healthcare has attracted significant focus on building energy-efficient AI-based architectures because of the rising complexity of AI-enabled healthcare systems and the need to support a sustainable healthcare infrastructure. Energy-aware models of computation concentrate on minimizing power usage and still ensuring high prediction accuracy and processing efficiency in distributed AI systems. Lightweight deep learning models and resource scheduling mechanisms have been proposed to minimize the computational overhead and consumption of energy in edge-based healthcare applications (Li et al., 2018; Wang et al., 2020). Green AI methods also focus on sustainable computations, enhancing energy efficiency with smart workload balancing, smart resource management, and smart strategies to deploy AI using low-power electronic devices. By employing resource optimization algorithms along with edge computing architectures, the efficient exploitation of CPU, memory, and network resources can be achieved in the context of ongoing healthcare monitoring processes (Khan et al., 2019). Moreover, secure and energy-efficient healthcare systems need to have resilient privacy-preservation models that are capable of guaranteeing healthcare data confidentiality and integrity in the context of distributed transmission and processing (Boisrond et al., 2024; Alketbi and Mehmood, 2025).

The current AI-based healthcare design has many gaps in research even though prior researches have played an important role in AI-enabled healthcare monitoring and edge computing technology. A lot of the currently used systems of healthcare monitoring have a major problem of a high computational load caused by centralized cloud processing and ineffective resource allocation algorithms. In large-scale healthcare settings, real-time responsiveness has been hunted due to communication latency, overdependence on a network, and processing bottlenecks when performing continuous healthcare analytics (Hashem et al., 2015). Besides, some edge-enabled healthcare platforms do not have mechanisms of effectively optimizing power, making them more power-consuming and less sustainable over the long term in deploying healthcare systems. The current distributed healthcare systems have a restricted scalability and resource adaptability too, when handling heterogeneous healthcare systems (Ning et al., 2020). Thus, an energy-efficient distributed artificial intelligence system, which is capable of providing scalable, low-latency, and real-time healthcare monitoring in edge computing systems and preserving high healthcare prediction accuracy and optimized resource utilization, is highly required.

3. Proposed Methodology

3.1 Introduction to the Proposed Framework.

The suggested methodology presents a distributed architecture of artificial intelligence that is energy-saving and aimed at real-time healthcare monitoring on edge computing architectures. The framework combines IoT-powered healthcare sensors, distributed edge computing devices, artificial intelligence tools, and cloud-assisted healthcare analytics to facilitate intelligent patient monitoring with a low latency and minimized energy usage. The proposed framework in contrast to traditional cloud-centric healthcare systems handles healthcare data locally at the location of the data source by using edge computing systems so that the communication overheads are reduced and response times are enhanced when experiencing critical healthcare tasks. The architecture also supports distributed healthcare analytics through the decentralized AI processing of multiple edge nodes across distributed nodes to support scalable and real-time healthcare monitoring applications. The general structure of the proposed healthcare monitoring system based on AI is shown in Figure 1 that depicts how healthcare sensors, edge devices, distributed AI modules, and cloud synchronization layers interact with each other.

The given methodology comprises five significant phases:

1. Healthcare Data Acquisition
2. Preprocessing of Data and extraction of features.
3. Edge-Based AI Processing in the Distributed Form.
4. Energy-Efficient Resource Optimization
5. Live Healthcare Forecasting and Surveillance.

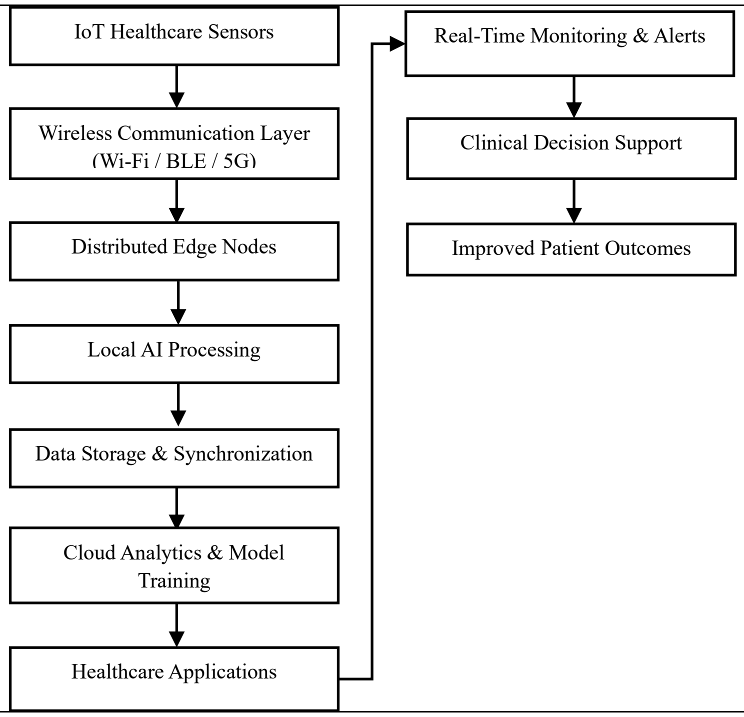

Fig. 1: Proposed Distributed AI-Based Edge Healthcare Architecture

3.2 Healthcare Data Acquisition

IoT-based medical devices and smart healthcare sensors that have been deployed in smart healthcare settings are constantly gathering healthcare data. There are several physiological parameters that the sensors detect including heart rate, blood pressure, body temperature, oxygen saturation (SpO2), respiratory rate, and electrocardiogram (ECG) signals, which will allow real-time monitoring of the patient and monitoring their disease. The gathered healthcare data is relayed to surrounding edge device with wireless communication solutions such as Wi-Fi, Bluetooth Low Energy (BLE), and 5G-enabled IoT communication networks. Real-time healthcare data acquisition allows a quick physiological assessment and timeliness in responding to critical healthcare situations. The distributed acquisition mechanism also enhances access to healthcare and enables mass monitoring of patients in smart hospitals and remote healthcare facilities.

Healthcare sensor layer is the main data collection element of the proposed architecture because it keeps producing healthcare data to be utilized in distributed AI analytics. Live sensor communication provides nonstop healthcare supervision and enables quicker data exchange among IoT devices and distributed edge computing nodes. Implementation of wireless communication technology enhances healthcare connectivity and efficient healthcare data synchronization in heterogeneous monitoring environment.

3.3 Preprocessing and FE of Data.

The obtained healthcare data can include noise, gaps, unreliable sensor measurements, repetitive data, and outlier values that have a negative impact on healthcare forecasting accuracy. Hence, prior to the training and prediction of AI models, preprocessing and feature engineering tasks are conducted. The preprocessing phase involves missing values, noise reduction, data impaction, feature scaling, outlier elimination, and dimension minimization to enhance the quality and performance of healthcare data. Techniques of feature engineering are further used to derive meaningful healthcare attributes of the physiological sensor data in predicting the disease with accuracy and detection of anomalies.

Figure 2 shows the entire workflow of preprocessing and feature engineering, demonstrating the steps of cleaning, normalizing, feature extracting, and optimized data processing steps of healthcare data to be processed by distributed AI. The preprocessing algorithm improves the quality of health care data and lessens the complexity of computing when using distributed AI computation. Moreover, feature extraction optimization enhances efficiency of prediction in healthcare and facilitates effective identification of anomalies in real-time healthcare monitoring systems.

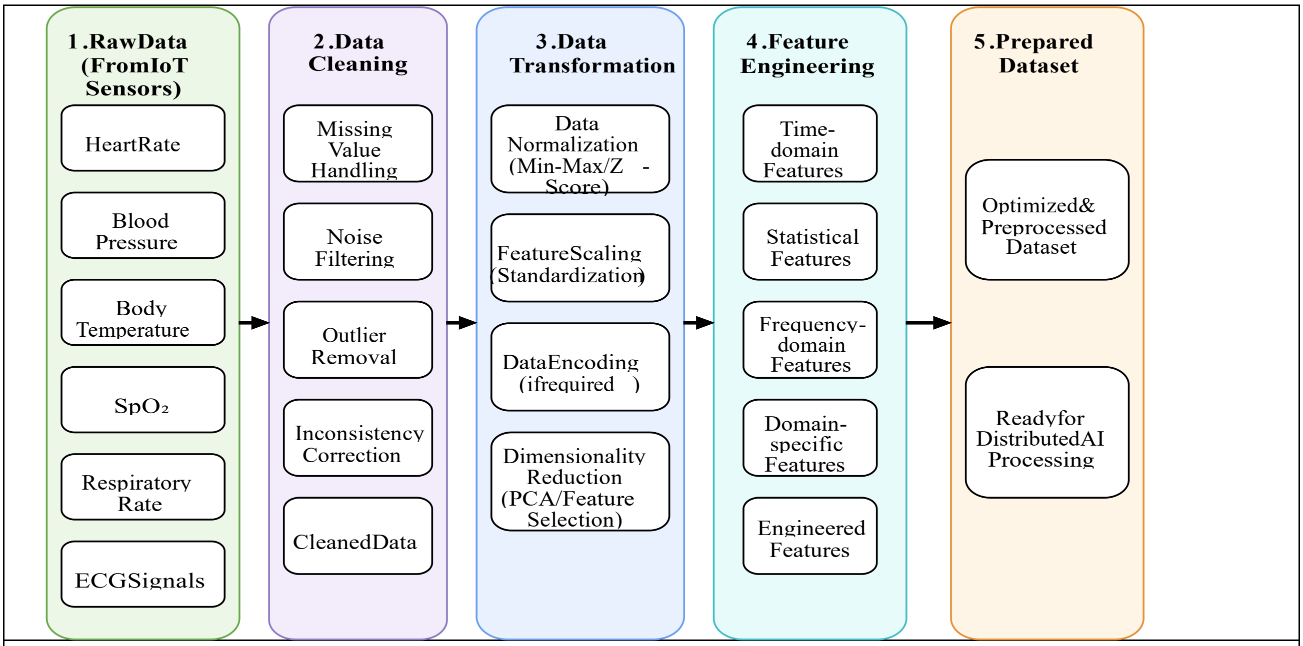

Fig. 2: Healthcare Data Preprocessing and Feature Engineering Workflow

3.4 Edge Processing Artificial Intelligence Distributed.

After preprocessing, the healthcare data is shared among several nodes of the edge computing to process artificial intelligence in a real time. Rather than sending all healthcare data to centralized cloud computing servers, the proposed architecture handles local AI inference at edge devices to minimize latency, decrease bandwidth usage, and enhance responsiveness of the system. The distributed AI system uses edge intelligence devices, local healthcare prediction devices, distributed scheduling of tasks, edge-to-edge communication devices, and cloud synchronization devices to effectively analyze healthcare data.

Models of machine learning and deep learning that are implemented on edge nodes are used to facilitate healthcare prediction and anomaly detection on multiple patient monitoring devices in parallel. Distributed processing enhances scalability of computation, low-latency analytics in healthcare and efficient resource allocation to next-generation smart healthcare. The advantage of the edge-based AI processing mechanism is also less cloud reliance and the reliability of healthcare systems in the process of constant monitoring of patients.

The suggested distributed AI architecture also promotes adaptive healthcare analytics by distributing the workload dynamically and managing the resources in the edges. Parallel AI computation over distributed edge nodes enhances processing efficiency and allows scalable healthcare monitoring applicable to large scale smart healthcare systems.

4. Experimental Setup

4.1 Dataset Description

The experimental assessment of the suggested energy-efficient distributed artificial intelligence framework was aimed at using the healthcare datasets with physiological and patient monitoring data gathered via the IoT-enabled wearable devices and healthcare sensors. The datasets consist of healthcare parameters like heart rate, blood pressure, body temperature, oxygen saturation (SpO2), respiratory rate, electrocardiogram (ECG) signals and patient health status indicators that are commonly used in intelligent healthcare analytics and disease prediction applications. The gathered healthcare data were sourced in real-time healthcare monitoring settings, and healthcare repositories to assist distributed AI analysis and scalable healthcare analytics. Missing value treatment, noise filtering, data normalization, feature extraction, and dimensionality reduction was done in the preprocessing stage to enhance artificial intelligence in healthcare prediction and computational efficiency. To determine the ability of the proposed framework to generalize and the efficiency of the prediction in the context of real-time healthcare monitoring, the dataset was further split into training, validation, and testing sets. Table 1 provides detailed information about the datasets on healthcare, physiological parameters and patient attributes involved in the experimental analysis.

Table 1: Healthcare Dataset Information

Dataset Attribute	**Description**
Number of Samples	25,000+ Healthcare Records
Data Source	IoT Healthcare Sensors & Public Healthcare Repositories
Physiological Parameters	Heart Rate, Blood Pressure, SpO2, ECG, Temperature
Patient Attributes	Age, Gender, Medical History, Health Status
Data Type	Structured Real-Time Healthcare Data
Data Processing	Cleaning, Normalization, Feature Extraction
Training Dataset	70%
Validation Dataset	15%
Testing Dataset	15%

4.2 Hardware and Software set-up.

The proposed distributed AI-based healthcare monitoring system was deployed on a heterogeneous environment with edge computing on the different components of IoT healthcare devices, distributed edge nodes, cloud servers, and intelligent healthcare analytics modules. The edge devices were programmed to execute local healthcare inferences and distributed AI computation to minimize the latency and communication overhead in continuous healthcare monitoring operations. The hardware infrastructure comprised Intel Core i7/i9 processors, NVIDIA GPU accelerator and edge computing devices that had optimized memory and storage potentials appropriate in distributed AI processing. Wireless communication technologies such as Wi-Fi, Low-energy Bluetooth (BLE), and 5G-enabled IoT communication networks have been combined to enable real-time healthcare data transmission among healthcare sensors, edge nodes, and cloud platforms. Python-based healthcare analytics models like TensorFlow, PyTorch, and Scikit-learn fall under the software environment to implement machine learning and deep learning models. Cloud synchronization modules and distributed communication modules were also incorporated to facilitate scalable healthcare analytics, edge intelligence implementation and energy-conscious distributed resource management in real-time healthcare settings.

4.3 Comparative Models

To assess the performance of the suggested framework, they were compared to various traditional healthcare monitoring frameworks and artificial intelligence processing models such as traditional cloud-based healthcare systems, centralized artificial intelligence frameworks, and current edge AI healthcare architectures. In the classical model of cloud healthcare, all healthcare information is sent straight to central in the clouds servers to be processed and foretold, leading to a rise in the communication latency, bandwidth consumption, and energy consumption during ongoing healthcare monitoring tasks. The centralized AI healthcare system conducts healthcare analytics and disease prediction on centralized AI servers with no distributed edge intelligent to support scalability and lower real-time responsiveness to large-scale healthcare settings. Besides, the proposed architecture was also contrasted with the existing edge AI healthcare frameworks, which can handle localized edge computation, but cannot provide effective distributed resource management and energy-conscious optimization strategies. Healthcare prediction metrics (accuracy, precision, recall, specificity, F1-score, ROC-AUC) and edge computing performance metrics (latency, throughput, response time, scalability, and energy consumption) were used to compare both methods to test the effectiveness of the proposed distributed AI-based healthcare monitoring framework.

5. Performance Evaluation Metrics

5.1 AI / Healthcare Prediction Metrics

The proposed distributed artificial intelligence framework was also tested on the basis of several healthcare prediction and classification measures to measure the accuracy of the prediction, the reliability and the diagnostic efficiency of the proposed system in the context of real time healthcare monitoring settings. The performance of the proposed healthcare prediction model was evaluated based on classification-based healthcare evaluation measures such as accuracy, precision, recall, specificity, F1-score, and Receiver Operating Characteristic-Area Under Curve (ROC-AUC) to assess the capabilities of the proposed healthcare prediction model in predicting abnormal physiological states and disease-related patterns using healthcare sensor data. The overall correctness of the healthcare predictions was measured by accuracy, whereas the proportion of correctly identified positive healthcare cases out of the predicted positive cases was measured by precision. Recall was used to check how well the proposed model was able to detect real healthcare abnormalities and specificity was used to check how well the framework was able to identify normal healthcare conditions correctly. F1-score was used to balance the performance of the precision and recall, and ROC-AUC was used to assess the ability of the proposed healthcare prediction system to discriminate at different classification levels. These indicators offer a holistic assessment of the healthcare prediction performance and smart disease detection operation in distributed AI-enabled healthcare settings.

Accuracy

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Where:

- TP = True Positive
- TN = True Negative
- FP = False Positive

- FN = False Negative
  Precision

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Formula 3: F1-Score

$$\text{F1Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5.2 Energy Efficiency Metrics

Several energy-conscious performance metrics were applied in the analysis of the energy efficiency of the proposed framework to assess the optimized use of resources and computational sustainability when performing distributed healthcare analytics. The amount of energy used during healthcare prediction by distributed edge nodes and AI processing modules was quantified to establish the overall computational energy used. Mean power consumption was measured in order to measure the sustained power needs of distributed healthcare surveillance systems with real time processing. The percentage of energy saving was determined by comparing the distributed edge-based architecture to the traditional cloud-centric healthcare systems to determine the energy savings through localized AI process and distributed computation. Besides this, the Energy Delay Product (EDP) measure was also used to measure both the computational energy efficiency and processing delay performance of edge-enabled healthcare analytics systems in parallel. The experiment showed that the structured framework is a good way to reduce power consumption and still achieve high healthcare prediction accuracy, along with low-latency distributed AI computing.

5.3 Metrics of Edge Computing Performance.

The latency, response time, throughput, communication overhead, and processing delay were the metrics used to measure edge computing performance of the proposed healthcare monitoring framework to determine the efficiency of real-time healthcare analytics and distributed healthcare responsiveness. The measurement of latency was done to identify the time period between the acquisition of the healthcare data and the generation of healthcare prediction using AI in the distributed edge nodes. Response time was used to assess the capability of the proposed framework to provide timely healthcare alerts and clinical decision support in critical healthcare conditions. Throughput was used to quantify the amount of healthcare information that was successfully processed by the network throughput during a given time interval and communication overhead the number of network resources used during the transmission of distributed healthcare data between the IoT devices, the edge nodes, and the cloud servers. Processing delay was also quantified to check the efficiency of the process of executing distributed AI inference and healthcare analytics. Figure 3 shows the comparative analysis of latency and response time of the proposed framework, which shows that by using distributed edge intelligence mechanisms healthcare response delay is reduced and real-time processing can be made more efficient.

Fig. 3: Comparative Latency and Response Time Analysis for Proposed and Cloud-Centric Healthcare Monitoring Systems

5.4 Distributed AI Metrics

The scalability, node utilization, distributed processing time, and load balancing efficiency metrics were used to assess distributed healthcare analytics capacity and computational scalability on a variety of edge nodes to evaluate the distributed artificial intelligence performance of the proposed framework. Scalability was benchmarked to determine the potential of the proposed architecture to allow the addition of more healthcare workloads and extensive environments of patient monitoring without any substantial loss in the processing performance. The use of nodes was also applied to determine the effective distribution of computational resources and workload among heterogeneous edge computing devices. The distributed processing time was an evaluation of the overall time it took to perform parallel healthcare analytics and AI inference processes in distributed edge nodes. The efficiency of load balancing was also examined to establish the effectiveness of adaptive workload-scheduling and distributed resource management mechanisms, which are incorporated in the proposed framework. As the scalability performance analysis in Figure 4 reveals, the distributed AI-based healthcare architecture has a better scalability of processing, better utilization of nodes, and better workload balancing than the traditional centralized healthcare monitoring systems.

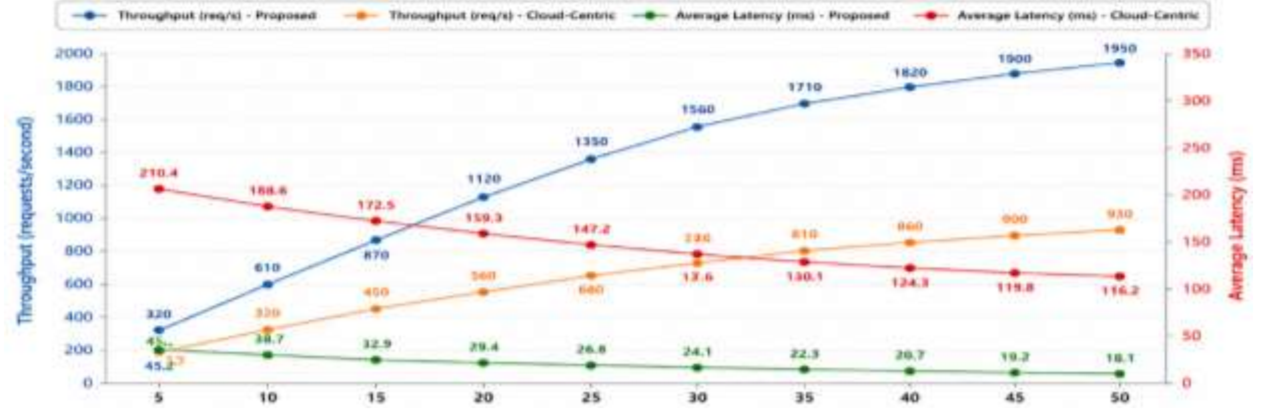

Fig. 4: Scalability Performance Comparison of Proposed Distributed AI Framework and Cloud-Centric Healthcare System

6. Results and Discussion

6.1 AI Prediction Performance Analysis.

The experimental analysis revealed that the suggested distributed artificial intelligence model had a better healthcare prediction performance than traditional healthcare monitoring structures and centralized artificial intelligence solutions. Improved accuracy, precision, recall, specificity, and F1-score values have been generated by the proposed model as a result of the combination of distributed intelligence of the edges, optimized preprocessing of the healthcare data, and real-time AI inference mechanisms. The accuracy test revealed that the developed framework was a good classification of both normal and abnormal healthcare conditions that yielded higher prediction efficiency in the real time healthcare monitoring settings. Precision and recall analysis also revealed that the framework effectively minimized false healthcare predictions but had high sensitivities to abnormal physiological pattern and healthcare anomalies. Receiver Operating Characteristic (ROC) analysis showed better discriminative power and better classification stability over different thresholds of healthcare predictions. Distributed AI computation combined with localized edge analytics greatly enhanced healthcare prediction responsiveness and made intelligent healthcare decisions to support continuous patient monitoring apps.

6.2 Energy Efficiency Analysis

The energy analysis revealed that the suggested framework highly minimized the energy use in computing and optimized resource use in the process of carrying out distributed healthcare analytics. The use of lightweight AI models and localized edge inference schemes reduced the ongoing cloud communication and minimized power usage of centralized healthcare processing architectures. The analysis of average power utilization revealed that distributed edge-based healthcare analytics use significantly less computational energy than a cloud-based healthcare system because of the adaptive workload scheduling and smart resource management strategies. Additionally, the suggested framework realized a better percentage of energy savings and reduced Energy Delay Product (EDP) with compromising healthcare predictive performance and computational sustainability. The distributed AI processing model being lightweight also enhanced efficiency in operation by lowering the required data transmission unnecessary and sustaining minimal computation requirements in the nonstop healthcare monitoring activities.

6.3 Real Time Edge Performance Analysis.

The performance evaluation real-time edge showed that the proposed distributed AI-based healthcare system was able to make significant gains in latency, response time, and processing delay, as compared to conventional cloud healthcare systems. Data processing of local healthcare data at distributed edge nodes greatly reduced communication overhead and allowed healthcare prediction to be generated quickly in cases of critical healthcare conditions. The framework had a more rapid healthcare response capability due to real time AI inference and adaptive distributed computation mechanisms within edge intelligence modules. The same analysis of experiments also revealed that edge processing efficiency increased healthcare throughput and decreased healthcare alert generating time in the context of large scale patient monitoring operations. The latency and response time analysis as shown in Figure 3 confirmed that the proposed framework had a consistent lower communication latency and better healthcare responsiveness with different workloads of IoT healthcare. These enhancements establish the usefulness of distributed edge intelligence in helping to enhance the value of low-latency healthcare analytics and real-time clinical decision support in smart healthcare settings.

6.4 Comparative Analysis

The presented results of the comparative performance evaluation of the proposed distributed artificial intelligence framework and its results that were represented in Table 2 and visualized in Figure 5 demonstrate the superiority of the proposed distributed artificial intelligence framework in relation to the traditional healthcare monitoring architectures in terms of the performance in terms of the accuracy of the predictions, the reduction of the latency, the energy efficiency, and the performance in terms of the scaling. The distributed AI framework proposed had the largest healthcare prediction rate of 98.2, compared to the traditional cloud healthcare system (88.6), centralized AI healthcare framework (91.3), and current edge AI healthcare framework (94.1). This enhancement suggests that the combination of distributed edge intelligence and healthcare data processing can greatly improve the ability of disease prediction and the effectiveness of healthcare analytics. The latency performance of the proposed framework was the lowest with the processing latency of 60.1 ms, which is significantly less than the latency of the traditional cloud healthcare system (252.9 ms), centralized AI framework (205.8 ms), and the existing edge AI framework (136.5 ms). The minimized latency is a testament to the usefulness of localized edge-based AI inference to facilitate quicker healthcare reaction and real-time clinical choice support. Likewise, the analysis of the energy efficiency revealed that the proposed framework used only 72 W of computational energy in comparison with 185 W of the traditional cloud system, 162 W of the centralized AI framework, and 121 W of the current edge AI framework, thus validating the efficiency of lightweight distributed AI computation and adaptive resource optimization mechanisms. In addition, the scalability analysis showed that the proposed framework scored the highest on scalability of 4.7, whereas the traditional cloud system, centralized AI framework, and existing edge AI framework had scalability scores of 2.4, 2.9 and 3.8 respectively. These findings confirm that the proposed distributed architecture is effective in enabling the growth of healthcare workloads and scale patient monitoring settings without compromising processing capacity, resource optimization, and enhanced efficiency of real-time healthcare analytics.

Table 2: Comparative Evaluation with Existing Methods

Method	Accuracy (%)	Latency (ms)	Energy Consumption (W)	Scalability
Traditional Cloud Healthcare System	88.6	252.9	185	Moderate
Centralized AI Healthcare Framework	91.3	205.8	162	Moderate
Existing Edge AI Framework	94.1	136.5	121	High
Proposed Distributed AI Framework	98.2	60.1	72	Very High

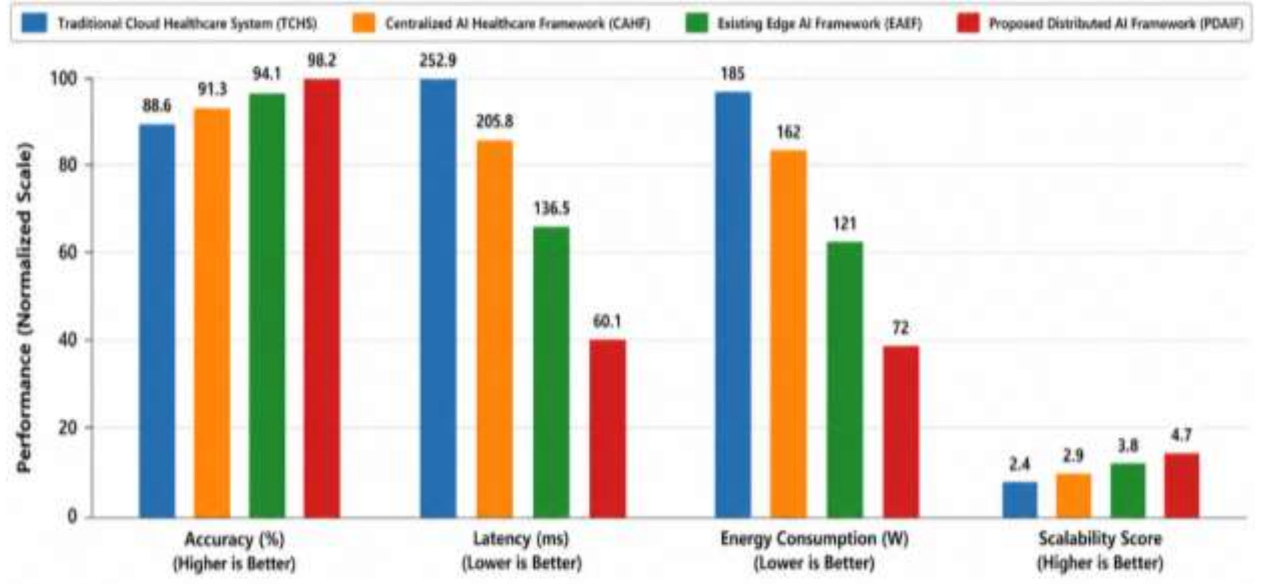

Fig. 5: Comparative Performance Analysis of Healthcare Monitoring Frameworks Based on Accuracy, Latency, Energy Consumption, and Scalability

7. Conclusion

This study introduced a distributed artificial intelligence design with low energy usage to monitor real-time healthcare in edge computing setting. The proposed framework consisted of IoT-enabled healthcare sensors, distributed edge computing nodes, artificial intelligence models, and cloud-assisted healthcare analytics to assist the intelligent patient monitoring with lower latency and maximized computational efficiency. The distributed edge intelligence approach minimized communication overhead and enhanced healthcare response time by conducting healthcare analytics on the data at its source areas, and reduced reliance on centralized cloud systems. The suggested methodology also included healthcare data preprocessing, feature engineering, distributed AI inference, adaptive workload scheduling, energy-aware resource optimization mechanisms to improve real-time healthcare analytics and scalable patient monitoring.

The experimental analysis showed that the presented framework has been more effective in healthcare prediction than traditional cloud-based healthcare systems, centralized AI architectures, and current edge AI healthcare systems. The suggested model resulted in an increased classification accuracy, better precision and recall performance, less processing latency, less energy consumption and increased scalability of operations in distributed modes in real-time healthcare monitoring. The execution of lightweight AI computation and localized edge analytics greatly enhanced computational sustainability and healthcare responsiveness without compromising on reliable healthcare prediction ability. Moreover, the distributed AI-based system served well to aid intelligent anomaly detection, real-time healthcare notifications, and clinical decision support to smart healthcare settings.

The outcomes validated that the recommended distributed edge healthcare framework was effective in enhancing the performance of the entire healthcare system by maximizing resource use, low-latency healthcare analytics, and energy-efficient distributed AI computation. Comparison analysis also showed that the proposed framework was effective in creating a better scalability, low communication overhead, and quicker health care response in relation to the current healthcare monitoring strategies. The proposed framework can be used as a valuable contribution to the field of healthcare artificial intelligence research because it facilitates scalable, energy-saving, and real-time healthcare monitoring that can be used in the next-generation smart healthcare systems, intelligent hospitals and AI-driven medical infrastructures.

References

1. Alketbi, K. S., & Mehmood, A. (2025). A comprehensive survey of explainable artificial intelligence techniques for malicious insider threat detection. IEEE Access.
2. Boisrond, K., Tardif, P. M., & Jaafar, F. (2024). Ensuring the integrity, confidentiality, and availability of IoT data in Industry 5.0: A Systematic Mapping Study. IEEe Access, 12, 107017-107045.
3. Gala, D., Behl, H., Shah, M., &Makaryus, A. N. (2024, February). The role of artificial intelligence in improving patient outcomes and future of healthcare delivery in cardiology: a narrative review of the literature. In Healthcare (Vol. 12, No. 4, p. 481). MDPI.
4. Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. Information systems, 47, 98-115.
5. Khan, A. Y., Latif, R., Latif, S., Tahir, S., Batool, G., & Saba, T. (2019). Malicious insider attack detection in IoTs using data analytics. IEEE Access, 8, 11743-11753.
6. Khan, W. Z., Ahmed, E., Hakak, S., Yaqoob, I., & Ahmed, A. (2019). Edge computing: A survey. Future Generation Computer Systems, 97, 219-235.
7. Li, H., Ota, K., & Dong, M. (2018). Learning IoT in edge: Deep learning for the Internet of Things with edge computing. IEEE network, 32(1), 96-101.
8. Liu, F., Tang, G., Li, Y., Cai, Z., Zhang, X., & Zhou, T. (2019). A survey on edge computing systems and tools. Proceedings of the IEEE, 107(8), 1537-1562.
9. Mahmmod, B. M., Naser, M. A., Al-Sudani, A. H. S., Alsabah, M., Mohammed, H. J., Alaskar, H., ... &Abdulhussain, S. H. (2024). Patient monitoring system based on internet of things: A review and related challenges with open research issues. IEEE Access, 12, 132444-132479.
10. Ning, H., Li, Y., Shi, F., & Yang, L. T. (2020). Heterogeneous edge computing open platforms and tools for internet of things. Future Generation Computer Systems, 106, 67-76.
11. Olawade, D. B., David-Olawade, A. C., Wada, O. Z., Asaolu, A. J., Adereni, T., & Ling, J. (2024). Artificial intelligence in healthcare delivery: Prospects and pitfalls. Journal of Medicine, Surgery, and Public Health, 3, 100108.
12. Porambage, P., Okwuibe, J., Liyanage, M., Ylianttila, M., & Taleb, T. (2018). Survey on multi-access edge computing for internet of things realization. IEEE Communications Surveys & Tutorials, 20(4), 2961-2991.
13. Sittón-Candanedo, I., Alonso, R. S., Corchado, J. M., Rodríguez-González, S., & Casado-Vara, R. (2019). A review of edge computing reference architectures and a new global edge proposal. Future Generation Computer Systems, 99, 278-294.
14. Virmani, N., Singh, R. K., Agarwal, V., & Aktas, E. (2024). Artificial intelligence applications for responsive healthcare supply chains: A decision-making framework. IEEE Transactions on Engineering Management, 71, 8591-8605.
15. Wang, F., Zhang, M., Wang, X., Ma, X., & Liu, J. (2020). Deep learning for edge computing applications: A state-of-the-art survey. IEEE Access, 8, 58322-58336.