Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery

WILEY

WIREs DATA MINING AND KNOWLEDGE DISCOVERY

ADVANCED REVIEW OPEN ACCESS

Edge Computing in Healthcare Using Machine Learning: A Systematic Literature Review

Amir Mashmool[1] | Giorgio Delzanno[2] | Hamid Saadatfar[3] | Aakash Ahmad[4] | Rainer Koschke[1] | Roohallah Alizadehsani[5] | U. Rajendra Acharya[6] | Daniele D'Agostino[2]

[1]Faculty of Mathematics and Computer Science, University of Bremen, Bremen, Germany | [2]Department of Computer Science and Technology, Bioengineering, Robotics and Systems Engineering, University of Genoa, Genoa, Italy | [3]Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran | [4]Data Science Research Centre, School of Computing, University of Derby, Derby, UK | [5]Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Geelong, Victoria, Australia | [6]School of Mathematics, Physics, and Computing, University of Southern Queensland, Springfield, Queensland, Australia

Correspondence: Amir Mashmool (mashmool.amir@uni-bremen.de)

Received: 15 June 2025 | **Revised:** 4 December 2025 | **Accepted:** 21 January 2026

Editor-in-Chief: Witold Pedrycz | **Associate Editor:** Sushmita Mitra

Keywords: edge AI | edge computing | healthcare systems | Internet of Medical Things | machine learning

ABSTRACT

Healthcare is rapidly evolving with the integration of machine learning (ML) and edge computing, which enables real-time data processing and improved patient care. Edge computing plays a critical role by reducing latency and enhancing data privacy, especially in patient monitoring systems. However, limitations such as device resource constraints and security issues persist. This study presents a systematic literature review (SLR) on using ML and edge computing in healthcare, identifying key benefits, challenges, and research trends. This SLR aimed to identify key benefits, challenges, and current research trends. We sourced relevant studies from databases such as IEEE Xplore, ScienceDirect, ACM Digital Library, and so forth. We applied inclusion and exclusion criteria. We also used the snowballing technique to find more relevant studies by checking selected papers' reference lists, ensuring we did not miss any important ones. Finally, 37 papers were selected and analyzed for their methodologies, algorithms, tools, frameworks, data sources, limitations, motivations, and challenges. Findings show a broad use of ML methods such as support vector machines, clustering, and deep learning, with a strong emphasis on data privacy and model performance; many studies employed federated learning and privacy-preserving techniques to support real-time decision-making. Overall, ML and edge computing integration promise to transform healthcare, though challenges remain. Future research should address resource limitations, enhance ML models for edge environments, and develop standardized protocols.

This article is categorized under:

Application Areas > Health Care

Technologies > Machine Learning

Technologies > Cloud Computing

1 | Introduction

While the Internet of Things (IoT) represented one of the key technologies for the fourth industrial revolution, the Internet of Medical Things (IoMT) remained only a great promise of the medicine of the future (El-Saleh et al. 2025). This technology represents a must-have for several reasons. One example is that when a pandemic like COVID-19 arises, it may not be feasible for all individuals to be hospitalized (Yang et al. 2020). Moreover, the number of older adults staying alone is increasing exponentially in several countries. Remote patient monitoring systems, often based on wearable devices (Adeniyi et al. 2021), can enable real-time detection of anomalies to alert caregivers in a timely way (Khan et al. 2021).

Traditional healthcare systems transmit large volumes of heterogeneous data, generated by medical devices, wearables, and patient monitoring systems, to centralized cloud infrastructures for analysis. Although this approach has enabled numerous breakthroughs, it often suffers from latency issues, bandwidth constraints, and the risk of data breaches. This aspect represents one of the key issues: medical data are subject to severe national and international laws because of its strict security and privacy requirements (Yaacoub et al. 2020; Hatzivasilis et al. 2019).

Edge computing represents a suitable solution to tackle these issues because it enables raw data processing at or near the source, reducing the amount of sensitive data to be transmitted and enhancing the responsiveness of healthcare applications (D'Agostino, Morganti, et al. 2019; D'Agostino, Quarati, et al. 2019). However, edge computing is insufficient if the goal is to create advanced services, for example, predictive diagnostics (Bellavista et al. 2020). Machine learning (ML) represents a suitable solution for deriving knowledge from the vast amounts of healthcare data (Habehh and Gohel 2021). ML models can be deployed directly on edge devices when integrated with edge computing, enabling real-time analysis and decision-making. This distributed approach accelerates computational processes and minimizes the risks of transmitting sensitive patient data to external servers.

The integration of edge computing and ML represents a paradigm shift in healthcare, promising to enhance efficiency, accessibility, and patient outcomes (Kishor and Chakraborty 2022). This paper aims to systematically review recent research on integrating these technologies into healthcare.

The findings of this systematic literature review (SLR) offer valuable insights for: (i) researchers aiming to explore the intersection of edge computing and ML in healthcare. These insights can support the formulation of new research hypotheses, the design of lightweight models suitable for edge environments, and the development of secure, privacy-preserving healthcare frameworks. (ii) Practitioners such as developers and healthcare technology providers can use this study to understand current trends in edge-based AI applications, including prevalent tools, frameworks, challenges, and implementation strategies relevant to real-world healthcare scenarios.

The structure of this paper is as follows. Section 2 presents a comprehensive overview of edge computing and cloud computing, establishing the conceptual foundation for the study. Section 3 describes the methodology employed in this SLR, including formulating research questions (RQs) and examining potential threats to validity. Sections 4 and 5 detail the study's findings: Section 4 outlines the demographic characteristics of the selected publications, while Section 5 synthesizes the key themes, tools, challenges, and data sources identified in the literature. Section 6 provides an in-depth review of primary studies. Section 7 discusses the findings in detail, limitations, challenges, and proposes future research directions. Finally, Section 8 concludes the paper.

2 | Background

Combining ML with edge computing has become a promising method for delivering real-time, innovative, and privacy-respecting solutions for patient care and medical diagnostics. This section presents a general understanding of edge and cloud computing, emphasizing their functions in today's data processing and connectivity, focusing on the healthcare sector.

2.1 | Edge and Cloud Computing

The computing industry has experienced significant progress in recent years, propelled by swift advancements in hardware miniaturization, affordability, computational performance, and network connectivity. These breakthroughs have contributed significantly to the extensive use of computing devices, which are essential for smooth data exchange and profoundly impact contemporary society (Lalis et al. 2007; Hammer 1984; Cox and Zeelenberg 1987).

As the number of devices connected to the Internet, such as personal computers (PCs), smartphones, and the wide range of devices that conserve energy on the IoT, has increased rapidly, the digital environment has changed significantly (Familiar 2015). The varying capacities of devices have created a multifaceted network environment, encompassing everything from high-speed fiber-optic and Wi-Fi links to long-distance, low-energy networks like LTE and 5G. Consequently, the rise in interconnected devices has resulted in an extraordinary increase in data production. Recent projections suggest that by 2025, the world's data volume could reach about 175 zettabytes, originating from numerous sources such as sensor-driven systems, mobile applications, and cloud services (Rydning et al. 2018).

Cloud computing has become crucial in healthcare (Shahbaz and Zahid 2022), as it could provide scalable computing power, ample storage, and various services through distant data centers. Such cloud infrastructure supports efficient data management, cost savings, and easy access to computational resources, which is advantageous for organizations managing large volumes of healthcare data. However, cloud computing also faces challenges such as increased latency, reliance on network connectivity, potential security risks, and considerable energy usage, especially for applications sensitive to time and with limited resources (El Kafhali et al. 2022).

Edge computing has emerged as an essential alternative to mitigate these constraints by relocating computational capabilities closer to the healthcare data source (Hartmann et al. 2022). In contrast to cloud computing, which depends on centralized cloud servers, edge computing processes data on the network periphery, utilizing local servers, gateways, or user devices. This approach reduces latency, protects privacy, and facilitates immediate decision-making. This model is particularly significant in sectors where promptness and data protection are paramount, such as healthcare, autonomous vehicles, and industrial automation. Implementing Edge AI, which uses ML models right at the edge, makes edge devices more capable, letting them run independently with less reliance on cloud-based infrastructure (Hua et al. 2023).

2.2 | Edge Computing in Healthcare

The healthcare industry is experiencing a considerable shift due to the integration of digital technologies such as cloud computing, IoT, and ML. These innovations have resulted in the development of innovative healthcare systems that use real-time data analytics to improve patient outcomes and hospital operations and efficiently manage resources (Hartmann et al. 2022). Nonetheless, conventional cloud-based healthcare solutions frequently encounter issues with data transmission delays, limited bandwidth, and regulatory compliance, especially with the handling of sensitive patient data.

Edge computing addresses these challenges and reduces reliance on centralized cloud infrastructures by enabling real-time data processing at the healthcare location (Putra et al. 2024). The IoMT, which includes wearable devices, remote monitoring equipment, and intelligent diagnostic systems, produces large amounts of data that require instant processing (He et al. 2025). Edge computing supports the rapid analysis of patient information, accelerating clinical decision-making and shortening emergency response times, as reported by Joo et al. (2024) for an IoT-based pulse oximeter.

Furthermore, protecting data privacy and security is crucial in healthcare settings due to the sensitivity of patient records and medical imaging information. Edge computing promotes data sovereignty by limiting the transmission of unprocessed patient data over public networks (Ganesan and Jagatheesaperumal 2024). This approach reduces the chance of data breaches. It supports adherence to strict regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR) (Ettaloui et al. 2023).

Although edge computing has many benefits, it also has many challenges. For instance, hardware limitations, lower computing power compared to cloud infrastructure, energy use problems, and the need for strong ML models optimized for edge devices are just a few of them (Hua et al. 2023). To overcome these challenges, novel approaches are required in areas such as model compression, efficient data transfer protocols, and federated learning, which enables machines to learn collaboratively without sharing raw patient data (Ganesan and Jagatheesaperumal 2024).

2.3 | Machine Learning in Edge Computing

ML has emerged as a fundamental component of contemporary computing, supporting predictive analytics, anomaly identification, and automated decision-making in many fields. Edge computing deploys ML algorithms on edge devices, enabling real-time inference (Murshed et al. 2021). This approach minimizes reliance on centralized cloud infrastructures by enabling faster, localized data processing and decision-making. This synergy, frequently termed edge AI, proves especially advantageous in resource-limited settings where minimizing latency and optimizing energy consumption are crucial (Alamouti 2025).

Edge computing, powered by ML, supports various healthcare applications. Wearable health monitors and biosensors that use ML algorithms make it easier to find diseases early by picking up on early signs of conditions like heart problems, diabetes, and neurological disorders (Ganesan and Jagatheesaperumal 2024). AI-powered edge devices process medical images locally, directly handling radiology, CT, and x-rays. By processing medical images locally, AI-powered edge devices reduce diagnostic delays and the need for cloud-based processing. Personalized treatment plans benefit from real-time patient monitoring, where ML models analyze historical and current data, offering adaptive treatment recommendations. Edge-enabled ML models have also made remote patient monitoring a lot better by being able to spot changes in vital signs and warn healthcare professionals of possible medical emergencies (Jagatheesaperumal et al. 2022).

Even with these benefits, implementing ML models on edge devices introduces numerous challenges. A primary concern is model optimization. Since conventional ML models often demand substantial computational resources, making them unsuitable for deployment on low-power edge devices (Kasula 2024). Techniques like quantization, pruning, and knowledge distillation are essential in optimizing models for edge deployment. Keeping data in sync is also challenging because edge devices work in spread-out environments (Anand et al. 2024). Edge devices are naturally distributed, so developers must create effective methods to synchronize local model updates with centralized systems. Also, ensuring that ML models are safe and reliable is very important because edge AI systems need to be able to handle attacks from other systems and privacy breaches (Nemec Zlatolas et al. 2024).

The integration of edge computing, ML, and healthcare represents an opportunity to create innovative, immediate, and privacy-focused healthcare solutions. Hardware acceleration progress, such as AI chips and Tensor Processing Units (TPUs), along with energy-saving algorithms and safe frameworks for edge computing, will be significant for the creation of intelligent healthcare systems in the future (Tripathi et al. 2022).

3 | Research Method

As part of our research method, we used the SLR method to carefully analyze and investigate the existing literature related to our RQs. Figure 1 presents an overview of the research methodology for this SLR.

3.1 | Planning the Review

3.1.1 | Specify Research Questions (RQs)

We present our RQs in Table 1.

3.1.2 | Identify Data Sources

In systematic reviews and mapping studies, electronic data sources (EDS) help automate the exploration process by using predefined and often customized search terms to find relevant scholarly works related to a specific topic (Chen et al. 2010).

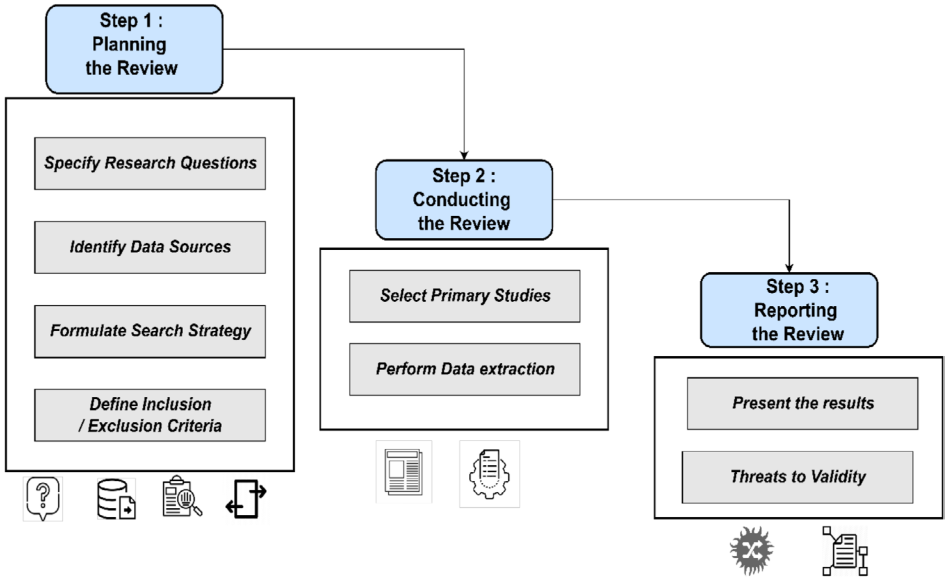

FIGURE 1 | An overview of the research methodology employed in this SLR.

TABLE 1 | Research questions of this SLR.

#	Research question	Rationale
	Demographic details of published research	
RQ1.1	What are the publication types and the frequency of published research on edge computing in healthcare using ML?	This research question seeks to identify the various publication types, such as journal articles, conference proceedings, and so forth, and emphasize how often they occur, measured by the number of publications in a given year. It aims to shed light on the evolution of research within the chosen topic by examining the types, such as journal articles, conference proceedings, and so forth, and the frequency of published research throughout the years.
RQ1.2	What types of research have been published on edge computing in healthcare using ML?	The type of research can be solution proposals, validation, and so on.
RQ1.3	What is the geographic distribution of the authors' affiliations?	We use the geographic distribution of authors' affiliations to ensure a diverse range of perspectives, enhance the generalizability of findings, and identify regional trends or biases in research.
	Themes, tools, proposed solutions, motivations, performance, and data sources	
RQ2.1	What are the prevalent research themes in edge computing applications within healthcare using ML techniques?	Identifying prevalent research themes helps in understanding the primary areas of focus, innovation, and potential gaps in knowledge within the field.
RQ2.2	What underlying motivations led the authors to explore edge computing in healthcare using machine learning?	Understanding researchers' motivations provides insight into the perceived benefits, societal impact, and industry demand driving the adoption of edge computing solutions in healthcare.
RQ2.3	What are the most prominent tools and technologies for edge computing in healthcare using machine learning?	Identifying the most prominent tools and technologies helps recognize the key technological advancements and resources that facilitate edge computing in healthcare.
RQ2.4	What types of data are used by machine learning applications?	Understanding whether applications utilize a single or multiple data sources provides insight into the data management strategies and complexities involved in edge computing for healthcare.
RQ2.5	What are the performance benchmarks of ML models?	To compare and analyze the performance of ML models reported in the primary studies, focusing on accuracy.

We followed the recommended guidelines for conducting a systematic search and identifying the most appropriate sources of information. We used seven EDS for an automated search. These databases include the ACM Digital Library, IEEE Xplore, ScienceDirect, SpringerLink, Wiley Online Library, Scopus, and Google Scholar, all well-known repositories of scholarly literature. Previous studies based on empirical evidence regarding SLRs have emphasized the importance and suitability of these five electronic repositories for finding relevant literature (H. Zhang et al. 2011; Chen et al. 2010).

3.1.3 | Formulate Search String

We used a search string to conduct our data search, which is presented in Box 1.

BOX 1 | Search string for literature search.

(("Edge Computing" AND ("Health" OR "Medical"))
AND
("Machine Learning" OR "Artificial Intelligence" OR "AI"))

The provided search string combines key terms using the "OR" and "AND" Boolean operators.

3.1.4 | Define Inclusion and Exclusion Criteria

Inclusion criteria:

- Research articles that specifically focus on the application of edge computing in the healthcare sector, utilizing ML techniques.
- Exclusively considering peer-reviewed and published research.
- Published before June 2024.
- Only peer-reviewed studies provide access to the full text.
- Studies written and published in the English language.

Exclusion criteria:

- Exclude the studies that do not discuss healthcare.
- Studies published in languages other than English.

3.2 | Conducting the Review

3.2.1 | Select Primary Studies

Searching for primary studies began by selecting digital repositories and searching using a specific string. The search began on February 10, 2024 and ended on June 20, 2024. We employed a snowballing approach to systematically expand our paper selection and identify additional relevant studies for inclusion in our research. Snowballing refers to identifying additional relevant studies by examining the references cited in pertinent already-known papers. The supervisors reviewed the selected articles in this SLR to ensure their relevance to the topic.

The selection process started with a comprehensive search across multiple digital libraries and online databases. Using the search string shown in Box 1 that combines terms like "Edge Computing," "Health/Medical," and "Machine Learning/Artificial Intelligence," the initial search pulls in a large number of studies from sources like IEEE Xplore, Springer Link, Science Direct, Wiley Online Library, ACM Digital Library, Scopus, and Google Scholar. This process yielded a broad initial set of 255 studies. Next, we carefully screen the titles, keywords, and abstracts of these studies to eliminate those that are less relevant, reducing the list to 127 studies. Next, we read the full text of each study and kept only those that met our criteria. After carefully applying our inclusion/exclusion criteria, we identified 34 studies meeting our requirements. The snowballing technique was employed to facilitate a more comprehensive review. Even looking at the reference list of selected studies for missed relevant studies in the initial search was part of the steps taken. This process revealed three additional vital studies, for a total of 37. The list of the selected studies included in this review is shown in Table A3 (see Appendix B). This approach added depth to the review, as snowballing uses the network of citations of scholarly papers to ensure that the literature base is comprehensive and robust. Figure 2 illustrates the process we used to select the studies for our review.

3.2.2 | Perform Data Extraction

In answer to the RQs, we set up a map of data extraction items (Table 2). Data items are the particular types of data extracted from each selected study that directly relate to the study's RQs.

3.3 | Reporting the Review

3.3.1 | Present the Results

We presented the findings of our SLR in Sections 4 and 5. As Table 1 shows, we structured our results in alignment with the defined RQs.

3.3.2 | Threats to Validity

Similar to empirical studies, SLRs are not immune to validity threats, which may affect the reliability of their findings (Cruzes and Dybå 2010). Based on Zhou et al. (2016) and Ampatzoglou et al. (2019), we identify threats to the validity of our SLR.

3.3.2.1 | Threats to the Identification of Primary Studies. The main goal of our search strategies was to retrieve as many relevant papers as possible to prevent any potential bias. One of the key challenges in addressing these threats was defining the scope of our study as it pertains to edge computing, healthcare, and artificial intelligence. To ensure comprehensive coverage and avoid bias, we searched for and incorporated commonly used terms into our search string (cf. Box 1). While this approach reduces bias, it significantly increases the complexity of the search

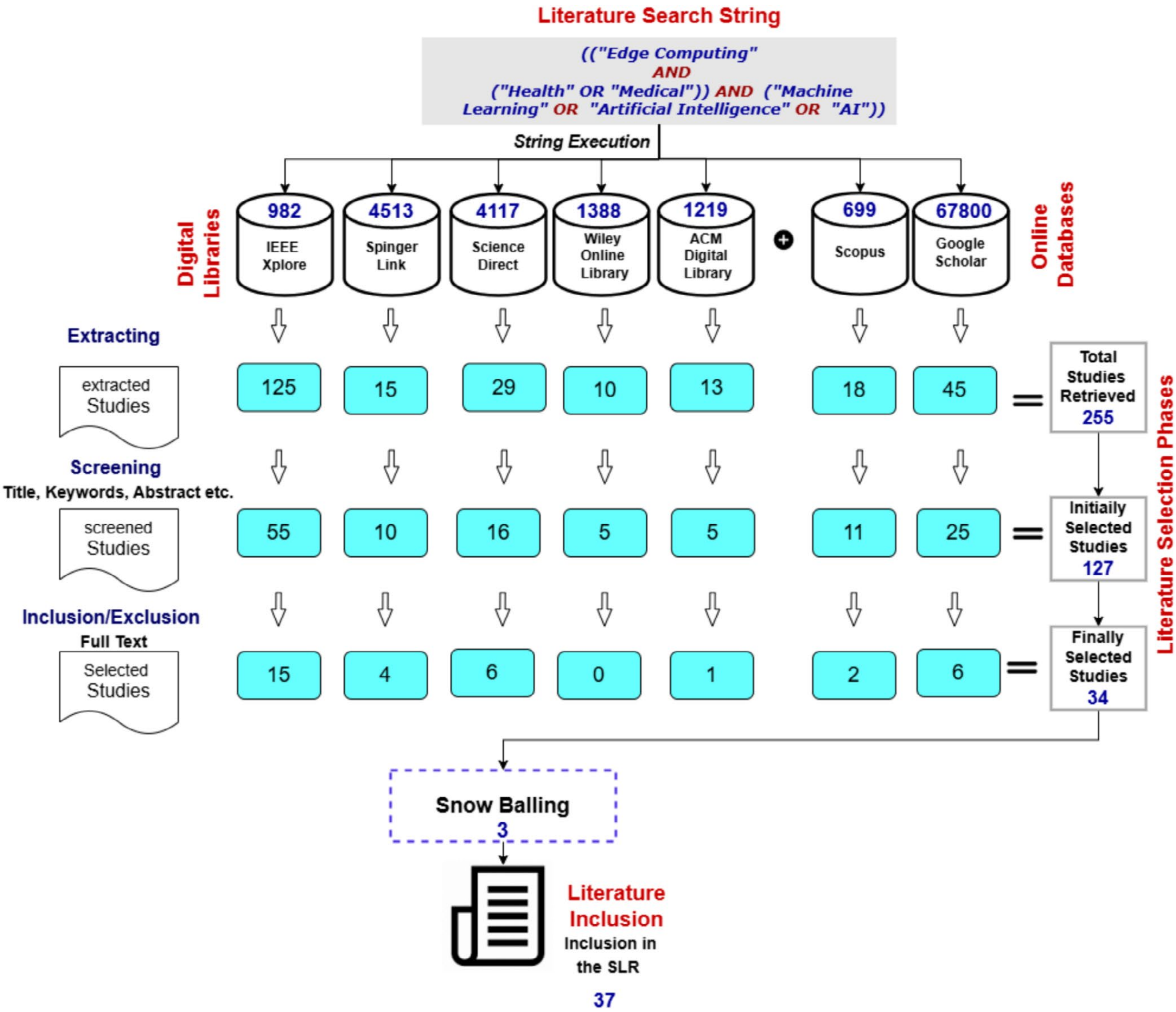

FIGURE 2 | Flowchart of the study selection process.

process. The selection of papers for review is critical, as the exclusion of relevant papers can substantially influence the results of the SLR. To tackle this problem, we utilized two techniques, that is, searching the online databases through search strings and the snowballing approach, to minimize the risk of missing relevant studies.

3.3.2.2 | Threats to the Selection of Primary Studies. The first author (Amir) compiled the selected articles into an Excel file to mitigate this risk. Subsequently, Daniele reviewed all the articles to verify their relevance to the study.

3.3.2.3 | Threats to the Quality of Studies. To mitigate this threat, we focused exclusively on articles that had undergone peer review, ensuring that they were legitimate research papers. These studies appeared in reputable conferences and journals indexed in Scopus.

3.3.2.4 | Threats to Researcher Bias. To address this threat, the authors (Amir and Daniele) conducted regular one-on-one meetings throughout the SLR process. They were engaged in a collaborative effort to address specific discrepancies and disagreements. They aimed to mitigate errors and reduce any potential for unconscious bias.

4 | Demographic Details of Published Research

4.1 | Types and Frequency of Publications (RQ1.1)

Classifying our selected primary studies by frequency and publication type is a vital workflow phase. This process aims to highlight the most important research directions in the field of interest and to draw attention from the academic community. Analyzing publication frequency over time and the publication type shown in Figure 3 provides a complete overview of the available research. Our investigation spans from the earliest publication in 2017 to the most recent in 2024. Notably, our analysis reveals that a significant proportion (82%, comprising 30 studies) of our selected studies were published within the last 5 years (from 2020 to February 2024). This finding underscores the growing importance of edge computing in healthcare

TABLE 2 | Relevant data items extracted from the selected studies.

Data item	Description	Related research question
Index	The study ID	Demographic
Study title	The full title of the primary study	Demographic
List of authors	The full names of all authors who contributed to the study	Demographic
Publication type	The publication categories include journal articles, conference papers, book chapters, and so forth	Demographic
Journal or conference name	The name of the journal or conference where the study was published or presented	Demographic
Journal or conference rank	The ranking or impact factor of the journal or conference, if applicable	Demographic
Name of publisher	The name of the organization or company that published the study	Demographic
Geographic distribution of authors	The authors' affiliation indicates the location of their research institution or university	Demographic
Publication year	The year of publication for each study	RQ1.1
Research type	The nature of the research includes validation research, proposal solutions, philosophical papers, and so forth	RQ1.2
Research themes	The main topics or areas of focus addressed by the study	RQ2.1
Paper's motivation	The underlying reasons or objectives that prompted the research	RQ2.2
List of tools	The specific tools and software used in the study	RQ2.3
Papers data source	The origin of the data used in the study, including primary data collected by the authors or secondary data obtained from existing sources	RQ2.4

research. It highlights the strong efforts within the research community to develop ML solutions for edge computing applications. Furthermore, our examination indicates that 73% (27 studies) of the primary research appeared in journals, while 27% (10 studies) were presented in conference proceedings. We must note that we exclude survey and review articles from our primary studies list.

Key findings of RQ1.1

Finding 1: Most primary studies ($n = 30$, 82%) were published between 2020 and 2024, indicating that edge computing in healthcare using machine learning is an emerging research area that has garnered significant attention from the research community.
Finding 2: Regarding publication type, the results highlight that journals are the preferred venue for publishing relevant studies, with 73% (27 studies) of publications appearing in journals compared to 27% (10 studies) in conferences.

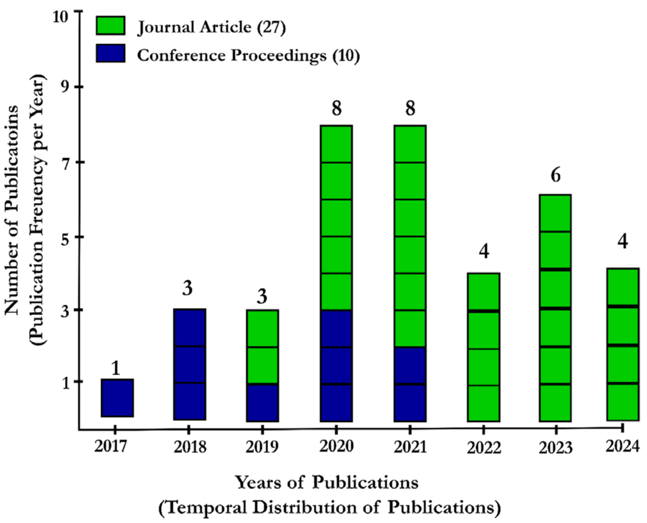

FIGURE 3 | Overview of frequency and types of publications.

4.2 | Types of Research (RQ1.2)

The selected publications are classified according to six established research types delineated by Wieringa et al. (2006): evaluation studies, solution proposals, validation research, philosophical papers, opinion papers, and personal experience papers. The proposed solution articles aim to introduce a novel approach or enhance existing techniques to address pertinent issues, prioritizing conceptual development over validation.

Validation research seeks to assess the quality attributes of the proposed solution, which has not yet been implemented in requirements engineering practice. Evaluation research assesses a problem or resolution in real-world applications using various empirical research methods, while philosophical papers outline perspectives or conceptual frameworks. Opinion papers express the author's perspectives on the strengths or weaknesses of a specific framework, model, or solution. In personal experience papers, the authors report their experiences with a particular project(s). The primary focus of these papers is on delineating the actions taken or events experienced ("what"), rather than delving into the underlying reasons or motivations ("why"). The process of thematic analysis aims to address RQ1.2 by categorizing the 37 chosen primary studies according to their research types (as depicted in Figure 4). Out of these, we found that 13 papers (35.14%) were related to proposed solutions, and six studies (16.22%) focused on validation research. Additionally, we pinpointed 6 (16.22%) studies that straddle both proposal of solution and validation research categories and 12 (32.43%) studies that encompass validation research, proposal of solution, and evaluation research.

These studies are categorized separately, as illustrated in Figure 4. We did not identify any papers fitting into the personal experience papers, opinion papers, or papers only focused on the evaluation research; hence, they were excluded from the mapping process. Most papers have used ML methods and performed validation and evaluation to prove their results or compare their work.

Key findings of RQ1.2

Finding 1: Examining the types of published research reveals that solution proposals constitute the most frequent publications. Thirteen studies (35.14%) fall into this category.
Finding 2: A notable portion of the articles (12 out of 37, 32.43%) simultaneously cover validation research, proposal solutions, and evaluation research. So, these articles are of high value.

4.3 | Geographic Distribution of the Authors' Affiliations (RQ1.3)

The geographic distribution of authors of this SLR is presented in Table 3, and the visualization with a map is presented in Figure 5. We considered the affiliations of all authors of the articles. To prevent label overlapping, we have not included the names of all countries on the map.

Key findings of RQ1.3

Finding 1: The United States (10 studies) and China (9 studies) are the most represented countries in the studies reviewed.
Finding 2: Researchers from 24 different countries have contributed to the studies, which shows that people from all over the world are interested in this topic and are working together.

5 | Themes, Tools, Challenges, and Data Sources

This section analyses the state of the research, including five RQs (RQ 2.1–RQ 2.5). We will answer these RQs in subsequent sections.

5.1 | Prevalent Research Themes (RQ2.1)

In this section, we concentrate on the results for RQ 2.1, which addresses the question: What are the prevalent research themes in edge computing applications within healthcare using ML techniques? To understand how all the existing research fits together, we first examined individual studies to uncover the main themes using a method called thematic analysis (Boyatzis 1998). We then organized these themes into a taxonomy illustrated in Figure 6 by reviewing some relevant studies (Medvidovic and Taylor 2000; Ahmad and Babar 2016). Following guidelines from the ACM Computing Classification System[1] and the Computing Research Repository,[2] we decided to classify them into two categories: (1) Generic Classification categorizes the reviewed studies into the following six distinct areas: (I) Applications and Solutions, (II) Techniques and Architectures, (III) Integration and Platforms, (IV) Specific Healthcare Challenges, (V) Frameworks and Methods, and (VI) Privacy and Security. (2) Thematic Classification extends the Generic Classification by incorporating additional dimensions derived from the primary research focus of the selected studies (Boyatzis 1998).

The Thematic Classification is as follows: (I) Applications and Solutions (8 studies, i.e., 21.6%), (II) Techniques and Architectures (10 studies, i.e., 27%), (III) Integration and Platforms (3 studies, i.e., 8.1%), (IV) Specific Healthcare Challenges (5 studies, i.e., 13.5%), (V) Frameworks and Methods (6 studies, i.e., 16.2%), and (VI) Privacy and Security (5 studies, i.e., 13.5%).

Key findings of RQ2.1

Finding 1: The Sunburst chart illustrates the taxonomy divided into six categories: Applications and Solutions, Techniques and Architectures, Integration and Platforms, Specific Healthcare Challenges, Frameworks and Methods, and Privacy and Security.
Finding 2: The taxonomy reveals that privacy and security are critical considerations in deploying ML and DL methods in edge computing for healthcare, emphasizing the need for robust frameworks and methodologies to protect sensitive health data.

5.2 | Motivations (RQ2.2)

Yu et al. (2018) pointed out that IoT devices in healthcare can monitor human body conditions and provide early warnings in case of danger. These services require low cost, high convenience, and low latency. Most existing solutions are based on cloud platforms that transfer data to the cloud and return the results. However, these methods are inefficient in the era of data explosion because IoT data has characteristics such as large scale, heterogeneity, and high noise, which create complexities

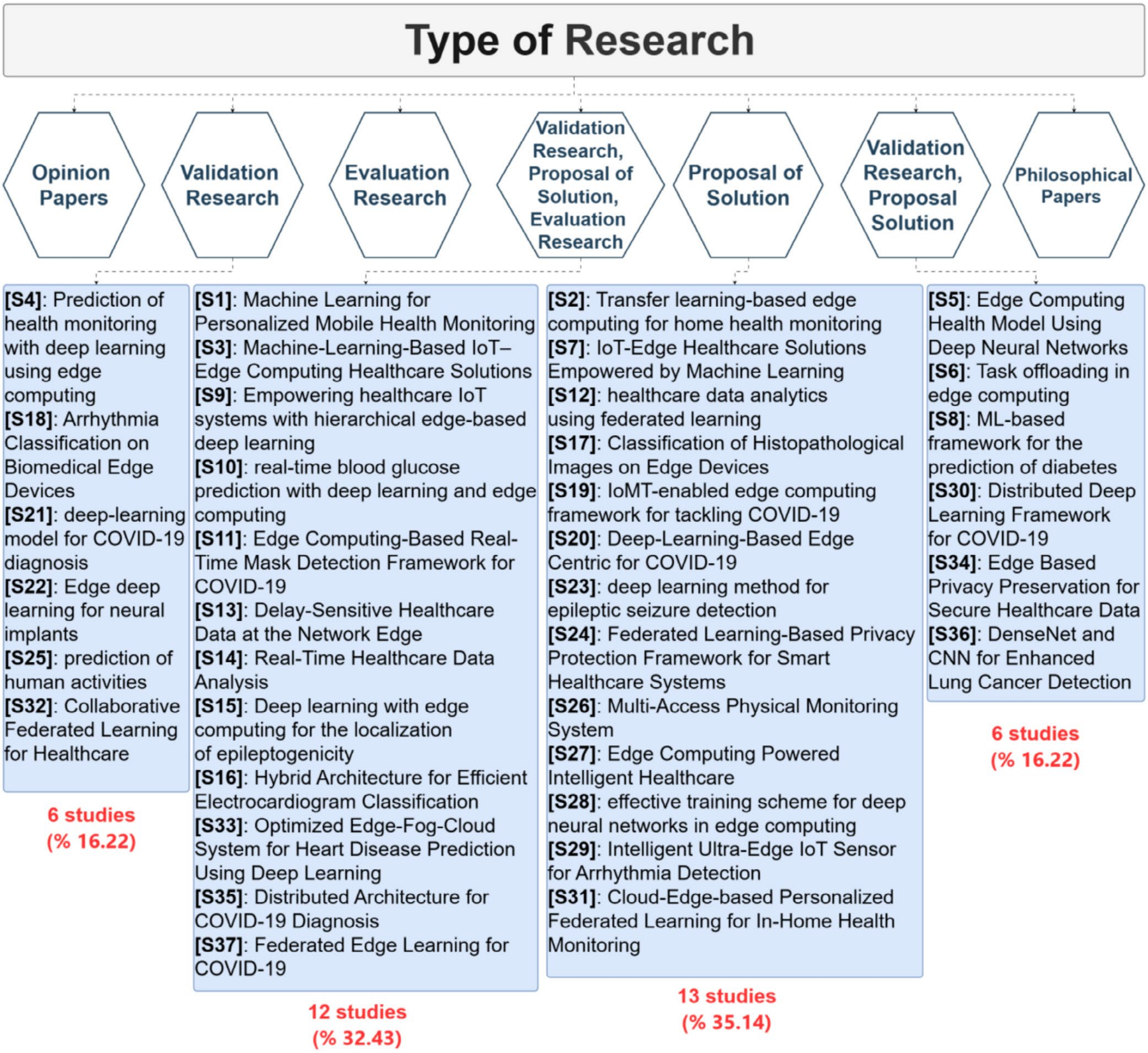

FIGURE 4 | Overview of types of research.

for processing. Other problems include the large volume of data overflow due to the high number of requests, user privacy issues, or the continuous data generation by IoT devices. Therefore, they proposed EdgeCNN, a hybrid architecture that can solve the problems of storing IoT data in the cloud platform.

The research of A. Kumar et al. (2021) is motivated by the limitations of histopathologic image classification (HIC) methods for breast cancer, especially those that rely on computationally intensive ML methods such as random forests and SVM. These models rely heavily on manually defined features, which makes them less compatible with modern applications. Researchers have recently investigated convolutional neural networks (CNNs) for automatic feature extraction. Still, due to their size and high computational cost, they are unsuitable for mobile or edge devices because the resources of these environments are limited. In edge computing, where cost-effective and efficient solutions are essential, these traditional deep learning models face significant challenges. To solve these issues A. Kumar et al. (2021) created a more efficient and lightweight CNN model using MobileNet, which is explicitly designed for edge devices. V. K. Singh and Kolekar (2022) stated in the motivation section of their research that getting tested for COVID-19 with methods like rRT-PCR can be slow, expensive, and hard to access, especially in rural or remote areas with limited medical resources. Plus, doctors manually checking chest CT scans for signs of COVID-19 takes a lot of time and can lead to mistakes. This is where AI and deep learning can help. These technologies can analyze medical data quickly and accurately, speeding up diagnosis. However, many AI models are complicated and need much computing power, which is not practical for real-time use or in places with limited resources. A hybrid edge-cloud system could solve this by reducing delays and making AI more accessible. Therefore, they set out to develop a simpler and faster deep learning model, based on MobileNet V2, to diagnose COVID-19 from chest CT scans. The model is small, efficient, and can analyze an image in just 43 ms.

TABLE 3 | Geographic distribution of authors.

Country name	Number of studies	Related studies
United States of America	10	S6, S7, S9, S15, S18, S22, S26, S30, S31, S33
China	9	S11, S14, S16, S27, S31, S33, S34, S36, S37
Saudi Arabia	7	S3, S12, S19, S20, S29, S30, S36
India	6	S2, S4, S17, S21, S26, S28
Japan	3	S1, S2, S13
Canada	3	S12, S18, S29
Qatar	3	S6, S23, S32
Australia	3	S8, S24, S26
South Korea	2	S5, S26
Malaysia	2	S6, S26
United Arab Emirates	2	S8, S36
Iran	2	S15, S35
Finland	1	S9
United Kingdom	1	S10
Norway	1	S25
Egypt	1	S26
Vietnam	1	S28
Singapore	1	S2
Bangladesh	1	S13
Belgium	1	S13
Russia	1	S28
Pakistan	1	S32
Ireland	1	S32
Thailand	1	S3

Security concerns are another motivation for research and development in healthcare and edge computing. Akter et al. (2022) conducted the study due to concerns about user data privacy in innovative healthcare systems that use the IoT. These systems often collect sensitive patient information, so there are many challenges in protecting these data. By using federated learning, a type of distributed learning, smart devices can process data locally and only share updated models with a central server, ensuring data privacy. However, despite these advantages, there are still problems with new cyber threats, such as privacy attacks. Akter et al. provided a new solution with a federated edge aggregator to solve these challenges and improve the performance and security of innovative healthcare systems.

Another research motivation is to overcome the limitations of cloud-based architectures in healthcare systems, especially when data synchronization or scalability are required. To address these problems, K. Wang et al. (2024) proposed a framework that uses deep learning and can take advantage of fog computing to reduce energy consumption and also improve communication bandwidth.

Key findings of RQ2.2

Finding 1: Security and privacy concerns were a primary focus in most studies, highlighting the need for solutions like federated learning and encryption to protect sensitive patient data in healthcare systems.
Finding 2: Edge devices have limited resources; thus, lightweight models are needed to run efficient healthcare applications.
Finding 3: Edge computing reduces latency and improves efficiency by processing data locally instead of relying on cloud systems.

5.3 | Common Tools, Frameworks, Software, and Technologies (RQ2.3)

RQ2.3 was designed to uncover the tools and frameworks utilized in supporting ML systems. While examining the selected studies, we observed that only a portion of the studies (5 studies) utilized their custom tools or developed frameworks (see Table A1 in Appendix A). We specifically considered executable frameworks. The source type classification denotes whether the tools are open source (OS) or closed source (CS). OS means copyright owners allow others to check, use, or change the tool, framework, or system. Figure 7 also shows the number of tools and frameworks. According to this figure, TensorFlow is one of the most popular tools.

Novel studies demonstrate an important gap in current Edge-AI deployment: the lack of integrated frameworks for managing the full lifecycle (e.g., remote updates, versioning, and monitoring) in real clinical practice. Full lifecycle management is a fundamental requirement for clinical reliability. This is also demonstrated by the fact that the IEC 62304 international standard has been established to ensure that medical device software is developed with a systematic, risk-driven approach that prioritizes safety and effectiveness (Jordan 2006). While popular runtimes like TensorFlow Lite, PyTorch Mobile, and OpenVINO support model compression and efficient on-device inference, they do not provide an integrated, end-to-end, regulatory-grade operational platform. Current edge runtimes (TensorFlow Lite, PyTorch Mobile, OpenVINO) are excellent at on-device inference and optimization, but none offer essential lifecycle capabilities such as secure over-the-air (OTA) model updates, signed and versioned deployments, continuous monitoring with drift detection, automated rollback, clinical-grade change control, or auditable model provenance. In practice, developers must assemble multiple additional components (model servers, device-management/OTA systems, monitoring pipelines, audit/logging infrastructure, and QMS processes) to achieve clinical reliability, which substantially increases engineering complexity

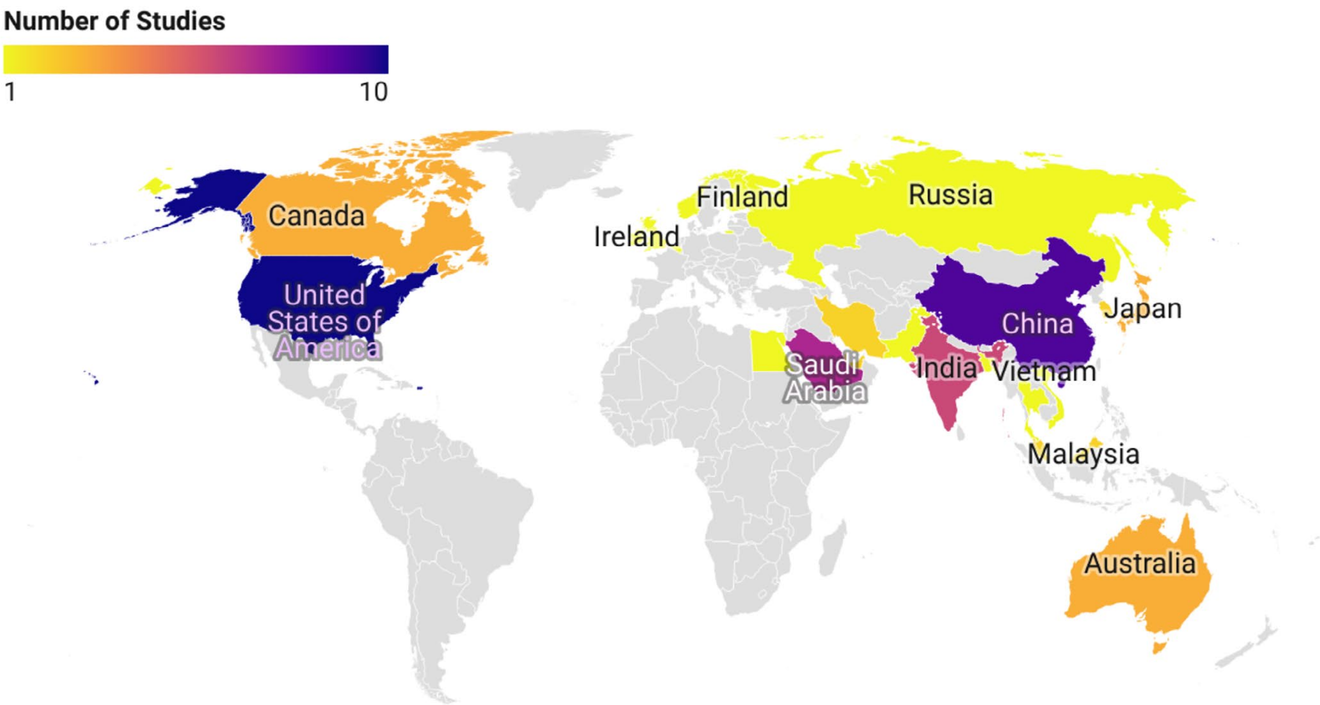

FIGURE 5 | Geographic distribution of authors.

and operational risk (Gentzen et al. 2025; Corbin et al. 2023; F. Wang and Beecy 2025; Chaturvedi et al. 2025; X. Wang, Tang, et al. 2025; Shivashankar et al. 2025). Research shows that when models are put into use without careful oversight throughout their lifecycle, they tend to become fragile and perform worse over time as real-world conditions change. This makes them harder to keep safe and reliable in healthcare settings (Elhanashi et al. 2024; Shivashankar et al. 2025). In addition, there are still unresolved engineering challenges around secure distributed version control and automated OTA update pipelines for edge medical devices (Gentzen et al. 2025). Research in TinyML also shows that lightweight runtimes currently lack built-in support for robustness testing or certification (Elhanashi et al. 2024), and other studies suggest that aggressive optimization techniques (such as quantization) can weaken system security (Ma et al. 2023). Together, these gaps highlight the need for integrated Edge-MLOps frameworks that can support safe, scalable, and continuously improving AI deployments in clinical practice.

Key findings of RQ2.3

Finding 1: We observed that the TensorFlow tool was used in nine studies (S10, S13, S16, S17, S19, S21, S22, S30, S32). This software has become widely used.
Finding 2: The data presented in Table A1 highlight the widespread adoption of open-source software, including well-known platforms/tools such as TensorFlow, PyTorch, and OpenVINO, within research studies.
Findings 3: For advancing Edge-AI runtimes, they must provide comprehensive lifecycle management capabilities, such as secure remote updates, robust version control, and continuous monitoring, to enable reliable and safe clinical deployment.

5.4 | What Types of Data Are Used by Machine Learning Applications? (RQ2.4)

This RQ aims to gather information about the datasets used in selected studies. Table A2 (see Appendix A) lists the datasets and the data type for each study.

Key findings of RQ2.4

Finding 1: Among the 37 studies included in this review, signal data emerged as the most commonly used data type, appearing in 20 studies. This category includes data from sources such as electrocardiograms (ECGs), electroencephalograms (EEGs), and other physiological signals. Image data were the second most frequently used, identified in 15 studies, while numerical data were employed in 13 studies. A smaller number of studies, specifically three, used video data. Studies utilized combinations of these data types in several cases, reflecting a trend toward multi-modal data integration to enhance diagnostic accuracy and system robustness. For example, Study S1 incorporated signal and numerical data, S2 used a combination of image and video, and S19 integrated signal, image, and numerical data. These combinations highlight the complexity and diversity of data sources used in edge-based healthcare applications.

5.5 | What Are the Performance Benchmarks of ML Models? (RQ2.5)

This RQ was designed to collect and analyze performance metrics reported in primary studies for ML models. The main focus in this section is on the metric "Accuracy" because it is the metric most widely reported among the selected studies. It should be noted that for this analysis, only studies that explicitly and clearly stated the accuracy number for their model were considered. Table 4 shows a summary of the performance (accuracy) of ML models in selected studies.

Key findings of RQ2.5

Finding 1: The accuracy of machine learning models reported in the reviewed studies is generally very high. Of the studies that reported accuracy, more than 60% achieved accuracy above 95%. This shows the great potential of integrating machine learning and edge computing for diagnostic and monitoring tasks in the health domain.
Finding 2: Models based on deep neural networks, especially convolutional neural networks (CNNs), are the most widely used and often most accurate machine learning architectures in this field. Many studies (e.g., S4, S9, S11, S14, S16, S17, S19, S20, S21, S23, S24, S29, S30, S36) have used different types of CNNs to achieve accuracy above 90% and even 99%.
Finding 3: Several studies have compared the performance of several models. For example, the S22 study compared the performance of DNN, CNN, and LSTM models on various seizure-related tasks and showed that LSTM and CNN consistently outperformed DNN. Such comparisons are invaluable to the research community in selecting the appropriate architecture for a particular application.

6 | Literature Review

6.1 | Studies Conducted in the Year 2017

Hosseini et al. (2017) proposed research to improve the detection and localization of epileptogenic zones within the human brain using the advanced autonomic edge computing methodology. The integrated platform for EEG and rs-fMRI features applies deep learning principles, extracting relevant features from both modalities. The edge computing approach allows for the processing and analysis of large-scale medical data in real-time, hence reducing latency and increasing data privacy by processing data locally using the edge gateway, which is supported by cloud servers for storage and intensive computation. The core contributions of this research include a general mobile edge cloud-based framework for real-time data processing, an optimized model for the localization of the epileptogenic zone, and an unsupervised feature extraction model combining CNNs and support vector machines.

6.2 | Studies Conducted in the Year 2018

Azimi et al. (2018) studied how to improve healthcare IoT systems using hierarchical edge-based deep learning. Their work focuses on overcoming challenges in remote health monitoring, especially the need for high availability and accuracy to care for patients in critical conditions. They proposed a hierarchical computing architecture integrating edge and cloud computing resources in such a way as to guarantee the optimization of the deep learning model deployment. The authors investigated, in particular, the possibility of deploying CNN-based classification models within this kind of architecture. This HiCH system, proposed in this paper, uses cloud servers to offload highly computationally intensive tasks and allows local decision-making at the edge, which grants high availability of a system not dependent on connectivity to the Internet. In this paper, a case study has been used to perform real-time health monitoring with the help of electrocardiogram data to show the efficacy of the approach. The evaluation was based on response time and accuracy. According to the results, a high level of availability and accuracy could be guaranteed by the HiCH architecture, which is very important in timely and reliable health monitoring.

Fadlullah et al. (2018) proposed an intelligent approach. Given the exponential growth of the IoT, there is a great need for efficient processing and analysis of health data. Conventional approaches transfer data from many biosensors to a central cloud for processing. The approach mentioned above suffers from huge latency due to network congestion and a large volume of data transmission. The authors presented a deep learning-based edge analytics approach to address the above challenge. This means that data will be processed closer to where it is generated or its origin, eliminating delays and increasing real-time capabilities for decision-making. Their contribution involves the application of CNNs at the edge in the analysis of healthcare data from sensors and wearables in patients' homes. This would be more useful in older patients and those requiring constant monitoring so that timely interventions could be instituted after analyzing critical health parameters like heart rate, blood sugar levels, and respiratory rates. Their simulations also show that their proposed method has extremely low loss rates, very high accuracy, and the least execution time.

Yu et al. (2018) propose a hybrid architecture, EdgeCNN, designed to improve real-time acquisition and analytics of healthcare data collected from IoT devices. One of the primary focuses is on creating a balance between edge computing and cloud computing to make the processing of healthcare-based data faster and more efficient. Subsequently, using EdgeCNN, it will be possible for the system to realize deep learning on edge devices. This, therefore, enables real-time evaluations, hence the advantages that accompany it, including lower delays, reduced network traffic, less burdening of cloud platforms, and lower maintenance costs. In this paper, the authors have presented a streamlined diagnosis model with an edge-based learning algorithm using CNNs that can identify and make real-time inferences of electrocardiograms. Experimental results prove that EdgeCNN, compared with the traditional cloud-based architecture, significantly reduces diagnosis delay, and network I/O, improves application usability, and shortens resource cost while enhancing the protection of user data privacy. EdgeCNN enables deep learning at edge devices, hence giving low latency for real-time diagnosis, consequently saving time and bringing efficiency and user privacy without relying on the cloud.

6.3 | Studies Conducted in the Year 2019

Ram et al. (2019) base their approach on challenges with an aging population and a shortage of healthcare workers by using cloud-based systems combined with ML at edge computing for

FIGURE 6 | A taxonomy of research themes on edge computing in healthcare using machine learning.

any mobile health monitoring and improving the accuracy in the tracking of human activities by using data from multiple sensors. The authors performed pre-processing to sanitize the dataset from sensor data and classify activities within it. They have used two ML algorithms, random forest, to compare them for the classification of activities. The results show that RF has an accuracy of approximately 99% compared to the 98% realized by SVM. For mismatched data analysis, they used a confusion matrix; this mismatching was mainly on the initial sensor values when the activities were being recorded. Further visualizations of the data were also presented to make it more understandable.

In 2019, Uddin (2019) proposed a different health monitoring approach using wearable sensors and edge computing. Their system uses RNNs to finally predict human activities from data originating from multiple sensors, which range from but are not limited to ECG, magnetometers, accelerometers, and gyroscopes. It is then processed by edge devices, PCs, or laptops and does not require cloud computing. It improves real-time processing by processing these vast amounts of healthcare data locally. Experiments proved that this system performed better than traditional methods on a publicly available dataset and turned out to have massive potential for continuous and inexpensive health monitoring without visiting hospitals. Besides, the employment of GPUs inside edge devices accelerates data processing, thereby rendering this system efficient and reliable for smart healthcare applications.

Manogaran et al. (2019) designed the Wearable IoT Smart-Log Patch system, incorporating an edge computing-based

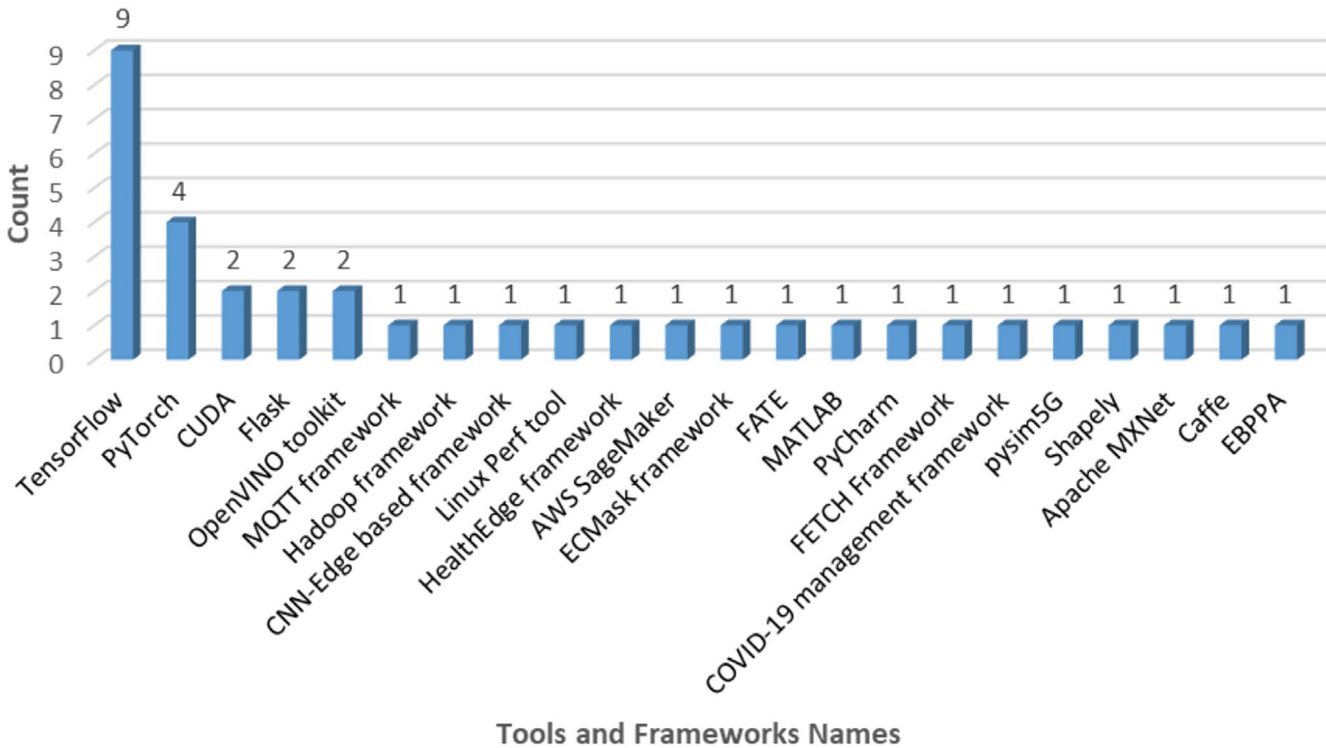

FIGURE 7 | Frequency analysis of tool and framework names.

Bayesian deep learning network for an improved physical Monitoring System. The state-of-the-art device gathers multi-dimensional physiological data like blood pressure, temperature, ECG, EMG, and EEG, among others, with the help of IoT sensors. Edge computing, along with the Bayesian deep learning network (EC-BDLN), is used to process that data and increase the accuracy and efficiency of health condition analysis. The high accuracy, mean residual error, delay, energy consumption, and overall strong solution in this study are supported by empirical evidence for real-time and continuous health monitoring. The research greatly supports the development of wearable health monitoring systems, leading to more accurate and constant health monitoring technology in the future.

6.4 | Studies Conducted in the Year 2020

Chang and Yoo (2020) have developed an approach where edge-computing with P2P is used to work for deep neural networks to improve health data processing. The model sends numerous issues, including neural net overfitting and high computational requirements, because multiple edge nodes are connected directly to modify deep neural networks; yet, the data processing load is distributed, and central servers are not overloaded. The measures of success gathered in their experiments showed an order-of-magnitude performance boost in response times compared to the corresponding server-based models they used to justify the model's role in real-time health monitoring and decision-making in IoT healthcare settings. This work offers a substantial contribution to health information processing in edge computing with the aid of P2P networking.

Hakak et al. (2020) proposed another approach to using edge computing and Federated Learning for improving healthcare analytics, where the data from the wearable devices update the local ML model, allowing the user's data to be private while reducing cloud resources required in the process. This framework provides personalized health insights and can be used for monitoring chronic diseases or for real-time health. The authors add challenges like data privacy, security, and edge device resource constraint issues that will show the great potential of this technology in healthcare.

He et al. (2020) proposed edge CNNs to enhance healthcare data processing from IoT devices. The system deals with noisy data, slow network speeds, and privacy concerns in remote health monitoring. Accordingly, a deep learning model working on edge devices for the rapid diagnosis of ECGs was proposed, along with a technique for augmenting ECG data through generative adversarial networks known as DCGAN. Their experiments prove that EdgeCNN works much better than the traditional cloud-based system in diagnostic accuracy, processing speed, and privacy protection.

Ali et al. (2020) suggested a system for epileptic seizure identification at the edge using deep learning and edge computing. In their proposal, the authors used a CNN to classify the brain wave data, which is a rather general approach. Most specifically, it decreases the amount of data sent after edge computing is done locally on the part of the data before it is transmitted to the cloud server, making the system faster and more efficient. Their approach gives an accuracy of 92% in the time domain and 99.22% in the frequency domain for seizure detection, while the system's complexity and response time remain low.

TABLE 4 | Performance of ML models.

Study ID	Task	Model	Accuracy
S1	Activity recognition	Random forest	~99%
S1	Activity recognition	SVM	~98%
S1	ECG classification	Random forest	~38%
S3	DDoS attack prediction	Hybrid voting classifier	99.7%
S4	Fall detection	Proposed CNN	99.23%
S7	Prediction of DDoS attacks	Hybrid machine learning model	99.7%
S8	Diabetes prediction (PIMA Indian Dataset)	Random forest (with feature selection)	78.27%
S8	Diabetes prediction (Sylhet Dataset)	Random forest	97.23%
S9	Arrhythmia detection	CNN and multilayer perceptron (MLP)	96%
S11	Detect and locate human faces in video frames (bus driver monitoring dataset with video restoration)	A custom anchor-based detector inspired by FaceBoxes architecture	97.98%
S11	Mask identification	MobileNet-V2 for binary classification (masked vs. nonmasked)	97.41%
S13	Edge analytics	Deep convolutional neural network (CNN)	94.44% (at batch size = 5, 150 epochs)
S14, S16	ECG classification	CNN	Normal: 87%, AF rhythm: 84%, other rhythm: 81%, noisy recordings: 60%
S17	Breast cancer classification	MobiHisNet (based on MobileNet)	91.21% (FP32, $\gamma = 0.5$, 200X)
			91.51% (FP16, $\gamma = 0.5$, 200X)
S18	Arrhythmia classification	Custom 1D convolutional neural network (1D-CNN)	DS1 (primary training set): 96.26%
S19	Cough sound detection	CNN	Training: 0.94, validation: 0.98
S19	Face mask detection	MobileNetV2 + CNN	Training: 0.98, validation: 0.97
S19	ECG classification	Custom CNN	~0.95 (from chart)
S19	Emotion recognition	CNN	~0.88 (from chart)
S20	Chest x-ray classification	ResNet101	97.5%
S20	CT scan classification	ResNet101	97.1%
S21	COVID-19 diagnosis	Fine-tuned MobileNet V2	96.40%
S22	Segment-based classification (10-fold CV)	DNN, CNN, LSTM	DNN: 64.55%
			CNN: 89.21%
			LSTM: 90.94%
S22	Event-based seizure detection	DNN, CNN, LSTM	DNN: 87.36%
			CNN: 96.70%
			LSTM: 97.61%
S22	Event-based seizure prediction	DNN, CNN, LSTM	DNN: 76.66%
			CNN: 90.66%
			LSTM: 90.72%

(Continues)

TABLE 4 | (Continued)

Study ID	Task	Model	Accuracy
S23	Seizure detection	A custom 1D convolutional neural network	99.22%
S24	Image classification	A two-layer sequential convolutional neural network	90%
S25	Activity prediction	Proposed RNN-based approach, hidden Markov model (HMM), deep belief network (DBN)	RNN: 99.69% HMM: 89.98% DBN: 92.01%
S26	Health monitoring	Edge computing-based Bayesian deep learning network (EC-BDLN)	~98%
S27	Heart disease diagnosis	GRU-MT-ABN (gated recurrent unit integrated multi-task learning with attention-based networks)	95%
S29	Arrhythmia classification	A 1-D convolutional neural network (CNN)	95.27%
S30	COVID-19 diagnosis	Distributed deep learning	98.17%
S31	Human activity recognition	Generative convolutional autoencoder (GCAE)	Balanced data: 95.87% Imbalanced data: 95.41%
S33	Heart disease prediction	Optimized cascaded convolutional neural network with hybrid PSO and GSO optimization (proposed), PSO-CCNN, DHOA-CCNN, WOA-CCNN, GSO-CCNN	Proposed: 87.4% PSO-CCNN: 82.5% DHOA-CCNN: 83.2% WOA-CCNN: 80.1%, GSO-CCNN: 84.2%
, S35	COVID-19 diagnosis	MobileCOVID-Net (proposed ensemble model), COVID-Net, MobileNet	MobileCOVID-Net: 93.8% COVID-Net: 92.4% MobileNet: 92%
S36	Lung cancer classification	DenseNet121 + CNN model	99.3%

Pustokhina et al. (2020) proposed a new method called ETS-DNN, which provisions for monitoring patient health from smart devices at the edge level. Their model collects health data from IoMT devices and analyzes them rapidly and precisely with advanced computing techniques. After that, they further optimized their DNN model with an optimization algorithm that was developed to make better predictions. They experimented with ETS-DNN; it worked much better compared to older methods, hence promising the improvement of healthcare services.

In their paper, Sakib et al. (2020) discuss adding smart features to IoT sensors for health monitoring. They focus on detecting heart arrhythmias using ECG signals with a new deep learning method. Their system is simple and accurate, achieving an accuracy of 95.27% without needing complex data processing steps. This method is better than older techniques like K-nearest neighbors and random forest, and it shows that using smart sensors can make health monitoring faster and more private by reducing the need for cloud computing.

Rahman et al. (2020) have investigated the potential of advanced 5G technology integrated into edge computing in fighting COVID-19. The authors suggest that the data be processed locally and, in the cloud, to achieve better diagnosis and management in the case of this nasty disease. It will have built-in fast, private, and scalable data handling, allowing the analysis of COVID-19 cases in real time. The authors have underlined the operationalization of deep learning models to make them more explainable and for experts to understand and trust those decisions. Their framework has performed well in several scenarios related to COVID-19, hence becoming one of the most useful tools in pandemic management.

Wu et al. (2020) propose FedHome, a system for in-home health monitoring that allows user data to be privacy-preserving by being stored locally. This uses a special balancing method, GCAE, to balance data and improve accuracy. FedHome thus handles data that are not evenly distributed and further cuts communication costs. It earned very high performance in human activity recognition with an accuracy of 95.41%, making it a strong option in smart healthcare while protecting privacy and improving efficiency.

6.5 | Studies Conducted in the Year 2021

Sufian et al. (2021) developed a solution for monitoring health at home with the help of smart technology. In this approach, researchers fine-tune pre-trained computer models with additional data for use on small devices at home. This approach ensures data privacy and security, reduces internet usage, and enables real-time monitoring. Their approach is constructive, especially during pandemics or in the care of older people, since monitoring their health can be both effortless and effective without transferring data outside the home.

Aazam et al. (2021) have explained how edge computing, when integrated with ML, creates value-added services in health care. The general problem they address is the inability of global devices, as simple as smartphones and smartwatches, to execute resource-intensive tasks because they lack the power and resources to perform them. They try to offload such tasks to other available nearby devices or systems, termed edge nodes, fog computing, or femto-clouds. The authors test this idea using ML algorithms, such as k-nearest neighbors, naive Bayes, and support vector classification on healthcare data, and the results turn out quite effective. This approach improves edge computing performance and enhances its capacity to support smart healthcare, notably COVID-19 symptom monitoring. A smart system on buses to monitor mask-wearing conditions in real-time has been developed by Kong et al. (2021). That is ECMask, applying innovative technologies such as edge computing and deep learning in processing video footage from bus cameras to improve video quality, detect faces, and identify whether one is wearing a mask. The team has tested it on various datasets, and this mask operates fast and accurately, making it very useful in keeping public transport safe during the COVID-19 pandemic.

A. Kumar et al. (2021) have developed MobiHisNet, an effective intelligent tool for analyzing medical images that could run seamlessly on mobile devices such as smartphones and Raspberry Pi. The authors knew conventional deep-learning models are too complicated and slow to run on such devices. Hence, MobiHisNet is lightweight and fast without compromising on accuracy. This model outperforms popular alternatives, such as VGG16 and ResNet50, especially in breast cancer diagnosis based on histopathological images. It is, hence, highly applicable and, therefore, will become very helpful to all in performing the most complicated medical image analyses on the go, enabling more powerful diagnostic tools for many more people.

The paper by Sakib et al. (2021) focuses on improving heart arrhythmia detection with deep learning. The traditional methods for its detection are intrinsically complex and, hence, hard to adopt on small medical devices. Sakib and his team took a much simpler approach with a one-dimensional CNN that simply skips over these complicated steps, hence becoming much easier to fit into those devices. The performance of this method was exceptionally great on four public datasets, showing that the work is quite applicable elsewhere. This work has also demonstrated that intelligent AI can be effectively performed on small medical devices with limited resources, improving the practicality and reliability of continuous health monitoring.

Rahman and Hossain (2021) have proposed a smart healthcare framework for health monitoring at home with IoMT and deep learning, accompanied by efficient strategies against COVID-19. The framework represents health data privacy and security by local processing at its source. He describes the designing and testing of the system, proving its practicality in real-life clinical trial scenarios. This solution is convenient during a pandemic, ensuring more reliable treatment from home. Muhammad and Hossain (2021) have developed an AI system to enhance COVID-19 screening and diagnosis using the latest 5G technology. It uses a mobile app to check vital signs like temperature and pulse. Further, medical images like x-rays and CT scans can be analyzed through it. The system processes data on-site with devices of higher power and provides privacy through blockchain to ensure quick and secure results. It is also adaptable to other diseases; it works to make healthcare more efficient and protects patient privacy.

Liu and Richardson (2021) of the University of Pittsburgh detail how their team plans to develop neural implants for real-time seizure detection and prediction in an IEEE Spectrum commentary. The group tried DNN, CNN, and LSTM models on a tiny microcontroller. Of the three, an LSTM model generally stood out: While it accurately picked up on a seizure 97.61% of the time and had the lowest rate of false alarms, it also tended to use more power. The CNN model struck a good balance, being accurate (96.70%) and efficient with memory. The DNN model was the quickest but was less accurate, identifying seizures 87.36% of the time with a higher rate of false alarms. Their work shows that deep learning in neural implants can significantly improve their effectiveness in treating epilepsy.

6.6 | Studies Conducted in the Year 2022

Zhu et al. (2022) developed a wearable device to help people with Type 1 diabetes track their blood sugar levels by predicting them in real-time. Their device uses an efficient deep learning model on a small and low-power chip. The Bluetooth-enabled wearable device could predict the blood sugar level and warn against hypoglycemia. They also developed a smartphone app to plot the trends in blood glucose and platforms to back up data and further improve the model. Their model was tested on 47 patients. The trials showed that the device helped reduce low blood sugar events and improved blood sugar control overall.

V. K. Singh and Kolekar (2022) provided a deep learning model to enhance the diagnosis of COVID-19 through chest CT scan images, optimized for mobile and edge devices, with an accuracy of 96.4% and achieved a faster diagnosis time. Their proposed system uses edge-cloud technology for real-time diagnosis. It helps remote/rural areas where access to quality healthcare infrastructure is poor and, hence, is one of the practical solutions for managing the pandemic.

Akter et al. (2022) designed a framework of a federated edge aggregator to protect healthcare data. The architecture processes patient information on IoT devices with additional privacy enhancement mechanisms before sharing, which lowers the risks of data breaches without affecting accuracy. This architecture demonstrated an accuracy of 90% on several benchmark datasets like MNIST, CIFAR10, STL10, and COVID-19 chest x-rays, outperforming traditional approaches. Qayyum et al. (2022) worked out a new technique utilizing edge computing and clustered federated learning to improve the diagnosis of COVID-19 using x-ray and ultrasound images. Their approach improved shortcomings such as security and privacy inherent in traditional cloud-centered healthcare systems. Its accuracy for diagnosis was enhanced by 16% based on the $F1$ score for x-rays and 11% for ultrasounds. This model further provides a way for safe data sharing between remote health centers without a breach of privacy.

6.7 | Studies Conducted in the Year 2023

Alnaim and Alwakeel (2023) provide the integration of ML, IoT, and edge computing in applications toward healthcare improvement. Edge computing will be used to handle huge volumes of data produced by medical sensors, improve real-time responses, and ensure the security of patients' data by using hard encryption. Their research shows that this approach may reduce time lags and enhance health system security and efficiency.

Gupta et al. (2023) developed a health monitoring system with deep learning and edge computing to process data quickly in 2023. In their case, IoTs collect and analyze health data with nearby servers. It attained a high accuracy of 99.23%. This is more effective and cost-efficient than other methods and works well for real-time fall detection.

The paper by Sistla and Konidena (2023) focuses on the issue of how to deal with large amounts of data coming from medical sensors with the help of ML to identify unusual patterns and reduce response time. Some of them include requesting the offloading of some of the operations to backend systems to extend the battery lifespan while at the same time enhancing the computational capacity of the appliances. The paper also highlights the need to improve opportunities for privacy and security as we successfully implement artificial intelligence and ML with cloud and edge computing for healthcare data and patient services.

Hennebelle et al. (2023) developed HealthEdge to predict Type 2 diabetes using data from medical sensors connected by IoT, edge, and cloud computing. According to the applied approach, a comparison between random forest and logistic regression ML algorithms is expected. RF showed an accuracy that was 6% better compared to LR in their study. HealthEdge looks to improve the prediction of diabetes and healthcare with advanced technology analysis of health data.

Chai and Guo (2024) proposed a new approach for processing cardiac data. They proposed a three-layer approach: edge, fog, and cloud computing, with deep learning using Particle Swarm Optimization (PSO) and Galactic Swarm Optimization (GSO) algorithms for optimization, which was more efficient than neural networks. Smart sensors gather patients' physiological data, and with the help of advanced feature extraction methods, critical information is captured from cardiac signals. PSO is used to enhance the feature selection stage, and GSO is used to tune the hyperparameters of the network architecture, all of which are then input into a cascade convolutional network (CCNN). The outcomes reveal that the planned architecture outperforms the previously available architectures in various aspects, such as reduced response time, energy usage, and bandwidth, and better performance, as measured in accuracy, recovery, and $F1$-score.

In a paper presented by Meng and Li (2023), a novel solution called Edge-Based Privacy Preserving Approach (EBPPA) is introduced to protect the security and privacy of medical data in smart health systems based on the IoMT. Homomorphic encryption, along with a robust XGBoost algorithm, allows the analysis of encrypted data without needing to decrypt it, which mitigates the risk of exposing sensitive information. Then, to protect any intermediate calculations, the operations are distributed between virtual edge nodes (VENs) that use a secret-sharing method to avoid passing the data to untrusted cloud servers. Such architecture results in lower latency, better resource management, and higher computational throughput. The experimental results illustrate that this model has less information leakage, lower communication overhead, and higher efficiency compared to existing approaches. EBPPA is a reliable approach to the security of patient data in a complex and high-risk cloud environment.

6.8 | Studies Conducted in the Year 2024

K. Wang et al.'s (2024) research tackles issues in cloud-based healthcare like data syncing and migration. They propose the FETCH framework, which uses edge computing and deep learning to make healthcare systems more efficient and scalable. Fog computing and the FogBus system reduce delays and power use while improving performance. This framework is helpful for real-time health monitoring and diagnosing heart disease, making healthcare faster and more accurate. Their work offers a better, more secure solution for modern healthcare needs.

A new method for COVID-19 diagnosis based on chest x-ray images was presented in a paper by Zamani and Sharifian (2024). In IoT-based e-health systems, this framework effectively integrates edge, fog, and cloud computing with deep learning based on ensemble methods to provide an accurate and efficient solution. In this system, the lightweight MobileNet model in the fog layer is responsible for initial diagnosis, and only when confidence is low is the image sent to the more accurate COVID-Net model in the cloud. These two models are used in the form of a hybrid structure called MobileCOVID-Net with an ensemble learning method to increase the accuracy of diagnosis. The results show that the proposed model performs significantly well and provides higher accuracy than similar systems.

In a study conducted by C. Zhang et al. (2024), a novel system was developed to detect lung cancer using CT images. This system combines CNNs and the DenseNet architecture. The system can improve the detection accuracy using edge computing and data fusion. Integrating data from different sources and fast processing at the system's edge enables real-time analysis of medical images. The input images are pre-processed in the analysis process, and then the CNNDenseNet hybrid model extracts and classifies essential features. With high accuracy, this model can detect three types of lung tissue, including healthy, benign, and malignant tumors. According to the experimental results, this method performed better than many previous models with an accuracy of 99%.

Huang et al. (2024) developed an innovative hybrid model that combines federated learning with edge computing to predict the clinical outcomes of COVID-19 patients during their hospital stay. Their model can provide accurate and timely results using 5G technology and the FATE framework. Also, it can prioritize patient privacy by securely managing sensitive medical data. By performing local processing at edge nodes and exchanging only training parameters with the central server, the model prevents the disclosure of sensitive information and improves its performance in heterogeneous data conditions. The results of this study show that the introduced model has higher accuracy than traditional methods, such as SVM and logistic regression, and shows high generalization ability even in diverse environments with incomplete data.

7 | Discussion

This systematic review has a structured approach that makes it different from many traditional reviews, which often provide overviews without following a clear methodological framework. Unlike studies such as Rancea et al. (2024) and Putra et al. (2024), which focus on specific aspects like privacy, federated learning, or resource management; this study presents a more comprehensive perspective by addressing multiple RQs. In addition to analyzing technological advancements, the geographical distribution of researchers and the most widely used development tools and frameworks in edge AI applications for healthcare are also examined. This multifaceted view allows for a multi-dimensional evaluation of previous research and offers a complete understanding of the field. The common finding across all studies is the growing importance of applying ML and edge computing to healthcare (Samarpita 2024).

A common trend observed across numerous review articles, such as Rancea et al. (2024) and Putra et al. (2024), as well as this study, is the considerable rise in edge computing research between 2020 and 2021. This increase aligns with the COVID-19 pandemic, which clarified the importance of AI-driven edge computing in healthcare. To be more prepared for similar global health crises that may happen in the future, the scientific community must continue developing this field. According to the studies reviewed, one of the most critical steps toward this goal is the establishment of standardized protocols for technology implementation, sensor integration, and software tools. Another common and essential issue discussed in most edge AI papers is the adoption of federated learning and pre-trained models to improve efficiency (Rancea et al. 2024; Sufian et al. 2020). These methods and blockchain technology (Rajagopal et al. 2024) enhance security and privacy by reducing data transmission needs, reducing the need for computing power on edge devices, and addressing the challenge of localized training data.

By enabling collaborative model updates without centralizing raw data, federated learning can be a promising approach for scalable, secure, and efficient AI deployment at the edge. This study categorizes previous research into architecture and methodologies, applications, security concerns, and optimization strategies. This categorization provides a better overview of this area and differentiates this study from other reviews, which often lack systematic segmentation. In addition, the most commonly used programming frameworks and ML libraries, such as TensorFlow and deep learning algorithms, are highlighted, which can lead to practical insights for researchers and developers. Also consistent with the findings of research such as Pereira et al. (2024), it should be stated that deep learning methods are the most widely used methods to obtain learning models in edge computing. While our systematic methodology has significant strengths, some limitations exist. Compared with studies such as S. M. Kumar and Majumder (2018) and Badidi (2023), which provide detailed analyses of IoMT applications and federated learning advantages, this study offers a more comprehensive but less specialized exploration of these topics. Additionally, some reviews such as Rocha et al. (2024) and Punugoti et al. (2023) employ real-world case studies, whereas our study primarily focuses on research trends and methodologies. In conclusion, this systematic review helps understand edge AI's evolving role in healthcare. This study also provides a good assessment of the field by trying to answer a broader set of RQs, identifying commonly used programming and learning tools, and emphasizing the importance of federated learning. Future research should focus on standardizing technology adoption, sensor integration, and system interoperability to ensure more adoption of AI-based healthcare at the edge. Our findings emphasize the need for continued advancements in security, efficiency, and collaborative learning approaches to enhance the effectiveness of edge AI in addressing real-world healthcare challenges. We reviewed previous studies to highlight their findings. Table 5 provides a summary of these studies and presents a comparison between related works and the findings of our review.

7.1 | Limitations and Challenges

Understanding the significant challenges during the implementation phase is essential for better planning, resource management, engaging stakeholders, and ensuring the overall success of projects across different fields. During the process of reviewing the selected articles, it was observed that researchers faced a set of challenges and limitations. In this section, we will examine the most critical challenges raised in the articles reviewed in this systematic review. As a result, Figure 8 provides a visual summary highlighting the key limitations and challenges identified in the reviewed studies.

Aazam et al. (2021) published a paper focusing on the chronic challenges of applying edge computing and ML in healthcare and emergency response situations. In this way, this paper addresses the challenges that occur in indoor and outdoor environments when we work with sensors and data. As an example, in indoor environments, you may have some constraints, such as being unable to access certain areas or the lack of sound or

TABLE 5 | Comparison of the present review with related systematic reviews.

Study	Main focus, contribution, and novelty	New findings
Rancea et al. (2024)	• A systematic review of edge computing and AI in healthcare • Classifying the literature into three key themes: privacy and security, AI-based optimization methods, and edge offloading techniques • Identifies research gaps and future directions for edge AI in healthcare	Edge computing offers the potential to substantially decrease latency, strengthen data security, and boost system responsiveness, all of which are vital for real-time applications in various healthcare environments The systematic review of 72 articles revealed that the primary research focus is on data privacy and security (39% of papers), followed by AI optimization techniques (35%) and edge offloading (26%)
Putra et al. (2024)	• A review of wearable IoMT for health monitoring using a Cloud-Edge AI approach • Exploring cloud-edge AI and edge federated learning to improve IoMT diagnostics and security • Introducing a novel federated cloud-edge AI blueprint for IoMT	The integration of Edge Federated Learning within a Cloud-Edge AI architecture is identified as a transformative solution for enhancing the security, efficiency, and diagnostic quality of wearable IoMT systems
Samarpita (2024)	• Examining edge intelligence and its role in smart healthcare • Synthesizing the architecture, advantages, and application scenarios of edge intelligence for healthcare • Proposing edge intelligence as a crucial framework to enhance smart healthcare systems	The integration of edge computing and artificial intelligence into edge intelligence creates a powerful framework that enhances smart healthcare by enabling real-time processing, reducing latency, and improving health monitoring and predictive analysis
Sufian et al. (2020)	• A survey on using Deep Transfer Learning and Edge Computing to mitigate the COVID-19 pandemic • Systematically exploring the potential and challenges of deep learning, deep transfer learning, and edge computing for pandemic response • Proposing a novel pipeline architecture for Deep Transfer Learning over Edge Computing (DTL-EC) as a future framework	The integration of Deep Transfer Learning and Edge Computing is a promising approach to overcoming data and computational limitations, offering an effective strategy for mitigating the COVID-19 pandemic
S. M. Kumar and Majumder (2018)	• Integrating cloud/edge computing and machine learning into a distributed IoT framework for healthcare • Exploring the use of backend machine learning • Proposing a hybrid intelligence model where the backend server learns data signatures to advise resource-constrained edge devices	Edge computing with cloud-based machine learning guidance enables real-time, intelligent healthcare monitoring by prioritizing critical data at the source
Badidi (2023)	• A review of the role of Edge AI in the early detection of chronic and infectious diseases • Exploring the synergy between AI and edge computing • Providing a holistic perspective on Edge AI in healthcare	Edge AI can transform early disease detection through real-time analysis and federated learning, but requires overcoming key challenges like privacy and data quality
Rajagopal et al. (2024)	• A survey on the integration of blockchain and federated learning for IoT-based healthcare in edge-fog-cloud systems • Identifying and analyzing research gaps in blockchain-based federated learning for medical IoT applications	The integration of blockchain and federated learning is a promising solution for enhancing security, privacy, and efficiency in smart healthcare applications

(Continues)

TABLE 5 | (Continued)

Study	Main focus, contribution, and novelty	New findings
Pereira et al. (2024)	• A systematic mapping of the literature on machine learning applied to edge computing and wearable devices for healthcare • Exploring and categorizing the current trends, techniques, and architectures	Over 32% of the studied applications focused on fall detection for elderly monitoring Neural networks, especially Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs), are the most frequently used machine learning algorithms Raspberry Pi and smartphones are the most common edge computing platforms in wearable health monitoring systems that use machine learning
Rocha et al. (2024)	• A literature review on the application of Edge AI and MLOps for the Internet of Medical Things (IoMT) • Proposing a novel Edge AI and MLOps architecture for heart anomaly detection	Edge AI combined with MLOps is a promising solution for developing efficient, secure, and real-time IoMT applications
Punugoti et al. (2023)	• A comprehensive survey of the impact of Edge Intelligence and IoT on healthcare systems • Exploring the applications, benefits, and challenges of integrating Edge Computing with IoT for healthcare systems	With a significant adoption rate of 32%, healthcare IoT emerges as the third most prevalent application among major IoT domains
Present review (2026)	• A systematic literature review on the integration of edge computing and machine learning in healthcare • Exploring the key benefits, challenges, research trends, tools, and data sources in this field by analyzing 37 primary studies	This paper presents more than 10 key and novel findings. Please refer to Sections 4 and 5, where multiple findings are provided for each research question

motion signals during emergencies, so with these limitations, detecting situations accurately will be hard. In contrast, in outdoor environments, we need to rely on sources like simple CCTV cameras, and we are still sensitive to uncertain conditions, like not being able to sense consciousness, where we fail to scan for proper emergencies on time. These drawbacks underscore the need for advancements in sensor technology and ML algorithms to enhance emergency identification.

In the research challenges section of this paper (Hakak et al. 2020), several key issues are highlighted, and they emphasize that their framework can address the challenges mentioned. Cyberattacks are one of the challenges that have increased markedly during the COVID-19 pandemic. Such attacks can be modeled as poisoning attacks and other threats to the system's security. The second challenge is data bias. Another challenge is the nature of edge devices. Edge devices, due to their constraints on resources such as storage and computational power, may face challenges in terms of computation and data storage. Finally, one of the most significant challenges is the user acceptance of the technology. For the system to function, users' data updates are required regularly, which requires their active participation and willingness to adopt the technology. The paper by Fadlullah et al. (2018) explains that the cloud-based analytics system cannot deliver near-real-time analytics for latency-sensitive healthcare data because of issues like network congestion in the IoT data delivery layer and computational overload in the analytics layer. This can make it challenging to respond rapidly to health risks for patients at home or the elderly who require rapid analysis. To solve this problem, the paper proposes IoT edge analytics with deep learning methods to make near-real-time analytics with high accuracy and a low loss rate.

Hosseini et al. (2017) identify the main challenges in developing automated systems for diagnosing neurological disorders: First, it is crucial to accurately identify different brain states, such as seizures and pre-seizures. Secondly, patient data must be used appropriately to improve the efficacy of the systems (learning from patient data). Transferring expert knowledge to automated systems is a challenge, as is specifying the mechanisms for therapy activation, such as RNS, which is also a challenge. Finally, there are challenges regarding the safety and stability of the systems for patient care and the processing of large amounts of data. In the paper of A. Kumar et al. (2021), developing models for edge devices has a primary challenge: the constraint of resources, for example, memory and processing. Because of these constraints, large parameter sizes of complex models are not feasible for resource-constrained devices. As a result, it is necessary to reduce inference time, memory consumption, and RAM consumption and make models efficient. They used pruning (deleting unnecessary connections) and quantization (lowering the precision of parameter representations) to tackle this challenge. These methodologies allow

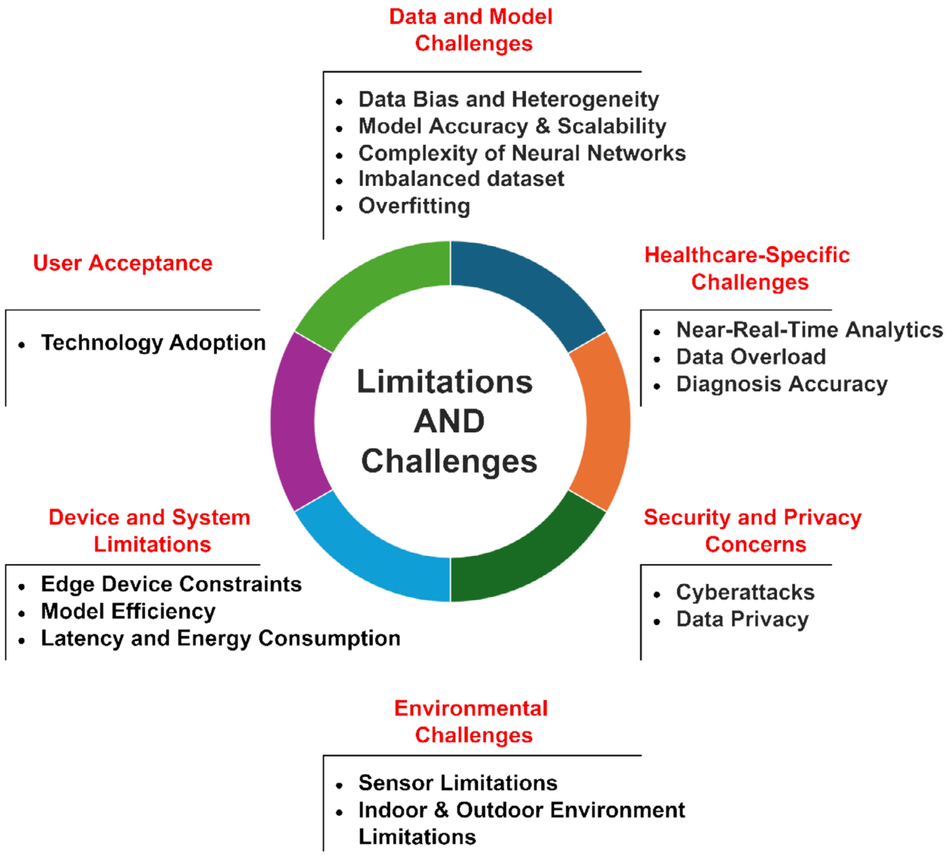

FIGURE 8 | Illustration of limitations and challenges.

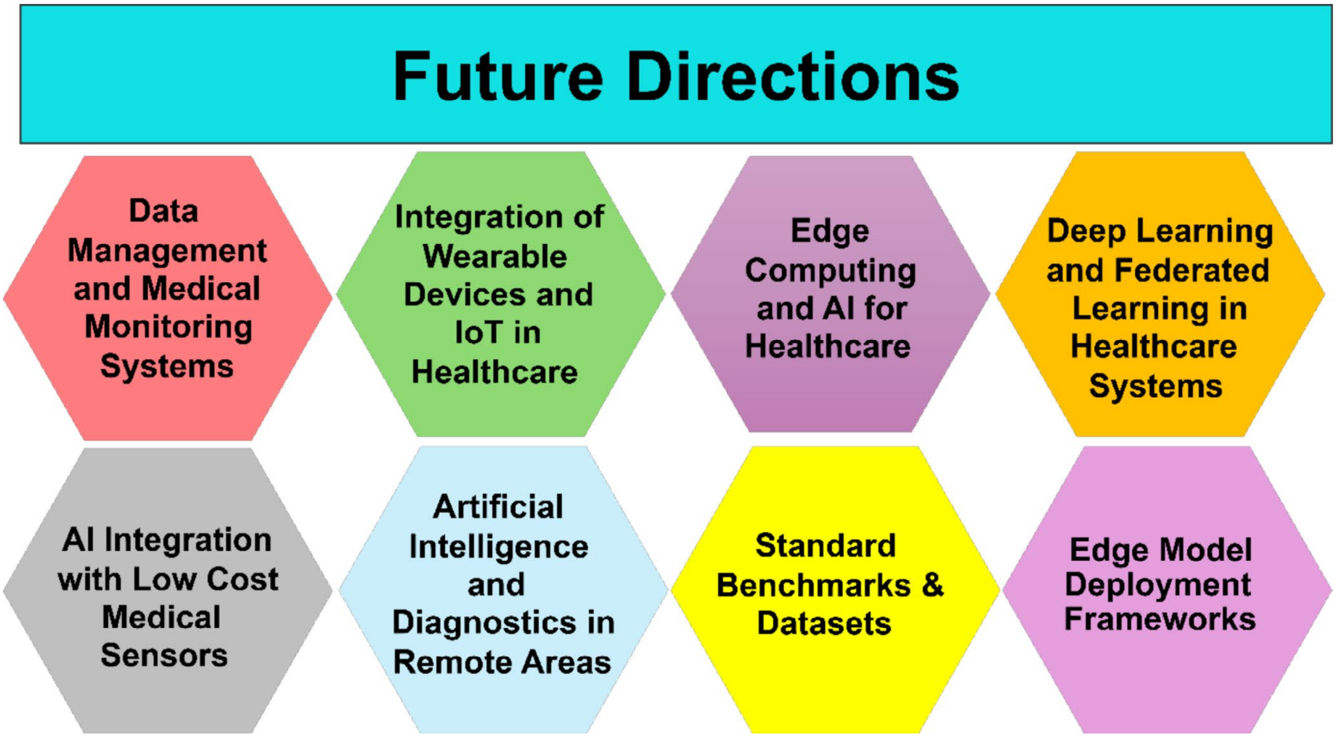

FIGURE 9 | Illustration of future directions.

for a decrease in the model size and increase the model performance to generate lower complexity, higher performance models for execution on edge devices.

Ali et al. (2020) in the methodology section of their paper stated that one of the challenges is determining the depth of a CNN in a way that produces both high accuracy and low computational complexity. Qayyum et al. (2022) systematized the pains of using ML over the edge into six categories: resource constraints, data heterogeneity, security and privacy, network instability, model accuracy, and scalability. Resource constraints refer to edge devices' limited processing power and storage, making it hard to deploy complex models. The devices collect a broad range of data that exhibit heterogeneity, making it challenging to integrate and process it for model training. Security and privacy concerns are also significant, as data breaches or attacks on sensitive information can open the door to serious inconveniences or not disastrous consequences. Poor networks lead to delays or interruptions in communication between devices and servers and are undesirable, especially for real-time applications. Due to these limited resources, model accuracy is more difficult to maintain and has consequences. Poor networks lead to delays or interruptions in communication between devices and servers and are undesirable, especially for real-time applications. Due to these limited resources, the model's accuracy is more challenging. Chai and Guo (2024) worked on a paper discussing the challenges of high latency and energy consumption in cloud computing. To solve these problems, they introduced an edge-fog-cloud model, which processes data closer to where it is created. This approach can reduce both latency and energy. They also faced issues handling large amounts of medical data and limited bandwidth. To improve this, they used edge and fog computing, which processes data near the source, making it more efficient and reducing the pressure on bandwidth.

Two significant challenges in Internet-based healthcare systems were identified in a paper by Meng and Li (2023). The first challenge comes from medical data security and cyberattacks around the hospitals, where the data will be compromised from such a breach. The second challenge is maintaining data privacy while executing computations on data held in untrusted cloud servers. To tackle these issues, a novel edge-based privacy-preserving approach (EBPPA) is proposed, which combines data protection with homomorphic encryption and the XGBoost algorithm. C. Zhang et al. (2024) highlighted an issue with deep learning models for CT image-based lung cancer diagnosis, which was that they were not accurate or interpretable. As a solution, they applied CNN and DenseNet to solve this problem, giving them greater accuracy in detection and fewer issues with steroids from data variety and model interpretability. In conclusion, model accuracy and interpretability are categorized as one of the main limitations and challenges. Hence, we need advanced ML models to improve precision and make the model's results understandable.

It is also necessary to highlight several other challenges. One of these challenges is the imbalanced dataset problem (S8). To address the data imbalance challenge, Wu et al. (S32) developed a Generative Convolutional Autoencoder (GCAE), aiming to provide accurate and personalized health monitoring by enhancing the model with a generated, class-balanced dataset derived from the user's personal data. Another challenge is overfitting. For instance, Gupta (S4) reported addressing this issue by using a pooling layer, which is inserted after the convolutional layer to reduce the size of the data and computational complexity of the images. Additionally, Studies S10, S18, S23, and S30 employed dropout layers to prevent overfitting. In Study S22, max pooling was used as a strategy to mitigate overfitting. Another challenge is personalizing models for multidimensional data, as discussed in Study S33. This approach helps improve the performance of both server and client models simultaneously.

We can conclude that this field presents some challenges that are common to other applicative fields and others that are specific. Challenges related to data and models are the same as in most AI-based systems, as well as security and privacy concerns that are common to most critical infrastructures, and device and system limitations are common to any application adopting the Edge-AI paradigm. For them, it is possible to adopt state-of-the-art solutions. On the contrary, user acceptance and the healthcare-specific challenges represent the key fields where research efforts should focus. At present, we can state that the field is not sufficiently mature because these limitations affect reliability, regulatory compliance, and trust in medical settings. In our opinion, one of the key challenges is that most methods have been tested on small or controlled datasets rather than in real-world situations, which raises doubts about their generalizability.

7.2 | Future Directions

Identifying potential future directions provides a comprehensive view of current advancements and possibilities in edge computing research within healthcare. Figure 9 presents a visual summary highlighting the key foundational issues of future directions identified during this study.

7.2.1 | Edge Computing and AI for Healthcare

Several papers highlight the potential of combining AI and edge computing in healthcare systems. Ram et al. (2019) note that their study is still in the early stages, and future work will concentrate on developing a framework for data semantics in intelligent analytics at the edge. A. Kumar et al. (2021) propose enhancing algorithm efficiency on edge devices and tackling problems related to multiclass classification in histopathological images. Similarly, the researchers behind (Gupta et al. 2023) plan to implement their proposed method on large datasets and experiment with different deep learning models to boost system accuracy and reduce classification costs. Additionally, the use of mobile edge computing offloading frameworks to manage large medical datasets is also expected to be explored further in Gupta et al. (2023). Incorporating federated learning and parallel processing on edge devices to improve system speed and efficiency is another key focus in papers Rahman and Hossain (2021) and Muhammad and Hossain (2021). Rahman and Hossain (2021) say they plan to make their apps more accurate. Once they are happy with the results, they will start clinical trials. They also want to use federated learning and end-to-end encryption to boost data security and improve the system (Muhammad and Hossain 2021). In the future, they will focus on making mobile deep learning models smaller and using parallel processing on edge devices to speed things up and make the system more effective.

7.2.2 | Integration of Wearable Devices and IoT in Healthcare

The fact that IoT and wearable devices are linked is driving future research in healthcare. In a recent publication, Zhu et al. (2022) suggest real-world clinical trials for evaluating wearable devices and propose the idea of performance improvement based on user feedback. They emphasize the importance of collaboration with IoMT device manufacturers to incorporate predictive algorithms into wearables. This direction will be critical in diabetes management, aiming at real-time data processing through edge computing. On the other hand, in the section emerging trends and future directions of C. Zhang et al. (2024), further research on the topic of lung cancer diagnosis by deep learning is expected to be focused on three different issues: ensuring the models can deliver good performance across different populations, utilizing different types of data to get a more accurate diagnosis, and helping doctors to understand how AI systems perform. Furthermore, C. Zhang et al. (2024) argue AI is becoming more and more contributing to patient care through disease management in an individual way, prognosis of patient recovery, and the decision of the surgical operation in real-time, which can also be done with the help of wearable devices and IoT.

7.2.3 | Data Management and Medical Monitoring Systems

Many researchers propose innovations in data management and healthcare monitoring systems. Sufian et al. (2021) suggest conducting pilot studies, creating datasets, and evaluating system performance through simulations. They also recommend combining home health monitoring systems with additional environmental sensors for enhanced performance. Hennebelle et al. (Hennebelle et al. 2023) want to enlarge the scope of their study, including the entire area of ML and deep learning algorithms possible for their HealthEdge system. Additionally, Yu et al. (2018) encourage open, deployable platforms on edge devices to create easy access to home medical monitoring equipment, thus enabling a comprehensive healthcare ecosystem. Correspondingly, K. Wang et al. (2024) cover the automation of diagnosis and treatment using AI, significantly improving diagnosis speed and accuracy, and integrating fog systems with other healthcare platforms for better resource allocation and system efficiency enhancement.

7.2.4 | Deep Learning and Federated Learning in Healthcare Systems

The progression of deep learning models in the healthcare sector is a common focus in many works. He et al. (2020) aspire to construct better ECG feature-diagnostic models for heart diseases and suggest stronger cooperation with the hospitals to resolve concerns like data imbalance. Rahman et al. (2020) look at applying deep learning techniques in clinical exercises and reframe the problem toward enhancing detection systems for adversarial attacks and constructing lightweight models suitable for hospital use. Similarly, Qayyum et al. (2022) propose using collaborative federated learning (CFL) for that purpose, enabling data sharing and model training across various institutions in a secure and privacy-preserving manner.

7.2.5 | Artificial Intelligence and Diagnostics in Remote Areas

Another direction for the future is AI's potential to improve processes in underserved regions (Pustokhina et al. 2020). Implementing their ETS-DNN model for disease monitoring and diagnosis in remote areas will enable the technology to expand the use of healthcare in remote areas (Akter et al. 2022). At the same time, it envisions the creation of microservices-based Federated Edge Aggregators that will enhance health and disease monitoring (this would make it more flexible to manage healthcare data).

7.2.6 | AI Integration With Low-Cost Medical Sensors

Another critical area for future research is the integration of AI with low-cost medical sensors. Sakib et al. (2021) mentioned that they will focus on AI for instrumentation, which could make low-cost medical sensors better performing and enable low-cost medical testing at scales suitable for mass production. In conclusion, federated learning and parallel processing on edge devices will be widely used to enhance healthcare systems' efficiency. Also, one possible research direction is to establish a remote patient monitoring system in which patients do not have to go to the hospital, and the doctor can use this system to manage the patient's status remotely.

7.2.7 | Standard Benchmarks and Datasets

One of the main problems in reviewing and comparing ML models in edge health environments is the absence of benchmarks and standards, as well as publicly available datasets. As shown in Table A2 and the answer to the RQ2.4 question, the data used in different studies are very different. Researchers usually use private or special data in their field, which makes it difficult to accurately compare the ability to generalize and the efficiency of models. This dispersion makes collective progress slow, and it becomes hard to evaluate new models under the same conditions. Therefore, future research should focus on creating and using a broad standard and multimedia dataset that reflects the real conditions of edge computing environments. These benchmarks should include various medical scenarios (e.g., heart monitoring, diagnosis of neurological disorders, or respiratory diseases) and, in addition to raw data, standard evaluation criteria such as processing time, energy consumption for each forecast, model volume, and communication cost. Developing such benchmarks will help innovate, reproduce research, and quickly test samples into reliable clinical solutions.

7.2.8 | Edge-Model Deployment Frameworks

Although there are tools and frameworks like TensorFlow Lite and PyTorch (see RQ2.3 and Figure 7), there is still a great need for integrated and comprehensive frameworks to deploy, manage, and operate ML models in edge health environments. The existing tools focus more on the initial optimization of the model and its conversion, and do not cover the full management of the model life cycle. Future research must create frameworks that automate and synchronize the entire process from cloud to edge devices. These frameworks should be able to update remote models and manage versions so that devices always use the latest valid models, without the need for manual intervention. Also, there is a need for function monitoring mechanisms and identifying the model behavior change in real time to activate the model's retraining process if the prediction quality is reduced. Given the importance of patients' health, these frameworks should also have the ability to explain the model security, efficiency in federated learning, and resistance to network interruption and connectivity. Creating OS and standard frameworks can make it easier to enter this field, reduce executive errors, and increase trust in AI solutions in edge environments.

7.3 | Implications for Patient Safety

Edge-based AI in healthcare has an impact on patient safety (Bates et al. 2021). This is because mistakes in diagnosis or monitoring can delay treatment or lead to wrong clinical choices. Recent research shows that on-device ML can improve accuracy and speed. However, most tests have taken place in labs, not in real clinical settings (Choudhury and Asan 2020; Rancea et al. 2024). As a result, these systems might not work as well in actual use. This can happen due to changes in data, sensor problems, limited hardware, unstable connections, or differences in how people work. All these issues can cause unsafe results (Channar et al. 2025). Also, edge-AI models can still fall victim to attacks and data tampering. These can change predictions and cause wrong diagnoses (Finlayson et al. 2019). Tests have shown that these attacks can trick medical imaging systems and tools that watch body signals. This poses big risks to patient safety if not dealt with. What is more, many current AI systems work like black boxes. This makes it hard to see how they work. It also makes it tough for doctors to understand or check decisions. This hurts trust and makes it harder to spot errors in places where safety is key (Bortsova et al. 2021; Houssein et al. 2025).

Healthcare systems that use edge computing include many connected parts, such as sensors, small devices that process data, communication networks, and cloud support. Because of this, keeping patients safe depends not only on the accuracy of AI models but also on the strength, security, and reliability of the whole system (Rancea et al. 2024). Recent studies show that bringing AI into medical work can create new safety problems, such as bias in decisions or failures that no one can predict. This means constant monitoring and clear rules for using these systems (Ferrara et al. 2024).

To keep patients safe, edge-AI deployments need extra layers that check performance in real time, for example, explainable-AI (Chaddad et al. 2023) and protection against attacks that attempt to fool the system. Standard testing and certification processes, similar to medical-device regulation, also help ensure reliable operation in real clinical settings.

7.4 | Research-to-Practice Gap

The gap between research and real-world deployment of Edge-AI is mainly due to the technical and infrastructural limitations. For instance, edge devices like wearables and IoT sensors have limited processing power, memory, and energy, which makes them incapable of running complex AI models in real time (R. Singh and Gill 2023). Besides, the complexity of operations and scalability requirements for clinical environments make it difficult to take these technologies off the lab bench and put them into practice (T. Wang, Guo, et al. 2025). Additionally, low-quality data, such as heterogeneous and incomplete data generated at the edge, leads to a decrease in model performance and, therefore, the limitation of models' usability in uncontrolled research settings (Belgoumri et al. 2024). Beyond technical issues, organizational and ethical concerns make it difficult for Edge-AI solutions to move from research labs into everyday clinical practice (Weiner et al. 2025).

While the reviewed studies have shown high performance, even at above 95% accuracy in experimental settings (e.g., S4, S9, S20, S21, S23, S36), the majority of edge-based AI systems have not reached a stage where they can be considered for practical deployment in clinical settings, despite efforts. This gap highlights a big divide between research and practice, where promising lab results don't make it into real medical settings.

8 | Conclusion

This paper provides a systematic review to gather, evaluate, critique, and synthesize previous publications on edge computing and ML in healthcare. This study addressed the pros, present challenges, and future technological trends in healthcare systems. By reviewing the selected articles, eight main RQs were raised in this study across three categories. These questions included identifying the most common research topics, the tools and frameworks, the challenges and limitations, and the types of data utilized in this field. The results indicate that recent research, particularly between 2020 and 2024, has focused on improving the efficiency of ML methods. This research includes designing and implementing optimized ML models for edge devices and adopting federated learning, transfer learning, and blockchain technologies.

Furthermore, these technologies face substantial obstacles due to hardware resource constraints in edge devices and ongoing security problems. Moreover, an effort was made to assist researchers in this field by extracting the software and data used in the selected articles. This allows for identifying the most efficient development packages, libraries, and commonly used datasets based on past research. Future studies should focus on optimizing ML models for resource-constrained devices, developing efficient methods to ensure data privacy, and creating new frameworks for managing complex and diverse data simultaneously with local processing. Moreover, the researchers should pay extra attention to federated learning and data encryption as two common ways to ensure security and privacy in healthcare edge computing. Finally, this research provides evidence of the widespread application of edge computing and ML in the future, which can lead to increased efficiency, diagnostic accuracy, and cost reduction in healthcare systems.

Author Contributions

Amir Mashmool: conceptualization (lead), data curation (lead), formal analysis (lead), funding acquisition (lead), investigation (equal),

methodology (equal), project administration (lead), resources (lead), software (lead), validation (equal), visualization (lead), writing – original draft (equal), writing – review and editing (equal). **Giorgio Delzanno:** methodology (equal), supervision (equal), writing – review and editing (equal). **Hamid Saadatfar:** writing – original draft (equal). **Aakash Ahmad:** methodology (equal), writing – review and editing (equal). **Rainer Koschke:** supervision (equal), writing – review and editing (equal). **Roohallah Alizadehsani:** writing – original draft (equal). **U. Rajendra Acharya:** supervision (equal), writing – review and editing (equal). **Daniele D'Agostino:** investigation (equal), methodology (equal), supervision (equal), writing – original draft (equal).

Acknowledgments

During the preparation of this work, the authors used Grammarly and ChatGPT to check grammar and improve text clarity. Afterward, they carefully reviewed and edited the content and took full responsibility for the final manuscript. Open Access funding enabled and organized by Projekt DEAL.

Funding

This work was supported by the Universität Bremen.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing does not apply to this article as no new data were created or analyzed in this study.

Related WIREs Articles

Remote patient monitoring using artificial intelligence: Current state, applications, and challenges

Endnotes

[1] http://www.acm.org/about/class/1998/.

[2] http://arxiv.org/corr/home.

[3] https://www.scimagojr.com/.

[4] https://portal.core.edu.au/conf-ranks/.

References

Aazam, M., S. Zeadally, and E. F. Flushing. 2021. "Task Offloading in Edge Computing for Machine Learning-Based Smart Healthcare." *Computer Networks* 191: 108019.

Adeniyi, E. A., R. O. Ogundokun, and J. B. Awotunde. 2021. "IoMT-Based Wearable Body Sensors Network Healthcare Monitoring System." In *IoT in Healthcare and Ambient Assisted Living*. G. Marques, A. K. Bhoi, V. H. C. de Albuquerque, and K. S. Hareesha. (eds) Studies in Computational Intelligence, vol 933, 103–121. Springer.

Ahmad, A., and M. A. Babar. 2016. "Software Architectures for Robotic Systems: A Systematic Mapping Study." *Journal of Systems and Software* 122: 16–39.

Akter, M., N. Moustafa, T. Lynar, and I. Razzak. 2022. "Edge Intelligence: Federated Learning-Based Privacy Protection Framework for Smart Healthcare Systems." *IEEE Journal of Biomedical and Health Informatics* 26, no. 12: 5805–5816.

Alamouti, S. 2025. "Quantifying Energy and Cost Benefits of Hybrid Edge Cloud: Analysis of Traditional and Agentic Workloads." arXiv Preprint arXiv:2501.14823.

Ali, Z. S., N. Subramanian, and A. Erbad. 2020. "Smart Health Monitoring for Seizure Detection Using Mobile Edge Computing." In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, 1903–1908. IEEE.

Alnaim, A. K., and A. M. Alwakeel. 2023. "Machine-Learning-Based IoT–Edge Computing Healthcare Solutions." *Electronics* 12, no. 4: 1027.

Ampatzoglou, A., S. Bibi, P. Avgeriou, M. Verbeek, and A. Chatzigeorgiou. 2019. "Identifying, Categorizing and Mitigating Threats to Validity in Software Engineering Secondary Studies." *Information and Software Technology* 106: 201–230.

Anand, A., S. Goel, and G. S. Panesar. 2024. "Energy-Efficient Edge Computing Architectures for AI Workloads: A Comparative Analysis in Cloud-Driven Environments." In *2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP)*, 614–620. IEEE.

Azimi, I., J. Takalo-Mattila, A. Anzanpour, A. M. Rahmani, J.-P. Soininen, and P. Liljeberg. 2018. "Empowering Healthcare IoT Systems With Hierarchical Edge-Based Deep Learning." In *Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies*, 63–68. ACM.

Badidi, E. 2023. "Edge AI for Early Detection of Chronic Diseases and the Spread of Infectious Diseases: Opportunities, Challenges, and Future Directions." *Future Internet* 15, no. 11: 370.

Bates, D. W., D. Levine, A. Syrowatka, et al. 2021. "The Potential of Artificial Intelligence to Improve Patient Safety: A Scoping Review." *Npj Digital Medicine* 4, no. 1: 54.

Belgoumri, M. D., M. R. Bouadjenek, S. Aryal, and H. Hacid. 2024. "Data Quality in Edge Machine Learning: A State-of-the-Art Survey." arXiv Preprint arXiv:2406.02600.

Bellavista, P., R. Della Penna, L. Foschini, and D. Scotece. 2020. "Machine Learning for Predictive Diagnostics at the Edge: An IIoT Practical Example." In *ICC 2020—2020 IEEE International Conference on Communications (ICC)*, 1–7. IEEE.

Bortsova, G., C. González-Gonzalo, S. C. Wetstein, et al. 2021. "Adversarial Attack Vulnerability of Medical Image Analysis Systems: Unexplored Factors." *Medical Image Analysis* 73: 102141.

Boyatzis, R. E. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development*. Sage.

Chaddad, A., J. Peng, J. Xu, and A. Bouridane. 2023. "Survey of Explainable AI Techniques in Healthcare." *Sensors* 23, no. 2: 634.

Chai, S., and L. Guo. 2024. "Edge Computing With Fog-Cloud for Heart Data Processing Using Particle Swarm Optimized Deep Learning Technique." *Journal of Grid Computing* 22, no. 1: 3.

Channar, N. A., M. Y. Koondhar, S. Shaikh, et al. 2025. "A Review on Edge Ai for Low-Latency Health Monitoring in Wearable IOT Devices: Challenges and Future Directions." *Spectrum of Engineering Sciences* 3, no. 7: 1424–1435. https://thesesjournal.com/index.php/1/article/view/749.

Chaturvedi, K., J. Gasthuber, and M. Abdelaal. 2025. "EdgeMLOps: Operationalizing ML Models With Cumulocity IoT and Thin-Edge.Io for Visual Quality Inspection." arXiv Preprint arXiv:2501.17062.

Chen, L., M. A. Babar, and H. Zhang. 2010. "Towards an Evidence-Based Understanding of Electronic Data Sources." In *14th International Conference on Evaluation and Assessment in Software Engineering (EASE)*. BCS Learning & Development Ltd.

Choudhury, A., and O. Asan. 2020. "Role of Artificial Intelligence in Patient Safety Outcomes: Systematic Literature Review." *JMIR Medical Informatics* 8, no. 7: e18599.

Chung, K., and H. Yoo. 2020. "Edge Computing Health Model Using P2P-Based Deep Neural Networks." *Peer-to-Peer Networking and Applications* 13, no. 2: 694–703.

Corbin, C. K., R. Maclay, A. Acharya, et al. 2023. "DEPLOYR: A Technical Framework for Deploying Custom Real-Time Machine Learning Models Into the Electronic Medical Record." *Journal of the American Medical Informatics Association* 30, no. 9: 1532–1542.

Cox, J. R., Jr., and C. Zeelenberg. 1987. "Computer Technology: State of the Art and Future Trends." *Journal of the American College of Cardiology* 9, no. 1: 204–214.

Cruzes, D. S., and T. Dybå. 2010. "Synthesizing Evidence in Software Engineering Research." In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, 1–10. ACM.

D'Agostino, D., L. Morganti, E. Corni, D. Cesini, and I. Merelli. 2019. "Combining Edge and Cloud Computing for Low-Power, Cost-Effective Metagenomics Analysis." *Future Generation Computer Systems* 90: 79–85.

D'Agostino, D., A. Quarati, A. Clematis, et al. 2019. "SoC-Based Computing Infrastructures for Scientific Applications and Commercial Services: Performance and Economic Evaluations." *Future Generation Computer Systems* 96: 11–22.

El Kafhali, S., I. El Mir, and M. Hanini. 2022. "Security Threats, Defense Mechanisms, Challenges, and Future Directions in Cloud Computing." *Archives of Computational Methods in Engineering* 29, no. 1: 223–246.

Elhanashi, A., P. Dini, S. Saponara, and Q. Zheng. 2024. "Advancements in TinyML: Applications, Limitations, and Impact on IoT Devices." *Electronics* 13, no. 17: 3562.

El-Saleh, A. A., A. M. Sheikh, M. A. Albreem, and M. S. Honnurvali. 2025. "The Internet of Medical Things (IoMT): Opportunities and Challenges." *Wireless Networks* 31, no. 1: 327–344.

Ettaloui, N., S. Arezki, and T. Gadi. 2023. "An Overview of Blockchain-Based Electronic Health Record and Compliance With GDPR and HIPAA." In *Artificial Intelligence, Data Science and Applications. ICAISE 2023*. Y. Farhaoui, A. Hussain, T. Saba, H. Taherdoost, and A. Verma. (eds) Lecture Notes in Networks and Systems, vol 838. Springer.

Fadlullah, Z. M., A.-S. K. Pathan, and H. Gacanin. 2018. "On Delay-Sensitive Healthcare Data Analytics at the Network Edge Based on Deep Learning." In *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, 388–393. IEEE.

Familiar, B. 2015. *Microservices, IoT, and Azure*. Springer.

Ferrara, M., G. Bertozzi, N. Di Fazio, et al. 2024. "Risk Management and Patient Safety in the Artificial Intelligence Era: A Systematic Review." *Healthcare* 12: 1–15.

Finlayson, S. G., J. D. Bowers, J. Ito, J. L. Zittrain, A. L. Beam, and I. S. Kohane. 2019. "Adversarial Attacks on Medical Machine Learning." *Science* 363, no. 6433: 1287–1289.

Ganesan, P., and S. K. Jagatheesaperumal. 2024. "Revolutionizing Emergency Response: The Transformative Power of Smart Wearables Through Blockchain, Federated Learning, and Beyond 5G/6G Services." *IT Professional* 25, no. 6: 54–61.

Gentzen, F., M. Bensalem, and A. Jukan. 2025. "Effective ML Model Versioning in Edge Networks." In *2025 International Wireless Communications and Mobile Computing (IWCMC)*, 1131–1136. IEEE.

Gupta, P., A. V. Chouhan, M. A. Wajeed, S. Tiwari, A. S. Bist, and S. C. Puri. 2023. "Prediction of Health Monitoring With Deep Learning Using Edge Computing." *Measurement: Sensors* 25: 100604.

Habehh, H., and S. Gohel. 2021. "Machine Learning in Healthcare." *Current Genomics* 22, no. 4: 291–300.

Hakak, S., S. Ray, W. Z. Khan, and E. Scheme. 2020. "A Framework for Edge-Assisted Healthcare Data Analytics Using Federated Learning." In *2020 IEEE International Conference on Big Data (Big Data)*, 3423–3427. IEEE.

Hammer, C. 1984. "Beyond the Data Processing Horizon." In *Proceedings of the 1984 Annual Conference of the ACM on the Fifth Generation Challenge*. ACM.

Hartmann, M., U. S. Hashmi, and A. Imran. 2022. "Edge Computing in Smart Health Care Systems: Review, Challenges, and Research Directions." *Transactions on Emerging Telecommunications Technologies* 33, no. 3: e3710.

Hatzivasilis, G., O. Soultatos, S. Ioannidis, C. Verikoukis, G. Demetriou, and C. Tsatsoulis. 2019. "Review of Security and Privacy for the Internet of Medical Things (IoMT)." In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 457–464. IEEE.

He, Q., Z. Xi, Z. Feng, et al. 2025. "Telemedicine Monitoring System Based on Fog/Edge Computing: A Survey." *IEEE Transactions on Services Computing* 18, no. 1: 479–498.

He, Y., B. Fu, J. Yu, R. Li, and R. Jiang. 2020. "Efficient Learning of Healthcare Data From IoT Devices by Edge Convolution Neural Networks." *Applied Sciences* 10, no. 24: 8934.

Hennebelle, A., H. Materwala, and L. Ismail. 2023. "HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System." *Procedia Computer Science* 220: 331–338.

Hosseini, M.-P., T. X. Tran, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh. 2017. "Deep Learning With Edge Computing for Localization of Epileptogenicity Using Multimodal Rs-fMRI and EEG Big Data." In *2017 IEEE International Conference on Autonomic Computing (ICAC)*, 83–92. IEEE.

Houssein, E. H., A. M. Gamal, E. M. Younis, and E. Mohamed. 2025. "Explainable Artificial Intelligence for Medical Imaging Systems Using Deep Learning: A Comprehensive Review." *Cluster Computing* 28, no. 7: 469.

Hua, H., Y. Li, T. Wang, N. Dong, W. Li, and J. Cao. 2023. "Edge Computing With Artificial Intelligence: A Machine Learning Perspective." *ACM Computing Surveys* 55, no. 9: 1–35.

Huang, R., Z. Wei, W. Feng, et al. 2024. "Edge Computing Model Based on Federated Learning for COVID-19 Clinical Outcome Prediction in the 5G Era." *KSII Transactions on Internet and Information Systems* 18, no. 4: 826–842.

Jagatheesaperumal, S. K., P. Mishra, N. Moustafa, and R. Chauhan. 2022. "A Holistic Survey on the Use of Emerging Technologies to Provision Secure Healthcare Solutions." *Computers and Electrical Engineering* 99: 107691.

Joo, M.-I., D.-Y. Kang, M.-S. Kang, and H.-C. Kim. 2024. "A Patient Management System Using an Edge Computing-Based IoT Pulse Oximeter." *Applied Sciences* 14, no. 1: 414.

Jordan, P. 2006. "Standard IEC 62304-Medical Device Software-Software Lifecycle Processes." In *2006 IET Seminar on Software for Medical Devices*, 41–47. IET.

Kasula, B. Y. 2024. "Optimizing Healthcare Delivery: Machine Learning Applications and Innovations for Enhanced Patient Outcomes." *International Journal of Creative Research in Computer Technology and Design* 6, no. 6: 1–7.

Khan, M. F., T. M. Ghazal, R. A. Said, et al. 2021. "An Iomt-Enabled Smart Healthcare Model to Monitor Elderly People Using Machine Learning Technique." *Computational Intelligence and Neuroscience* 2021, no. 1: 2487759.

Kishor, A., and C. Chakraborty. 2022. "Artificial Intelligence and Internet of Things Based Healthcare 4.0 Monitoring System." *Wireless Personal Communications* 127, no. 2: 1615–1631.

Kong, X., K. Wang, S. Wang, et al. 2021. "Real-Time Mask Identification for COVID-19: An Edge-Computing-Based Deep Learning Framework." *IEEE Internet of Things Journal* 8, no. 21: 15929–15938.

Kumar, A., A. Sharma, V. Bharti, A. K. Singh, S. K. Singh, and S. Saxena. 2021. "MobiHisNet: A Lightweight CNN in Mobile Edge Computing for Histopathological Image Classification." *IEEE Internet of Things Journal* 8, no. 24: 17778–17789.

Kumar, S. M., and D. Majumder. 2018. "Healthcare Solution Based on Machine Learning Applications in IOT and Edge Computing." *International Journal of Pure and Applied Mathematics* 119, no. 16: 1473–1484.

Lalis, S., A. Savidis, A. Karypidis, J. Gutknecht, and C. Stephanides. 2007. "Towards Dynamic and Cooperative Multi-Device Personal Computing." In *The Disappearing Computer: Interaction Design, System Infrastructures and Applications for Smart Environments*. N. Streitz, A. Kameas, and I. Mavrommati. (eds) Lecture Notes in Computer Science, vol 4500, 182–204. Springer.

Liu, X., and A. G. Richardson. 2021. "Edge Deep Learning for Neural Implants: A Case Study of Seizure Detection and Prediction." *Journal of Neural Engineering* 18, no. 4: 046034.

Ma, H., H. Qiu, Y. Gao, et al. 2023. "Quantization Backdoors to Deep Learning Commercial Frameworks." *IEEE Transactions on Dependable and Secure Computing* 21, no. 3: 1155–1172.

Manogaran, G., P. M. Shakeel, H. Fouad, et al. 2019. "Wearable IoT Smart-Log Patch: An Edge Computing-Based Bayesian Deep Learning Network System for Multi Access Physical Monitoring System." *Sensors* 19, no. 13: 3030.

Medvidovic, N., and R. N. Taylor. 2000. "A Classification and Comparison Framework for Software Architecture Description Languages." *IEEE Transactions on Software Engineering* 26, no. 1: 70–93.

Meng, L., and D. Li. 2023. "Novel Edge Computing-Based Privacy-Preserving Approach for Smart Healthcare Systems in the Internet of Medical Things." *Journal of Grid Computing* 21, no. 4: 66.

Muhammad, G., and M. S. Hossain. 2021. "A Deep-Learning-Based Edge-Centric COVID-19-Like Pandemic Screening and Diagnosis System Within a B5G Framework Using Blockchain." *IEEE Network* 35, no. 2: 74–81.

Murshed, M. S., C. Murphy, D. Hou, N. Khan, G. Ananthanarayanan, and F. Hussain. 2021. "Machine Learning at the Network Edge: A Survey." *ACM Computing Surveys* 54, no. 8: 1–37.

Nemec Zlatolas, L., T. Welzer, and L. Lhotska. 2024. "Data Breaches in Healthcare: Security Mechanisms for Attack Mitigation." *Cluster Computing* 27, no. 7: 8639–8654.

Pereira, C. V. F., E. M. de Oliveira, and A. D. Souza. 2024. "Machine Learning Applied to Edge Computing and Wearable Devices for Healthcare: Systematic Mapping of the Literature." *Sensors* 24, no. 19: 6322.

Punugoti, R., V. Dutt, A. Anand, and N. Bhati. 2023. "Exploring the Impact of Edge Intelligence and IoT on Healthcare: A Comprehensive Survey." In *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, 1108–1114. IEEE.

Pustokhina, I. V., D. A. Pustokhin, D. Gupta, A. Khanna, K. Shankar, and G. N. Nguyen. 2020. "An Effective Training Scheme for Deep Neural Network in Edge Computing Enabled Internet of Medical Things (IoMT) Systems." *IEEE Access* 8: 107112–107123.

Putra, K. T., A. Z. Arrayyan, N. Hayati, C. Damarjati, A. Bakar, and H.-C. Chen. 2024. "A Review on the Application of Internet of Medical Things in Wearable Personal Health Monitoring: A Cloud-Edge Artificial Intelligence Approach." *IEEE Access* 12: 21437–21452.

Qayyum, A., K. Ahmad, M. A. Ahsan, A. Al-Fuqaha, and J. Qadir. 2022. "Collaborative Federated Learning for Healthcare: Multi-Modal Covid-19 Diagnosis at the Edge." *IEEE Open Journal of the Computer Society* 3: 172–184.

Rahman, M. A., and M. S. Hossain. 2021. "An Internet-Of-Medical-Things-Enabled Edge Computing Framework for Tackling COVID-19." *IEEE Internet of Things Journal* 8, no. 21: 15847–15854.

Rahman, M. A., M. S. Hossain, N. A. Alrajeh, and N. Guizani. 2020. "B5G and Explainable Deep Learning Assisted Healthcare Vertical at the Edge: COVID-I9 Perspective." *IEEE Network* 34, no. 4: 98–105.

Rajagopal, S. M., M. Supriya, and R. Buyya. 2024. "Blockchain Integrated Federated Learning in Edge/Fog/Cloud Systems for IoT-Based Healthcare Applications: A Survey." In *Federated Learning*, 237–269. Taylor & Francis Group.

Ram, S. S., B. Apduhan, and N. Shiratori. 2019. "A Machine Learning Framework for Edge Computing to Improve Prediction Accuracy in Mobile Health Monitoring." In *Computational Science and Its Applications–ICCSA 2019: 19th International Conference, Saint Petersburg, Russia, July 1–4, 2019, Proceedings, Part III 19*. Lecture Notes in Computer Science, vol 11621. Springer.

Rancea, A., I. Anghel, and T. Cioara. 2024. "Edge Computing in Healthcare: Innovations, Opportunities, and Challenges." *Future Internet* 16, no. 9: 329.

Rocha, A., M. Monteiro, C. Mattos, et al. 2024. "Edge AI for Internet of Medical Things: A Literature Review." *Computers and Electrical Engineering* 116: 109202.

Rydning, D. R.-J. G.-J., J. Reinsel, and J. Gantz. 2018. *The Digitization of the World from Edge to Core*. Vol. 16, 1–28. International Data Corporation.

Sakib, S., M. M. Fouda, and Z. M. Fadlullah. 2021. "A Rigorous Analysis of Biomedical Edge Computing: An Arrhythmia Classification Use-Case Leveraging Deep Learning." In *2020 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, 136–141. IEEE.

Sakib, S., M. M. Fouda, Z. M. Fadlullah, and N. Nasser. 2020. "Migrating Intelligence From Cloud to Ultra-Edge Smart IoT Sensor Based on Deep Learning: An Arrhythmia Monitoring Use-Case." In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, 595–600. IEEE.

Samarpita, S. 2024. "Edge Intelligence and Its Healthcare Application." In *Reconnoitering the Landscape of Edge Intelligence in Healthcare*, 3–18. Apple Academic Press.

Shahbaz, M., and R. Zahid. 2022. "Probing the Factors Influencing Cloud Computing Adoption in Healthcare Organizations: A Three-Way Interaction Model." *Technology in Society* 71: 102139.

Shivashankar, K., G. S. A. Hajj, and A. Martini. 2025. "Scalability and Maintainability Challenges and Solutions in Machine Learning: Systematic Literature Review." arXiv preprint arXiv:2504.11079.

Singh, R., and S. S. Gill. 2023. "Edge AI: A Survey." *Internet of Things and Cyber-Physical Systems* 3: 71–92.

Singh, V. K., and M. H. Kolekar. 2022. "Deep Learning Empowered COVID-19 Diagnosis Using Chest CT Scan Images for Collaborative Edge-Cloud Computing Platform." *Multimedia Tools and Applications* 81, no. 1: 3–30.

Sistla, S. M. K., and B. K. Konidena. 2023. "Iot-Edge Healthcare Solutions Empowered by Machine Learning." *Journal of Knowledge Learning and Science Technology* 2, no. 2: 126–135.

Sufian, A., A. Ghosh, A. S. Sadiq, and F. Smarandache. 2020. "A Survey on Deep Transfer Learning to Edge Computing for Mitigating the COVID-19 Pandemic." *Journal of Systems Architecture* 108: 101830.

Sufian, A., C. You, and M. Dong. 2021. "A Deep Transfer Learning-Based Edge Computing Method for Home Health Monitoring." In *2021 55th Annual Conference on Information Sciences and Systems (CISS)*, 1–6. IEEE.

Tripathi, J. N., B. Kumar, and D. Junjariya. 2022. "Hardware Accelerator Design for Healthcare Applications: Review and Perspectives." In *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1367–1371. IEEE.

Uddin, M. Z. 2019. "A Wearable Sensor-Based Activity Prediction System to Facilitate Edge Computing in Smart Healthcare System." *Journal of Parallel and Distributed Computing* 123: 46–53.

Wang, F., and A. Beecy. 2025. "Implementing AI Models in Clinical Workflows: A Roadmap." *BMJ Evidence-Based Medicine* 30, no. 5: 285–287.

Wang, K., S. Kong, X. Chen, and M. Zhao. 2024. "Edge Computing Empowered Smart Healthcare: Monitoring and Diagnosis With Deep Learning Methods." *Journal of Grid Computing* 22, no. 1: 30.

Wang, T., J. Guo, B. Zhang, G. Yang, and D. Li. 2025. "Deploying AI on Edge: Advancement and Challenges in Edge Intelligence." *Mathematics* 13, no. 11: 1878.

Wang, X., Z. Tang, J. Guo, et al. 2025. "Empowering Edge Intelligence: A Comprehensive Survey on On-Device Ai Models." *ACM Computing Surveys* 57, no. 9: 1–39.

Weiner, E. B., I. Dankwa-Mullan, W. A. Nelson, and S. Hassanpour. 2025. "Ethical Challenges and Evolving Strategies in the Integration of Artificial Intelligence Into Clinical Practice." *PLOS Digital Health* 4, no. 4: e0000810.

Wieringa, R., N. Maiden, N. Mead, and C. Rolland. 2006. "Requirements Engineering Paper Classification and Evaluation Criteria: A Proposal and a Discussion." *Requirements Engineering* 11: 102–107.

Wu, Q., X. Chen, Z. Zhou, and J. Zhang. 2020. "Fedhome: Cloud-Edge Based Personalized Federated Learning for In-Home Health Monitoring." *IEEE Transactions on Mobile Computing* 21, no. 8: 2818–2832.

Yaacoub, J.-P. A., M. Noura, H. N. Noura, et al. 2020. "Securing Internet of Medical Things Systems: Limitations, Issues and Recommendations." *Future Generation Computer Systems* 105: 581–606.

Yang, T., M. Gentile, C.-F. Shen, and C.-M. Cheng. 2020. *Combining Point-of-Care Diagnostics and Internet of Medical Things (IoMT) to Combat the COVID-19 Pandemic*. Vol. 10, 224. Multidisciplinary Digital Publishing Institute.

Yu, J., B. Fu, A. Cao, Z. He, and D. Wu. 2018. "EdgeCNN: A Hybrid Architecture for Agile Learning of Healthcare Data From IoT Devices." In *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, 852–859. IEEE.

Zamani, M., and S. Sharifian. 2024. "Distributed Edge to Cloud Ensemble Deep Learning Architecture to Diagnose Covid-19 From Lung Image in IoT Based e-Health System." *Journal of Supercomputing* 80, no. 13: 18492–18520.

Zhang, C., M. Aamir, Y. Guan, et al. 2024. "Enhancing Lung Cancer Diagnosis With Data Fusion and Mobile Edge Computing Using DenseNet and CNN." *Journal of Cloud Computing* 13, no. 1: 91.

Zhang, H., M. A. Babar, X. Bai, J. Li, and L. Huang. 2011. "An Empirical Assessment of a Systematic Search Process for Systematic Reviews." In *15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011)*, 56–65. IET.

Zhou, X., Y. Jin, H. Zhang, S. Li, and X. Huang. 2016. "A Map of Threats to Validity of Systematic Literature Reviews in Software Engineering." In *2016 23rd Asia-Pacific Software Engineering Conference (APSEC)*, 153–160. IEEE.

Zhu, T., L. Kuang, J. Daniels, P. Herrero, K. Li, and P. Georgiou. 2022. "IoMT-Enabled Real-Time Blood Glucose Prediction With Deep Learning and Edge Computing." *IEEE Internet of Things Journal* 10, no. 5: 3706–3719.

Appendix A

TABLE A1 | List of identified tools.

Study ID	Tool/framework	Source type	Type	Main aim/focus/support
S4	CNN-Edge–based framework	Closed source	Custom	Transferring information from IoT devices to the doctors
S6	Linux Perf tool	Open source	Existing	Computing the number of CPU cycles required by the execution of each algorithm
S8	HealthEdge framework	Closed source	Custom	Diabetes prediction
S10	AWS SageMaker	Open source	Existing	Deploying the deep learning models in the cloud
S10	TensorFlow	Open source	Existing	Deploying the deep learning models
S11	ECMask framework	Closed source	Existing	Identifying the condition of mask-wearing
S11	OpenVINO toolkit	Open source	Existing	Model optimization, inference engine, distributed computing
S11	PyTorch framework	Open source	Existing	Implementation
S11	CUDA	Open source	Existing	Implementation
S13	TensorFlow	Open source	Existing	Performance evaluation
S16	TensorFlow	Open source	Existing	Data training
S17	TensorFlow	Open source	Existing	Implementation (execution time, maximum memory consumption)
S19	TensorFlow	Open source	Existing	Data training
S19	PyTorch	Open source	Existing	Data training
S19	Flask Server	Open source	Existing	Deployment (Docker container)
S19	CUDA	Open source	Existing	To access the GPU over the container
S21	TensorFlow	Open source	Existing	Implementation (fine-tuning the model in terms of its size and latency)
S22	MATLAB	Closed source	Existing	Data handling
S22	TensorFlow	Open source	Existing	Data training
S24	PyTorch	Open source	Existing	Development
S24	PyCharm	Open source	Existing	Development
S27	FETCH Framework	Closed source	Custom	Automated diagnosis
S30	COVID-19 management framework	Closed source	Custom	Testing
S30	Pysim5G	Open source	Existing	Implementation
S30	Shapely	Open source	Existing	Implementation
S30	Flask	Open source	Existing	Implementation
S30	TensorFlow	Open source	Existing	Testing
S30	PyTorch	Open source	Existing	Testing
S30	Apache MXNet	Open source	Existing	Testing
S30	Caffe	Open source	Existing	Testing
S30	OpenVino toolkit	Open source	Existing	Testing
S32	TensorFlow	Open source	Existing	Implementation
S34	EBPPA	Closed source	Custom	Data privacy and security
S37	FATE framework	Open source	Existing	Privacy preserving

TABLE A2 | Data source analysis for selected studies.

Study ID	Dataset name	Data type
S1	ECG dataset, multi-modal sensors dataset	Numerical, signal
S2	Sensors data	Image, video
S3	N/A	N/A
S4	Real dataset	Image
S5	Heart rate data (collected using a Zigbee biosensor)	Signal
S6	PAMAP2 dataset	Numerical
S7	N/A	N/A
S8	PIMA Indian, Sylhet	Numerical
S9	MIT Arrhythmia dataset	Signal+
S10	OhioT1DM dataset, ABC4D dataset, ARISES dataset	Numerical
S11	Bus drive monitoring dataset, FDDB dataset, WIDER FACE dataset	Video, Image
S12	N/A	N/A
S13	MIT-BIH database	Numerical
S14	ECG dataset	Signal
S15	ECoG dataset, rs-fMRI data	Signal, image
S16	PhysioNet/CinC2017, MIT-BIH dataset	Signal, numerical
S17	BreakHis dataset	Image
S18	MIT-BIH Supraventricular Arrhythmia database, MIT-BIH Arrhythmia Database, INCART 12-lead Arrhythmia database, Sudden Cardiac Death Holter database	Signal
S19	Cough Sound Dataset, UTA Drowsiness Dataset, Face Mask Detection Dataset, MIT-BIH Arrhythmia dataset, Emotion Detection Dataset, Fever Detection Dataset (Kaggle dataset), EEG Signal Dataset (Kaggle dataset)	Signal, image, numerical
S20	Chest x-ray images, CT scan images, lung ultrasound images, ocular surface images, vital signs data	Signal, image, numerical
S21	SARS-CoV-2 CT-scan dataset, ImageNet dataset	Image
S22	CHB-MIT database	Signal
S23	EEG signals dataset	Signal
S24	MNIST database, CIFAR10 dataset, STL10 dataset, COVID19 Chest X-RAY image dataset	Image
S25	MHEALTH dataset	Signal, numerical
S26	IoT sensor datasets	Signal, numerical
S27	ECG data, wearable sensor data	Signal, numerical
S28	UCI repository data	Numerical
S29	MIT-BIH Supraventricular Arrhythmia Database, MIT-BIH Arrhythmia Database, St Petersburg INCART 12-lead Arrhythmia Database	Signal
S30	Collected datasets (CT-scan images, RNA sequencing data, thermal images, x-ray images)	Image, numerical
S31	MobiAct dataset	Signal, numerical
S32	Chest x-ray dataset, chest ultrasound images dataset	Image, video-based image
S33	Manually collected dataset	Numerical, signal
S34	N/A	N/A
S35	Medical chest x-ray images	Image
S36	CT scan image dataset	Image
S37	Real-world COVID-19 dataset	Numerical, signal

TABLE A3 | List of selected studies.

Study ID	Publication title and reference	Authors name	Major contributions	Journal/conference name	Journal/ conference rank	Publisher	Year
S1	A machine learning framework for edge computing to improve prediction accuracy in mobile health monitoring (Ram et al. 2019)	Sigdel Shree Ram, Bernady Apduhan And Norio Shiratori	Evaluating machine learning algorithms for personalized mobile health monitoring systems using mobile health data	Conference—Computational Science and Its Applications—ICCSA	C	Springer	2019
S2	A deep transfer learning-based edge computing method for home health monitoring (Sufian et al. 2021)	Abu Sufian, Changsheng You, and Mianxiong Dong	Proposing a transfer learning-based edge computing method for home health monitoring	Conference—Conference on Information Sciences and Systems (CISS)	N/A	IEEE	2021
S3	Machine-learning-based IoT–edge computing healthcare solutions (Alnaim and Alwakeel 2023)	Abdulrahman K. Alnaim and Ahmed M. Alwakeel	Exploring the integration of cloud, edge computing, and machine learning in a distributed-edge-computing-based Internet-of-Things (IoT) framework for medical data analysis	Journal—*Electronics*	Q2	MDPI	2023
S4	Prediction of health monitoring with deep learning using edge computing (Gupta et al. 2023)	Piyush Gupta, Ajay Veer Chouhan, Mohammed Abdul Wajeed, Shivam Tiwari, Ankur Singh Bist, and Shiv Charan Puri	Integrating IoT, edge computing, and CNN-based deep learning for efficient fall detection in smart healthcare systems	Journal—*Measurement: Sensors*	Q3	Elsevier	2023
S5	Edge computing health model using P2P-based deep neural networks (Chung and Yoo 2020)	Kyungyong Chung and Hyun Yoo	Proposing an edge computing health model using P2P-based deep neural networks for efficient processing of health big data	Journal—*Peer-to-Peer Networking and Applications*	Q2	Springer	2020
S6	Task offloading in edge computing for machine learning-based smart healthcare (Aazam et al. 2021)	Mohammad Aazam, Sherali Zeadally, and Eduardo Feo Flushing	Enabling intelligent healthcare services through edge computing and machine learning-based task offloading	Journal—*Computer Networks*	Q1	Elsevier	2021
S7	IoT-edge healthcare solutions empowered by machine learning (Sistla and Konidena 2023)	Sistla Sai Mani Krishna, and Bhargav Kumar Konidena	Optimizing real-time response and security in medical sensor networks through edge computing and machine learning-based analytics	Journal—*Journal of Knowledge Learning and Science Technology*	N/A	N/A	2023
S8	HealthEdge: a machine learning-based smart healthcare framework for prediction of type 2 diabetes in an integrated IoT, edge, and cloud computing system (Hennebelle et al. 2023)	Hennebelle Alain, Huned Materwala, and Leila Ismail	Developing HealthEdge: a machine learning-based smart healthcare framework for Type 2 diabetes prediction in integrated IoT-edge-cloud computing systems, and evaluating it with the most commonly used machine learning algorithms and real-life diabetes datasets	Journal—*Procedia Computer Science*	N/A	Elsevier	2023

(Continues)

TABLE A3 | (Continued)

Study ID	Publication title and reference	Authors name	Major contributions	Journal/conference name	Journal/ conference rank	Publisher	Year
S9	Empowering healthcare IoT systems with hierarchical edge-based deep learning (Azimi et al. 2018)	Azimi Iman, Janne Takalo-Mattila, Arman Anzanpour, Amir M. Rahmani, Juha-Pekka Soininen, and Pasi Liljeberg	Enhancing remote health monitoring using hierarchical computing architecture and convolutional neural networks: a feasibility study and performance evaluation	Conference—International Conference on Connected Health: Applications, Systems and Engineering Technologies	N/A	IEEE/ACM	2018
S10	IoMT-enabled real-time blood glucose prediction with deep learning and edge computing (Zhu et al. 2022)	Zhu Taiyu, Lei Kuang, John Daniels, Pau Herrero, Kezhi Li, and Pantelis Georgiou	A novel deep learning model employing an attention-based evidential recurrent neural network is proposed for real-time blood glucose prediction and hypoglycemia detection. This model is embedded into a low-cost, low-power IoMT device, leveraging Bluetooth connectivity and edge computing for efficient performance	Journal—*Internet of Things Journal*	Q1	IEEE	2022
S11	Real-time mask identification for COVID-19: an edge-computing-based deep learning framework (Kong et al. 2021)	Kong Xiangjie, Kailai Wang, Shupeng Wang, Xiaojie Wang, Xin Jiang, Yi Guo, Guojiang Shen, Xin Chen, and Qichao Ni	Introducing ECMask: an edge computing-based real-time mask detection framework for public health precautions	Journal—*Internet of Things Journal*	Q1	IEEE	2021
S12	A framework for edge-assisted healthcare data analytics using federated learning (Hakak et al. 2020)	Hakak Saqib, Suprio Ray, Wazir Zada Khan, and Erik Scheme	Proposing an edge-assisted federated learning framework for leveraging user-generated data from wearable devices in healthcare	Conference—International Conference on Big Data	B	IEEE	2020
S13	On delay-sensitive healthcare data analytics at the network edge based on deep learning (Fadlullah et al. 2018)	Fadlullah Zubair Md, Al-Sakib Khan Pathan, and Haris Gacanin	Proposing deep learning-based IoT edge analytics for near real-time smart healthcare: addressing delay-sensitive healthcare data at the network edge	Conference—International Wireless Communications & Mobile Computing Conference (IWCMC)	B	IEEE	2018
S14	EdgeCNN: a hybrid architecture for agile learning of healthcare data from IoT devices (Yu et al. 2018)	Yu Jian, Bin Fu, Ao Cao, Zhenqian He, and Di Wu	Introducing EdgeCNN: a hybrid architecture for real-time healthcare data analysis using edge and cloud computing	Conference—International Conference on Parallel and Distributed Systems (ICPADS)	B	IEEE	2018
S15	Deep learning with edge computing for localization of epileptogenicity using multimodal rs-fMRI and EEG big data (Hosseini et al. 2017)	Hosseini Mohammad-Parsa, Tuyen X. Tran, Dario Pompili, Kost Elisevich, and Hamid Soltanian-Zadeh	Developing autonomic edge computing solutions for epilepsy management: integrating noninvasive and invasive methods for monitoring and regulation of the epileptic brain	Conference—International Conference on Autonomic Computing (ICAC)	B	IEEE	2017

(Continues)

TABLE A3 | (Continued)

Study ID	Publication title and reference	Authors name	Major contributions	Journal/conference name	Journal/conference rank	Publisher	Year
S16	Efficient learning of healthcare data from IoT devices by edge convolution neural networks (He et al. 2020)	He Yan, Bin Fu, Jian Yu, Renfa Li, and Rucheng Jiang	Introducing EdgeCNN: a hybrid architecture for efficient electrocardiogram classification in smart healthcare applications	Journal—*Applied Sciences*	Q2	MDPI	2020
S17	MobiHisNet: a lightweight CNN in mobile edge computing for histopathological image classification (A. Kumar et al. 2021)	Kumar Abhinav, Anshul Sharma, Vandana Bharti, Amit Kumar Singh, Sanjay Kumar Singh, and Sonal Saxena	Developing an efficient, lightweight CNN model based on MobileNet for the classification of histopathological images on edge devices	Journal—*Internet of Things Journal*	Q1	IEEE	2021
S18	A rigorous analysis of biomedical edge computing: An arrhythmia classification use-case leveraging deep learning (Sakib et al. 2021)	Sakib Sadman, Mostafa M. Fouda, and Zubair Md Fadlullah	Efficient arrhythmia classification on resource-constrained biomedical edge devices using customized 1-D CNN model	Conference—International Conference on Internet of Things and Intelligence Systems (IoTaIS)	N/A	IEEE	2021
S19	An internet-of-medical-things-enabled edge computing framework for tackling COVID-19 (Rahman and Hossain 2021)	Rahman Md Abdur, and M. Shamim Hossain	Affective computing framework for in-home COVID-19 symptom management using edge GPUs architecture; crafting DL libraries tailored for diverse IoMT gadgets, perfect for crafting health-oriented applications at the edge. These IoMT units are equipped to handle DL-based CNN, aided by edge TPUs or GPUs, developing a collection of Quality of Life (QoL) monitoring apps leveraging IoMT edge learning	Journal—*Internet of Things Journal*	Q1	IEEE	2021
S20	A deep-learning-based edge-centric COVID-19-like pandemic screening and diagnosis system within a B5G framework using blockchain (Muhammad and Hossain 2021)	Muhammad Ghulam and M. Shamim Hossain	Edge-centric AI-enabled COVID-19 screening and diagnosis system utilizing beyond 5G network with blockchain-based secure data transmission	Journal—*IEEE Network*	Q1	IEEE	2021
S21	Deep learning empowered COVID-19 diagnosis using chest CT scan images for collaborative edge-cloud computing platform (V. K. Singh and Kolekar 2022)	Singh Vipul Kumar, and Maheshkumar H. Kolekar	Proposing a novel deep-learning model for COVID-19 diagnosis that is readily deployable on edge devices	Journal—*Multimedia Tools and Applications*	Q1	Springer	2022

(Continues)

TABLE A3 | (Continued)

Study ID	Publication title and reference	Authors name	Major contributions	Journal/conference name	Journal/conference rank	Publisher	Year
S22	Edge deep learning for neural implants: a case study of seizure detection and prediction (Liu and Richardson 2021)	Liu Xilin and Andrew G. Richardson	Developing Edge DL models for clinical neural implants: a study on epileptic seizure detection; utilizing three prevalent deep learning architectures (DNN, CNN, and LSTM) and enhancing these models for efficient deployment on hardware with limited resources	Journal—*Journal of Neural Engineering*	Q1	Iop Science	2021
S23	Smart health monitoring for seizure detection using mobile edge computing (Ali et al. 2020)	Ali Zien Sheikh, Nandhini Subramanian, and Aiman Erbad	Proposing a deep learning method for epileptic seizure detection using mobile edge computing	Conference—International Wireless Communications and Mobile Computing (IWCMC)	B	IEEE	2020
S24	Edge intelligence: federated learning-based privacy protection framework for smart healthcare systems (Akter et al. 2022)	Akter Mahmuda, Nour Moustafa, Timothy Lynar, and Imran Razzak	Proposing a three-fold federated edge aggregator architecture for privacy-preserving smart healthcare systems	Journal—*Journal of Biomedical and Health Informatics*	Q1	IEEE	2022
S25	A wearable sensor-based activity prediction system to facilitate edge computing in smart healthcare system (Uddin 2019)	Uddin Md Zia.	This paper investigates a system for predicting human activities through a multimodal approach, leveraging multiple wearable healthcare sensors. This involves employing deep recurrent neural networks (RNNs) to train on data sourced from the publicly available MHEALTH dataset and to model 12 distinct human activities	Journal—*Journal of Parallel and Distributed Computing*	Q1	Elsevier	2019
S26	Wearable IoT smart-log patch: an edge computing-based Bayesian deep learning network system for multi access physical monitoring system (Manogaran et al. 2019)	Manogaran Gunasekaran, P. Mohamed Shakeel, Hassan Fouad, Yunyoung Nam, S. Baskar, Naveen Chilamkurti, and Revathi Sundarasekar	Designing IoT-enabled wearable patch with edge computing for accurate physical activity monitoring	Journal—*Sensors*	Q1	MDPI	2019
S27	Edge computing empowered smart healthcare: monitoring and diagnosis with deep learning methods (K. Wang et al. 2024)	Wang Kemeng, Shurui Kong, Xuezheng Chen, and Min Zhao	This paper presents a FETCH framework that leverages fog computing to complement cloud computing for more efficient healthcare systems. It uses FogBus middleware to optimize resource allocation, latency, and power consumption in the processing of medical data	Journal—*Journal of Grid Computing*	Q1	Springer	2024
S28	An effective training scheme for deep neural network in edge computing enabled Internet of Medical Things (IoMT) (Pustokhina et al. 2020)	Pustokhina Irina Valeryevna, Denis Alexandrovich Pustokhin, Deepak Gupta, Ashish Khanna, Kannan Shankar, and Gia Nhu Nguyen	ETS-DNN: systems effective training scheme for deep neural networks in edge computing enabled IoMT Systems	Journal—*IEEE Access*	Q1	IEEE	2020

(Continues)

TABLE A3 | (Continued)

Study ID	Publication title and reference	Authors name	Major contributions	Journal/conference name	Journal/ conference rank	Publisher	Year
S29	Migrating intelligence from cloud to ultra-edge smart IoT sensor based on deep learning: an arrhythmia monitoring use-case (Sakib et al. 2020)	Sakib Sadman, Mostafa M. Fouda, Zubair Md Fadlullah, and Nidal Nasser	Intelligent ultra-edge IoT sensor for arrhythmia detection in mobile health scenarios	Conference—International Wireless Communications and Mobile Computing (IWCMC)	B	IEEE	2020
S30	B5G and explainable deep learning assisted healthcare vertical at the edge: COVID-I9 perspective (Rahman et al. 2020)	Rahman Md Abdur, M. Shamim Hossain, Nabil A. Alrajeh, and Nadra Guizani	Distributed deep learning framework for COVID-19 management with mobile edge computing	Journal—*IEEE Network*	Q1	IEEE	2020
S31	FedHome: cloud-edge based personalized federated learning for in-home health monitoring (Wu et al. 2020)	Wu Qiong, Xu Chen, Zhi Zhou, and Junshan Zhang	FedHome: a novel cloud-edge federated learning framework for privacy-preserving in-home health monitoring	Journal—*IEEE Transactions on Mobile Computing*	Q1	IEEE	2020
S32	Collaborative federated learning for healthcare: multi-modal COVID-19 diagnosis at the edge (Qayyum et al. 2022)	Qayyum Adnan, Kashif Ahmad, Muhammad Ahtazaz Ahsan, Ala Al-Fuqaha, and Junaid Qadir	Edge computing empowered intelligent diagnosis of COVID-19: a clustered federated learning approach	Journal—*IEEE Open Journal of the Computer Society*	Q1	IEEE	2022
S33	Edge computing with fog-cloud for heart data processing using particle swarm optimized deep learning technique (Chai and Guo 2024)	Chai Sheng and Lantian Guo	Introduction of an integrated edge-fog-cloud computing model for heart data processing	Journal—*Journal of Grid Computing*	Q1	Springer	2023
S34	Novel edge computing-based privacy-preserving approach for smart healthcare systems in the Internet of Medical Things (Meng and Li 2023)	Meng Lingbin and Daofeng Li	Introduction of the edge-based privacy-preserving approach (EBPPA), which combines homomorphic encryption and the XGBoost algorithm to enhance data privacy, security, and system efficiency in healthcare applications	Journal—*Journal of Grid Computing*	Q1	Springer	2023
S35	Distributed edge to cloud ensemble deep learning architecture to diagnose Covid-19 from lung image in IoT based e-Health system (Zamani and Sharifian 2024)	Zamani Mohammadreza and Saeed Sharifian	Development of a distributed edge-to-cloud ensemble deep learning architecture for diagnosing COVID-19	Journal—*Journal of Supercomputing*	Q2	Springer	2024
S36	Enhancing lung cancer diagnosis with data fusion and mobile edge computing using DenseNet and CNN (C. Zhang et al. 2024)	Zhang Chengping, Muhammad Aamir, Yurong Guan, Muna Al-Razgan, Emad Mahrous Awwad, Rizwan Ullah, Uzair Aslam Bhatti, and Yazeed Yasin Ghadi	Enhancing lung cancer diagnosis using DenseNet and CNN; Utilization of mobile edge computing to process and analyze lung cancer CT scan images closer to the data source	Journal—*Journal of Cloud Computing*	Q1	Springer	2024
S37	Edge computing model based on federated learning for COVID-19 clinical outcome prediction in the 5G era (Huang et al. 2024)	Huang Ruochen, Zhiyuan Wei, Wei Feng, Yong Li, Changwei Zhang, Chen Qiu, and Mingkai Che	Development of a federated learning-supported edge computing model that predicts COVID-19	Journal—*KSII Transactions on Internet & Information Systems*	Q3	Korea Society of Internet Information	2024

Appendix B

Journal and Conference Ranking: When discussing how prestigious academic publications are, we must understand the criteria we use to rank them. One significant factor is the journal's impact factor. This factor tells us how influential the journal is in its field. We also use quartile classification to organize journals into different tiers based on prestige. So, there are four tiers: Q1, Q2, Q3, and Q4. Q1 is like the top tier; it includes the most respected journals. Then we have Q2, Q3, and Q4, representing progressively less prestigious journals in their respective fields. It is a way of categorizing them based on their standing in the academic world. We used SCImago Journal Rank [3] to determine the ranking of our selected studies. In the realm of conference categorization, they're sorted like journals. There are four main groups: A*, A, B, and C. It is sort of like the quartile system for journals. A* is the top tier, the top-ranking conferences. Then come A, B, and C, each one a step down in how prestigious they are. We used the ICORE Conference Portal[4] to understand the ranking of conferences.

Selected Studies: The Study ID column is attached to each study for ease of referral and listing. Title and Reference contain the article's full title, volume, issue, and page details. Author's name lists the authors' names; for papers with one or two authors, list all names, but for papers with three or more authors, use the first author's name followed by "et al." (e.g., Amir Mashmool et al.). The Major Contributions column shows all paper's significant contributions or findings, including the most important results and theories developed or improvements made. The name of the journal/conference is crucial, as it tells us where the research work was published and how widely the results could have been circulated. Rank provides information on the ranking or impact factor of the journal or conference. The publisher names the entity responsible for the publication, often adding further detail about the source type. Then there is the column headed by year, which refers to the year of publication.