Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 265 (2025) 149–156

Procedia Computer Science

www.elsevier.com/locate/procedia

The 22nd International Conference on Mobile Systems and Pervasive Computing (MobiSPC)
August 4-6, 2025, Leuven, Belgium

Enhancing Privacy Transparency in Remote Patient Monitoring with Explainable AI

Jolly Trivedi[a], Jouni Isoaho[a], Tahir Mohammad[a,*]

[a]*University of Turku, Turku 20014, Finland*

Abstract

This paper presents an Explainable Privacy-Preserving Intelligent System for Monitoring (X-PRISM) framework designed to enhance transparency and privacy in AI-based remote patient monitoring (RPM) systems. The framework addresses the critical demand for explainable AI (XAI) in healthcare by integrating explainability techniques, such as SHapley Additive exPlanations (SHAP), to provide clarifications and reasoning behind AI-driven decisions based on key patient metrics such as heart rate and body temperature. X-PRISM implements pseudonymization and encryption to improve privacy and secure sensitive healthcare data while ensuring compliance with global regulations such as the General Data Protection Regulation (GDPR). Federated learning plays a vital role in the framework by enabling decentralized training of AI models across multiple healthcare nodes without directly sharing patient data. The layered architecture of X-PRISM enables seamless data collection from wearables and IoT devices, preprocessing in Azure Cloud, and AI model development using TensorFlow. X-PRISM is designed to offer transparent decision-making, protect patient data through decentralized training, and enable real-time interpretable feedback in RPM applications. Although the framework is not empirically tested in this study, it is presented as a foundation for future research and development in ethical AI deployment in healthcare. Existing gaps in RPM systems and synthesizing best practices in AI explainability and data privacy are reviewed in this study, which lays the foundation for developing secure, interpretable, and regulatory-compliant RPM systems. X-PRISM establishes the groundwork for advanced ethical, scalable, and trustworthy AI applications in RPM by addressing the dual goals of privacy and explainability, thereby enhancing patient trust and healthcare outcomes.

Keywords: Explainable AI ; XAI; Privacy; Transparency; Remote patient monitoring; RPM; Ambient systems; Healthcare IoT; Data privacy; AI decision-making; Privacy policies Healthcare Compliance

1. Introduction

The quick acceptance of ambient systems in healthcare, particularly in RPM, has revolutionized patient care by enabling continuous monitoring of vital health parameters through interconnected devices. AI-powered solutions enhance the precision and timeliness of medical interventions, reduce costs, and improve decision-making [1]. Talati [2] highlighted that AI integration in RPM enables personalized treatment, reduced hospitalizations, and early inter-

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: tahir.mohammad@utu.fi

10.1016/j.procs.2025.07.167

vention. Certain challenges of these solutions include standardization, mobility issues, heterogeneous networks, and quality-of-service concerns [3]. Their reliance on sensitive patient data introduces significant concerns regarding privacy protection, transparency, and trust. Most of the time, healthcare providers and patients are unable to understand how decisions are made by AI-based RPM systems, as they operate as opaque "black boxes." This raises concerns regarding data misuse, accountability, and bias. Addressing these issues is critical for adopting and effectively utilizing XAI in ambient healthcare systems. XAI ensures transparency by elucidating the rationale behind AI-driven decisions, and thus helps promote accountability and informed decision-making in critical healthcare contexts. Moreover, explainability is vital to meet regulatory and ethical requirements, such as those outlined in the GDPR, which emphasize the importance of fairness and transparency in data processing. The adoption of XAI in RPM further facilitates the identification and mitigation of biases in AI models, enhancing the reliability and equity of patient care. This study explores the role of XAI in enhancing privacy transparency within ambient systems for RPM. XAI techniques aim to provide interpretable insights into AI decision-making processes, enabling healthcare providers and patients to understand, validate, and trust the system's output. XAI improves AI-human interaction and decision acceptability [4].

The motivations behind the implementation of XAI in RPM include the need to enhance transparency in AI decision making, address privacy concerns, ensure regulatory compliance, enhance the adoption of explainable AI, and resolve scalability issues in RPM systems. To support the adoption of XAI, this study proposes an Explainable Privacy-Preserving Intelligent System for Monitoring (X-PRISM) framework that integrates XAI and privacy-preserving technologies to enhance transparency and data security in ambient RPM systems. By implementing federated learning, X-PRISM ensures that sensitive patient data is not exposed during model training, while still providing interpretable and actionable outputs. By highlighting the twin objectives of explainability and privacy protection, this approach builds user trust and facilitates the adoption of AI-powered RPM systems. This study illustrates how XAI might close the gap between AI-driven decision-making and privacy transparency in healthcare through X-PRISM.

2. Background and Related Work

RPM systems enable continuous, real-time monitoring of patient health parameters like as heart rate, glucose levels, blood pressure, and oxygen saturation without having patients to be physically present in a clinical setting. There are benefits of RPM in controlling chronic diseases, detecting health abnormalities early, and reducing hospital readmissions, as highlighted by Malasunghe et al. [5]. Ambient intelligence (AmI) has the potential to revolutionize RPM by seamlessly integrating sensors and AI into the environment to collect and analyze patient data without requiring explicit user interaction. According to Hristoskova et al. [6], AmI has emerged as a viable solution for RPM, providing personalized healthcare and early detection of exacerbations in chronic illnesses. By passively monitoring vital signs, activity levels, and environmental factors, AmI systems can provide continuous, unobtrusive health monitoring. However, the widespread adoption of AmI-based RPM systems is hindered by several challenges such as lack of transparency, privacy concerns, and algorithmic bias [7]. As AI algorithms become increasingly complex, it becomes difficult to understand their decision-making processes, leading to a lack of trust and accountability.

The importance of transparency, explainability, and fairness in AI systems has been underscored by established guidelines and frameworks. The EU GDPR provides stringent standards for data protection and privacy, emphasizing the need for organizations to implement mechanisms that ensure the responsible handling of personal data. GDPR mandates transparency in how data is collected, stored, and processed, along with ensuring individuals' rights to understand automated decision-making processes [8].

RPM systems encounter plethora of challenges in privacy and transparency. Privacy concerns are a significant issue for healthcare practitioners using RPM technologies [9]. Additionally, the rapid adoption of AI in RPM systems introduces new privacy concerns, particularly in federated learning for personalized patient monitoring [10]. It has also been highlighted by Shaik et al. [10] that the use of Internet of Things (IoT) wearable devices and sensors and other technologies in RPM architectures introduces additional privacy and transparency challenges that need to be carefully managed. RPM systems typically gather vast amounts of sensitive health data, including personal health records, biometric measurements, and behavioral data. In this context, many patients remain unaware of how their health data is used, who has access to it, and how it is shared. This lack of transparency has contributed to a growing

distrust of AI-based healthcare technologies, with patients expressing concerns about potential misuse of their data and the opacity of AI models driving decision-making.

XAI can address these challenges by making AI models and systems interpretable and by providing clear, understandable explanations for their decision-making processes. In healthcare, XAI can help bridge the trust gap between patients and AI systems by offering insights into how and why a particular decision, recommendation, or alert is generated. This transparency allows patients and healthcare professionals to enhance their understanding of the reasons for AI-driven decisions and improve patient engagement. XAI can help identify patient data points that have the greatest impact on the decision. It also illustrates how a decision would have changed under different conditions or data inputs, allowing patients to understand the boundaries of decision-making in the system. Various XAI methods have been applied to healthcare. Sadeghi et al. [11] identified challenges in applying XAI to healthcare, such as the trade-off between model complexity and interpretability, and emphasize the need for collaboration between data scientists and medical experts.

2.1. XAI Algorithms for AI-based RPM systems

Various XAI algorithms, including SHAP, LIME, and Saliency Maps, offer different interpretability approaches, each with unique strengths and limitations. To protect privacy, several techniques have been developed in addition to enabling meaningful data analysis and decision-making. Saifullah et al. [12] investigate the impact of privacy-preserving machine learning (PPML) on XAI aiming to understand how different privacy techniques affect the quality of explanations generated by XAI methods. Federated learning and hybrid approaches are prominent privacy-preserving methods for AI-based healthcare applications [13]. XAI techniques such as saliency maps and rule-based explanations can help interpret complex AI models [8].

SHAP is a model-agnostic XAI method and an excellent solution for providing both local explanations for individual predictions and global explanations for overall feature importance. This dual capability makes the SHAP highly effective for RPM systems. It can help clinicians both explain a critical event for a specific patient and understand the general patterns across the dataset. However, the computational cost of SHAP for complex models can be challenging in real-time scenarios. LIME locally approximates the predictions of a complex AI model around a specific instance by creating a surrogate interpretable model. For RPM, LIME is beneficial for explaining individual patient alerts, but falls short for comprehensive feature importance analysis, which is often critical in medical contexts. Saliency maps are gradient-based techniques that are designed primarily for neural networks. Saliency maps are gradient-based techniques that are typically used for visual data. Table 1 compares the strengths of the three XAI algorithms.

Table 1. Comparison of XAI Algorithms for RPM

Feature	**LIME**	**SHAP**	**Saliency Maps**
Model Agnostic	Yes	Yes	No, specific neural networks
Interpretation Scope	Local	Local + Global	Local
Data Type	Tabular, Text, Images	Tabular, Text, Images	Primarily images
Precision	Approximation-based	Game-theory	Gradient-based

2.2. Privacy-Preserving Techniques in Healthcare AI

In addition to transparency, privacy preservation is a critical concern for healthcare AI. The sensitive nature of health data necessitates strong safeguards to prevent unauthorized access or misuse. Several techniques have been developed to protect privacy while still enabling meaningful data analysis and decision making. Saifullah et al. [12] investigate the impact of privacy-preserving machine learning (PPML) on XAI aiming to understand how different privacy techniques affect the quality of explanations generated by XAI methods. Privacy preservation in healthcare is crucial due to the sensitive nature of patient data and increasing security threats. Various techniques have been developed to address this challenge. Federated learning and hybrid approaches are prominent privacy-preserving methods for AI-based healthcare applications [13].

Anonymization of patient data before sharing with healthcare providers or third parties ensures that personally identifiable information (PII) is not exposed. Encrypting patient data during transmission and storage prevents unauthorized access. Chong *et al.* [14] has focused on the recently proposed schemes based on data anonymization and differential privacy approaches in the protection of healthcare data privacy, highlighted the strengths and limitations of these two approaches and discussed some promising future research directions in this area. Abbas and Khan [15] reviewed various cryptographic and non-cryptographic privacy-preserving approaches used in e-health clouds to address patient data privacy concerns.

Another technique is Federated Learning (FL), in which AI models are trained across decentralized data sources, where sensitive data never leave the patient's device. This approach allows models to learn from patient data without compromising privacy. Supporting this, Differential Privacy (DP) is a technique that ensures that any data shared or analyzed does not compromise the privacy of individual patients by introducing noise into the data, making it difficult to trace data back to any specific individual. DP is critical for maintaining patient confidentiality [16]. However, balancing privacy and explainability remains a key research question for healthcare AI.

3. Contributions

This paper introduces the conceptual framework of X-PRISM, addressing critical challenges in the domain of RPM.

1. Integration of XAI for Enhanced Transparency in RPM Systems: This paper proposes the integration of XAI techniques, particularly SHAP, into the RPM systems. By providing interpretable explanations for AI model predictions, the framework ensures transparency in the decision-making processes and supports GDPR Compliance.
2. Privacy-Preserving Data Handling Through Federated Learning and Pseudonymization: X-PRISM minimizes the risks of data breaches by decentralizing model training and anonymizing patient data. This dual-layered approach supports compliance with privacy standards and enhances the data security. Thus, X-PRISM is particularly suitable for sensitive healthcare applications.
3. Scalable RPM systems with Cloud-Based Solutions: X-PRISM leverages Cloud Services for a complete lifecycle, including data collection, preprocessing, AI model training, explanation generation, and feedback.

4. Proposed Framework: X-PRISM (Explainable Privacy-Preserving Intelligent System for Monitoring)

The main goal of this study was to address the black-box nature of AI-based decision systems. Shaik et al. [10] explained the importance of AI in RPM but failed to provide clear explanations for decisions. Multiple studies support the importance of AI in healthcare, but explainability needs urgent attention. In response to the increasing demand for trust and transparency in RPM systems, this study proposes the X-PRISM framework. Based on the comparison in Table 1, SHAP is primarily used to enhance transparency of the decisions. The X-PRISM framework offers a complete solution for striking a balance between privacy and transparency in RPM systems by integrating both characteristics, guaranteeing that patients may trust the technology while still having control over their personal health information. This paradigm is a crucial step toward the moral implementation of RPM systems in contemporary healthcare, because it is intended to increase user trust and make it easier to comply with privacy laws. Figure 1 shows the four core layers in the proposed X-PRISM framework.

4.1. Core Layers of the X-PRISM Framework for Remote Patient Monitoring

The X-PRISM framework consists of four main layers, each of which plays a crucial role in ensuring privacy, security, and transparency during the decision-making process.

1. **Data Collection and Preprocessing Layer**: The process starts with wearable healthcare devices worn by patients. These devices collect real-time patient data such as heart rate, temperature, and blood pressure. The collected data are transmitted to the Azure IoT Hub after processing it locally and sending only model updates. Data

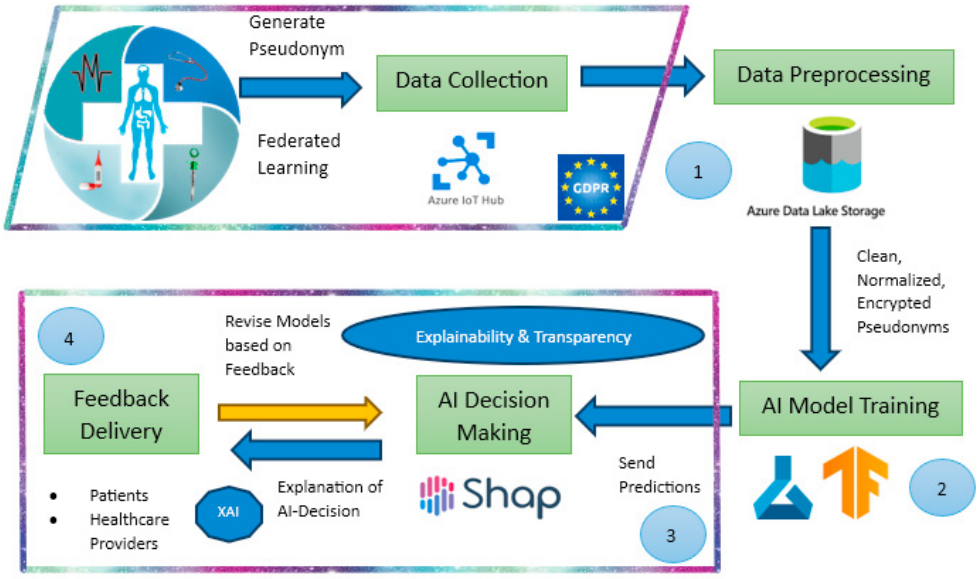

Fig. 1. Process Flow in X-Prism Framework

cleaning is done using Azure Databricks. For privacy compliance with GDPR, Azure Data Factory handles this pseudonymization of patient data and stores encrypted data in Azure Data Lake Storage.

2. **AI Model Development**: Using TensorFlow Federated, an AI model is trained on the preprocessed data to predict patient health status or detect anomalies in real time. TensorFlow Federated allows models to be trained across multiple devices without data leaving the device.
3. **AI Decision Layer**: Once the global model is trained through federated learning, it is deployed for real-time predictions on new, incoming patient data. The decision layer can use the updated global model to make predictions for individual patients. XAI techniques such as SHAP are utilized to provide clear, human-understandable explanations of model predictions.
4. **Feedback layer**: Azure Application Insights is leveraged to track user interactions with the system, such as clinician input or corrections to AI predictions. This facilitates continuous improvement of the model's accuracy and interoperability. It also enables patients to provide feedback on explanations and adjust privacy settings.

Each layer plays a significant role in ensuring the appropriate handling of sensitive health data while delivering clear, understandable, and actionable insights. The X-PRISM framework is designed to address the privacy and transparency challenges inherent in ambient systems for RPM.

4.2. Effectiveness of Feedback Layer

The statistical analysis of the feedback layer can be done to quantify how clinician feedback loops reduce false positives and/or negatives in AI predictions over multiple iterations. This layer can help reduce False Positives (FPs) and False Negatives (FNs) in AI predictions over multiple iterations. In this process a synthetic RPM dataset containing features like heart rate and temperature can be utilized. This data is annotated with ground truth labels - "Normal" or "Anomaly". Furthermore, the feedback layer will incorporate corrections from clinicians or healthcare professionals who label predictions as correct or incorrect. The true labels will be fed back into the training process to improve the model iteratively for incorrect predictions (FPs or FNs). The model is retrained after each feedback cycle using corrected data. At each iteration, measurements of the changes in the FPs and FNs are captured. In order to visualize how FP rate and FN rate decrease with each feedback iteration, a line plot can be created.

4.3. Transparency and Explainability Using SHAP

The Explainability and Transparency are the heart of X-PRISM, providing transparency to the AI system's decision-making process. The explanations provided by these techniques refer to pseudonymized identifiers instead of actual patient data. One of the examples to give more explanation to the patient can be, "Your recent increase in heart rate (by 20 bpm) is the primary reason behind the high-risk alert for arrhythmia." A simple algorithm is shown in Figure 2 to understand the implementation of SHAP or the explainability of AI-based decisions in the RPM. This study leverages source code from an open-source repository [17], which provides implementation details for SHAP.

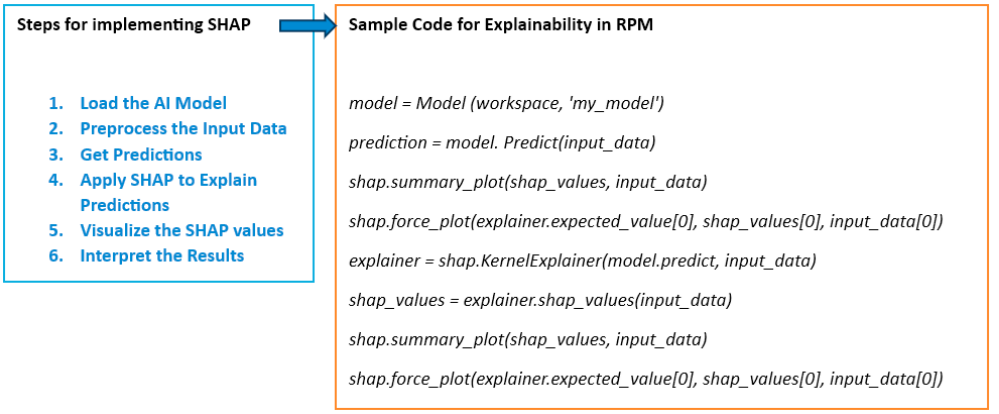

Fig. 2. SHAP Implementation

The interpretation of the results was based on the SHAP summary plot, which shows the features ordered by importance. Features with larger SHAP values had a greater influence on the output of the model. The force plot shows how the feature values contributed to the prediction, showing whether they pushed the model's output towards a certain class. For example, if the heart rate feature has a strong positive SHAP value, it suggests that a higher heart rate significantly contributed to the patient being classified as "at risk." The DeepExplainer is most suitable for the AI-based RPM system built using deep learning models, for time-series patient data.

5. Ethical and Technical Considerations

Despite the benefits, integrating XAI and privacy-preserving mechanisms in ambient systems for RPM raises both ethical and technical challenges. Hence, it is critical to address these challenges to ensure that the proposed framework, X-PRISM, operates responsibly while achieving its goals of privacy transparency and trustworthy decision making. One of the primary ethical concerns in RPM systems is maintaining the privacy and ownership of sensitive patient data. Ethically, patients must retain control over their data with clear policies defining data usage, sharing, and storage.

Guleria et al. [18] propose the use of XAI in the prediction of cardiovascular diseases but do not adequately address privacy concerns, potentially exposing sensitive patient information. X-PRISM emphasizes strong data encryption and anonymization techniques, ensuring patient privacy is prioritized during data collection and analysis. Thalpage [8] emphasizes that one of the important aspects of responsible AI deployment is to ensure accountability. XAI techniques enable stakeholders to trace the decision-making process and to understand the factors that lead to specific outcomes. The development of the X-PRISM framework adheres to the key principles outlined in international guidelines, such as the GDPR and standards for ethical AI. It is always a possibility that AI models may unintentionally reinforce biases that are present in the training data, leading to unfair or inaccurate predictions for specific demographic groups. Ensuring fairness in predictions is an important ethical obligation. Techniques such as fairness-aware machine learning and regular auditing of model performance can be used to mitigate the bias in X-PRISM. Barocas et al. [19] emphasized that addressing fairness in machine learning is essential for ensuring equitable outcomes in AI systems, particularly in sensitive applications, such as healthcare.

Technically, one critical consideration is scalability, because RPM systems generate vast amounts of real-time patient data. X-PRISM leverages federated learning, a decentralized approach in which data remain on edge devices, while models are trained collaboratively across multiple nodes. This approach reduces the computational burden on centralized servers and enhances privacy. However, as noted by McMahan et al. [20], maintaining the accuracy and robustness of federated models comparable to centralized systems is a key technical challenge. Another consideration is to ensure resilience against adversarial attacks to protect AI models from malicious actors that may manipulate data. Goodfellow et al. [21] highlight that adversarial training and robust evaluation mechanisms are necessary to mitigate such risks and ensure reliable AI performance in healthcare applications. Additionally, the resource constraints of ambient systems, such as wearables and IoT devices, necessitate energy-efficient and lightweight AI algorithms. Sze et al. [22] provide comprehensive guidelines on optimizing deep neural networks for low-power environments, which can be applied to enhance X-PRISM's performance on resource-constrained devices. These considerations are critical to ensure the practicality and reliability of X-PRISM in real-world healthcare scenarios. A systematic review by Aziz et al. [23] discussed the use of XAI in healthcare. Most of the existing frameworks lack feedback loops. X-PRISM encompasses a feedback loop that allows healthcare providers and users to contribute insights based on their experiences with the AI system, and thus helps in the continuous refinement of the model's accuracy and interoperability.

6. Future Research Directions

X-PRISM provides a strong foundation for privacy transparency and decision making explainability in RPM systems. In the present work, pseudonymization is employed. Future work can explore the use of Differential Privacy and Homomorphic encryption to improve the privacy and security of patient data as they can reduce the risk of re-identification. Techniques other than SHAP can be considered to improve the interpretability in various scenarios. Real-time integration of the feedback layer will improve the responsiveness of the framework. Future research could also include fairness metrics to identify and mitigate potential biases. Federated Learning and Azure Cloud both require a considerable amount of computational resources. Hence, future research should focus on improving energy efficiency and sustainability.

The current X-PRISM framework provides transparency but does not fully address the challenge of real-time explainability. It is very important that the integration of AI-based decision-making explanations is timely, accurate, and without delay, especially in scenarios like emergency alerts or critical health events. Future research could focus on scaling XAI without compromising its speed and usability, especially in systems that handle thousands of patients. Research should explore ways to protect explainable AI systems from adversarial attacks such as data poisoning or model inversion attacks, which could lead to erroneous or misleading explanations. Finally, future research can focus on extending the X-PRISM framework to other healthcare domains beyond the RPM, Clinical Decision Support Systems, and Telemedicine.

7. Conclusion

This research emphasizes the importance of XAI in improving the privacy transparency inside ambient systems for RPM. By making data processing and AI decision-making more transparent, XAI can increase patient trust, ensure privacy compliance, and improve system usability. This study presents an Explainable Privacy-Preserving Intelligent System for Monitoring (X-PRISM) framework designed to enhance transparency, privacy, and trust in AI-based RPM systems. This framework addresses critical challenges in healthcare monitoring by integrating XAI, privacy-preserving mechanisms, and an iterative feedback loop that incorporates clinician expertise.

Using the layered architecture of X-PRISM, this study demonstrates how data are collected from healthcare wearables and IoT sensors, pseudonymized, preprocessed, and fed into an AI model developed using TensorFlow. The AI decision layer, supported by XAI tools, such as SHAP, provides explainability by identifying and quantifying the most influential features in predictions, such as heart rate and temperature. These insights are then utilized with Azure Cloud Services to deploy, process, and retrain the AI model iteratively, thereby creating a robust and scalable system. The feedback layer is a cornerstone of this framework, which enables the continuous refinement of AI predictions through clinician inputs. The proposed method for statistical evaluation shows that this iterative approach effectively

reduces false positives and false negatives, with an improvement in the overall model accuracy over multiple iterations. X-PRISM can redefine the role of AI in healthcare monitoring and ensure safe, reliable, and patient-centric solutions for smart medical systems.

References

[1] Ayushmaan Dubey and Anuj Tiwari. Artificial intelligence and remote patient monitoring in us healthcare market: a literature review. *Journal of Market Access & Health Policy*, 11(1):2205618, 2023.

[2] Dhruvitkumar Talati. Telemedicine and ai in remote patient monitoring. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, 2(3):254–255, 2023.

[3] Thanveer Shaik, Xiaohui Tao, Niall Higgins, Lin Li, Raj Gururajan, Xujuan Zhou, and U Rajendra Acharya. Remote patient monitoring using artificial intelligence: Current state, applications, and challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13 (2):e1485, 2023.

[4] Christoforos N Spartalis, Theodoros Semertzidis, and Petros Daras. Balancing xai with privacy and security considerations. In *European Symposium on Research in Computer Security*, pages 111–124. Springer, 2023.

[5] Lakmini P Malasinghe, Naeem Ramzan, and Keshav Dahal. Remote patient monitoring: a comprehensive study. *Journal of Ambient Intelligence and Humanized Computing*, 10:57–76, 2019.

[6] Anna Hristoskova, Vangelis Sakkalis, Giorgos Zacharioudakis, Manolis Tsiknakis, and Filip De Turck. Ontology-driven monitoring of patient's vital signs enabling personalized medical detection and alert. *Sensors*, 14(1):1598–1628, 2014.

[7] Albert Haque, Arnold Milstein, and Li Fei-Fei. Illuminating the dark spaces of healthcare with ambient intelligence. *Nature*, 585(7824): 193–202, 2020.

[8] N Thalpage. Unlocking the black box: Explainable artificial intelligence (xai) for trust and transparency in ai systems. *Journal of Digital Art & Humanities*, 4(1):31–36, 2023.

[9] Luiza Palmieri Serrano, Karla C Maita, Francisco R Avila, Ricardo A Torres-Guzman, John P Garcia, Abdullah S Eldaly, Clifton R Haider, Christopher L Felton, Margaret R Paulson, Michael J Maniaci, et al. Benefits and challenges of remote patient monitoring as perceived by health care practitioners: a systematic review. *The Permanente Journal*, 27(4):100, 2023.

[10] Thanveer Shaik, Xiaohui Tao, Niall Higgins, Lin Li, Raj Gururajan, Xujuan Zhou, and U Rajendra Acharya. Remote patient monitoring using artificial intelligence: Current state, applications, and challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13 (2):e1485, 2023.

[11] Zahra Sadeghi, Roohallah Alizadehsani, Mehmet Akif Cifci, Samina Kausar, Rizwan Rehman, Priyakshi Mahanta, Pranjal Kumar Bora, Ammar Almasri, Rami S Alkhawaldeh, Sadiq Hussain, et al. A brief review of explainable artificial intelligence in healthcare. *arXiv preprint arXiv:2304.01543*, 2023.

[12] Saifullah Saifullah, Dominique Mercier, Adriano Lucieri, Andreas Dengel, and Sheraz Ahmed. The privacy-explainability trade-off: unraveling the impacts of differential privacy and federated learning on attribution methods. *Frontiers in Artificial Intelligence*, 7:1236947, 2024.

[13] Nazish Khalid, Adnan Qayyum, Muhammad Bilal, Ala Al-Fuqaha, and Junaid Qadir. Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, 158:106848, 2023.

[14] Kah Meng Chong. Privacy-preserving healthcare informatics: a review. In *ITM Web of Conferences*, volume 36, page 04005. EDP Sciences, 2021.

[15] Assad Abbas and Samee U Khan. A review on the state-of-the-art privacy-preserving approaches in the e-health clouds. *IEEE journal of Biomedical and health informatics*, 18(4):1431–1441, 2014.

[16] Steven M Williamson and Victor Prybutok. Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in ai-driven healthcare. *Applied Sciences*, 14(2):675, 2024.

[17] GitHub - shap/shap: A game theoretic approach to explain the output of any machine learning model. — github.com. https://github.com/shap/shap, 2024. [Accessed 17-01-2025].

[18] Pratiyush Guleria, Parvathaneni Naga Srinivasu, Shakeel Ahmed, Naif Almusallam, and Fawaz Khaled Alarfaj. Xai framework for cardiovascular disease prediction using classification techniques. *Electronics*, 11(24):4086, 2022.

[19] Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and machine learning: Limitations and opportunities*. MIT press, 2023.

[20] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[21] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[22] Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, and Joel S Emer. Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12):2295–2329, 2017.

[23] Noor A Aziz, Awais Manzoor, Muhammad Deedahwar Mazhar Qureshi, Muhammad Atif Qureshi, and Wael Rashwan. Explainable ai in healthcare: Systematic review of clinical decision support systems. *medRxiv*, pages 2024–08, 2024.