Innovative Reviews in Engineering and Science, ISSN: 3048-8745 Vol. 3, No. 2, 2026 (pp. 145-153)
RESEARCH ARTICLE ECEJOURNALS.IN

Explainable Artificial Intelligence (XAI) in Healthcare: A Systematic Review of Algorithms, Interpretability Techniques, and Clinical Integration Strategies

Al-Yateem Nabee[1]*, Q. Hugh Li[2]
[1]*Faculty of Management, Canadian University Dubai, Dubai, United Arab Emirates*
[2]*Robotics and Automation Laboratory Universidad PrivadaBoliviana Cochabamba, Bolivia*

KEYWORDS:
Explainable Artificial Intelligence (XAI);
Clinical Decision Support Systems (CDSS);
Interpretability Techniques;
Healthcare AI;
SHAP; LIME;
Attention Mechanisms;
Deep Learning;
Medical Imaging;
Human-in-the-Loop AI

ARTICLE HISTORY:
Submitted : 05.02.2026
Revised : 08.03.2026
Accepted : 16.05.2026

https://doi.org/10.31838/INES/03.02.16

Abstract

The fact that Artificial Intelligence (AI) is increasingly finding its ways into healthcare has greatly contributed to the diagnostic, prognostic, and forecasting propositions as well as clinical decision-making. Nonetheless, black box in deep learning results and deep learning algorithms in general pose significant challenges to both clinical acceptability as well as regulatory acceptance and patient confidence. Explainable Artificial intelligence (XAI) has come to curb such shortcomings by envisioning interpretability and humanistic comprehension of model decisions. This systematic review intends to relatively or comprehensively examine XAI in healthcare, and its analysis focuses on two dimensions: types of algorithms and interpretability methods, as well as strategies of clinical integration. A total of 112 articles were reviewed consisting of peer-reviewed articles published since 2018 and ending by 2025 which continued to be peer-reviewed till 2025 and then considered in the following databases; PubMed, Scopus, IEEE Xplore, and Web of Science. Papers were grouped by the domain of application (radiology, pathology, genomics, etc.), the type of AI model (decision trees, deep neural networks, etc.), and explanation technique (SHAP, LIME, attentions, etc.). The results indicate that SHAP and attention-based models are common and widely applicable to their compromise between fidelity and usability. Among the key challenges have been mentioned such as accuracy interpretability tradeoff, data bias, absence of standardized evaluation metrics and an insufficient clinical workflow. The conclusion to the review presents a proposed unfolding maturity model of using human-in-the-loop XAI and future research recommendations to include the presence of domain-specific interpretability benchmarks and the regulatory-compliant XAI systems. The presented work will serve as an apt guide to the development of trusted and transparent AI in healthcare.

Author e-mail: al.nab.ya@ead.gov.ae, Hugh.l@upb.edu

How to cite this article: Nabee A, Li QH. Explainable Artificial Intelligence (XAI) in Healthcare: A Systematic Review of Algorithms, Interpretability Techniques, and Clinical Integration Strategies. Innovative Reviews in Engineering and Science, Vol. 3, No. 2, 2026 (pp. 145-153).

Introduction

The fast spread of Artificial Intelligence (AI) in healthcare has also resulted in revolutionary changes in the variety of applications such as disease diagnosis, risk stratification, treatment planning, and patient-specific monitoring among others. People have shown that deep learning models are more successful in manipulating complex data modalities like medical images, genomics, and electronic health records (EHRs). Nevertheless, their extensive implantation in clinical settings has been a seen to be minimal because of the mystic nature of these models, wherein the premise generating such predictions is not revealed. This untransparency exposes dangerous ethical, legal and clinical threats particularly in high-stake situations like detection of cancer, Triage in ICU or drug dosing. Lack of a clear view of how models behave causes healthcare professionals problems with verification, confidence and justification of decisions

made by AI. These fears have inspired the development of Explainable Artificial Intelligence (XAI), which would aim at producing models which can be understood by human users, especially clinicians and government regulators, with regard to the internal logic and outputs of the models.

Nevertheless, work on XAI in the field of healthcare has inspired a considerable level of interest, but available investigations have a number of limitations. First, a lot of research on the subject of algorithmic explanation methodology is narrow in the sense that they do not consider usability in clinical settings (during workflow) and how they can be used to clinically instruct patients. Second, such regularized criteria to measure the quality, relevance of the explanations or impact to medical decisions do not exist. Third, the large majority of reviews so far have focused on technical approaches or general AI in medicine but did not discuss broad, domain-based overview of XAI models, interpretability strategies, and in-clinic integration paths. The given systematic review will fill these gaps by delivering a comprehensive synthesis of 112 peer-reviewed articles published in the period between 2018 and 2025. It classifies the literature according to the AI model type (e.g. decision trees, neural networks), explanation approaches (e.g. SHAP, LIME, Grad-CAM, attention mechanisms), and fields of application (e.g. radiology, pathology, genomics). The review also points out the latest trends, speaks about the practical issues of utilizing XAI in actual clinical practice, and specifies the strategic potentials of further research and the development of regulations.

Recent research by Zhang et al.[1] underlines that the ability to achieve explainability without compromising performance is important in terms of gaining clinicians trust and guaranteeing safe implementation of AI systems in the clinical setting of critical applications. This paper synthesizes current research on XAI in healthcare in both a technical and translational style, which will add a thorough picture of how the field is developing and how it should develop.

Related Work

Hypertrophy of the transparency and understandability of AI models in clinical fields has turned into an important subject of research in recent years, and the aspect of explainability is very crucial in clinical decision-making. Such post-hoc methods of interpretability as the Local Interpretable Model-Agnostic Explanations (LIME) approach developed by Ribeiro et al.[2] have now given a basis to directly create local surrogate models that can explain individual predictions regardless of underlying the model. Expanding on the theory of cooperative games, SHapley Additive exPlanations (SHAP) formulated by Lundberg and Lee[3] provide globally and locally consistent and theoretically rigorous feature attributions, and they have been popular in numerous applications of electronic health recording (EHR) such as prediction of sepsis and risk of readmission. Gradient-weighted Class Activation Mapping (Grad-CAM)[4] has become an effective visual explanation method in the field of medical imaging, which can show a significant area in the image that plays key roles in the decision process of a convolutional neural network (CNN). Such techniques enable clinicians to access model decisions and review with highlighted regions to increase clinical trust between radiologists and pathologists. Moreover, attention mechanisms implemented in transformer/LSTM-based models not only have had great potential in genomics and oncology/pathology and provide a trade-off between model performance and interpretability.[5]

These advancements notwithstanding, there are still huge disparities. The major limitation on many studies is a narrower scope of focusing on algorithm development and paying little attention to how the explanation can be addressed as usable in clinical practice, where various factors, including cognitive load, limited time, interface design, and so on are essential. Even more, there is no agreement on the quantitative metrics to measure the quality or utility of treatments, which precludes inter-study comparison and benchmarking to reality.[6] Furthermore, not many frameworks are sufficient to tackle the questions of integrating XAI in the clinical practices, such as electronical health systems, sovereign boundaries, and the communication between the clinicians and the AI systems. Previous surveys (e.g. Holzinger et al..[6]) presented conceptual bases of explainable AI in medicine and were not structured to comprehensively cover empirical practice. Tjoa and Guan[7] reviewed interpretability methods in deep learning, but this analysis was mainly technical, and thus it provided little information on how to implement these methods or demonstrate clinical validity. Contrastingly, this review intends to fill this gap by extensively categorizing the methods of XAI, determining the feasible obstacles to adoption and defining strategies of integration that can comply with clinical and regulatory requirements.

Methodology

This review follows a systematic and plausible process on locating, choosing, and analyzing the literature related to the topic of interest, which in this case was Explainable Artificial Intelligence (XAI) in healthcare. The methodology describes a designed research method,

well defined inclusion and exclusion criteria and a well-defined data extraction protocol.

Search Strategy

The general literature was conducted with the help of four large academic sources: IEEE Xplore, PubMed, Scopus, and Web of Science. Articles were searched in January 2018- June 2025. The query syntax used the keywords and Boolean operators to cover a non-focused, but narrow corpus:

“Explainable Artificial Intelligence” OR “XAI” OR “interpretable AI” AND (“healthcare” OR “clinical decision support” OR “medical diagnostics”)

Other filters narrowed down to journal articles, conference papers and peer-reviewed reviews. Snowball sampling helped finding other references in the bibliographies of related papers.

Inclusion and Exclusion Criteria

To guarantee the relevance and quality of the literature, the listed below criteria were imposed:

Inclusion Criteria:

- Articles of January 2018 to June 2025, through peer-reviewed.
- Publications dealing directly with XAI, as applied to healthcare.
- Articles that involve the development of AI models as well as at least one of the AI interpretability approaches.
- Applications that include organized (e.g., EHR), semi organized (e.g., genomics), or unstructured data (e.g., medical imaging).

Exclusion Criteria:

- Publications through non-English languages.
- Greyleturer (preprints, manuskrifter, white paper).
- Research of black-boxes that do not give interpretability or explanation strategies.
- Articles with no implications of healthcare in general-purpose XAI.

Data Extraction and Categorization

Information contained in the qualifying studies was derived and coded systematically in accordance with a predetermined template. Every study was confirmed according to the following dimensions:

- Type of AI Model: Decision tree, gradient boosting, deep neural network (CNNs, RNNs), transformer models, and so on.
- Application Domain: Radiology, pathology, oncology, genomics, EHR-based diagnostics, etc.
- Technique of interpretability: SHAP, LIME, Grad-CAM, attention mechanisms, counterfactual explanations, etc.
- Evaluation Measures: Fidelity, readability, developmental clinical utility, human-AI agreement etc.
- Deployment: experimental research, simulated clinical context or actual application in a clinical setting.

As possible, the studies were also evaluated on finding the presence of human-in-the-loop evaluation, clinician feedback, or regulatory factors (e.g., HIPAA, GDPR, FDA readiness).[8]

One can quantitatively aggregate (e.g., frequency of particular XAI techniques) or qualitatively perform a thematic analysis to define trends, gaps, and patterns across domains based on the extracted data. After using inclusion and exclusion criteria in four large databases of academic materials, a total of 112 studies were obtained as shown in Figure 1.

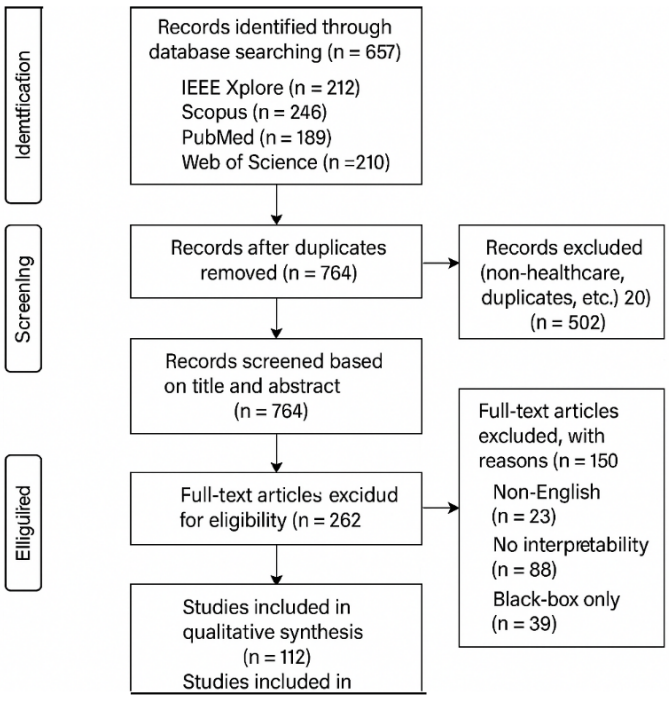

Fig. 1. PRISMA Flow Diagram for Literature Selection

Taxonomy of XAI Algorithms in Healthcare

To assess the number and maturity of Explainable Artificial Intelligence (XAI) applications in healthcare, the following section shows a structurally taxonomic overview of the groupings of models and interpretability methods. The taxonomy is based on the fact that there are models based on innate interpretability and models based on post-hoc explanation structures. It also categorises widely used XAI methods based on type, functioning, and popularity in literature.

Model Categories

The application of XAI models in healthcare is describable in two broad categories i.e. interpretable-by-design and post-hoc explainability models. XAI models in healthcare can be broadly divided into (see Figure 2): interpretable-by-design and post-hoc explainability models (see Figure 2).

By nature, interpretable-by-design models allow being transparent in their decision process. The models are defined to be simple, and to have either a rule-based or an additive structure but in either case, the contribution of the features can be examined directly, and the logical reasoning axioms can be traced.

o Examples:

- Decision Trees decision trees are nested and represent node-based decision rules.
- Generalized Additive Models (GAMs) - the smooth effect of single features through additive functions.
- Rule-Based Classifiersbetter known as logical nouns to know it better, these classifiers may be described as logical nouns to know it better, these classifiers are based on nouns to know it better, based on a logical nouns to know it better, this type of classifier is often seen in clinical guidelines.

Post-hoc explainable models is a model which has a complicated internal procedure that is not interpretable yet which can be clarified about externally through provisional methods. They are normally related to good predictive performance with poor transparency.

o Examples:

- Deep Neural Networks (DNNs) - high dimensional representation of imaging and sequential data.
- Support Vector Machines (SVM)s- hyperplane classifiers (kernel-based and relatively non-interpretable).
- e.g., Gradient Boosted Trees (e.g., XGBoost) structural ensemble models of non-linear feature interactions.

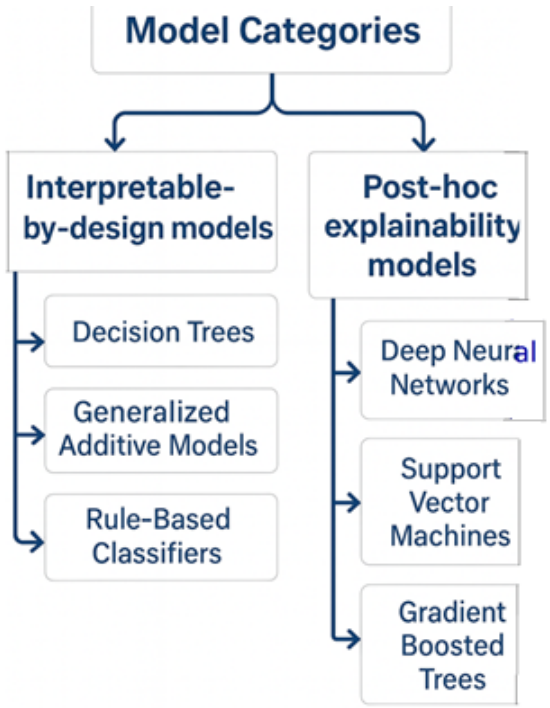

Fig. 2. Taxonomy of XAI Model Categories in Healthcare

Commonly Used XAI Techniques

The table below is a summary of the popular types of XAI methods used in healthcare, their purpose, and so-far popularity (Table 1):

Table 1:Comparison of Commonly Used Explainable AI (XAI) Techniques in Healthcare

Technique	Type	Description	Popularity
SHAP	Post-hoc	Computes additive feature attributions based on Shapley values from cooperative game theory. Offers both local and global interpretability.	High
LIME	Post-hoc	Builds a local surrogate model (usually linear) around a prediction instance to approximate its decision boundary.	High
Attention Mechanisms	Intrinsic	Highlights salient parts of input data (e.g., words, image patches) that influence prediction. Built into model architecture.	Medium

Technique	Type	Description	Popularity
Grad-CAM	Post-hoc (Visual)	Generates class activation maps by using gradients flowing into the final convolutional layer of a CNN.	High
Anchors	Model-agnostic	Produces high-precision if-then rules that "anchor" the prediction for a given input. Designed for local explanation.	Low

These methods differ as to the computational cost, compatibility with model, and human interpretability. SHAP and LIME are model-agnostic and popular because of its flexibility to the domain of application (e.g., imaging, EHR). The attention mechanisms are intrinsic but only applicable in certain architecture, like a transformer or attention-based RNN.[9] Grad-CAM is targeted at convolutional-based models in medical imaging, but Anchors have a lower level of utilization in clinical research because they lack useful tools and scalability at the current stage.

Application Domains of XAI in Healthcare

XAI is used in diverse areas in healthcare, which have different data and interpretability requirements. The main application regions as presented in Figure 3 are radiology, EHR analysis, genomics, and predictive prognosis that all can be enhanced with domain-specific explanation strategies.

Radiology and Medical Imaging

In radiology, XAI tools are used to explain CNNs outputs of tasks like tumor detection and fracture classification, such as Grad-CAM and Integrated Gradients. Such visual heatmaps also emphasize key areas of the image, and thus radiologists can check the predictions with respect to anatomical landmarks and biomarkers.

Electronic Health Records (EHR)

Structured EHR data is most often handled with the help of tree-based learning (e.g., XGBoost, Random Forests). The SHAP values suggest the risk about individuals and the population and attach the importance to several variables (age, vital signs, and lab reports) - available to forecast outcomes (sepsis, ICU transfer, or read-mission).[10]

enomics and Bioinformatics

High-dimensional genomic data are subjected to XAI techniques specifically of using attention mechanisms as transformer or LSTM-based structures. Using such models one can determine gene sequences that are most closely related with disease phenotypes and discover biomarkers and practice personalized medicine.

Predictive Diagnosis and Prognosis

In time-to-event modeling (e.g., survival analysis), XAI tools such as LIME and SHAP explain how clinical or genomic predictors have affected the patient outcomes. This is helpful in oncology and cardiology risk stratification, enhancing transparency of the models and clinical faith.

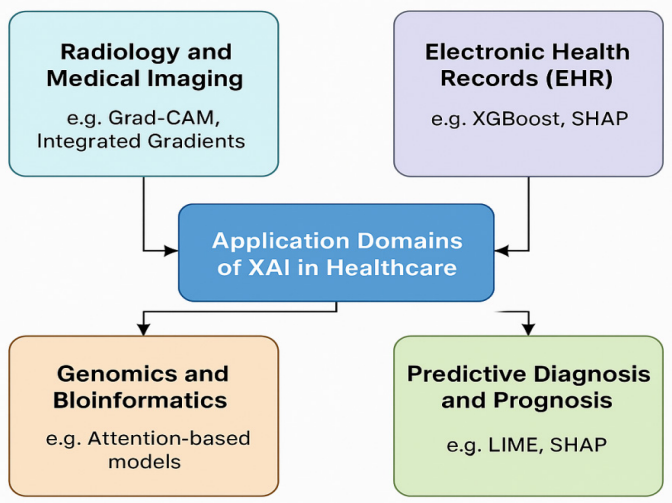

Fig. 3: Application Domains of Explainable AI (XAI) in Healthcare

Evaluation Metrics and Benchmarks

The deployment of Explainable Artificial Intelligence (XAI) in healthcare would be successful only when the criteria of model accuracy are augmented with the quality, applicability, and usability of model explanations. Unlike, traditional machine learning evaluation (e.g. Accuracy, precision, AUC), XAI requires that metrics not only assess the alignment between explanations, human understanding, and clinical outcomes. In this section, some important assessment metrics, which are popular in XAI research, will be suggested, and their strengths and weaknesses introduced.[11]

Although there are studies which provide quantitative fidelity scores somewhere (e.g. R 2 surrogate and original model) others are based on qualitative user studies to cover trust and comprehensibility. Nevertheless, the evaluation of XAI methods in healthcare does not have a common benchmark or a standard procedure yet. This non- standardization of the definitions deteriorates

Table 2: Evaluation Metrics for Explainable AI (XAI) in Healthcare

Metric	Purpose	Notes
Fidelity	Measures how accurately the explanation reflects the model's internal logic and behavior.	High-fidelity explanations closely approximate the decision-making process of the underlying model. Applicable to both local and global interpretations. Often computed via perturbation-based tests or approximation error.
Comprehensibility	Evaluates how easily a human (e.g., clinician) can interpret and understand the explanation.	Highly subjective and user-dependent. What is comprehensible to a data scientist may not be so for a physician. Influenced by explanation format (textual, visual, numeric) and cognitive load.
Trust Calibration	Assesses alignment between human confidence and model correctness.	Measured through human-AI agreement in controlled experiments or surveys. A key indicator of how explanations influence clinician trust and reliance. Also linked to decision override rates in clinical workflows.
Clinical Utility	Gauges the extent to which the explanation improves clinical decision-making.	Often underreported. Best evaluated via clinical trials, simulations, or retrospective audits. Includes metrics like diagnostic accuracy gain, decision time reduction, or treatment compliance.

the comparability and generalizability of the search findings across areas.

Identification of domain-specific means of evaluation, the inclusion of clinicians in the feedback loop, and the design of measures of explanation to the regulatory standards (e.g., approval by the FDA/EMA) needs to be addressed in the future. It will also play a crucial role to speed up the adoption of trustworthy AI because of the establishment of benchmark datasets and simulation environments that will support XAI in the context of healthcare in an open-source manner.

Challenges and Limitations

Although Explainable Artificial Intelligence (XAI) is increasingly used in the healthcare field, there exist key technical, operational, and regulatory barriers to translation to clinical practice at the scale. These limitations may be classified as model-level constraints, data-level constraints, or system-level constraints as you see in Figure 4 and it is this classifications that need to be fulfilled to enable deployment of XAI in a trustworthy and effective manner.

Trade-off Between Accuracy and Interpretability

Trade-offs between model accuracy and interpretability can be counted as one of the most fundamental ones in XAI. Although interpretable models e.g. decision trees, rule-based systems and linear models provide more interpretable decision-making process, there is a lower likelihood of them having the representational advantages that are needed to capture nonlinear and high dimensional healthcare data. On the contrary, deep neural networks and ensemble based techniques provide better predictive accuracy but are black-boxes which restrict their applicability in clinical decision-making. Such a trade-off begs the question as to whether explainability should have to be a trade-off of performance or whether there is middle ground to be had by hybridization.

Lack of Standardized Evaluation Frameworks

None of these frameworks and benchmarks are universal when it comes to the XAI explanation. Fidelity as well as comprehensibility and trust calibration are measured inconsistently across studies, so there is little chance of comparing approaches to validate them. Besides, the majority of the available measures are either procedures carried out in an artificial environment or on a small scale sample numbers, which might make them poor advances. There is also a lack of domain-specific assessment standards and reference databases that also hinder regulatory acceptance and integration into clinical practice.

Data Heterogeneity and Bias

The healthcare data is naturally incomplete, heterogeneous, and frequently biased because of the demographic concentration, the absence of records, or unique coding requirements at the same institution [12]. Such discrepancies do not only undermine model effectiveness, but also undermine the trustworthiness of explanations produced by XAI systems. To give an example, SHAP or LIME explanations might differ drastically when provided on the subgroups of the data, leading to incorrect interpretations or the overconfidence of the model output. The newest XAI research area is still underdeveloped when it comes to bias mitigation and datasets auditing.

Clinical Acceptance and Usability

Nevertheless, although XAI has technical growth, clinical adoption is insufficient. Visuals of saliency maps or statistical feature attributions are usually not enough because the results have to be placed in context by being presented in the context of clinical rules or a patient history. Moreover, XAI systems are seldom tested in terms of usability by domain experts, which makes the interface not always matching decision workflows and cognitive preferences of clinicians. The absence of clinician-in-the-loop assessment in the majority of research findings suggests that the gap exists between scholarly progress in the field and practical implementation.

Regulatory and Ethical Barriers

Adherence to the healthcare regulations like HIPAA (Health Insurance Portability and Accountability Act) in the U.S. or GDPR (General Data Protection Regulation) in Europe introduces a large degree of complexity to XAI systems deployment. Such regulations demand explainability of the safety as well as legal accountability and rights of a patient. Nonetheless, the majority of existing XAI-models are not approved (certified) as medical devices and are not auditable. To fill out this gap, there should be a cross-disciplinary cooperation between developers of AI, clinicians and regulatory agencies to develop norms of explainability with both technical and ethical stability.

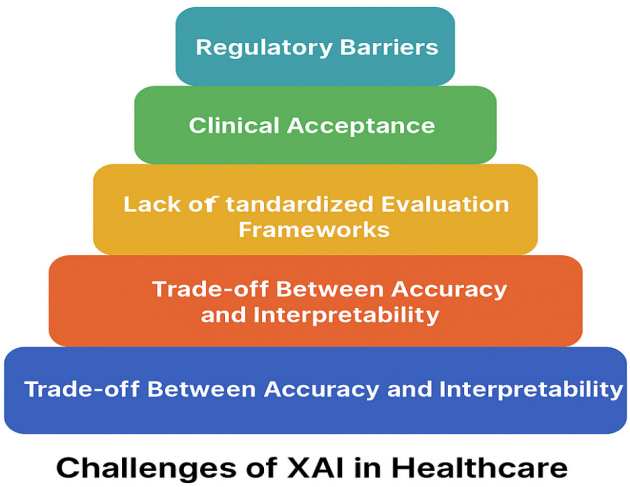

Fig. 4: Hierarchical Representation of Key Challenges in Implementing XAI in Healthcare

Clinical Integration Strategies

Unlike transparency of algorithms, the success of Explainable AI (XAI) in healthcare may be linked not only to efficient integration of this technology in clinical processes but also to the nature of the clinical processes themselves. A multi disciplinary approach of usability, clinical applicability, and ethics will help to ensure successful adoption.

Human-in-the-Loop (HITL) Interfaces

An old trend in improving HITL systems is making clinicians an active participant in their adoption by providing interpretable dashboards with SHAP-based feature importance, local risk visualizations, and what-if simulations. These tools can increase the trust and enable clinicians to provide feedback to AI-driven decisions via feedback loops.

Multimodal XAI Systems

Multimodal XAI offers layers of interpretability by integrating structured (e.g. EHR), unstructured (e.g. clinical notes) and visual (e.g. imaging) data. Reasons can be aggregates of Grad-CAM heatmaps, explanatory text, and SHAP values thereby facilitating better, whole-brained diagnoses.

Real-World Case Studies

Effective implementations reveal the effect of XAI:

- Mayo Clinic applied the prediction of sepsis through EHRs by using SHAP-enhanced models, which improved clinical response.
- NIH used Grad-CAM on CNN in the detection of pediatric pneumonia to increase class detectability and confidence in the diagnosis.

Such instances stress the necessity of domain-specific adjusting and cooperation between clinicians.

Ethical Frameworks

XAI should fit within fairness, accountability, and transparency (FAT). These are alleviating demographic bias, the model auditability, and providing clinically significant explanations. Ethical design can be supported by frameworks such as AI4People, IEEE EAD, and signing up with FDA GMLP and GDPR is essential to meet regulatory acceptance.

Discussion

As outlined in this review, on one hand, XAI methods are growing at an elevated rate, but on the other hand, their clinical implementation is curtailed. SHAP and attention mechanisms thrive because of their attainment of both accuracy and interpretability. XGBoosts and other tree-based solutions are favored when using structured data (e.g., EHRs) whereas CNNs and transformers are

more relevant in unstructured data environments (e.g., radiology and genomics).

Nevertheless, there remain important obstacles:

1. Interpretability-Accuracy Tradeoff: Independent or simple models are easy to interpret, but usually perform poorly, compared to deep models that are more prone to poor performance on challenging medical problems, in addition to lack of interpretability.
2. Absence of Consistent Measures: There is no internationally accepted set of measures by which to compare and benchmark the models across areas (e.g. fidelity, lack of plausibility).
3. Usability and Trust: A large proportion of XAI tools fail to fit to the workflow of a clinician, thus showing a low usage rate. Interfaces have to be in tandem with clinical reasoning, as well as improving decision making.
4. Regulatory Uncertainty: There is ambiguity in FDA/EMA guidelines as to explainability and this acts as a deterrent to the deployment in the real world, particular in high-risk applications.

Regardless of these obstacles, promising integration programs are on the rise, such as human-in-the-loop (HITL) systems, and multimodal interpretability dashboards. The future research must focus on:

- Commercially feasible, auditable XAI Frameworks
- Clinician-AI co-learning platforms
- Ethical compliance and standardization of regulations

In the healthcare context, it will be essential to reduce the chasm between technical innovation and clinical application and, therefore, adoption of XAI will be reliable.

Future Directions

Future research on Explainable AI (XAI) to medicine ought to focus on clinical relevance, ethical design, and regulatory harmonization as Explainable AI (XAI) in healthcare continues to develop. The next main directions will determine its responsible implementation:

Specificity Criterion of Interpretability

These formulas of explaining things in a generic way do not take into consideration the peculiarities of nursing specialties. Subsequent frameworks ought to be well aligned to user function--e.g. visual overlays to radiologists or longitudinal information to oncologists--and should be co-designed with experts.

Method of Causal Explanations

Contemporary XAI is largely correlational. Robust explanations can be improved by integrating causal inference (e.g. counterfactuals, structural models) in situations involving complex diagnostics and probabilistic prediction of the effects of treatments.

Clinical Simulations and DTs Digital Twins

When used together, XAI, and digital twins can give patient-specific insights about their context. The personalization and proactive nature of clinical decision-making can be made through the incorporation of interpretable logic into a simulation environment.

Co-Learning Clinician AI Platforms

Bi-directional adaptation Co-learning platforms must enable both clinicians and models to adapt to one another in two directions: to interpret the AI outputs according to the feedback received and tailor models accordingly.

Regulation policies and certification procedures

The standardization of the XAI certification according to FDA, EMA, and ISO/IEC standards is required to eliminate the barriers to deployment. These must contain the standards of interpretability, bias reduction, and the human usability.

These methods, in combination, will allow XAI to develop into a system that is clinically effective, ethically responsible, and legal-ready network of smart healthcare systems.

Conclusion

In healthcare, Explainable Artificial Intelligence (XAI) has become one of the significant pillars of the safe, ethical and efficient application of AI technologies. With the rise of AI into clinical decisions, diagnosis and treatment design, transparency, accountability and humanity-centered design have become essential criteria. XAI gives answers to these concerns, as it provides interpretable explanations of the model behavior, thus encouraging trust, enhancing usability, and making regulatory compliance easier. The presented systematic review has demonstrated the taxonomy of XAI models, techniques, and application areas in the context of healthcare that revealed the widespread application of such methods as SHAP, LIME, Grad-CAM,

and attention mechanisms. It has also discussed the issues that stand in the way of clinical implementation such as the accuracy-interpretability tradeoff, absence of standardized measures of evaluation, heterogeneity of data, and minimal clinician input. Moreover, it described feasible clinical associated integration plans and showed future directions like causal reality, computerized twins, and XAI confirmation systems.

Although the technical development of XAI has been significant, it will still likely require connecting the algorithm development world to the healthcare needs world before it can make any real clinical difference. This will involve multidisciplinary cooperation, design area specific, as well as ethical governance, and regulatory preparedness, and user-centric examination. To sum up, the predictive accuracy is not the only factor defining the future of AI in healthcare, but there is also the transparency, fairness, clinical relevance of explanations. The development of XAI into the clinical decision-making setting at the bedside is not just a technology target but a moral obligation in the age of smart medicine.

References

1. Zhang, Y., Wang, L., & Liu, H. (2023). Data-driven battery management for electric vehicles using deep learning and edge computing: A review. *IEEE Transactions on Industrial Informatics, 19*(2), 1501–1514. https://doi.org/10.1109/TII.2022.3189127
2. Ribeiro, M. T., Singh, S., &Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778
3. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)* (pp. 4765–4774).
4. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 618–626). https://doi.org/10.1109/ICCV.2017.74
5. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... &Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)* (pp. 5998–6008).
6. Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems, 32*(11), 4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314
7. Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2019). What do we need to build explainable AI systems for the medical domain? *Review of Machine Learning in Healthcare, 1*(1), 1–35. https://doi.org/10.1016/j.patrec.2017.04.008
8. Srilakshmi, K., Preethi, K., Afsha, M., Pooja Sree, N., & Venu, M. (2022). Advanced electricity billing system using Arduino Uno. *International Journal of Communication and Computer Technologies, 10*(1), 1–3.
9. Papalou, A. (2023). Proposed information system towards computerized technological application – Recommendation for the acquisition, implementation, and support of a health information system. *International Journal of Communication and Computer Technologies, 8*(2), 1–4.
10. Monisha, S., Monisha, M., Deepa, P., & Sathya, R. (2019). An android application for exhibiting statistical chronicle information. *International Journal of Communication and Computer Technologies, 7*(1), 7–9.
11. Usikalu, M. R., Okafor, E. N. C., Alabi, D., & Ezeh, G. N. (2023). Data Distinguisher Module Implementation Using CMOS Techniques. *Journal of VLSI Circuits and Systems, 5*(1), 49–54. https://doi.org/10.31838/jvcs/05.01.07
12. Kondam, R. R., Sekhar, P. C., & Boya, P. K. (2024). Low power SoC-based road surface crack segmentation using UNet with EfficientNet-B0 architecture. *Journal of VLSI Circuits and Systems, 6*(1), 61–69. https://doi.org/10.31838/jvcs/06.01.11