



EXPLAINABLE ARTIFICIAL INTELLIGENCE IN HEALTHCARE

A Literature Review

Jul 2026

Abdullah S. Basaad

Table of contents:

1. Definition Of Terms
2. Research Papers's Summary
3. Critical Analysis
4. Synthesis Matrix For Literature Review
5. References

1.Definition of terms	
Term	Description
XAI	Explainable Artificial Intelligence
SHAP	Shapley Additive explanations
LIME	Local Interpretable Model-Agnostic Explanations
Grad-CAM	Gradient-weighted Class Activation Mapping
AIoMT	Artificial Intelligence of Medical Things
AI	Artificial Intelligence
Random Forest	Artificial Intelligence Algorithm
XGBoost	eX- treme Gradient Boosting

2. RESEARCH PAPERS'S SUMMARY

1. Explainable AI Meets Healthcare: A Study on Heart Disease Dataset

The paper explains how Explainable Artificial Intelligence (XAI) can improve the accuracy of machine learning models used for heart disease prediction. It states that even though AI models can achieve high accuracy, their hidden way of work makes it difficult for healthcare professionals to understand and trust the predictions. To overcome this issue, the study applies several XAI techniques to a heart disease prediction model built using a heart disease dataset and a classification algorithm. The paper uses many explanation methods, such as LIME and SHAP, Anchors, Counterfactual Explanations, Counterfactuals with Prototypes, Contrastive Explanation Method (CEM), Kernel SHAP, and Integrated Gradients. Using these methods, the study shows how different patient features contribute to prediction outcomes and provides explanations for the model's decisions. The results show that factors such as the number of blocked blood vessels, chest pain type, cholesterol level, and thalassemia have the greatest influence on predicting heart disease. Integrating techniques with machine learning models can increase transparency, helps clinicians to better understand AI predictions, and increase trust in using AI for medical decisions.

2. Explainable Artificial Intelligence for Healthcare Applications using Random Forest Classifier with LIME and SHAP

The paper provides an overview of Explainable Artificial Intelligence (XAI) and describes its application in healthcare using a Random Forest classifier combined with LIME and SHAP explainability techniques. The paper highlights the increasing adoption of AI in healthcare and emphasize that many machine learning and deep learning models operate as black boxes making it difficult for users to understand how predictions are generated. To address this issue, the paper reviews XAI concepts including explainability, transparency, interpretability, trustworthiness, and quality evaluation criteria. The experimental study uses a publicly available diabetes dataset from Kaggle, where a Random Forest classifier is employed to predict diabetes, and SHAP and LIME are applied to explain model behavior. The study demonstrates that integrating XAI techniques with machine learning models improves transparency, interpretability, and trust without affecting prediction quality. The authors conclude that explainable AI has significant potential to support healthcare decision-making by allowing clinicians to better understand AI-generated predictions. They also suggest future research should explore more explainability techniques, and enhance the trustworthiness and practical deployment of explainable AI systems in real-world healthcare environments.

3. Explainable AI in Healthcare: A Systematic Review of XAI Use Cases in Imaging, Diagnostics, and Rehabilitation

The paper answers the question of how Explainable Artificial Intelligence (XAI) is being used to increase transparency and trust in healthcare AI systems. The authors examined 36 studies in the last 5 years on three healthcare domains: medical imaging, clinical diagnosis, and

rehabilitation. The study emphasizes that traditional machine learning and deep learning models operate mysteriously which leads to complicate the understanding of the reasoning behind predictions for healthcare professionals. To overcome this challenge, XAI techniques such as SHAP, LIME, and Grad-CAM can be used to describe model decisions. The findings have proofed that Grad-CAM is commonly used in medical imaging to generate visual heatmaps, but SHAP and LIME are primarily used in diagnosis and rehabilitation to explain individual patient predictions. Several studies have combined multiple explanation methods to improve reliability and clinician confidence. However, the paper also mentions several limitations such as limited external validation, inconsistent evaluation of explanation quality, reliance on single-center datasets, and the lack of standardized reporting and faithfulness metrics. It then suggests that future research should focus on developing standardized frameworks for evaluating XAI, generating dataset diversity and integrating explainability into real clinical workflows.

4. An Explainable AI Framework for Artificial Intelligence of Medical Things (AIoMT)

The paper presents a new Explainable Artificial Intelligence (XAI) framework intended to increase the transparency, reliability, and accuracy of AI-driven healthcare systems, especially for brain tumor detection using MRI images. The proposed framework combines three popular XAI techniques: LIME, SHAP, and Grad-CAM to deliver understandings for the model's predictions, allowing healthcare professionals to understand why an analysis was made instead of depending on a unexplained system. The architecture introduces cloud and edge computing within an Artificial Intelligence of Medical Things (AIoMT) environment, allowing efficient processing of large medical datasets while supporting simultaneous diagnosis. The ensemble model combines predictions from five convolutional neural networks, leading to improved performance and reduced prediction errors. Experimental evaluation using the Br35H brain tumor MRI dataset demonstrated outstanding performance, achieving 99% training accuracy, 98% validation accuracy, and high precision, while visual explanations generated by SHAP, LIME, and Grad-CAM emphasized the image regions and features influencing each prediction. These explanations increase clinician trust, improve interpretability, and support more informed medical decision-making. The paper showed that integrating explainable AI with AIoMT has significant potential to enhance healthcare outcomes by providing accurate, transparent, and trustworthy diagnoses. It also directs future work to focus on extending the framework to other diseases, improving explainability methods and incorporating real-time patient data.

5. The Role of Explainable Artificial Intelligence in Disease Prediction

This paper defines how Explainable Artificial Intelligence (XAI) is moving transparency and trust in AI-based disease prediction systems to higher levels. Using the PRISMA methodology, 30 studies published in the last years were analyzed, covering cardiovascular diseases among other diseases. In result, SHAP and LIME were found as the most used XAI techniques because they assist clinicians to understand how AI models get to their predictions, making diagnostic decisions more transparent and trustworthy. Traditional AI has proved its accuracy in predicting diseases, however, clinicians find it hard to give their full trust to it as they do not know how it really works. The paper numbers many key challenges, including

limited dataset diversity, reliance on single data modalities, and rarity of user-friendly interfaces for healthcare professionals. To avoid these challenges, the paper suggests the following: standardized XAI frameworks, more diverse datasets, improved clinician training, and continued researching into transparent AI systems that can be integrated into everyday medical practice.

6. Explainable AI for Healthcare: Developing Transparent and Interpretable Models for Medical Diagnosis

This paper explores the growing importance of Explainable Artificial Intelligence (XAI) in making healthcare AI systems more transparent, trustworthy, and clinically useful. While advanced machine learning and deep learning models achieve high diagnostic accuracy in applications, their "black-box" nature limits physician trust and practical adoption. The paper reviews model-agnostic explanation methods, including LIME, SHAP, and LORE and Model-specific such as decision trees and Attention Mechanisms, explaining how these techniques help clinicians understand the reasoning behind AI predictions. Through a comprehensive review of existing studies and real-world healthcare applications, the authors demonstrate that explainable AI improves clinical confidence, patient safety while maintaining high accuracy diagnostic performance. Although XAI models may experience a slight reduction in predictive accuracy compared to traditional black-box models, the increased transparency enhances their acceptance in clinical practice. The paper concludes that explainability should be considered a fundamental requirement rather than an optional feature in medical AI, as it enables healthcare professionals to validate AI-generated recommendations, supports ethical and legal standards, and promotes patient-centered decision-making. The authors recommend researching AI models that combine the predictive strength of deep learning with explainability techniques to accelerate the responsible adoption of AI in healthcare.

3. CRITICAL ANALYSIS

3.1 Strengths

All papers have proved high prediction accuracy when finding diseases using AI with improved transparency and trust through XAI. They have agreed on the effective use of SHAP and LIME as XAI explainability methods, with some using also Grad-CAM. Well-established machine learning models. Most papers have used publicly available datasets, in particular UCI Cleveland Heart Disease Dataset, allowing for other researchers to verify findings and redo the experiments. Most papers followed a structured and systematic research methodology. The explainability methods (e.g., SHAP and LIME) were also clearly described, making it easier to understand how predictions were interpreted. Moreover, the usefulness of XAI for supporting clinical decision-making was proved by enabling healthcare professionals to understand how a model reached into an outcome instead of only providing a prediction. SHAP and LIME help clinicians to make sure that AI's reasoning is medically meaningful, therefore, increasing clinicians' confidence in AI diagnosis.

3.2 Weaknesses and Limitations

Most of the papers used the UCI Cleveland Heart Disease Dataset, which is a well-established and widely accepted dataset, however, relying on a single dataset may limit the diversity of patient populations, different ethnic groups, and patients from different geographical regions. The papers also were developed using small datasets which may lead to unstable predictions when applied to larger populations. Few papers validated their systems using data collected directly from hospitals or healthcare institutions, but most papers did not evaluate their models in real clinical environments. This caused the papers to miss clinicians' point of view, who are the ones who will actually be using the models. Also, technical cost such as execution time and memory requirements were not discussed in the papers.

3.3 Connections

- All papers shared the objective of increasing transparency and trust in AI-based medical prediction.
- The papers recognized that traditional AI models are difficult for clinicians to trust without clear explanations.
- XAI was presented as a solution for improving confidence on AI-generated diagnosis.
- SHAP and LIME were the most frequently adopted explainability techniques across the papers.
- All of them also agreed that XAI allows clinicians to better understand AI predictions.
- Most papers suggested use explainability in future healthcare AI systems instead of relying on predictions only.

4. SYNTHESIS MATRIX FOR LITERATURE REVIEW

Paper	Authors	Year & Research Design	Findings	Methodology	Limitations
Explainable AI meets Healthcare: A Study on Heart Disease Dataset	Devam Dave, Het Naik, Smiti Singhal, Pankesh Patel	2020-Experimental	SHAP and LIME improved prediction transparency; key heart disease features identified	XGBoost with SHAP, LIME, UCI Heart Disease dataset.	Only using data from a single dataset and no actual clinical experiemnt.

Explainable Artificial Intelligence for Healthcare Applications using Random Forest Classifier with LIME and SHAP	Mrutyunjaya Panda, Soumya Ranjan Mahanta	2023- Experimental	SHAP and LIME increased model interpretability while maintaining prediction performance.	Random Forest classifier with SHAP and LIME on Kaggle diabetes dataset.	Using a single dataset with no real-world clinical evaluation
Explainable AI in healthcare: a systematic review of XAI use cases in imaging, diagnostics, and rehabilitation	Apoorva Aravindkumar, Marimuthu Ramadoss, Saqhibuddeen Ahmed Fakhruddin Ahmed, Vidhya Sampath and Kishor Lakshminarayanan	2026- Review	XAI improves trust and transparency across healthcare applications.	Review of 23 XAI studies used SHAP, LIME, and Grad-CAM	No experimental validation, only study review.
An Explainable AI Framework for Artificial Intelligence of Medical Things	Al Amin, Kamrul Hasan, Saleh Zein-Sabatto, Deo Chimba, Imtiaz Ahmed, Tariqul Islam	2024- Experimental	Ensemble model with XAI achieved high accuracy and transparent diagnosis.	Ensemble with SHAP, LIME, and Grad-CAM on MRI dataset.	Focused on one disease and dataset.
The role of explainable artificial intelligence in disease prediction	Razan Alkhanbouli, Hour Matar Abdulla Almadhaani, Farah Alhosani and Mecit Can Emre Simsekler	2025- Reivew	SHAP and LIME are the most widely used XAI techniques for disease prediction.	PRISMA-based systematic review of 30 studies.	Review only; lacks experimental implementation.
Explainable AI for Healthcare: Developing Transparent and Interpretable Models for Medical Diagnosis	Jonathan Kenigson	2025- Review	XAI improves clinician trust, transparency, and patient safety.	Literature review of XAI methods and healthcare applications.	No original experiments; limited quantitative comparison.

4. 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>
- 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>
- Varadakumar, 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>