



EXPLAINABLE ARTIFICIAL INTELLIGENCE IN HEALTHCARE

Problem Statement and Research Objectives

Jul 2026

Abdullah S. Basaad

Table of contents:

1. Problem Statement
2. Research Objectives
3. References

1. PROBLEM STATEMENT

Artificial Intelligence (AI) became an essential technology in modern healthcare, supporting applications such as disease diagnosis, medical imaging, patient monitoring, and clinical decision-making. The recent improvements in machine learning and deep learning have improved the accuracy of healthcare prediction models. However, even though with high performance, many AI systems operate as black-box models, making it difficult for healthcare professionals to understand how predictions are generated. Aravindkumar et al. stated that numerous studies have shown that models can achieve accurate predictions despite being based on incorrect assumptions (2026). This lack of transparency reduces clinicians' confidence in AI systems and limits their adoption in real-world healthcare environments (Dave et al., 2020; Kenigson, 2025).

Existing Explainable Artificial Intelligence (XAI) techniques, including SHAP, LIME, and Grad-CAM, is proposed to improve the interpretability of AI models by providing explanations for their predictions. These methods enable clinicians to identify the factors influencing diagnostic decisions and improve trust in AI-assisted healthcare systems (Panda et al., 2023; Amin et al., 2024). Although these techniques have shown good results, several challenges remain. Most existing studies rely on publicly available datasets with limited patient diversity, while only a small number evaluate their models using real clinical data or involve healthcare professionals in assessing the usefulness of AI explanations (AlKhanbouli et al., 2025).

Developing trustworthy and interpretable AI systems is becoming important because healthcare decisions directly affect patient safety and cure outcomes. Explainable AI can improve transparency, support clinical decision-making, reduce diagnostic errors, and encourage healthcare professionals to adopt AI technologies with greater confidence. Furthermore, transparent AI systems help satisfy ethical, legal, and regulatory requirements for medical applications, making explainability an essential component of future healthcare AI systems (Kenigson, 2025).

Several studies have attempted to address the black-box problem by integrating explainability methods with machine learning and deep learning models. SHAP and LIME have become the most widely adopted techniques for explaining structured healthcare data, while Grad-CAM is commonly applied to medical imaging applications. Although these approaches improve model transparency, most studies still focus on prediction accuracy rather than evaluating the quality and clinical usefulness of the generated explanations. Additionally, there is currently no standardized framework for evaluating explainability across different healthcare applications (Aravindkumar et al., 2026).

Therefore, more researching is needed to investigate the current state of Explainable Artificial Intelligence in healthcare by reviewing existing explainability techniques, identifying their strengths and limitations, and examining how they can be effectively integrated into clinical practice. Such research will help identify current knowledge gaps and provide recommendations for developing more transparent, trustworthy, and clinically applicable AI systems that better support healthcare professionals in making informed medical decisions (AlKhanbouli et al., 2025; Aravindkumar et al., 2026).

2. RESEARCH OBJECTIVES

The primary objective of this research is to investigate the current applications of Explainable Artificial Intelligence in healthcare and evaluate how explainability techniques improve the transparency and trustworthiness of AI-based medical decision-making.

The specific objectives of this research are:

- **To identify the Explainable Artificial Intelligence techniques currently used in healthcare applications.**
 - Will help finding strengths and weakness for each technique, and therefore exploring ways to improve.
 - Differentiate between the techniques will help to compare between them.
- **To examine the machine learning and deep learning models integrated with Explainable AI methods.**
 - To help validating the outcome of XAI.
- **To compare the strengths and limitations of existing Explainable AI techniques used in healthcare.**
 - To help deciding which technique is most suitable in application cases.
 - To help avoiding the use a particular technique in some cases and prefer others.
- **To evaluate the research methodologies and datasets employed in previous studies.**
 - To make sure results were actually generated from accurate assumption and valid data.
 - To ensure covering all possible population and variabilities.
- **To identify the current challenges and research gaps affecting the adoption of Explainable Artificial Intelligence in healthcare.**
 - To guide the way for future improvements.
- **To recommend future research directions for developing more transparent, interpretable, and trustworthy AI systems for healthcare applications.**
 - To provide practical solutions that could close gaps found.

REFERENCES

- Amin, A., Hasan, K., Zein-Sabatto, D., Chimba, S., Ahmed, I., & Islam, T. (2024). *An explainable AI framework for Artificial Intelligence of Medical Things*. arXiv. <https://arxiv.org/abs/2403.04130>
- Aravindkumar, A., Ramadoss, M., Ahmed, S., Ahmed, F., Sampath, V., & Lakshminarayanan, K. (2026). *Explainable AI in healthcare: A systematic review of XAI use cases in imaging, diagnostics, and rehabilitation*. *Healthcare*, 14(1), 92. <https://doi.org/10.3390/healthcare14010092>
- Dave, D., Naik, H., Singh, S., & Patel, P. (2020). *Explainable AI meets healthcare: A study on heart disease dataset*. arXiv. <https://arxiv.org/abs/2011.03195>
- Kenigson, J. (2025). *Explainable AI for healthcare: Developing transparent and interpretable models for medical diagnosis*. ResearchGate. https://www.researchgate.net/publication/394917443_explainable_ai_for_healthcare_developing_transparent_and_interpretable_models_for_medical_diagnosis
- Panda, M., Mahanta, S. R., & Mahanta, R. (2023). *Explainable artificial intelligence for healthcare applications using Random Forest classifier with LIME and SHAP*. arXiv. <https://arxiv.org/abs/2311.05665>
- Rajan AlKhanbouli, R., Almadhaani, H. M., Alhosani, F., & Simsekler, M. C. E. (2025). *The role of explainable artificial intelligence in disease prediction: A systematic review*. *BMC Medical Informatics and Decision Making*, 25. <https://doi.org/10.1186/s12911-025-02944-6>