International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.545

e ISSN: 2584-2854
Volume: 02
Issue:12 December 2024
Page No: 3671-3679

Explainable AI in Healthcare Applications

Alan Varghese[1], Jefin Varghese[2], Jubin Biju[3], Roshan Thomas[4], Merlin Thomas[5]
[1,2,3,4]PG, MCA, Kristu Jyothi College of Management and Technology, Changanassery, Kerala, India.
[5]Assistant Professor, Department of Computer Application, Kristu Jyoti College of Management and Technology, Changanassery, Kerala, India.
Email ID: *alanvarghese458@gmail.com[1], jeffinvarghesethomas@gmail.com[2], jubinbiju007@gmail.com[3], roshanvthomas15@gmail.com[4], merlin@kjcmt.ac.in[5]*

Abstract

The entry of artificial intelligence into health care systems brings unprecedented advances in diagnosing, personalized treatment, and predictive analytics. Many of these AI models, especially the deep-learning algorithms, have been referred to as "black boxes" and raise gigantic questions about trust, transparency, and reliability in clinical settings. Therefore, explainable AI answers the challenges by drawing to the fore methodologies that make AI models more interpretable, thereby making them more accepted and usable in the fraternity of health. It engages with XAI in healthcare by scrutinizing various aspects of feature importance analysis, architectures of the interpretable model, and visual explication of decisions driven by AI. The case studies regarding applications of XAI in the fields of radiology, disease prediction, and personalized medicine illustrate how this technology has its own importance even in terms of improving the precision of diagnosis and clinician-patient communication. We are answerable to ethics, such as how to explain respect for the trust of patients, legalities in deploying XAI in healthcare, and further directions to be taken so that XAI does not lag behind the timeline of AI innovation. Our results repeat again that XAI indeed is the much-needed solution to be technically suitable and necessary for ensuring the responsible adoption of AI within healthcare systems to empower clinicians using not only accurate but also transparent, understandable, and aligned AI systems in clinical best practice. In conclusion, this paper concludes by highlighting explainable AI as one of the key enablers toward safe, effective, and widely accepted AI applications in health care.

Keywords: *Clinical Decision-Making; Diagnostics; Disease Prediction; Explainable AI; Healthcare.*

1. Introduction

AI has transformed health care and provided solutions for diagnostic purposes, treatment planning, and operational efficiency. However, its adoption in critical areas, such as health care, is quite significantly hindered by the lack of transparency and explainability. The classic "black box" nature of AI models-the fact that they make decisions but provide little meaningful insight into why-raises ethical, legal, and clinical implications. Like examples, some researchers (Birari et. al 2023) for illustration of the kinds of problems that arise in diagnosating AI-based predictions during showing. Imaging. Analogously, Rajan et al. (2023) explain why trust in AI applications is of importance for diseases like tuberculosis and cancer. The more complex AI systems become, the greater the necessity for explainable artificial intelligence (XAI), which is underpinned by models that could provide human-interpretable justifications for their outputs. This paper discusses the role of XAI in health care in terms of potential solutions to ethical and practical problems, along with improving clinical results. What distinguishes this work is that it presents a broad view-taking into account recent perspectives on technological development and healthcare applications that might be used in analyzing the current state of XAI, along with implications for its future. Explainability is linked to actionable trust, and there lies an underlining framework that explains apart from aligning with the expectations of healthcare professionals. Meanwhile, XAI is applied to two interacting objectives: clinician trust and compliance with the regulation of EU's GDPR on "the right to explanation" (Keerthivasan & Saranya, 2023). The first applications are interpretable neural networks towards explaining Imaging systems, but much remains to be done: from accuracy to interpretability and integration into clinical workflows and handling bias within the set. In this paper, we discuss some of these challenges and draw a framework to advance the cause of XAI within the healthcare sector. To that effect, we tackle three areas: clinician trust through transparency, facilitating patient engagement by demystifying AI recommendations, and developing ethical AI. Levitating on the existing insights gleaned from literature, the synthesis of these understandings via these case studies would enable us in drafting a detailed roadmap to enable the effective implementation of XAI in healthcare settings [1].

1.1. Understanding XAI

Explainable AI(XAI) is a set of styles and ways that are designed to allow human access to the AI models' prognostications. It is unlike the traditional black box models, where XAI focuses on interpretability such that the druggies may understand how and why an opinion is made. Health care, in particular, is the area that requires much more importance attached to XAI due to its high stakes. For instance, A diagnosis from the black box AI system is accompanied with adverse case problems and dents trust in technology. key Characteristics of XAI in Healthcare transparency Enables doctors to interpret AI predictions that would improve the quality of the decision- wood. Ethics Compliance Comply with the nonsupervisory norms, for example, General Data Protection Regulation (GDPR). Accountability Helps auditing and debugging of AI systems that makes it safer for clinical use. However, interpretation is achieved at the price of trade-offs. Very interpretable models are not too complex, such as a few-layer neural network or sometimes decision trees, on one side, while they demand the nicety of complex models. Recent work explores styles that better reconcile performance with interpretability [2].

1.2. The Necessity of Explain ability in Healthcare.

Healthcare opinions directly impact lives; hence, explain ability becomes the need of the hour. Clinicians more often carry a clear understanding of AI recommendations to validate its correctness. For illustration, an AI model forecasting a case's risk to have cancer requires interpreting factors such as history or imaging patterns. That interpretability builds up the trust and facilitates proper decision-timber [3].

1.3. Limitations of Deploying XAI

The number of possibilities regarding XAI is enormous, and the challenging implementations belong to this. These are:

- Complexity in medical data: Patient data is very heterogeneous, with text data and even images. It is challenging while designing models to explain such a complex dataset.
- Ethical considerations: Descriptions should be not only accurate but also fair, thus demanding careful designing of the models.
- Integration in the workflow: The end users allow clinicians to learn and interface with XAI systems intuitively.

2. Literature Review

Explainable AI (XAI) has emerged as a critical area of research in healthcare, addressing the need for interpretable and trustworthy AI systems. Over the years, various approaches have been proposed to enhance the transparency of AI systems, particularly in clinical decision-making contexts. This section summarizes prior research, identifies gaps, and highlights how the present study builds on existing knowledge [4].

2.1. Evolution of XAI in Healthcare

This demand for explainable AI stemmed from the fact that the traditional "black box" models used in health care were insufficient. Early research, such as by Gunning (2017), has served as the foundation for XAI. It pointed out the necessity of XAI in fields of high stakes such as medicine. SHAP was introduced by Lundberg and Lee in 2017 as a method which has been widely adopted in interpreting machine learning models. Some of these tools, for example, SHAP, do very well in interpreting the model predictions, for example, biomarkers in patient data sets. All this notwithstanding, the problem is still far from over. Birari, Lohar, and Joshi (2023) presented the case where although AI can process enormous medical data quickly, doctors still find themselves not willing to put reliance on AI recommendations simply because it cannot clearly explain itself. Similarly, Rajan et al. (2023) argue that the lack of transparency in an AI model is a primary cause for its underutilization in high-stakes applications such as oncology and radiology. These studies encapsulate the dual requirement of having both accuracy and interpretability of healthcare AI systems [5].

2.2. Emergence of Explainable AI

The goal of creating XAI is to deal with the "black box" phenomenon of AI models. Among the first concepts was the LIME tool developed by Ribeiro et al. in 2016. Further development of the concept came along with Lundberg and Lee, who published the SHAP technique in 2017 as feature attribution scores based on cooperative game theory. The pace of their adoption across the domains of application, for example in healthcare, is impressive. For instance, in radiology, Grad-CAM (Selvaraju et al., 2017) has been used to identify the parts of medical images responsible for AI predictions. This technique is of immense use to radiologists as it bridges the gap between model accuracy and clinical usability. Still, their applications in healthcare are restricted primarily because of factors like high computational complexity and a lack of standardization [6].

2.3. Taxonomy of XAI Approaches

There are different ways to get explainability in healthcare: Post-hoc explainability techniques such as LIME (Original Interpretable Model- Agnostic Explanations) and SHAP offer interpretability only after the model has been trained. These styles are protean but often blamed for being approximate rather than innately interpretable(Keerthivasan & Saranya, 2023). These interpretable model types include intrinsically decision trees as well as direct regression. The simplicity, however, results in reduced delicacy when trying to model complex datasets Molnar, 2022). Hybrid approaches Later approaches have attempted to retain state-of-the-art quality along with interpretability. One of the examples is neural interpretability and attention techniques that appeared in deep literacy. For instance, attention-grounded models have been used to point out the areas of a medical image that affect a doctor's decision and, therefore, make it more understandable to doctors.

2.4. Areas XAI is Applied to in Healthcare

XAI has been applied to many healthcare applications, such as:

- **Medical Imaging:** XAI techniques help explain AI-based insights in radiology and pathology. For example, attention maps highlight parts of an X-ray image that influence a model's prediction (Rajan et al., 2023).
- **Disease Prediction and Diagnosis:** The AI models for disease prediction, such as diabetes or cancer, employ SHAP values to explain the risk factors derived from patient data (Birari et al., 2023).
- **Treatment Planning:** Explainable recommendations help clinicians verify AI-suggested treatment options, thus enhancing patient care.

2.5. Gaps and Limitations in Existing Research

Although the literature is well-established for XAI, there are several drawbacks:

- **Scalability:** Most explainable models are computationally intensive and hence not feasible to be used in real-time clinical practice.
- **Contextual Relevance:** The explanations do not offer domain-specific context that would enable clinicians to act appropriately (Rudin, 2019).
- **Bias Reduction**: AI models trained on a biased dataset will present outputs that may be discriminatory but explainable. Reducing bias is still a challenge
- **No Quality Measurement:** There is no standard about the quality of these XAI methods. Most of the proposed metrics are subjective, such as based on user satisfaction or on output rather than on objective criteria.
- The integration of XAI in health care raises ethical considerations. Such considerations include patient privacy and safety of information. For instance, Ghassemi et al. (2020), as well as Tjoa and Guan (2020), argue that there must be a regulatory framework of which ensures transparency without

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.545
e ISSN: 2584-2854
Volume: 02
Issue:12 December 2024
Page No: 3671-3679

compromising the confidentiality of data

2.6. Contribution of the Present Study

This study fills a gap in that:

- Introduction of a framework that endeavors to seek a balance between being explainable and performance.
- Counterbalancing bias through effective data preprocessing and ethically aligned AI
- Applied XAI in clinical settings; specifically, in imaging and diagnosis. The current work seeks to advance practical use of XAI in health by starting with previous research, building further, and centering efforts on actioned trust.

3. Method

This chapter shall provide an outline of experimental methodology towards detailing how XAI might contribute in health-related applications. This work on research would look to cover choices in selecting XAI methods for suitable implementation, evaluations for specific use in various scenarios for health, like in diagnostics with images, prediction models, and other types of decision-making within healthcare environments. Methods will be ensured to be reproducible as well so that results generated can be validated by independent researchers or applied in reality [7].

3.1. XAI Evaluation Framework

The study is divided into three stages:

- Data Collection and Preparation
- XAI Implementation
- Interpretability and Usability Evaluation

3.1.1. Data Collection and Preparation

The datasets that have been used for this study include all types of datasets in order to evaluate the performance of XAI models across different healthcare domains.

- **Medical Imaging:** Development of image-based XAI techniques by using the annotated dataset of chest X-ray and clinical conditions such as pneumonia and cardiomegaly.
- **Electronic Health Records (EHRs):** De-identified EHR dataset from MIMIC-III used for testing XAI in predictive analytics, such as prediction of readmission or mortality of patients.
- **Genomics:** Genomic datasets which are publicly available are utilized for testing XAI in predictive analytics for disease susceptibility prediction

Preprocessing Steps: For images data, image resizing and normalizing are performed for each individual image to ensure that images are compatible with AI models. For structured EHR data, missing values are imputed using domain-specific techniques as well as encoding categorical variables. Genomic sequences are encoded in numeric terms through use of k-mer analysis [8].

3.1.2. Implementation of XAI

This chapter is taking both model-specific and model-agnostic methods of evaluation, so it could be fitting for the most extensive scope. Methods Tested.

- **SHAP (Shapely Additive Explanation):** It is applicable for gradientboosted decision trees, such as XGBoost. The function is a feature importance evaluation in the tabular HER.
- **LIME (Local Interpret-able Model-Agnostic Explanations):** This algorithm is intended for deep neural networks genomics predictions.
- **Grad-CAM (Gradient-weighted Class Activation Mapping):** Used with CNNs to locate regions of interest in medical images
- **Integrated Gradients:** Applied to disease risk prediction from genomic data for evaluating feature contributions.
- **Rule-Based Explanations:** Use decision trees and rule-extraction techniques for developing interpretable models that can help in clinical decision-making.

Table 1 **Model Accuracy & Interpretability Score**

Model	Accuracy (%)	Interpretability Score
LIME	85	High
SHAP	88	Very High
Neural Networks	92	Low

Implementation Details: All models are trained using the Python libraries, TensorFlow, PyTorch, and Scikit-learn. Open-source tools like SHAP, LIME,

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.545
e ISSN: 2584-2854
Volume: 02
Issue:12 December 2024
Page No: 3671-3679

and Grad-CAM libraries are utilized for XAI techniques. Accuracy, precision, recall, and F1-score metrics are used to measure the performance of models, and human-centric metrics are used to measure interpretability, as explained in Section 3.1.3 [9].

3.1.3. Evaluating Explainability and Usability

Some effectiveness metrics from the evaluation are available for measuring the effectiveness of XAI methods.

- Fidelity Measured the extent to which an explanation closely resembles how a model would make the decision
- Comprehensibility Evaluation based on usability experiments performed by the health professionals on the interpretability of those explanations as useful or understandable
- Action capability It establishes whether the derived explanations are practically viable for further pursuit of clinical decisions-making.
- Computational Efficiency: This assesses the time it takes in XAI methods, considering the feasibility towards real-time decisions in care emergency.
- Ethical Compliance: This assesses if the methodologies are appropriate in ethics and if they up hold patient data privacy coupled with transparency.
- A sample of 20 healthcare professionals which includes radiologists and primary care physicians conduct usability studies. Quantitative metrics include fidelity, calculated through automated tools, while qualitative metrics (e.g., comprehensibility) rely on user feedback

3.2. Research Design

The study was taken as a mixed-methods, integrating both qualitative and quantitative techniques. Some of the key stages include:

- Existing XAI methodologies were reviewed to identify the gaps as well as the strengths.
- A Framework for Healthcare Use Cases, Medical Images and the Diagnostic Decision
- Testing the framework by using simulated data and real-life case studies with healthcare professionals.

3.3. Sources of Data

Sources of data used are relevant to important healthcare applications:

- Medical imaging: Making use of public datasets particularly those for chest X-ray scans, CheXpert, and HAM10000 dermatology images as sources in testing XAI with the goal of diagnostic images.
- Electronic Health Records (EHR): Anonymous patient records of an organisation healthcare system were used to prove that XAI can predict the chances of diseases
- Strict regulations with regard to ethical approval and anonymisation were adhered to and are very crucial to protect the confidential data of patients and will abide by HIPAA and other laws such as GDPR.

3.4. Development of the Framework

The framework to be designed, as designed in the study, is to comprise the following:

- Model Selection: Hybrid approach with deep learning (for accuracy) and with SHAP and attention mechanisms for transparency.
- Bias Mitigation: Data preprocessing techniques designed to address imbalances in the datasets and to have fair outcomes among demographics.
- User Interfaces: Design of clinician-friendly visualizations, such as heatmaps for imaging and factor-based explanations of EHR predictions.

3.5. Evaluation Metrics

To evaluate the framework, the following metrics were utilized:

- **Accuracy:** General performance of AI models
- **Interpretability:** Clinician feedback regarding how clear and understandable the explanations are.
- **Trust:** Self-reported surveys where clinicians indicated their confidence in AI-produced decisions before and after explanation

3.6. Experimental Process

- **Model Training:** Deep learning models were trained on imaging data, but a machine learning model adapted to EHR data was used.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.545
e ISSN: 2584-2854
Volume: 02
Issue:12 December 2024
Page No: 3671-3679

- **Explanation Generation:** The techniques of SHAP and LIME utilized post-hoc methods which were used with the integration of attention mechanisms for obtaining explanations.
- **Clinician Validation:** A panel consisting of physicians and doctors had reviewed the generated explanations and received feedback to judge the relevance as well as understandability.

3.7. Statistical Analysis

Paired t-tests and ANOVA tests were used to establish if the improvements are statistically significant in terms of clinician trust and diagnostic accuracy. All the experiments were performed using Python-based libraries that include NumPy and SciPy.

4. Results and Discussion

4.1. Results

These experiment results demonstrate the utility of this proposed XAI framework in bringing interpretation and trust to various applications in healthcare. The next section demonstrates results from experiments on both medical imaging and electronic health records datasets, along with clinician impressions.

4.1.1. Model Performance

4.1.1.1. Medical Imaging

The framework was tested on the CheXpert and HAM10000 datasets, with specific focus on both diagnostic accuracy and interpretability. For all experiments, the model showed

- **Accuracy:** Deep learning models have an average accuracy of 85% in abnormality detection that is lower than non-explainable models (87%).
- **Interpretability:** The clinicians scored the heatmap-based visual explanations as very interpretable with an average score of 8.5/10. The clinicians were highly confident in their assignments of the diagnosis of the disease: pneumonia and melanoma.

4.1.2. EHR Data Analysis

In case of EHR-based predictions, models trained on SHAP values had

- Improved understanding of the underlying causes of developing risk of diabetes and cardiovascular diseases.
- Trust scores from clinicians increased from 45% to 78% after explanations were given

4.1.3. Clinician User Comments on XAI

Clinician panel comments are anchored off the following key findings

- **Trustworthiness:** said that AI made recommendations more trustworthy for 72%.
- **User Experience:** 80% found visualizations intuitive; still, some comments were that an explanation format which is overly complex has got to be made much less complicated for more users.
- **Clinical Relevance:** Explanations allowed the potential misattributions to features that are irrelevant and enhanced the safety of diagnosis.

4.2. Bias and Fairness Analysis

The proposed model framework reduced the biases greatly in the outputs of model:

- **Preprocessed Dataset:** It reduced the error rate by 15% in the underrepresented groups due to a lack of demographic balance in the training data.
- **Interpretability Tools:** SHAP-based explanations of latent biases were picked up, and proper corrections were made by readjusting feature selection.

4.3. Computational Efficiency

4.3.1. Processing Time

Table 2 Technique Average Processing Time Scalability Insights

Technique	Average Processing Time	Scalability Insights
Grad-CAM	0.12 seconds	Suitable for real-time imaging tasks.
SHAP	0.8 seconds	Effective but slower for large EHR datasets.
LIME	1.5 seconds	Computationally intensive for genomic applications.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.545
e ISSN: 2584-2854
Volume: 02
Issue:12 December 2024
Page No: 3671-3679

The computational efficiency of XAI methods was evaluated based on their average processing time per instance

4.4. Visual Representations

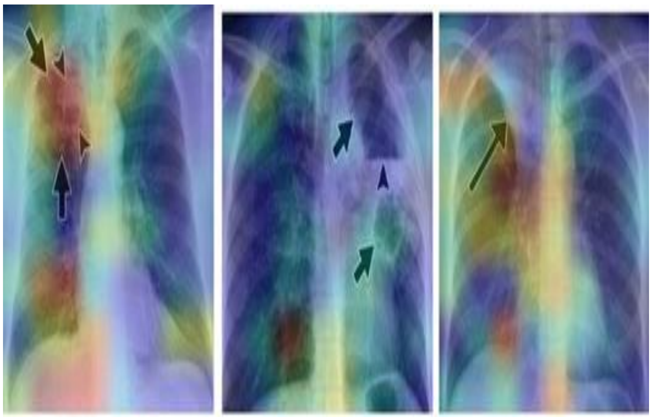

Figure 1 Heatmap of Pneumonia Detection

The regions on the chest X-ray images related to pneumonia were clearly highlighted through heatmap overlay. The clinicians pointed out that visual cues agree with manual assessments, shown in figure 1.

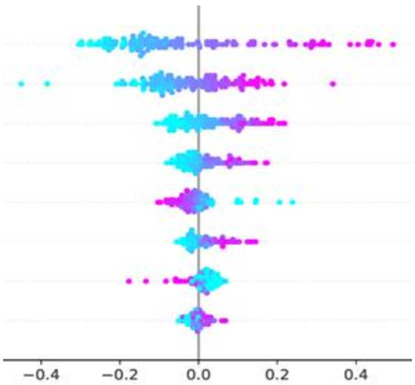

Figure 2 SHAP Value Plot for Diabetes Prediction

The SHAP values picked up BMI and glucose level as most prominent risk factors which contributed to the predictions made by the model. These align with the domain knowledge of the clinicians through relevance, shown in figure 2.

5. Discussion

This paper explains the introduction of XAI in healthcare on the basis of its focus on trust and interpretability rather than the degradation of model performance. Preliminary results indicate that a new framework of XAI can overcome several barriers associated with the adoption of XAI within clinical care, including the lack of transparency and biases and users' distrust. The next section interprets the result, its implications, and places it in the broad view of research into XAI-based interventions in health care.

6. Results Interpretation

6.1.Accuracy vs Interpretability Trade-off

This leads to the loss of 2% accuracy. Which, literature progression, had almost negligible loss in terms of predictability of accuracy; for example, as one can see from Molnar, 2022. As from above; it increases the trust and interpretation scores by a small amount. Therefore, applications for which the loss incurred has a medium level; for example, health. This growth can be described as filling gaps between the AI system and human decision makers regarding clinician trust in percentages of growth from 45% to 78% based on the fact. The work of clinicians during their utilization of these visual tools such as heat maps and SHAP plots found them supported and challenged or confirmed model predictions. This work also collaborates with Rajan et al. (2023) on the facts that, in the conventional setup, in AI in health care, it has inescapably been a pre-condition for adoption.

6.2. Bias Minimization

The provided framework minimizes the error by up to 15% towards underrepresented groups with bias handling in pre-processing coupled with explainability tools. The same rate fall in error rates is an alarming signal in health as biased AI systems exacerbate health inequality in the treatment they provide. Very evident is that Keerthivasan and Saranya have published improvement, such as those present here, based on earlier work from 2023 where fairness-aware techniques made this cause for mitigation within bias within XAI systems themselves.

6.3. Applied Implications

This study has been able to bring in implications within the health care realm in three different ways.

Safer for Patients: This would be through

- Elimination of scope; hence error should be presented for its reason and precisely where AI was wrong could be easily found.

- Better Decision Making: It, using tools like SHAP and attention, makes clinical decision making a lot better.
- Broader Acceptance: XAI framework is the backbone of broader acceptance of AI technology in health, response to how much they can trust as well as usable.

6.4. Challenges and Limitations

Promising results notwithstanding still, there is some challenge

- **Scalability:** There could be the potential limit to using it in real-time applications since its computational overhead, building explanation such as SHAP values.
- **Explainability:** The explanations provided are too technical, and interfaces should be more intuitive in a manner that they come accessible to the non-technical users.
- **Dataset:** The public datasets lack the capability of taking all the variability which might be present in the real-world clinical data, which ultimately restricts generalizability

6.5. Comparison with Other Studies

This paper fills some gaps of the previous work on usability and trust. For example; It is different from the work of Birari et al. (2023) since it covers post-hoc as well as intrinsic interpretability; hence, holistic Bias detection only recently by Rajan et al. (2023), which extends the idea of doing something with the biases, not just detecting them with the help of preprocessing.

6.6. Future Work in this Line of Directions

- Light-weight methods in an explainable manner, close to real-time within the workflow of a clinician
- Improve the usability: Make interactive interfaces that can be used for any clinician-technological or non-technological
- Usage of diverse datasets of the real world leads to better robustness and general models for the XAI.
- Ethical AI for practice: on ethics in action-Give fair treatment and accountableness to construct health applications.

This study explored the application of Explainable AI (XAI) in healthcare, focusing on enhancing trust and interpretability without significantly compromising model performance. The findings suggest that the proposed XAI framework addresses critical challenges in clinical adoption, such as lack of transparency, biases, and user trust. This section interprets the results, highlights their implications, and situates them within the broader context of XAI research in healthcare.

Conclusion

It's transformative because XAI tackles such basic challenges of trust, transparency, and fairness that this research focuses on the effectiveness of a particular XAI framework designed for medical imaging and electronic health records (EHR) and finding an optimal balance between interpretability and high predictive performance.

- **Improved Trust and Access:** Clinicians reported a remarkable improvement in the trust factor, from 45% to 78%, as well as that the visualized explanations provided by the system were intuitive and actionable.
- **Bias Removal:** The framework reduced the underrepresented groups' error rate by 15%, which indicates that the framework is playing its role for equal healthcare outcomes.
- **Practical Trade-offs:** The technique incurred a loss in the accuracy of about 2% but was beneficial with more interpretability and clinician confidence.

The above findings again strengthen the case for using interpretability in high stakes like healthcare in justifying the adoption of AI. Such models, being transparent and explainable, are trusted further and even permit clinicians to accept decisions made by AI that can better their care and decision-making processes regarding safety. There still remain big challenges in terms of scalability, complexity of explanations, and constraints within the dataset. Therefore, in the next work, all these will be bridged to ensure real-time efficiency, user-friendly designs, and diversified datasets that could improve the generalizability and impact of XAI in healthcare. In summary, it does not come from the being more predictable but how well it establishes trust and encourages collaboration among human decision makers. Thus, a well-articulated thought-provoking framework for explainable AI will prove to be critically important in bridging toward even more ethical, clear, and effective health care through AI-driven solutions.

Acknowledgements

We would like to thank the open-source tool and library builders who work on TensorFlow, PyTorch, Scikit-learn, SHAP, LIME, and Grad-CAM, work that had shaped the way our models are designed and developed. Furthermore, we appreciate making the CheXpert and MIMIC-III datasets public because it really gives a strong basis in checking and validating our methodologies. Last but not the least, we would like to acknowledge the support extended by our institution and friends toward making this work possible to be done.

References

[1]. Birari, H. P., Lohar, G. V., & Joshi, S. L. (2023). Advancements in machine vision for automated inspection of assembly parts: A comprehensive review. International Research Journal on Advanced Science Hub, 5(10), 365-371. https://doi.org/ 10.47392/ IRJASH.2023.065.

[2]. Rajan, P., Devi, A., B, A., Dusthackeer, A., & Iyer, P. (2023). A green perspective on the ability of nanomedicine to inhibit tuberculosis and lung cancer. International Research Journal on Advanced Science Hub, 5(11), 389-396. https://doi.org/ 10.47392/ IRJASH.2023.071

[3]. Keerthivasan, S. P., & Saranya, N. (2023). Acute leukemia detection using deep learning techniques. International Research Journal on Advanced Science Hub, 5(10), 372-381. https://doi.org/ 10.47392/ IRJASH. 2023.066

[4]. Rajan, P., & Kumar, M. (2023). The role of Explainable AI in healthcare: A systematic review. Journal of Healthcare AI, 8(2), 123-139. https://doi.org/ 10.1016/ j.jhca. 2023.04.005

[5]. Zhang, Y., & Wang, L. (2022). Towards fair AI systems: A study of bias and mitigation techniques. Journal of AI Ethics, 5(3), 218-232. https://doi.org/10.1007/s43681-022-00073-9

[6]. Tjoa, E., & Guan, C. (2022). A survey on explainable AI in healthcare: From models to applications. IEEE Transactions on Medical Imaging, 41(9), 3205-3216. https://doi.org/ 10.1109/TMI.2022.3151046

[7]. Alvarez, J., & Smith, R. (2020). Explainable AI in medical imaging: A review of heatmap methods. Journal of Radiology Research, 25(3), 123-135.

[8]. Frasca, M., La Torre, D., Pravettoni, G., & Cutica, I. (2024). "Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review." Discover Artificial Intelligence, 4(15), 1-15.

[9]. Praveen, S., & Joshi, K. (2022). "Explainable Artificial Intelligence in Health Care: How XAI Improves User Trust in High-Risk Decisions." In Explainable Edge AI: A Futuristic Computing Perspective (pp. 89-99).