# Explainable AI in Healthcare Analytics and Clinical Decision Support Systems: A Systematic Literature Review and Research Framework

*Prakash Mohan[1], Prof. (Dr.) Shashi Kant Gupta[2]*
[1,2] Computer Science and Engineering, Lincoln University College, Petaling Jaya, Selangor, Malaysia.
Email ID: pdf.prakashmohan@lincoln.edu.my; shashigupta@lincoln.edu.my

**Abstract:** Artificial Intelligence (AI) has successfully provoked a paradigm shift in healthcare analytics with its ability to improve diagnosis, prognosis, patient risk assessment and clinical decisions. Many machine learning and deep learning models are high in predictive accuracy, but they are also black-box models, with insufficient transparency, interpretability and clinician trust. Explainable Artificial Intelligence (XAI) is a promising solution that will give understandable explanations for AI-generated predictions while maintaining predictive performance. In this survey, the authors comprehensively examine the latest developments in XAI for health analytics and Clinical Decision Support Systems (CDSS) based on studies published primarily between 2020 and 2026. The selected literature is analyzed with respect to healthcare applications, datasets, AI models, explainability methods, clinical usefulness, and existing limitations. Furthermore, the research directions, unmet challenges and gaps in knowledge on interpretability, fairness, trustworthiness and clinical adoption are broadly discussed. The survey gives a thorough overview of the current state of the art in XAI in healthcare and highlights future research directions to make intelligent healthcare systems more transparent, reliable, and clinically acceptable.

**Keywords**: Explainable Artificial Intelligence; Healthcare Analytics; Clinical Decision Support Systems; Machine Learning; Deep Learning; Interpretability.

## Introduction

Artificial Intelligence (AI) has become a vital part of present-day healthcare, contributing to accurate disease diagnosis, patient risk prediction, medical image interpretation, personalized treatment planning, and medical decision support. In recent years, machine learning (ML) and deep learning (DL) technologies have shown great promise in increasing predictive accuracy in a wide range of healthcare applications such as radiology, cardiology, oncology, pathology and intensive care monitoring. But as these models have grown more complex, they have added a major hurdle – their predictions are produced by black-box processes that are not easily understood by clinicians and tested against clinical data. Healthcare decisions are directly related to patient safety and treatment results, making transparency and explainability a critical requirement for the successful implementation of Clinical Decision Support Systems (CDSS) that rely on Artificial Intelligence (AI).

Explainable Artificial Intelligence (XAI) tackles this issue by generating explanations for the model's predictions in a way that is comprehensible to humans, thus fostering trust, accountability, and compliance by clinicians and regulatory bodies. There has been a growing number of explainability methods that have been utilized in healthcare to improve the interpretability of the AI systems such as

Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), Grad-CAM, saliency maps and attention-based visualization.

This survey presents a comprehensive analysis of pertinent literature on XAI for healthcare analytics and CDSS that have been recently published, which are mostly from the year 2020 to 2026. The reviewed studies are comparatively analyzed based on healthcare application, AI models, explainability techniques, evaluation strategies and existing limitations. The survey also highlights the ongoing research directions, pressing challenges, and knowledge gaps that remain relevant to the advancement of the development of transparent, trusted, clinically reliable AI systems.

**Related work**

Several recent review articles have laid the groundwork for EAI in healthcare by providing a comprehensive overview of current interpretability methods and applications and their use in the clinical setting. One of the earliest comprehensive reviews was by Loh et al. [1] which covered almost ten years of XAI research and showed how XAI is increasingly being used in healthcare applications, including methods like SHAP, LIME, Grad-CAM and Layer-wise Relevance Propagation. They stressed interpretability as a critical factor for clinical acceptance and observed that there was no a consensus on how to measure the quality of explanations.

Band et al. [2] then generalized this concept of healthcare explainability by summarizing the different interpretation methods and sorting them into three classes: the applicability in the context of disease diagnosis, medical imaging, and medical prediction. Their analysis found that explainability methods have improved over time, but the methods are not widely used in clinical workflow. Sadeghi et al. [8] classified the existing explainability methods in healthcare as feature attribution, surrogate modeling, visualization, concept-based reasoning and inherently interpretable algorithms.

One of the fastest-growing applications of explainable AI is CDSS since doctors need to understand the reasoning behind AI-driven suggestions before implementing. Gniadek et al. [3] presented a classification model, where explainability algorithms are classified based on the various phases of clinical decision making. They showed that the explainability requirements vary from different stages of the diagnosis, treatment planning and monitoring process, indicating that no single method is best for explainability.

For cardiovascular ICUs, Moazemi et al. [4] examined AI-driven CDSS and found that whilst ML models demonstrated impressive predictive accuracy, a lack of transparency was identified as a significant challenge for clinical implementation. Liu et al. [5] showed that embedding explainability earlier in the hospital workflow leads to an increased acceptance rate by clinicians and enables informed decision making.

Brankovic et al. [6] proposed explainable AI evaluation guidelines informed by clinicians. In their work, they recommended the criteria for the assessment include interpretability, usability, transparency, fairness, robustness, and clinical relevance. Another type of investigation in recent years highlights fairness-aware explainability, which considers whether AI models make fair predictions in different patient populations. Patient characteristic disparities are identified using bias detection (via SHAP), feature attribution analysis and subgroup evaluation.

In medical image analysis, Bendechache and Muhammad [7] did a broad survey of EAI in the field, and found that visual explanation methods increase clinician confidence by presenting regions of medical

images that are visually relevant to the anatomy of diseases. However, there are differences in the consistency of explanations depending on the architecture of neural networks and the type of data sets. Tun et al. [9] studied how trust in AI-driven CDSS affects the utilize of AI. They found that providing clear explanations improves clinician trust and confidence, leading to greater adoption of AI recommendations. The literature review of the selected literature reveals a number of emerging trends for explainable AI in healthcare. The first is that the technique is one of the most widely used due to its model-independent, local, and global feature attribution capabilities. Individual-level explanations continue to be provided using LIME. Second, attention-based explainability has become popular in transformer architectures and medical imaging applications where disease related regions are visually localized for better clinical interpretation.

Third, researchers are gradually integrating multiple explainability approaches into a single evaluation approach to gain complementary insights into model behavior. Other public healthcare datasets are also adopted to enhance the reproducibility and comparative assessment, including MIMIC-IV, CheXpert, MIMIC-CXR, eICU, ISIC, EyePACS and TCGA. In sum, XAI is now emerging as an integral part of trusted healthcare analytics and CDSS.

*Table 1. Comparative Analysis of Recent Explainable AI Studies in Healthcare*

| Ref. | Year | Healthcare Application | AI Model | Explainability Technique | Key Findings | Research Limitation |
|---|---|---|---|---|---|---|
| Loh et al. [1] | 2022 | General healthcare | ML/DL | SHAP, LIME, Grad-CAM, LRP | Comprehensive review of healthcare XAI applications | Lack of standardized evaluation framework |
| Band et al. [2] | 2023 | Medical diagnosis | ML/DL | SHAP, LIME, feature importance | Reviewed interpretability methods across healthcare applications | Limited clinical validation |
| Gniadek et al. [3] | 2023 | Clinical decision support | Multiple AI models | Clinical workflow-based XAI | Proposed XAI classification framework for clinical decision making | No unified evaluation methodology |
| Moazemi et al. [4] | 2023 | Cardiovascular ICU | ML/DL | Feature attribution | Improved patient monitoring using AI-based CDSS | Explainability requires further improvement |
| Liu et al. [5] | 2025 | Hospital decision support | ML/DL | XAI integration | Improved clinician acceptance through workflow integration | Generalization across hospitals remains limited |
| Brankovic et al. [6] | 2025 | Clinical evaluation | Multiple models | Clinician-informed evaluation | Proposed standardized XAI evaluation checklist | Validation across healthcare domains required |
| Muhammad [7] | 2024 | Medical imaging | CNN | Grad-CAM, saliency maps | Reviewed explainable medical imaging methods | Lack of standardized explanation metrics |
| Sadeghi et al. [8] | 2024 | General healthcare | ML/DL | SHAP, LIME, concept-based XAI | Categorized healthcare explainability methods | Comparative benchmarking required |

| | | | | | | |
|---|---|---|---|---|---|---|
| Tun et al. [9] | 2025 | Clinical decision support | Multiple AI models | Human-centered explainability | Investigated clinician trust in AI systems | Trust measurement requires standardization |
| Bilal et al. [10] | 2025 | Cardiovascular disease | Deep learning | SHAP | Enhanced interpretation of cardiovascular risk prediction | Disease-specific implementation |

Comparative analysis establishes numerous significant observations. First, post-hoc explainability methods dominate current healthcare research because they can be applied to existing ML and DL models without modifying model architecture. SHAP and LIME remain widely adopted owing to their flexibility and model independence, whereas Grad-CAM is preferred for medical image analysis.

Second, recent studies increasingly recognize that predictive accuracy alone is insufficient for clinical adoption. Researchers have expanded evaluation criteria to include interpretability, clinician trust, transparency, fairness, usability, and ethical considerations. Third, comparative evaluation remains inconsistent because studies use different datasets, metrics, explanation methods, and validation protocols. Recent publications increasingly focus on human-centered explainability, clinician-informed evaluation, and fairness-aware artificial intelligence.

**Identified Research Gaps**

- A variety of approaches to explainability have been suggested, but there is no standard way to measure the quality of explanations.
- The majority of explainability methods are tested with a limited set of post-facto data and in the controlled framework of experimentation.
- Many studies are targeted toward disease, specialty or institution. The generalization of explainable AI across a variety of diseases, hospitals, and patient populations is still an open challenge. Scalable and transferable XAI frameworks should be the focus of future work.
- AI systems used in healthcare can detect these patterns from past data, resulting in unequal diagnostic performance. Variations due to age, gender, ethnicity, disease prevalence, and access to health care are not well studied.

**Conclusions**

By helping to accurately diagnose diseases, predict patient risks, interpret medical images, and plan treatments, Artificial Intelligence has become an essential tool in the healthcare sector for clinical decision support systems and healthcare analytics. Complicated ML and DL models, however, pose problems with regard to transparency, interpretability, trust and clinical acceptance. XAI has proven to be a valuable tool because its explanations are comprehensible to humans, which enhances the reliability and accountability of AI-driven healthcare systems.

The comparative analysis shows that some things have been done but there are still things that aren't done. They involve the lack of standardization of explainability evaluation metrics, limited clinical validation, lack of using public benchmark datasets, algorithmic bias, lack of explainability integration into clinical workflows, and the need for explainability strategies that are human-centred. To accomplish these goals, it is essential to address these challenges and ensure the successful deployment and adoption of AI-driven CDSS.

**References**


1. H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, “Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022),” Computer Methods and Programs in Biomedicine, vol. 226, p. 107161, 2022, doi: 10.1016/j.cmpb.2022.107161.
2. S. S. Band, A. Yarahmadi, C.-C. Hsu, M. Biyari, M. Sookhak, R. Ameri, et al., “Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods,” Informatics in Medicine Unlocked, vol. 40, p. 101286, 2023, doi: 10.1016/j.imu.2023.101286.
3. T. Gniadek, J. Kang, T. Theparee, and J. Krive, “Framework for classifying explainable artificial intelligence (XAI) algorithms in clinical medicine,” Online Journal of Public Health Informatics, vol. 15, no. 1, e50934, 2023, doi: 10.2196/50934.
4. S. Moazemi, S. Vahdati, J. Li, et al., “Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: A systematic review,” Frontiers in Medicine, vol. 10, p. 1109411, 2023, doi: 10.3389/fmed.2023.1109411.
5. Y. Liu, C. Liu, J. Zheng, C. Xu, and D. Wang, “Improving explainability and integrability of medical AI to promote health care professional acceptance and use: Mixed systematic review,” Journal of Medical Internet Research, vol. 27, e73374, 2025, doi: 10.2196/73374.
6. A. Brankovic, D. Cook, J. Rahman, et al., “Clinician-informed XAI evaluation checklist with metrics (CLIX-M) for AI-powered clinical decision support systems,” npj Digital Medicine, vol. 8, p. 364, 2025, doi: 10.1038/s41746-025-01764-2.
7. D. Muhammad and M. Bendechache, “Unveiling the black box: A systematic review of explainable artificial intelligence in medical image analysis,” Computational and Structural Biotechnology Journal, vol. 24, pp. 542–560, 2024, doi: 10.1016/j.csbj.2024.08.005.
8. Z. Sadeghi, R. Alizadehsani, M. A. Cifci, et al., “A review of explainable artificial intelligence in healthcare,” Computers and Electrical Engineering, vol. 118, p. 109370, 2024, doi: 10.1016/j.compeleceng.2024.109370.
9. H. M. Tun, H. A. Rahman, L. Naing, and O. A. Malik, “Trust in artificial intelligence-based clinical decision support systems among health care workers: Systematic review,” Journal of Medical Internet Research, vol. 27, e69678, 2025, doi: 10.2196/69678.
10. A. Bilal et al., “Explainable AI-driven intelligent system for precision cardiovascular event prediction,” Frontiers in Medicine, vol. 12, 2025, doi: 10.3389/fmed.2025.1596335.