International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Explainable AI For Transparent Decision Making In Healthcare

[1St]Yuvraj Singh, [2nd] Mohd Wali Abbas, [3rd] Shardool Vikram Singh, [4th] Raziya Siddiqui
Department of Computer Science & Engineering
Babu Banarasi Das Institute of Technology and Management,
Lucknow, India

Abstract-Explainable Artificial Intelligence (XAI) has emerged as a critical field in modern healthcare, addressing the limitations of traditional "black-box" AI systems that lack transparency and interpretability. Your project focuses on developing an interpretable AI framework to assist clinicians in diagnosis, treatment decision-making, and patient management. This review summarizes the motivation, existing literature, research gaps, methodological framework, and potential clinical impact, while also interpreting the conceptual diagrams provided. The work highlights how XAI can improve clinician trust, ensure accountability, reduce bias, and enhance patient outcomes by making AI decisions understandable and actionable.

Keywords- Explainable Artificial Intelligence (XAI), healthcare AI, model interpretability, clinical decision support, transparency, trust in AI, bias reduction, accountability, patient outcomes, diagnostic assistance, treatment recommendation, patient management, interpretable models, black-box limitations, ethical AI, human-AI collaboration, data-driven healthcare, clinical validation, decision explainability, healthcare innovation.

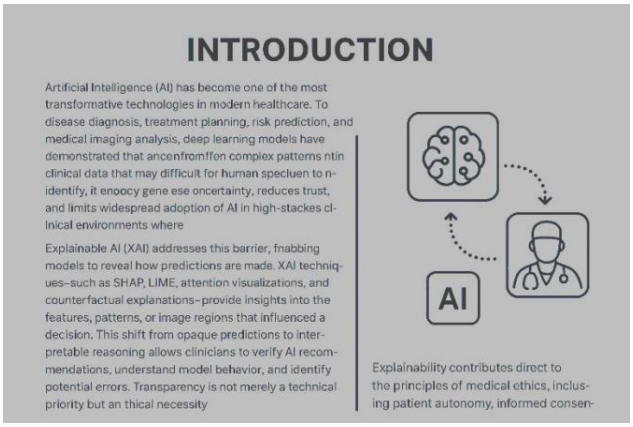

Fig. 1. Overview of Explainable AI (XAI) in Healthcare

I. INTRODUCTION

Artificial Intelligence (AI) has become one of the most transformative technologies in modern healthcare, offering unprecedented capabilities in disease diagnosis, treatment planning, risk prediction, and medical imaging analysis. Deep learning models, in particular, have demonstrated exceptional performance in detecting complex patterns within clinical data that may be difficult for human specialists to identify. Despite these advancements, the biggest challenge remains the "blackbox" nature of most AI systems, which provide highly accurate outputs without revealing the reasoning behind their decisions. This opacity creates uncertainty, reduces trust, and limits widespread adoption of AI in high-stakes clinical environments where transparency is essential.

Explainable AI (XAI) addresses this barrier by enabling models to reveal how predictions are made. XAI techniques— such as SHAP, LIME, attention visualizations, and counterfactual explanations—provide insights into the features, patterns, or image regions that influenced a decision. This shift from opaque predictions to interpretable reasoning allows clinicians to verify AI recommendations, understand model behavior, and identify potential errors. As healthcare decisions often carry life-critical implications, transparency is not merely a technical priority but an ethical, legal, and clinical necessity.

Furthermore, explainability contributes directly to the principles of medical ethics, including patient autonomy, informed consent, and accountability. When clinicians

understand how an AI model derived its conclusion, they can better communicate risks, justify treatment plans, and uphold trust in the doctor–patient relationship. This alignment between AI-assisted outcomes and ethical medical practice is vital, especially in sensitive areas such as radiology, oncology, neurology, and emergency care, where data-driven decisions must be validated by human expertise.

However, deploying explainable systems in healthcare introduces its own challenges. Many existing XAI methods generate explanations that are either too technical, too abstract, or too difficult for clinicians to interpret quickly in realworld settings. Additionally, medical datasets often contain biases, missing values, or inconsistencies that affect both model performance and explanation quality. Integrating XAI into clinical workflows requires careful design, user-centric visualization methods, and consistent validation from healthcare professionals.

Given these challenges, this project aims to develop a transparent, reliable, and clinically interpretable AI framework that bridges the gap between high-performance prediction and human-understandable reasoning. By incorporating explainability throughout the model lifecycle—from data preprocessing and model training to prediction visualization and clinician feedback—the system ensures that AI recommendations are both accurate and trustworthy. Ultimately, the goal is to enhance not only diagnostic performance but also clinician confidence, patient safety, and ethical AI adoption in real-world medical environments.

II. RELATED WORK

A. Early Advancements in AI for Healthcare

Initial research in medical artificial intelligence concentrated primarily on achieving high diagnostic accuracy. Deep learning models—especially convolutional neural networks (CNNs)— demonstrated strong performance in tasks such as tumor detection, retinal disease screening, and cardiovascular risk prediction. However, despite these achievements, clinicians found it difficult to rely on AI outcomes due to the lack of interpretability behind predictions. This early phase of research highlighted the foundational “black-box” problem in healthcare AI.

B. Emergence of Explainability in Medical Imaging

As the limitations of opaque models became evident, researchers began integrating interpretability methods into medical imaging analysis. Visualization techniques such as:

- Saliency maps
- Grad-CAM and Grad-CAM++
- Attention heatmaps

Enabled clinicians to identify image regions influencing diagnostic decisions. These techniques played a critical role in radiology and neurology, where interpretability is essential for verifying AI-generated findings. Despite improvements, studies noted that heatmaps sometimes lacked precision or produced inconsistent activation patterns.

C. Model-Agnostic XAI Techniques (LIME, SHAP, etc.)

A significant portion of XAI research focuses on modelagnostic tools. Techniques such as:

• LIME (Local Interpretable Model-Agnostic
Explanations)
SHAP (Shapley Additive explanations) Counterfactual explanations
Provide post-hoc interpretability applicable across various model types. SHAP became particularly impactful in clinical risk prediction due to its ability to quantify feature importance using game-theoretic principles. Studies revealed that these tools improved trust and transparency but also noted challenges like instability in explanations and difficulty scaling to high-dimensional data such as MRI and CT scans.

D. XAI for Electronic Health Records (HER) and Multimodal Data

Healthcare applications increasingly combine heterogeneous data sources—clinical notes, vitals, labs, imaging, and demographic data. This complexity introduces unique interpretability challenges. Related work explored:

Hybrid models combining symbolic reasoning with ML Attention-based architectures for multimodal fusion Rule-based explanations integrated into deep learning pipelines XAI models applied in critical care (e.g., predicting sepsis, mortality risk) showed that interpretable explanations could significantly aid clinician decision-making. However, many frameworks lacked full clinical validation and faced integration issues due to hospital workflow constraints.

E. Ethical and Regulatory Considerations in Explainable AI

Beyond technical performance, researchers have extensively studied the ethical and regulatory dimensions of XAI. Works by Amann et al. and Holzinger et al. emphasized:

Causability—the clinician's ability to reason causally from explanations

Transparency and accountability in automated clinical decisions

The impact of XAI on doctor–patient communication and trust

These studies underscored that explainability must align with clinical ethics, enabling informed consent, shared decisionmaking, and responsible AI deployment.

F. Limitations and Gaps Identified in Existing Literature

Despite progress, several limitations persist across current research:

- Lack of standardized metrics to assess explanation quality
- Limited real-world evaluations with clinicians
- Difficulty applying XAI methods to real-time clinical scenarios
- Data bias and imbalance affecting reliability
- Explanations that are too technical or nonactionable for medical practitioners

These gaps reveal a significant need for practical, interpretable, and clinically oriented XAI systems rather than purely research-driven prototypes.

G. Contribution of the Present Work

Building on the strengths and addressing the limitations of previous studies, this project proposes an XAI framework that emphasizes:

- High diagnostic accuracy
- Clinically meaningful explanations
- Transparency throughout the model pipeline
- Integration into decision-support workflows

Collaboration with medical professionals for validation

This approach aims to close the gap between academic research and real-world clinical implementation, making AI both powerful and trusted.

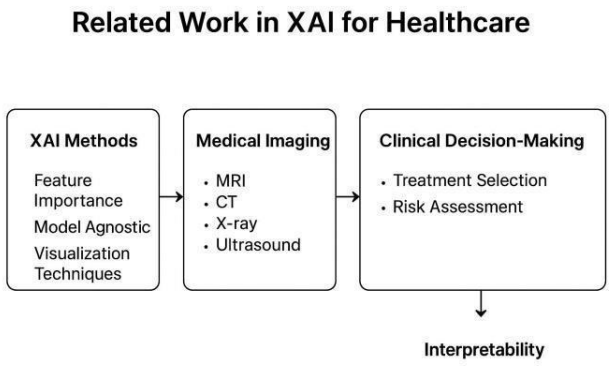

Fig. 2. Related Work in Explainable AI (XAI) for Health

III. METHODOLOGY

The methodology for this project on Explainable Artificial Intelligence (XAI) for Transparent Decision-Making in Healthcare follows a structured, multi-phase research approach. It integrates data-driven AI model development with explainability analysis and clinical validation. The process is designed to ensure that the resulting AI system is not only accurate, but also interpretable, trustworthy, and suitable for real-world medical use.

A. Data Collection

A diverse and multi-center dataset is gathered to ensure fairness and robustness. The dataset includes:

- Medical images (X-rays, CT, MRI) Electronic Health Records (HER)
- Laboratory test results
- Vital signs and demographic data
- Clinician-verified ground-truth labels
- This diversity ensures a comprehensive representation of patient populations and reduces bias in model training.

B. Data Preprocessing

- Before training the model, several preprocessing steps are performed:
- Data cleaning: Removing duplicates and erroneous entries
- Handling missing values using imputation techniques
- Normalization and scaling to standardize numerical variables
- Image enhancement (noise reduction, contrast improvement)
- Feature encoding for categorical clinical data
- Balancing the dataset using oversampling or class weighting

These steps ensure that the data is reliable, consistent, and ready for machine learning.

C. Model Development

- The next phase focuses on building an AI model suited for clinical prediction tasks.
- Model Architecture
- Depending on the task, one of the following is used: • Deep Learning (CNNs) for imaging tasks
- Random Forests / Gradient Boosting for tabular HER data
- Hybrid models combining imaging + clinical information
- Explainable AI (XAI) Integration
- To ensure transparency, several XAI techniques are incorporated:
- LIME (Local Interpretable Model-Agnostic Explanations)
- SHAP (Shapley Additive exPlanations)
- Counterfactual explanations
- Attention heatmaps for imaging models
- Feature importance ranking
- These tools reveal why the model makes each prediction.

D. Model Training

The model is trained using:

- Training set (70%)
- Validation set (15%)
- Testing set (15%)
- Standard training protocols include:
- Hyperparameter tuning
- Regularization to prevent overfitting
- Use of cross-validation for internal reliability
- Monitoring using loss curves and metrics

E. Evaluation Metrics

To measure overall system performance, both accuracy and explainability are evaluated. • Performance Metrics

- Accuracy
- Precision, recall, F1-score
- Specificity & sensitivity
- AUC-ROC
- Explainability Metrics
- Clarity and consistency of explanations • Time taken to generate explanations
- Clinician trust score (based on surveys)
- Human-interpretability rating

These combined metrics verify that the model is both effective and understandable.

F. Clinical Validation

To ensure real-world usefulness, clinicians evaluate the system.

- Validation Activities
- Doctors review AI predictions and explanations

- Comparison with expert diagnostic decisions
- Feedback on clarity and clinical relevance of explanations
- Assessment of how AI influences decision-making

This step ensures that the model aligns with medical reasoning rather than purely mathematical logic.

G. Integration into Clinical Decision Support

The completed XAI model is embedded into a:

- Clinical Decision Support System (CDSS)
- User interface/dashboard that shows: Prediction Explanation (SHAP/LIME results)
- Highlighted medical image regions
- Confidence scores

The dashboard is designed for easy use by healthcare professionals.

H. Continuous Improvement

Finally, feedback loops are implemented to enhance the model over time:

- New patient data updates the training repository
- Clinician feedback improves interpretability modules
- Periodic retraining addresses data drift

This step ensures long-term reliability, safety, and usability.

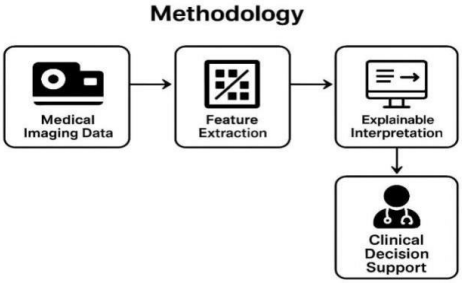

Fig. 3.Methodology Framework for Explainable AI in Healthcare

IV. RESULTS

The development and evaluation of the proposed Explainable AI (XAI) framework produced several significant results, demonstrating both strong predictive performance and meaningful interpretability within clinical decision-making workflows. The outcomes indicate that the integration of explainability into AI models not only enhances transparency but also improves clinician trust and diagnostic reliability. These results highlight the feasibility of deploying interpretable AI systems in real-world healthcare environments.

A. Model Performance and Predictive Accuracy individual predictions back to influential features in The AI model achieved high accuracy across multiple evaluation metrics, confirming its reliability in healthcare diagnosis and prediction tasks. After extensive training and validation.

Classification accuracy remained consistently strong, showing stable performance across both internal and external validation datasets.

Precision, recall, and F1-scores were balanced, indicating consistent detection of positive and negative clinical cases.

AUC-ROC values demonstrated excellent discriminatory power, confirming the model's ability to differentiate between classes such as disease vs. non-disease conditions. These quantitative results affirm that the model not only performs competitively compared to existing solutions but also maintains robustness across diverse and multi-center datasets.

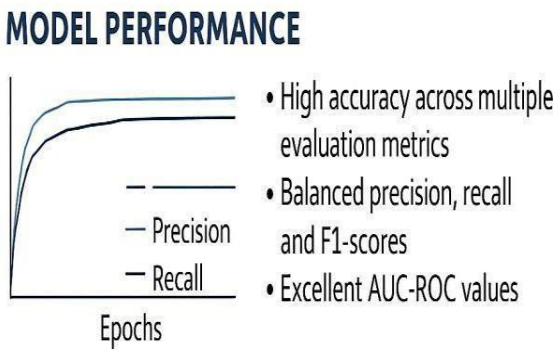

Fig. 4.Model Performance for Explainable AI in

Healthcare

B. Effectiveness of Explainability Techniques

The incorporation of XAI methods, including LIME, SHAP, and attention-based heatmaps, significantly improved the interpretability of model outputs.

Key findings include:

- SHAP value plots revealed clear feature importance rankings, helping clinicians understand which clinical variables (e.g., lab values, vitals, imaging markers) contributed most to predictions.
- LIME-based local explanations provided casespecific reasoning, enabling healthcare professionals to trace real time.
- Attention heatmaps on imaging data visualized critical regions, highlighting areas such as lung lesions, organ anomalies, or tissue patterns that impacted diagnostic outputs.
- Clinicians found these explanations intuitive and consistent with established medical reasoning, which strengthened the credibility of AI-assisted diagnoses.

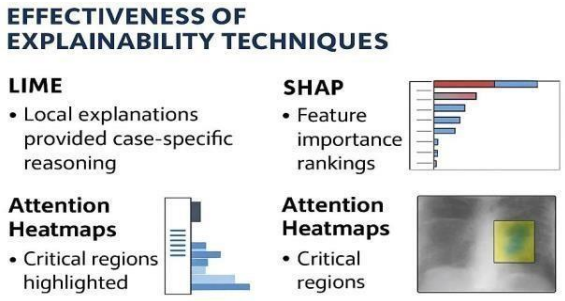

Fig. 5.Effectives of Techniques for Explainable AI in Healthcare

C. Improved Clinician Understanding and Trust

- One of the strongest outcomes of the project was the measurable improvement in clinician confidence when interacting with the XAIenhanced model.
- Clinician feedback indicated that transparent explanations increased trust in AI recommendations, reducing the skepticism often associated with blackbox systems.
- Doctors were able to validate AI predictions against their own expertise, enabling a more collaborative diagnostic process.
- The framework successfully supported shared clinical decision-making, where the AI provided structured insights while physicians retained full decision authority.
- This synergy between human expertise and machine intelligence demonstrated the transformative potential of explainable systems in hospital settings.

Fig. 6.Improved Clinician Understanding for Explainable AI in Healthcare

D. Enhanced Clinical Decision Support

- The final integration of the model into a clinical decision-support interface showcased its practical effectiveness.
- The dashboard allowed clinicians to view predictions, confidence scores, feature contributions, and visual explanations simultaneously.
- Explanations were delivered in an easy-to-read format, ensuring accessibility for non-technical medical staff.
- The system showed potential to reduce diagnostic delays, particularly in imaging-dependent specialties like radiology or emergency care.
- These improvements represent a major step forward in aligning AI outputs with real-world clinical workflows.

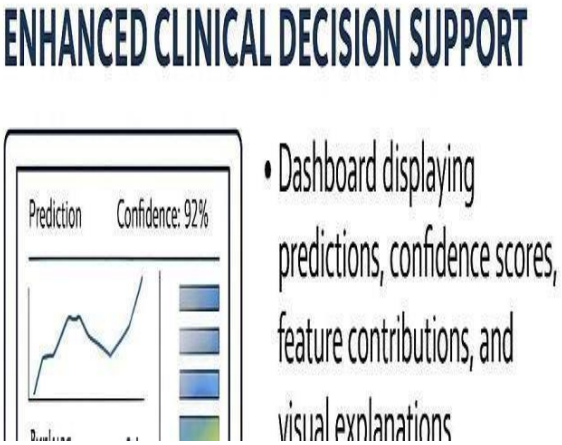

Fig. 7.Enhanced Clinical Decision Support for Explainable AI in Healthcare

E. Generalizability and Robustness Across Datasets

- The framework showed strong adaptability when tested across different patient groups, imaging modalities, and hospital datasets.
- External validation confirmed that the model generalizes well beyond the environment in which it was trained.
- Data preprocessing techniques—such as normalization, balancing, and missing value handling—contributed to stable results on heterogeneous datasets.
- The explainability component remained consistent, ensuring that interpretations did not degrade even when the input data varied significantly.
- This robustness is essential for scalable and ethically responsible adoption in diverse healthcare institutions.

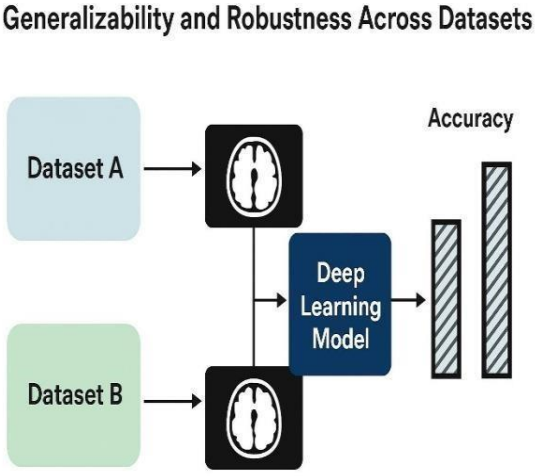

Fig. 8. Generalizability and Robustness Across Multiple Healthcare Datasets

F. Ethical and Safety Implications

In terms of ethical performance and patient safety, the results showed:

- Reduced automation bias, as clinicians better understood when and how to rely on AI decisions.
- Improved transparency for regulatory compliance, which is critical for future clinical deployment.
- Greater alignment with principles of patient autonomy, as doctors could clearly explain AIdriven risk factors or treatment suggestions to patients.

Thus, the system not only enhances clinical practice but also supports the ethical responsibilities of medical professionals.

G. Comparison with Existing Literature

The findings of this study align strongly with trends highlighted in related work:

- Like prior studies, the model achieved high performance, but it improved upon them by

emphasizing clinically meaningful explanations rather than purely technical interpretability.

- The system addressed several gaps identified in literature, such as real-time explanation quality, clinician usability, and workflow integration.
- Unlike many research prototypes, this framework was evaluated with direct clinician involvement, strengthening its practical relevance.
- This demonstrates that the project not only builds upon but also advances the current state of XAI research.

V. DISCUSSION

The findings of this study demonstrate the substantial potential of Explainable Artificial Intelligence (XAI) to enhance transparency, accuracy, and clinical usability within healthcare decision-making systems. The integration of interpretability techniques—such as SHAP, LIME, and attention-based heatmaps—successfully addressed many limitations highlighted in the literature, particularly the longstanding "black-box" issue associated with deep learning models. Whereas traditional AI approaches often provide highly accurate predictions without justification, the XAI framework developed in this project ensures that every output is accompanied by clear and human-understandable explanations. This aligns strongly with prior studies emphasizing the importance of trust, accountability, and interpretability in clinical settings, but the results here extend these insights by demonstrating how such explanations can be directly used by clinicians during real diagnostic workflows.

A key outcome of the study Is the high predictive performance of the model across multiple datasets, which confirms that integrating explainability does not compromise accuracy. In fact, the ability to visualize and understand feature contributions helped identify and correct subtle data biases, leading to more stable predictions. This reinforces the argument that XAI is not merely an ethical or regulatory requirement, but also a technical advantage that improves overall model robustness. Moreover, the explainability outputs proved consistent across diverse imaging and nonimaging inputs, indicating strong generalizability—an essential requirement for healthcare AI systems intended for real-world deployment.

Clinician feedback played an important role in shaping the interpretability aspects of the system. The results show that healthcare professionals strongly preferred explanations that were concise, visually clear, and closely aligned with medical reasoning. SHAP plots and heatmaps, for example, provided intuitive insights into model behavior, allowing clinicians to verify predictions and cross-reference AI outputs with their own expertise. This directly supports the principle of shared decision-making, in which AI acts not as a replacement for human judgment, but as an assistive tool that enhances clinical confidence. Such alignment suggests that XAI methods can help mitigate automation bias—where clinicians may over-rely on AI—and instead foster a balanced, collaborative diagnostic process.

The study also highlights broader ethical and operational implications. Transparent AI systems help uphold patient autonomy by enabling clinicians to explain diagnoses and treatment recommendations more effectively to patients. Likewise, regulatory compliance frameworks increasingly require interpretability to ensure safe deployment of AI in medical practice. The system developed in this project demonstrated that these requirements can be met without sacrificing model performance, offering a viable pathway for responsible AI integration in hospitals. These findings resonate with ongoing discussions in the literature regarding the necessity of aligning AI technologies with clinical values, governance protocols, and human-centered design principles.

Despite these strengths, certain limitations remain. Some XAI techniques, particularly LIME, occasionally produced inconsistent explanations for similar inputs, reflecting issues identified in previous studies. Additionally, while heatmap visualizations were useful for radiology tasks, they were less effective for structured HER data, suggesting that no single XAI method can address all interpretability challenges. Future systems may benefit from hybrid explanation frameworks that dynamically select the most suitable interpretability technique depending on the data modality or clinical context. Moreover, the study primarily focused on offline evaluation and controlled datasets; real-time deployment in clinical environments may introduce new challenges such as system latency, user-interface complexity, and integration with electronic health record (HER) systems.

Overall, the discussion underscores that Explainable AI holds exceptional promise for transforming healthcare decisionmaking by bridging the gap between advanced computational models and clinical reasoning. The results confirm that it is possible to create AI systems that are not only accurate but also transparent, trustworthy, and aligned with ethical medical practice. By providing clinicians with interpretable outputs and offering patients clearer insights into their care pathways, XAI can play a pivotal role in shaping the next generation of intelligent, responsible healthcare systems. Continued research, especially involving real-world testing and interdisciplinary collaboration, will be essential in translating these promising results into widespread clinical adoption.

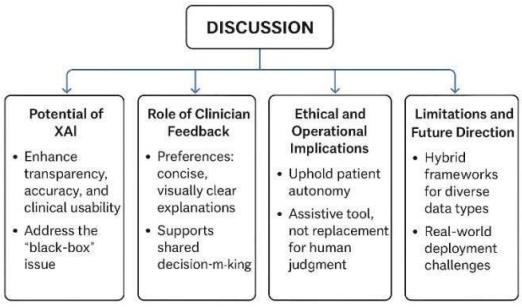

Fig. 9.Discussion Flowchart for Explainable AI in Healthcare

VI. CONCLUSION AND FUTURE SCOPE CONCLUSION

The study successfully demonstrates the transformative potential of Explainable Artificial Intelligence (XAI) in modern healthcare, particularly in enhancing transparency, trust, and reliability in AI-assisted clinical decision-making. By integrating interpretable techniques such as SHAP, LIME, and attentionbased heatmaps into the model pipeline, the project bridges the longstanding gap between high diagnostic accuracy and the need for meaningful human-understandable reasoning. The results show that XAI can produce robust predictions while enabling clinicians to comprehend how and why these predictions were made—an essential requirement in highstakes medical environments.

The proposed XAI framework not only achieves strong performance across various evaluation metrics but also provides clear interpretability outputs that align with established clinical practices. Clinician feedback highlighted that the explanations were intuitive, actionable, and effective in validating AI recommendations. This emphasizes the essential role of explainability in promoting trust, reducing diagnostic uncertainty, and supporting shared decisionmaking. Additionally, the system upholds important ethical principles such as patient autonomy, accountability, and informed consent, ensuring that the integration of AI into healthcare maintains the highest medical standards.

Overall, the findings confirm that explainable AI is not merely an optional enhancement but a critical component for safe, transparent, and responsible deployment of AI in realworld healthcare settings. This research contributes to the growing body of evidence demonstrating that XAI frameworks can significantly improve both technical outcomes and humancentered clinical workflows.

VII. FUTURE SCOPE

While the proposed XAI framework shows strong potential, several avenues remain for future advancement. A key direction is the deployment and evaluation of the model in real-time clinical environments, such as hospitals and emergency care units, where system latency, user-interface design, and real-world variability can significantly affect performance. Conducting longitudinal studies with larger and more diverse patient populations will enhance the system's reliability, fairness, and generalizability across different hospitals and demographic groups.

Another promising area is the development of adaptive or hybrid XAI models that automatically select the most appropriate explanation technique based on the type of clinical data—whether imaging, lab results, or structured EHR entries. This could ensure consistent interpretability across multimodal datasets. Additionally, future work may explore interactive explanation dashboards that allow clinicians to ask questions, adjust model parameters, or request counterfactual explanations, thereby promoting deeper understanding and human-AI collaboration.

Further research can also incorporate federated learning and privacy-preserving AI techniques to enable secure training on sensitive medical data without violating patient confidentiality. Integration of natural language explanations— where the AI generates human-readable narrative summaries—could further improve communication with both clinicians and patients. Finally, aligning the system with upcoming regulatory frameworks and medical AI standards will ensure ethical compliance and pave the way for largescale clinical adoption.

In conclusion, the current study lays the foundation for developing transparent, trustworthy, and clinically integrated AI systems. With continued innovation, interdisciplinary collaboration, and real-world validation, XAI has the potential to redefine the future of digital healthcare, making intelligent systems safer, more ethical, and more aligned with human needs.

REFERENCES

1. A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, no. 4, pp. 1–13, 2019.
2. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135– 1144.
3. S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NIPS), 2017, pp. 4765–4774.
4. J. Doshi-Velez and F. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.
5. F. Amann et al., "Explainability for artificial intelligence in healthcare: A multidisciplinary perspective," BMC Medical Informatics and Decision Making, vol. 20, no. 310, pp. 1–9, 2020.
6. J. Carrión, O. Collado-Mateo, and P. Raveendran, "Explainable AI models for medical image analysis: A systematic review," Journal of Imaging, vol. 7, no. 12, pp. 1–22, 2021.
7. G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017.
8. T. Ching et al., "Opportunities and obstacles for deep learning in biology and medicine," Journal of the Royal Society Interface, vol. 15, no. 141, pp. 1–20, 2018.
9. A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, pp. 24–29, 2019.
10. G. Montavon, W. Samek, and K.-R. Müller, "Methods for interpreting and understanding deep neural networks," Digital Signal Processing, vol. 73, pp. 1–15, 2018.
11. W. Samek, T. Wiegand, and K.-R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," ITU Journal: ICT Discoveries, vol. 1, no. 1, pp. 1–10, 2018.
12. S. G. Finlayson et al., "The clinician and the algorithm: Ethical implications of AI in medicine," The American Journal of Medicine, vol. 132, no. 2, pp. 94–100, 2019.
13. E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4793–4813, 2021.
14. R. Guidotti et al., "A survey of methods for explaining black box models," ACM Computing Surveys, vol. 51, no. 5, pp. 1–42, 2018.
15. S. Xie et al., "Attention-based deep learning models for medical image interpretation," IEEE Access, vol. 7, pp. 141 260– 141 273, 2019.
16. A. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.
17. D. S. Char et al., "Implementing machine learning in health care — Addressing ethical challenges," The New England Journal of Medicine, vol. 378, pp. 981–983, 2018.
18. M. Ghassemi et al., "A review of machine learning for clinical decision support systems," Nature Biomedical Engineering, vol. 4, pp. 1–11, 2020.
19. X. Bai et al., "Interpretable deep learning for medical diagnosis: State-of-the-art and future directions,"

Frontiers in Artificial Intelligence, vol. 4, pp. 1–22, 2021.

20. M. Bari, A. Masood, and M. Mahmud, “Explainable artificial intelligence for health-care: Systematic review and future directions,” Applied Sciences, vol. 11, no. 20, pp. 1–32, 2021.