International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 9 Issue 6, Nov-Dec 2025 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Explainable Artificial Intelligence (XAI) in Healthcare: A Comprehensive Review

Kamble V. B, Dr. Halgare N. M, Mohammed Aejaz Tumkur

Department of Computer Science and Engineering, M.S. Bidve Engineering College, Latur, Maharashtra, India

ABSTRACT

Artificial Intelligence (AI) has become integral to healthcare, enabling disease prediction, medical imaging analysis, and personalized treatment planning. However, the opaque or "black-box" nature of many AI models-particularly deep learning-creates significant challenges for trust, safety, ethics, and regulatory acceptance. Explainable Artificial Intelligence (XAI) offers solutions by making AI decisions interpretable, transparent, and accountable.

This paper presents a comprehensive study of XAI methods, tools, and applications specifically in the healthcare domain. Feature-based, concept-based, surrogate models, pixel-based explanations, and human-centric XAI approaches are explored in detail. Common XAI frameworks such as SHAP, LIME, ELI5, IBM AIF360, and the What-If Tool are evaluated for their role in clinical interpretability and fairness.

Applications of XAI in Parkinson's disease detection, cancer diagnostics, Alzheimer's prediction, and COVID-19 risk assessment are reviewed, along with broader use cases in cardiovascular diagnostics and treatment planning. Key challenges such as interpretability– performance trade-offs, data bias, workflow integration, and ethical concerns are also analyzed.

The paper concludes that XAI is essential for bridging the gap between AI technology and clinical decision-making. Future research directions include human-centered explainability, regulatory frameworks, real-time EHR integration, and next-generation interpretable deep learning architectures.

How to cite this paper: Kamble V. B | Dr. Halgare N. M | Mohammed Aejaz Tumkur "Explainable Artificial Intelligence (XAI) in Healthcare: A Comprehensive Review" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-9 | Issue-6, December 2025, pp.275-277, URL: www.ijtsrd.com/papers/ijtsrd98807.pdf

IJTSRD98807

Copyright © 2025 by author (s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (http://creativecommons.org/licenses/by/4.0)

KEYWORDS: *Explainable AI, Healthcare, SHAP, LIME, Medical Diagnosis, Transparency, Clinical Decision Support.*

1. INTRODUCTION

Artificial Intelligence is transforming healthcare through predictive analytics, medical imaging interpretation, patient risk stratification, and precision medicine. Machine learning models, especially deep neural networks, demonstrate exceptional accuracy in diagnosis and prognosis. However, most high-performing models function as "black boxes," providing predictions without justifying how decisions were reached.

For healthcare-where decisions affect human life-explainability is essential. Doctors, patients, regulators, and hospital administrators require clarity and reasoning, not just accurate outputs. Explainable AI (XAI) aims to make AI decisions understandable, reliable, and ethically acceptable.

1.1. Need for XAI in Healthcare

XAI is important because:

- **Trust:** Clinicians accept AI decisions only when the rationale is clear.
- **Safety:** Identifying incorrect or biased predictions prevents medical harm.
- **Accountability:** Regulatory frameworks demand transparent AI behavior.
- **Ethical AI:** Explanations help detect discrimination or unfair outcomes.

1.2. Objectives

This research aims to:

- Review XAI methods applicable to healthcare.
- Analyze widely used XAI tools and frameworks.
- Evaluate real-world applications of XAI across diseases.
- Examine challenges and future research directions.

2. Explainable AI Methods

XAI methods can be broadly classified into:

2.1. Feature-Oriented Methods

Explain how individual features influence model predictions.

Examples: SHAP values, feature importance, partial dependence plots.

2.2. Global vs Local Explanations

- **Global:** Describe overall model behavior.
- **Local:** Explain a single prediction for an individual patient.

Both are essential for clinical decision support.

2.3. Concept-Based Explanations

Translate features into human-interpretable medical concepts, such as:

- "Tumor boundary"
- "Lung opacity"
- "Hippocampal shrinkage"

2.4. Surrogate Models

Simplify black-box behavior using interpretable models like decision trees or linear models.

2.5. Pixel-Based Explanations

Used primarily for medical imaging.

Examples: Grad-CAM, saliency maps.

2.6. Human-Centric Explanations

Provide natural-language or interactive explanations designed for doctors and patients.

3. Tools for Explainable AI

Several open-source tools facilitate interpretability:

3.1. SHAP (Shapley Additive Explanations)

Uses game theory to quantify each feature's influence on predictions.

Widely used for disease risk assessment.

3.2. LIME

Generates local explanations by approximating complex models with interpretable ones.

3.3. ELI5

Simple Python framework for visualizing model weights, feature contributions, and errors.

3.4. What-If Tool

Interactive UI for exploring how different variables affect AI predictions.

3.5. IBM AI Fairness 360

Focuses on detecting and mitigating bias in healthcare datasets and models.

4. Role of XAI for Healthcare Decision-Makers

4.1. Trust and Adoption

Doctors trust AI more when:

- Explanations are clear
- Predictions align with clinical knowledge
- Key contributing factors are visible

4.2. Accountability

XAI provides audit trails, preventing malpractice and enabling responsible AI deployment.

4.3. Transparency and Ethics

Transparent predictions help meet ethical and regulatory requirements like HIPAA, GDPR, and medical device guidelines.

4.4. Clinical Decision Support

XAI enhances:

- Treatment recommendations
- Risk stratification
- Early disease detection
- Scenario simulation ("what-if" analysis)

5. Healthcare Applications of XAI

5.1. Parkinson's Disease

AI models analyze gait, tremor frequency, and voice recordings.

XAI highlights which symptoms contribute most to the prediction.

5.2. Cancer Diagnostics

Deep learning models identify tumors in MRI, CT, and histopathology images.

Pixel-based XAI explains exact regions influencing predictions.

5.3. Alzheimer's Disease

Explainable models analyze brain scans and cognitive tests.

Tools like LIME show the influence of hippocampal atrophy.

5.4. COVID-19 Diagnosis & Risk Assessment

XAI explains predictions from:

- Chest X-ray models
- Vital-sign based prediction models
- ICU admission risk models

5.5. Cardiovascular & Diabetes Prediction

Feature-based XAI highlights clinical factors such as:

- BP, cholesterol
- HbA1c, BMI
- Lifestyle patterns

These explanations validate AI decisions for clinicians.

6. Challenges in Implementing XAI

6.1. Interpretability vs Performance

Deep learning is accurate but complex; simpler models are interpretable but less powerful.

6.2. Data Quality and Bias

Biased or incomplete datasets reduce both accuracy and trust.

6.3. User-Centric Explanation Design
Doctors often prefer visual/narrative explanations over mathematical ones.

6.4. Clinical Workflow Integration
XAI tools must integrate with hospital systems like EHRs.

6.5. Legal & Ethical Concerns
Regulations require traceability, fairness, and patient safety.

7. Conclusion
Explainable AI is essential for trustworthy, ethical, and effective AI-driven healthcare. XAI bridges the gap between complex algorithms and human clinical expertise. By combining interpretability techniques, robust tools, and user-centric design, healthcare can benefit from safe and transparent AI deployment.

Future Work Includes:
- Interpretable deep learning architectures
- XAI-enabled EHR integration
- Standardized regulatory frameworks
- Human-in-the-loop XAI systems
- Real-time explainability dashboards in hospitals

XAI will play a foundational role in the future of medical AI, ensuring that predictions are not only accurate but also understandable, fair, and clinically meaningful.

References
[1] Adadi A., Berrada M., "Peeking inside the black-box: A survey on explainable AI," IEEE Access, 2018.

[2] Samek W., Wiegand T., Müller K., "Explainable AI: Interpreting deep learning models," 2019.

[3] Ribeiro M.T., Singh S., Guestrin C., "Why should I trust you? Explaining classifiers," SIGKDD, 2016.

[4] Lundberg S., Lee S., "A unified approach to interpreting model predictions," NeurIPS, 2017.

[5] Ghassemi N., Oakden-Rayner A., Beam A., "False hope of current XAI in healthcare," Lancet Digital Health, 2020.

[6] IBM AI Fairness 360 Toolkit.

[7] Gilpin L., Bau D., Yuan B., "Explaining explanations: Interpretability in ML," DSAA, 2018.

[8] Holzinger A., et al., "Explainable AI for medical domain," arXiv, 2017.

[9] Google What-If Tool.

[10] Mohseni S., Ragan-Kelley P., "Explainable AI in healthcare: A review," arXiv, 2020.