Int. J. Integr. Stud. 2026; 2(7): 47-59

ISSN: 3049-3277

International Journal of Integrative Studies (IJIS)

Journal homepage: www.ijis.co.in

RESEARCH ARTICLE

Explainable Artificial Intelligence in Healthcare: Progress, Challenges, and Future Directions

Dr. S. Rasheed Mansoor Ali[1*], Dr .A. Meher Nisha[2], K Anwar Sathik[3], Dr. K. Syed Kousar Niasi[4]

[1]Assistant Professor, Department Of Computer Science, Jamal Mohamed College (Autonomous), Affiliated To Bharathidasan University, Tiruchirappalli, Tamil Nadu, India. srasheedmansoor@gmail.com
[2]Assistant Professor, Department Of Biotechnology, Jamal Mohamed College (Autonomous), Affiliated To Bharathidasan University, Tiruchirappalli, Tamil Nadu, India. mehernisha@gmail.com
[3]Assistant Professor, Department Of Computer Science, Jamal Mohamed College (Autonomous), Affiliated To Bharathidasan University, Tiruchirappalli, Tamil Nadu, India. anwarsathik1584@gmail.com
[4]Assistant Professor, Department Of Computer Science, Jamal Mohamed College (Autonomous), Affiliated To Bharathidasan University, Tiruchirappalli, Tamil Nadu, India. skn@jmc.edu

ABSTRACT

This study aims to illustrate the significance of Abstract Science Intelligence (XAI) in the medical field, outlining its role in enhancing the interpretability of traditional "black-box" AI models.The goal of this research is to shed light on the significance of Abstract Science Intelligence (XAI) in the medical sector, elucidating its function in bolstering the interpretability of conventional "black-box" Artificial Intelligence models. In medical imaging, disease diagnosis, clinical decision support, and precision medicine, deep learning techniques have proven to be extremely successful; however, the inability to explain and understand how deep learning models work raises a variety of challenges for clinician trust, patient safety, ethical accountability, and regulatory compliance. This review provides a comprehensive overview of recent developments in XAI for healthcare, focusing on key methods for achieving interpretability, such as intrinsically interpretable models, and post-hoc explanation methods such as SHAP, LIME, Grad-CAM, attention mechanisms, surrogate models, and counterfactual explanations. A detailed review of the use of these methods in a variety of clinical areas such as radiology, oncology, cardiology, genomics, electronic health records and drug discovery is also given. Furthermore, the paper examines technical issues concerning explanation fidelity, computational complexity, robustness, scalability, and model validation, as well as ethical issues such as fairness, transparency, privacy, bias, and governance. Other research trends are also discussed, such as causal explainability, human-centered XAI, models that are uncertain, human-in-the-loop systems, and evaluation frameworks. In summary, the review highlights that explainability is not just a technical aspect but a key component in creating AI systems that are both trustworthy and clinically sound and can assist in safe and effective health care decision-making.

Keywords: *Explainable Artificial Intelligence (XAI); Healthcare AI; Interpretability; Clinical Decision Support; Medical Imaging; Machine Learning; Deep Learning; SHAP; LIME; Grad-CAM; Transparency; Trustworthy AI; Explainability; Ethical AI; Precision Medicine.*

INTRODUCTION

However, the use of AI in the clinical workflow requires interpretability to guarantee patient safety and generate trust in high stakes environments like diagnostic decisions and settings where opaque “black-box” models are not enough (Chaddad et al., 2023).

To address this technical and ethical need, this review aims to critically discuss the effectiveness of existing AI interpretability methods across various clinical data modalities (Martino & Delmastro, 2022; Nasarian et al., 2024). This article situates itself in a uniquely multidisciplinary space, examining current taxonomies in the context of the new demands for regulatory compliance and practitioners' utility (Amann et al., 2020; Mienye et al., 2024).

[1]Assistant Professor, Department Of Computer Science, Jamal Mohamed College (Autonomous), Affiliated To Bharathidasan University, Tiruchirappalli, Tamil Nadu, India
Email-id: srasheedmansoor@gmail.com

Corresponding Author: Dr. S. Rasheed Mansoor Ali
Assistant Professor, Department Of Computer Science, Jamal Mohamed College (Autonomous), Affiliated To Bharathidasan University, Tiruchirappalli, Tamil Nadu, India
DOI: https://doi.org/10.63856/ijis/v2i7/07

How to cite this article: Ali, S. R. M., (2026). Explainable Artificial Intelligence in Healthcare: Progress, Challenges, and Future Directions. *International Journal of Integrative Studies*, 2(7), 47-59.

Source of support: Nil

Conflict of interest: None.
Received: 25//062026 **Revised:** 30/06/2026 **Accepted:** 15/07/2026
Published: 28/07/2026

This work builds on prior work which mainly discusses algorithmic explainability methods, while this paper highlights the importance of contextual alignment between methods for technical explainability and critical needs for human-centric clinical decision support (Sadeghi et al., 2024; Yang et al., 2023). Specifically, we analyse the changing landscape of explainability with a focus on clinician-patient interaction, and explain the risks associated with automated diagnostic inference and how they can be reduced with model transparency (Aziz et al., 2024; Sun et al., 2024). A systematic literature search was conducted following the PRISMA 2020 guidelines to identify, screen, and select relevant studies on Explainable Artificial Intelligence (XAI) in healthcare. The detailed study selection process is presented in

Figure: 1.

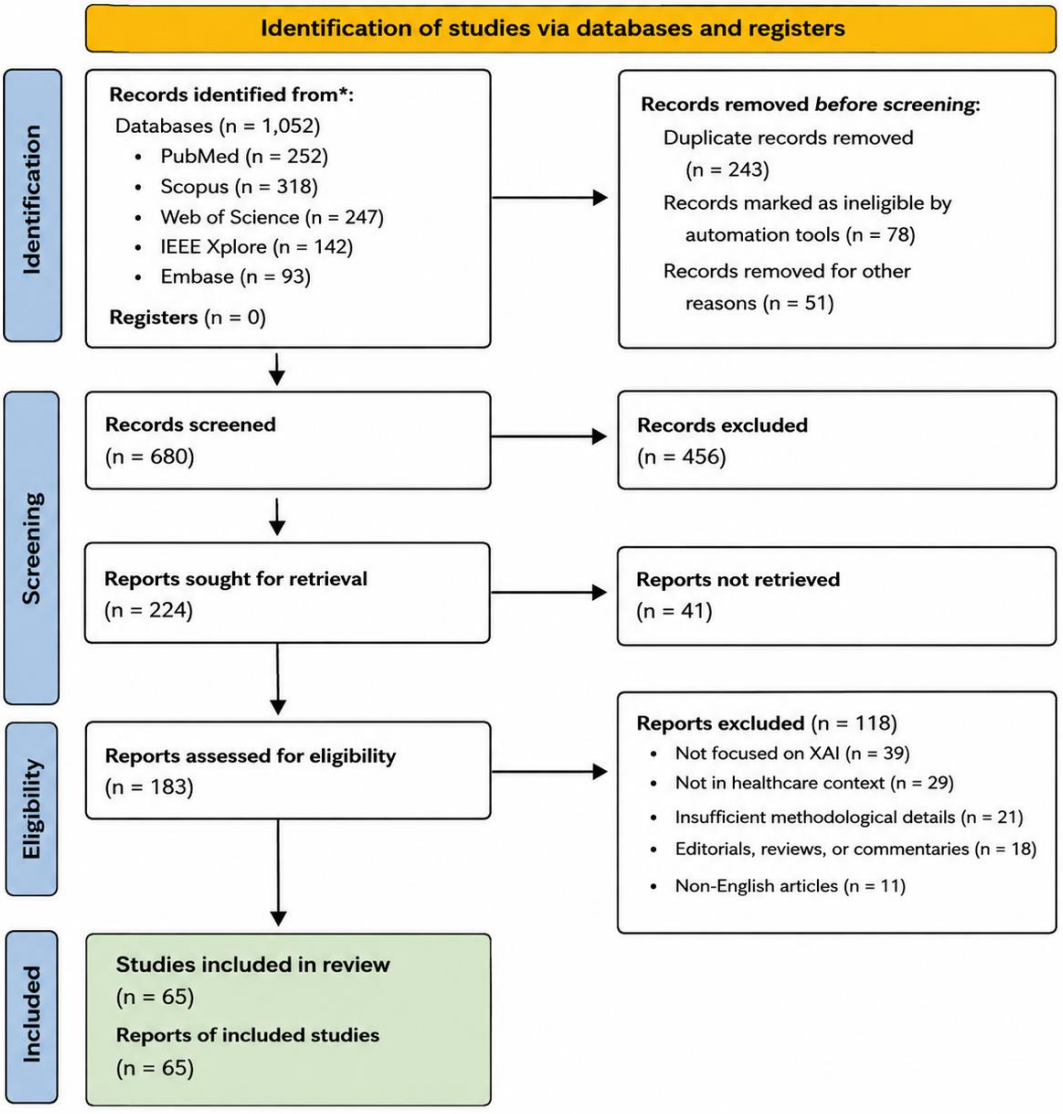

Figure 1. PRISMA 2020 flow diagram illustrating the identification, screening, eligibility assessment, and inclusion of studies reviewed in this article.

Motivation and Significance

In high-stakes healthcare settings, like oncological screening or critical care monitoring, there are often disadvantages to using opaque deep learning models because the resulting diagnostic outputs from these automated systems are often opaque (Hassija et al., 2023; Pal et al., 2026). This inherent lack of interpretability makes it very difficult to guarantee patient safety, since the reason for a treatment suggestion based on models might not be understandable to the clinician (Markus et al., 2020). Thus, a crucial step is to create a connection between AI-generated technical output and clinical reasoning that is easy for the clinician to follow, to build trust in the institution and ensure interventions are clinically sound and accountable (Ennab & Mcheick, 2024; Hossain et al., 2023). Moreover, it is essential to develop such transparency in order to get beyond mere predictive performance metrics to provide strong, interoperable frameworks that meet regulatory requirements and functional needs on the bedside of the clinician (Sadeghi et al., 2023; Shiddik, 2026). This paper's approach differs from prior research that focuses on algorithmic

performance benchmarks and instead proposes to assess the explainability of an algorithm as an ethical obligation to reduce the impact of diagnostic error (Din et al., 2026; Graziani et al., 2022).

Review Objectives

The goal of this review is to classify current XAI approaches and establish a unified assessment system focusing on clinical utility and the interpretability of the approach by the stakeholders involved (Zhang et al., 2026). These are the reasons that, by combining these different approaches (from surrogate modeling to attention models), we offer a systematic diagnostic analysis of the impact of technical transparency on the adoption of diagnostics in the real world (Bharati et al., 2023; Gambetti et al., 2025). Moreover, this synthesis provides a clear understanding of the enduring problem of achieving prediction accuracy and interpretability in models (Frasca et al., 2024; John et al., 2025). Last but not least, this work summarises the potential research paths to be explored to establish generalised validation procedures that guarantee the fidelity of the models and introduce methods to build system-biased and algorithmically accountable environments for clinical applications (Adeniran et al., 2024; Chaddad et al., 2023).

This course introduces the key principles and challenges of Explainable AI.

To develop a common analytical tool, the concepts of interpretability (the level at which the internal functioning of a model is understood by the observer) and explainability (the ability to postpone give human-readable explanations of the reasons that support the specific predictions generated by an algorithm) should be separated (Scarpato et al., 2024; Tjoa & Guan, 2020). Explainability in machine learning can be thought of as a set of post hoc methods used to explain complex and high-performing black-box models; models that are inherently transparent due to their simplicity are more likely to be interpretable (Salih et al., 2024; Tonekaboni et al., 2019). In addition to these basic distinctions are terms like "comprehensibility" and "mental fit" that refer to the effectiveness of the clinician's translation of model logic to known physiological or pathological processes (Stiglic et al., 2020). Additionally, these definitions have not been agreed upon by developers and clinicians, which has been a major challenge to have a common set of expectations in routine clinical practice (Ossa et al., 2022).

Tracing the history of XAI Methods

The classification of XAI techniques can be done on two main grounds: transparency in the model architecture (intrinsically interpretable models) and transparency through external explanations (post-hoc methods) (Thakur et al., 2025; Verhoeven et al., 2025). As part of this, model-specific techniques take advantage of the intrinsic characteristics of the model (e.g., the decision nodes in a shallow tree or coefficients in a linear model) to make it transparent, while model-agnostic approaches separate the explanation task from the underlying model (Palama et al., 2026). Feature attribution techniques, such as SHAP for measuring variable importance, surrogate modelling for modelling complex logic, and example-based methods for finding influential training examples, are some of the common methods used (Fahim et al., 2025). Unlike interpreting the decision logic within complex models like intrinsically interpretable architectures such as linear regression or decision trees (Amann et al., 2022), post-hoc methods aim to uncover the decision logic of a complex model without needing to access the internal parameters of the model (Mersha et al., 2024).

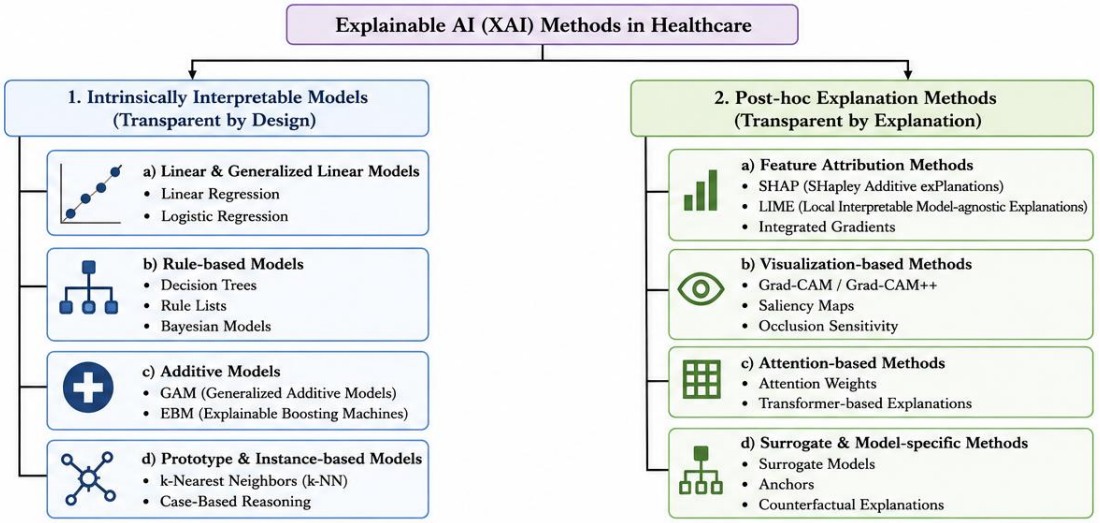

Figure 2. Taxonomy of Explainable Artificial Intelligence (XAI) Methods in Healthcare

Source: Adapted from Graziani et al. (2023), *A Global Taxonomy of Interpretable AI: Unifying the Terminology for the Technical and Social Sciences*, *Artificial Intelligence Review*, 56, 3473–3504. https://doi.org/10.1007/s10462-022-10256-8. Licensed under **CC BY 4.0**.

Interpretability versus Performance

The traditional view is that transparency in a model is likely to come at the cost of loss of prediction accuracy (Nasarian et al., 2023). Recent progress in high-performing models with inherent interpretability, however, indicates that this is not an absolute requirement – specialized architectures can sometimes accommodate both model precision and interpretability (Ramachandram et al., 2025). In high-stakes clinical settings, however, systems may be designed with a compromise in place that will be a "grey-box," attempting to achieve the balance between advanced predictive performance and enough transparency for diagnostic confirmation (Nasarian et al., 2023, 2024). Safeguarding areas where a small error can result in a high degree of morbidity or mortality, stakeholders may justifiably put emphasis on model interpretability over improvements in prediction accuracy alone (Jain, 2024).

Current Healthcare Applications

Explainability has become a key success in the field of diagnostic imaging and predictive risk modelling, where doctors need to gain insights into any anomalies at the pixel level or risk factors specific to the patient (Ali et al., 2023; Kinde & Xu, 2026). In radiology, for example, gradient-weighted Class Activation Mapping is widely used to select the specific characteristics of the medical image that are key to the diagnosis and subsequently used to classify the image automatically (Band et al., 2023). In the same vein, in the field of oncology, post-hoc approaches like SHAP and LIME are used to clarify the logic used in determining therapeutic recommendations, as they determine the relative importance of specific genomic markers and patient parameters across time (Özdemir & Fatunmbi, 2024). Moreover, in cardiology, integrated surrogate models enable clinicians to gain understandable explanations for arrhythmia detection, which helps enhance the understanding between signal classification and clinical decision-making (Sewada et al., 2023; Shafik et al., 2024). At the same time, feature importance techniques have been widely employed in the medical field of neurology to explain neural network predictions for a variety of neurodegenerative diseases, highlighting their applicability across the medical field (Noor et al., 2025). Besides diagnostic applications, XAI can be applied in antimicrobial multidrug resistance forecasting and personalized nutrition recommendation systems, where communicative interfaces aid in understanding complex algorithmic results by helping clinicians and patients interpret them (Chen et al., 2024; Haque et al., 2022).

As summarized in **Table 1**, each explainable AI approach offers unique advantages and limitations, and the selection of an appropriate method depends on the clinical application, data characteristics, and required level of interpretability.

Table 1. Comparative Overview of Major Explainable AI Methods in Healthcare

XAI Method	Explanation Type	Common Healthcare Applications	Major Strength	Key Limitation
SHAP	Feature attribution	Clinical prediction, EHR analysis	Accurate feature importance	Computationally expensive
LIME	Local explanation	Disease diagnosis	Model-agnostic and easy to interpret	Limited global consistency
Grad-CAM	Visual explanation	Radiology, pathology	Highlights diagnostic image regions	Mainly applicable to CNNs
Attention Mechanisms	Model-based explanation	Medical imaging, NLP	Improves interpretability during prediction	Attention weights may not fully explain decisions
Decision Trees	Intrinsically interpretable	Risk prediction, diagnosis	Highly transparent	Lower predictive performance for complex data
Counterfactual Explanations	Example-based	Clinical decision support	Supports "what-if" clinical reasoning	May generate unrealistic scenarios

Medical imaging and Diagnostics

Saliency maps and attention mechanisms have become key elements in this field for tasks like localization of pathological features in mammography, such as the detection of malignant lesions, or in thoracic radiography, particularly for identifying interstitial opacities (Abbas et al., 2025; Hsieh et al., 2024). These methods allow to image specific regions of the image, which contribute to the algorithmic inference, thus affording radiologists the opportunity to compare the algorithmic results to clinical markers to validate the model (Khadse & Puri, 2026). Furthermore, attention mechanisms have been found to be useful in identifying critical spatial patterns in lesions, as they dynamically adjust their weighting of relevant information based on the spatial location of the lesion in the image, thereby matching deep learning predictions to clinically relevant spatial information (Pappala et al., 2025), (Phan et al., 2026). Such methods are also instrumental in improving the decision support as they present information about the outcome of image-based diagnosis at a regional level, which can be compared with information about the ground truth of image acquisition in

the form of tissue histopathology (Mandala, 2023).

Clinical Decision Support

In this space, XAI is increasingly being used to give explanations and justifications, such as alerts, for electronic health record (EHR) systems, helping providers to better understand why automated coding and clinical documentation suggestions are generated. In addition, risk prediction models are now being improved with transparent visualization of the contribution of the different features in the model, enabling the practitioner to pinpoint specific historical health information—including laboratory trends or comorbidity indices—that are associated with the clinical events in the future (Barmak et al., 2024). Last but not least, these explainable outputs are increasingly being embedded in clinician-facing interfaces, ensuring that decision-support metrics are displayed coherently and integrated into decision-making workflows (Babita & Goel, 2026).

Drug discovery and Genomics.

XAI methods are crucial in drug discovery for identifying potential molecular targets and understanding the structural aspects that are crucial for these complex biochemical processes (Kumar et al., 2026). These approaches can help shed light on the trustworthiness of molecular property prediction and aid in the discovery of pharmaceutical candidates, while also providing a measure of uncertainty for particular molecular structures (Askr et al., 2022). Additionally, in the context of omics data analysis, interpretable models can help to uncover biological signatures (Dey et al., 2022) that are useful for discovering biomarkers from high dimensionally expressed gene expression data. These pipelines facilitate the identification of known genes associated with specific disease outcomes, further enriching precision medicine strategies and enabling physicians to make decisions regarding a patient's specific molecular profile (Qadri et al., 2025; Sharma et al., 2025).

Involving technical and practical challenges

Although the methodologies hold promise, there are still a number of technical issues that have yet to be addressed, especially when it comes to deriving comprehensible explanations from existing XAI frameworks from very complex model architecture (Yang et al., 2023). Moreover, current approaches tend to perform poorly when applied to different patient groups, resulting in poor performance in cases where data distributions differ from the training data distribution (Alharthi et al., 2024). Moreover, descriptors of high dimensional clinical data are often unable to keep up with the complexity of these data, making clinical validation difficult (Grover & Dogra, 2024). Furthermore, some popular post-hoc interpretability methods require substantial computational complexity hindering their application in clinical settings where time-sensitive inferences are needed, such as in real-time clinical scenarios (Talukder et al., 2025).

The Explanation Quality is evaluated by this

The current assessment frameworks are greatly limited by the absence of standardized measures to capture the fidelity of explanations, making it difficult to compare the ones reported in the literature (Hu et al., 2024; Vasanthageethan, 2025). In addition, there is no consensus on which type of evaluation paradigm – human-grounded or functional – is more useful in the field; often technical proxy indicators are used instead of genuine clinical understanding (Payrovnaziri et al., 2020). The fact that many of these studies have been conducted at a single centre and that there has been no good external validation of these studies makes it difficult to translate the results of these evaluations across contexts, with high reproducibility problems in the published assessments of XAI (Cascarano et al., 2023). The "false hope" is that current techniques by themselves instill clinical trust in the model, which in turn can lead to "misconceptions" that are seemingly correct but not actually true based on the underlying domain knowledge that is still incomplete (Ghassemi et al., 2021; Rosenbacke et al., 2024).

Scalability and Robustness

Generating local explanations, e.g., SHAP or LIME, can be quite heavy in computational cost, especially for large scale, high dimensional clinical data sets (Ansari et al., 2026; Watson et al., 2019). Moreover, these approaches often lack robustness to attack and data distribution changes, bringing to the fore some serious concerns about the secure use of these approaches in high-stakes settings (Khaled et al., 2025; Martino & Delmastro, 2022). Furthermore, instability in these explanation algorithms increases fragility, with explanations produced by different algorithms often conflicting (e.g., different feature ranks in LIME and SHAP).In addition, frequent disagreement between explanation algorithms leads to undesirable opacity in the clinical workflow (Viswan et al., 2023). In addition to these technical challenges, the issue of closed system architecture can be a major obstacle because it can require significant back-end development and regulatory compliance to achieve interoperability (Alsentzer et al., 2025; RAFIQ et al., 2025).

Ethical and Regulatory Considerations are explored.Ethical and Regulatory Considerations are discussed.

The need for explainability has fundamental ethical grounds and is necessary to maintain patient autonomy and informed consent, since opaque “black box” systems directly hinder the patient's ability to understand the clinical rationales underlying their treatment (Bouderhem, 2024; González-Gonzalo et al., 2021). Moreover, such ambiguity makes it difficult to see the full extent to which clinical outcomes might be influenced by underlying biases, undermining the principle of medical equity and fairness (Chinta et al., 2024). The use of explainability

mandates, therefore, requires consideration of the changing governance landscape of digital health, where more and more the digital artefacts of algorithms need to be examined and understood to avoid institutionalization of systematic bias (Ferlito et al., 2024).

All three are trust, bias and fairness.

Although many researchers have pointed out the connection between XAI and algorithmic fairness, it is important to be aware of the multidimensionality of fairness and of the fact that many such connections are misleading and do not tackle the multidimensionality of fairness (Ziethmann et al., 2025). In fact, it is possible that explanations can make it sound like they provide a transparent reason, when in reality, they are just a way to cover up algorithmic biases which underlie the predictions they make (Ueda et al., 2023). To reduce this: fairness auditing will need to shift towards using the interpretable components of the model, which naturally represent causal clinical relationships, rather than relying on rationalisations based on post hoc reasoning, without any underlying mechanistic basis (Kostick & Gerke, 2022; Siebra et al., 2024).

The Policy And Governance Frameworks In Place Are Adequate

The regulatory framework today is highly fractured and makes it difficult to set a common standard for AI explainability for different jurisdictions and cultures (Rane et al., 2024). In particular, the lack of accountability for algorithmic errors presents a challenge to determining liability, as the difference between "after-the-fact" explanations and real-world phenomena leads to a need for well-defined mechanisms of recourse for errors that occur with these tools (Nannini et al., 2024; Raz et al., 2025). In addition, the rise of ‘governance-by-design' requires a continuous accountability process through post-market surveillance, enhanced by transparency requirements (Bailo et al., 2026). In this context, informed consent needs to be rethought and clearly outlined to consider the limitations of algorithmic reasoning to make sure patients understand how much automated rationales may impact their individualized treatment plan (Sabet et al., 2025).

Future Directions

Research needs to further progress the field by transitioning from an isolated and fragmented approach to developing standardised, multi-disciplinary assurance frameworks that incorporate ethical, legal and clinical aspects throughout the entire model lifecycle (Harris et al., 2022; Sharma et al., 2024). However, future inquiry needs to explicitly consider the creation of comprehensive, interdisciplinary frameworks which specify standardized measures of the effectiveness of explanations, and thus close the current gap between the performance of technical models and the needs of various stakeholders (Jung et al., 2023). To ensure verifiable causality, and long-term algorithmic accountability, there is need to pay attention to architectures that are inherently interpretable rather than relying on later fitting a model to the data. (Solanki et al., 2022)

Emerging Methodologies

Research needs to focus on causal and counterfactual structure to [move beyond] static "what-would-have-been" post-hoc approximations that are useful for advancing clinical interpretability and enable practitioners to test "what if" scenarios which capture the underlying mechanisms of medical decision-making (Holzinger, 2020). At the same time, integrating formal clinical logic directly into neural networks is a promising direction for embedding medical knowledge in the network, ensuring that prediction results are constrained by verifiable medical knowledge (Bruckert et al., 2020). Moreover, uncertainty-aware foundation models can also help clinicians get calibrated confidence intervals for the generated explanations to better assess the reliability of the model in cases where there is uncertainty. (Mirbabaie et al., 2021; Raees et al., 2024) To ensure that these methodological advances can be practically implemented in clinical workflows without creating extra cognitive burden, it is vital to complement them with empirical studies in real clinical settings.Complementary to these methodological advances, it is essential to conduct empirical studies in real clinical settings to make sure the tools can be effectively integrated into the current clinical workflow without adding extra cognitive burden (Neveditsin et al., 2024).

The study will take place in two phases: Integration and Translation Pathways

To achieve success those who translate must shift to a human-in-the-loop approach for assessing the explanation, and designs for explanation interfaces must be based on the user's cognitive needs and clinical workflows. Developers can create interfaces that generate explanation-relevant outputs that are both relevant to the specific diagnostic task and actionable for the clinician by co-designing with them to ensure that explainable outputs are relevant to the current situation and don't present unnecessary or unhelpful data (Cutillo et al., 2020). Furthermore, prospective validation strategies should be undertaken and clinical practices should be monitored over time to maintain explanations from environmental drift effects and changing clinical practices (Ebadi & Wong, 2026). Lastly, developing cooperative research projects involving both causal AI technologies and symbolic approaches will help connect clinicians to algorithmic output and clinical evidence, thus overcoming the current disconnect (Kelly et al., 2019; Martino & Delmastro, 2022). Future studies should focus on creating standardized evaluation tools that provide quantifiable measures of the clinical utility and practical applicability of such explanations in various medical settings, to ensure stability and trust in the

explanations over time (Aziz et al., 2024).

Concluding Remarks

The maturation of XAI will depend on making the systems both technically interpretable and clinically plausible, to ensure they can be a trusted partner in patient care, not a black box (Badawy, 2023; Bienefeld et al., 2023). Transparency is an essential aspect of trustworthy clinical AI, laying the groundwork for ethical and responsible systems that go beyond performance metrics to demonstrate accountability. Overall, building a foundation of explainability is a key component in creating trustworthy clinical AI systems, ensuring they are more ethically sound and accountable beyond just performance metrics. (Babita & Goel, 2026; Pal et al., 2026). When this is achieved, stakeholders will be able to effectively transition from experimental performances to a sustainable, human-centric, and healthcare-integrated approach (Mienye et al., 2024; Zając et al., 2024). Addressing this gap necessitates a comprehensive and multi-layered approach that combines user-centered design with robust quantitative assessments, leading to the creation of XAI as an integral and trustworthy part of contemporary medical practice (Ahamed & Jabez, 2025; Markus et al., 2020).

References:

1. Abbas, Q., Jeong, W.-M., & Lee, S. W. (2025). Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges. *Healthcare*, *13*(17), 2154–2154. https://doi.org/10.3390/healthcare13172154
2. Adeniran, A. A., Onebunne, A. P., & William, P. (2024). Explainable AI (XAI) in healthcare: Enhancing trust and transparency in critical decision-making. *World Journal of Advanced Research and Reviews*, *23*(3), 2447–2658. https://doi.org/10.30574/wjarr.2024.23.3.2936
3. Ahamed, S. M., & Jabez, J. (2025). *Bridging the Gap* (pp. 349–374). https://doi.org/10.1002/9781394249312.ch16
4. Alharthi, A., Alqurashi, A. A., Alharbi, T. A., Alammar, M. M., Aldosari, N., Bouchekara, H. R. E. H., Sha'aban, Y., Shahriar, M. S., & Ayidh, A. A. (2024). The Role of Explainable AI in Revolutionizing Human Health Monitoring: A Review. In *arXiv (Cornell University)*. Cornell University. https://doi.org/10.48550/arxiv.2409.07347
5. Ali, S., Akhlaq, F., Imran, A. S., Kastrati, Z., Daudpota, S. M., & Moosa, M. (2023). The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. *DiVA (Linnaeus University)*, *166*, 107555–107555. https://doi.org/10.1016/j.compbiomed.2023.107555
6. Alsentzer, E., Charpignon, M., Chen, B. B., Dsouza, N., Fries, J., Jiang, Y., Kashyap, A., Kim, C., Lee, S., Mandyam, A., Mbilinyi, A. C., Mehandru, N., Nagesh, N., Nuwagira, B., Pierson, E., Pillai, A. S., Sano, A., Syeda-Mahmood, T., Yadav, S., … Thakoor, K. A. (2025). Reflections from Research Roundtables at the Conference on Health, Inference, and Learning (CHIL) 2025. In *ArXiv.org*. https://doi.org/10.48550/arxiv.2510.15217
7. Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*(1), 310–310. https://doi.org/10.1186/s12911-020-01332-6
8. Amann, J., Vetter, D., Blomberg, S. N. F., Christensen, H. C., Coffee, M., Gerke, S., Gilbert, T. K., Hagendorff, T., Holm, S., Livne, M., Spezzatti, A., Strümke, I., Zicari, R. V., Madai, V. I., & initiative, on behalf of the Z.-I. (2022). To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems. *PLOS Digital Health*, *1*(2). https://doi.org/10.1371/journal.pdig.0000016
9. Ansari, Z. A., Kumar, K. K., Fatima, S. S., Siddiqui, S., & Mohsin, S. W. (2026). Explainable artificial intelligence for cross domain evaluation of predictive models in multi-disease diagnosis. *Discover Computing*, *29*(1). https://doi.org/10.1007/s10791-026-10061-9
10. Askr, H., Elgeldawi, E., Ella, H. A., Elshaier, Y. A. M. M., Gomaa, M. M., & Hassanien, A. E. (2022). Deep learning in drug discovery: an integrative review and future challenges. *Artificial Intelligence Review*, *56*(7), 5975–6037. https://doi.org/10.1007/s10462-022-10306-1
11. Aziz, N. A., Manzoor, A., Qureshi, M. D. M., Qureshi, M. D. M., Qureshi, M. A., Qureshi, M. A., & Rashwan, W. (2024). Unveiling Explainable AI in Healthcare: Current Trends, Challenges, and Future Directions. In *medRxiv*. https://doi.org/10.1101/2024.08.10.24311735
12. Babita, Ms., & Goel, Dr. B. M. (2026a). Integrating Explainable Artificial Intelligence In Healthcare: Models, Applications, And Challenges. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.19218726
13. Babita, Ms., & Goel, Dr. B. M. (2026b). Integrating Explainable Artificial Intelligence In Healthcare: Models, Applications, And Challenges. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.19218727

14. Badawy, M. (2023). Integrating Artificial Intelligence and Big Data into Smart Healthcare Systems: A Comprehensive Review of Current Practices and Future Directions. *Artificial Intelligence Evolution*, 133–153. https://doi.org/10.37256/aie.4220232980
15. Bailo, P., Nittari, G., Pesel, G., Basello, E., Spasari, T., & Ricci, G. (2026). Governing Healthcare AI in the Real World: How Fairness, Transparency, and Human Oversight Can Coexist: A Narrative Review. *Sci*, *8*(2), 36–36. https://doi.org/10.3390/sci8020036
16. Band, S. S., Yarahmadi, A., Hsu, C.-C., Biyari, M., Sookhak, M., Ameri, R., Dehzangi, A., Chronopoulos, A. T., & Liang, H. (2023). Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods. *Informatics in Medicine Unlocked*, *40*, 101286–101286. https://doi.org/10.1016/j.imu.2023.101286
17. Barmak, O., Крак, Ю., Yakovlev, S., Manziuk, E., Radiuk, P., & Kuznetsov, V. (2024). Toward explainable deep learning in healthcare through transition matrix and user-friendly features. *Frontiers in Artificial Intelligence*, *7*, 1482141–1482141. https://doi.org/10.3389/frai.2024.1482141
18. Bharati, S., Mondal, M. R. H., & Podder, P. (2023). A Review on Explainable Artificial Intelligence for Healthcare: Why, How, and When? *arXiv (Cornell University)*, *5*(4), 1429–1442. https://doi.org/10.1109/tai.2023.3266418
19. Bienefeld, N., Boss, J. M., Lüthy, R., Brodbeck, D., Azzati, J., Blaser, M., Willms, J., & Keller, E. (2023). Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. *Npj Digital Medicine*, *6*(1). https://doi.org/10.1038/s41746-023-00837-4
20. Bouderhem, R. (2024). Shaping the future of AI in healthcare through ethics and governance. *Humanities and Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-024-02894-w
21. Bruckert, S., Finzel, B., & Schmid, U. (2020). The Next Generation of Medical Decision Support: A Roadmap Toward Transparent Expert Companions. *Frontiers in Artificial Intelligence*, *3*, 507973–507973. https://doi.org/10.3389/frai.2020.507973
22. Cascarano, A., Mur-Petit, J., Hernández-González, J., Camacho, M., Eadie, N. de T., Gkontra, P., Chadeau-Hyam, M., Vitrià, J., & Lekadir, K. (2023). Machine and deep learning for longitudinal biomedical data: a review of methods and applications. *Artificial Intelligence Review*, *56*, 1711–1771. https://doi.org/10.1007/s10462-023-10561-w
23. Chaddad, A., Lu, Q., Li, J., Katib, Y., Kateb, R., Tanougast, C., Bouridane, A., & Abdulkadir, A. (2023). Explainable, Domain-Adaptive, and Federated Artificial Intelligence in Medicine. *IEEE/CAA Journal of Automatica Sinica*, *10*(4), 859–876. https://doi.org/10.1109/jas.2023.123123
24. Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of Explainable AI Techniques in Healthcare. *Sensors*, *23*(2), 634–634. https://doi.org/10.3390/s23020634
25. Chen, X., Xie, H., Tao, X., Wang, F. L., Leng, M., & Lei, B. (2024). Artificial intelligence and multimodal data fusion for smart healthcare: topic modeling and bibliometrics. *Artificial Intelligence Review*, *57*(4). https://doi.org/10.1007/s10462-024-10712-7
26. Chinta, S. V., Wang, Z., Avash, P., Zhang, X., Kashif, A., Smith, M. A., Jun, L., & Zhang, W. (2024). AI-Driven Healthcare: A Review on Ensuring Fairness and Mitigating Bias. In *arXiv (Cornell University)*. Cornell University. https://doi.org/10.48550/arxiv.2407.19655
27. Cutillo, C. M., Sharma, K. R., Foschini, L., Kundu, S., Mackintosh, M., Mandl, K. D., Group, M. in H. W. W., Beck, T., Collier, E., Colvis, C. M., Gersing, K., Gordon, V., Jensen, R. E., Shabestari, B., & Southall, N. (2020). Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. *Npj Digital Medicine*, *3*(1), 47–47. https://doi.org/10.1038/s41746-020-0254-2
28. Dey, S. K., Chakraborty, P., Kwon, B. C., Dhurandhar, A., Ghalwash, M., Suárez, F., Ng, K., Sow, D., Varshney, K. R., & Meyer, P. (2022). Human-centered explainability for life sciences, healthcare, and medical informatics. *Patterns*, *3*(5), 100493–100493. https://doi.org/10.1016/j.patter.2022.100493
29. Din, S. U., Kemna, R., Ket, J. C. F., Iqbal, M., Bohoudi, O., Hoogendoorn, M., Beretta, E., & Lisowska, A. (2026). Which explainable AI methods in medical imaging are clinically impactful? A systematic literature review addressing the clinician's perspective. *Frontiers in Artificial Intelligence*, *9*, 1819422–1819422. https://doi.org/10.3389/frai.2026.1819422
30. Ebadi, A., & Wong, C. W. (2026). Artificial intelligence-powered biomedical imaging: Recent achievements and challenges. *Current Opinion in Biomedical Engineering*, *38*, 100650–100650. https://doi.org/10.1016/j.cobme.2026.100650
31. Ennab, M., & Mcheick, H. (2024). Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions. *PubMed Central*, *11*, 1444763–1444763. https://doi.org/10.3389/frobt.2024.1444763

32. Fahim, Y. A., Hasani, I. W., Kabba, S., & Ragab, W. M. (2025). Artificial intelligence in healthcare and medicine: clinical applications, therapeutic advances, and future perspectives. *European Journal of Medical Research*, *30*(1), 848–848. https://doi.org/10.1186/s40001-025-03196-w
33. Ferlito, B., Segers, S., Proost, M. D., & Mertes, H. (2024). Responsibility Gap(s) Due to the Introduction of AI in Healthcare: An Ubuntu-Inspired Approach. *Science and Engineering Ethics*, *30*(4), 34–34. https://doi.org/10.1007/s11948-024-00501-4
34. Frasca, M., Torre, D. L., Pravettoni, G., & Cutica, I. (2024). Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. *Discover Artificial Intelligence*, *4*(1). https://doi.org/10.1007/s44163-024-00114-7
35. Gambetti, A., Han, Q., Shen, H., & Soares, C. (2025). A Survey on Human-Centered Evaluation of Explainable AI Methods in Clinical Decision Support Systems. In *ArXiv.org*. https://doi.org/10.48550/arxiv.2502.09849
36. Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, *3*(11). https://doi.org/10.1016/s2589-7500(21)00208-9
37. González-Gonzalo, C., Thee, E. F., Klaver, C. C. W., Lee, A., Schlingemann, R. O., Tufail, A., Verbraak, F. D., & Sánchez, C. I. (2021). Trustworthy AI: Closing the gap between development and integration of AI systems in ophthalmic practice. *UvA-DARE (University of Amsterdam)*, *90*, 101034–101034. https://doi.org/10.1016/j.preteyeres.2021.101034
38. Graziani, M., Dutkiewicz, L., Calvaresi, D., Amorim, J. P., Yordanova, K., Vered, M., Nair, R., Abreu, P. H., Blanke, T., Pulignano, V., Prior, J. O., Lauwaert, L., Reijers, W., Depeursinge, A., Andrearczyk, V., & Müller, H. (2022). A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences. *Artificial Intelligence Review*, *56*(4), 3473–3504. https://doi.org/10.1007/s10462-022-10256-8
39. Grover, V., & Dogra, M. (2024). Challenges and Limitations of Explainable AI in Healthcare. In *Advances in healthcare information systems and administration book series* (pp. 72–85). IGI Global. https://doi.org/10.4018/979-8-3693-5468-1.ch005
40. Haque, A. B., Islam, A. K. M. N., & Mikalef, P. (2022). Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *arXiv (Cornell University)*, *186*, 122120–122120. https://doi.org/10.1016/j.techfore.2022.122120
41. Harris, S., Bonnici, T., Keen, T., Lilaonitkul, W., White, M., & Swanepoel, N. (2022). Clinical deployment environments: Five pillars of translational machine learning for health. *Frontiers in Digital Health*, *4*, 939292–939292. https://doi.org/10.3389/fdgth.2022.939292
42. Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2023). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, *16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8
43. Holzinger, A. (2020). Explainable AI and Multi-Modal Causability in Medicine. *I-Com*, *19*(3), 171–179. https://doi.org/10.1515/icom-2020-0024
44. Hossain, Md. I., Zamzmi, G., Mouton, P. R., Salekin, M. S., Sun, Y., & Goldgof, D. B. (2023). Explainable AI for Medical Data: Current Methods, Limitations, and Future Directions. *ACM Computing Surveys*, *57*(6), 1–46. https://doi.org/10.1145/3637487
45. Hsieh, W. C., Bi, Z., Jiang, C., Liu, J., Peng, B., Zhang, S., Pan, X., Xu, J., Wang, J., Chen, K., Feng, P., Wen, Y., Song, X., Wang, T., Liu, M., Yang, J., Li, X., Jing, B., Ren, J., … Liang, C. X. (2024). A Comprehensive Guide to Explainable AI: From Classical Models to LLMs. In *arXiv (Cornell University)*. Cornell University. https://doi.org/10.48550/arxiv.2412.00800
46. Hu, J. F., Li, Y., Jameel, S., Long, Y., & Papanastasiou, G. (2024). From explainable to interpretable deep learning for natural language processing in healthcare: How far from reality? In *Computational and Structural Biotechnology Journal* (Vol. 24, pp. 362–373). Elsevier BV. https://doi.org/10.1016/j.csbj.2024.05.004
47. Jain, R. (2024). Transparency in AI Decision Making: A Survey of Explainable AI Methods and Applications. *Advances in Robotic Technology.*, *2*(1), 1–10. https://doi.org/10.23880/art-16000110
48. John, W., Ogunbode, A., & Iqbal, M. (2025). *Human-Centric Explainable AI in Healthcare: A Review of Methods, Evaluation Metrics, and Ethical Governance (Preprint)*. https://doi.org/10.2196/preprints.89088
49. Jung, J., Lee, H., Jung, H., & Kim, H. (2023). Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. *Heliyon*, *9*(5). https://doi.org/10.1016/j.heliyon.2023.e16110

Kelly, C., Karthikesalingam, A., Suleyman, M., Corrado, G. S., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*(1), 195–195. https://doi.org/10.1186/s12916-019-1426-2

50. Khadse, C., & Puri, C. (2026). *Explainable AI in Healthcare: Bridging the Gap Between Doctors and Models*. 814–819. https://doi.org/10.1109/icipcn67432.2026.11438566
51. Khaled, A., Sabir, M., Qureshi, R., Caruso, C. M., Guarrasi, V., Xiang, S., & Zhou, S. K. (2025). Leveraging MIMIC Datasets for Better Digital Health: A Review on Open Problems, Progress Highlights, and Future Promises. In *ArXiv.org*. https://doi.org/10.48550/arxiv.2506.12808
52. Kinde, A., & Xu, J. (David). (2026, January 1). Explainable AI: Characteristics, Intended and Unintended Consequences and Mitigation Strategies in Decision-Making in Healthcare. *Proceedings of the ... Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences*. https://doi.org/10.24251/hicss.2026.450
53. Kostick, K. M., & Gerke, S. (2022). AI in the hands of imperfect users. *Npj Digital Medicine*, *5*(1), 197–197. https://doi.org/10.1038/s41746-022-00737-z
54. Kumar, S., Pal, A., & Chatterjee, S. (2026). Unlocking the potential of computational phenotypic drug discovery: methods, challenges, and future directions. *Npj Systems Biology and Applications*, *12*(1). https://doi.org/10.1038/s41540-026-00676-5
55. Mandala, S. K. (2023). XAI Renaissance: Redefining Interpretability in Medical Diagnostic Models. In *arXiv (Cornell University)*. Cornell University. https://doi.org/10.48550/arxiv.2306.01668
56. Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2020). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics*, *113*, 103655–103655. https://doi.org/10.1016/j.jbi.2020.103655
57. Martino, F. D., & Delmastro, F. (2022a). Explainable AI for clinical and remote health applications: a survey on tabular and time series data. In *arXiv (Cornell University)*. Cornell University. https://doi.org/10.48550/arxiv.2209.06528
58. Martino, F. D., & Delmastro, F. (2022b). Explainable AI for clinical and remote health applications: a survey on tabular and time series data. *Artificial Intelligence Review*, *56*(6), 5261–5315. https://doi.org/10.1007/s10462-022-10304-3
59. Mersha, M. A., Lâm, K. N., Wood, J., AlShami, A. K., & Kalita, J. (2024). Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *arXiv (Cornell University)*, *599*, 128111–128111. https://doi.org/10.1016/j.neucom.2024.128111
60. Mienye, I. D., Obaido, G., Jere, N., Mienye, E., Aruleba, K., Emmanuel, I. D., & Ogbuokiri, B. (2024). A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges. *Informatics in Medicine Unlocked*, *51*, 101587–101587. https://doi.org/10.1016/j.imu.2024.101587
61. Mirbabaie, M., Hofeditz, L., Frick, N., & Stieglitz, S. (2021). Artificial intelligence in hospitals: providing a status quo of ethical considerations in academia to guide future research. *AI & Society*, *37*(4), 1361–1382. https://doi.org/10.1007/s00146-021-01239-4
62. Mohammed, B. (2023). A Review on Explainable Artificial Intelligence Methods, Applications, and Challenges. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, *11*(4). https://doi.org/10.52549/ijeei.v11i4.5151
63. Nannini, L., Manerba, M. M., & Beretta, I. (2024). Mapping the landscape of ethical considerations in explainable AI research. *Ethics and Information Technology*, *26*(3). https://doi.org/10.1007/s10676-024-09773-7
64. Nasarian, E., Alizadehsani, R., Acharya, U. R., & Tsui, K. (2023). Designing Interpretable ML System to Enhance Trustworthy AI in Healthcare: A Systematic Review of the Last Decade to A Proposed Robust Framework. *Research Square*. https://doi.org/10.21203/rs.3.rs-3626164/v2
65. Nasarian, E., Alizadehsani, R., Acharya, U. R., & Tsui, K. (2024). Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. *Information Fusion*, *108*, 102412–102412. https://doi.org/10.1016/j.inffus.2024.102412
66. Nasarian, E., Alizadehsani, R., Acharyac, U. R., & Tsui, D. (2023). Designing Interpretable ML System to Enhance Trust in Healthcare: A Systematic Review to Proposed Responsible Clinician-AI-Collaboration Framework. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2311.11055
67. Neveditsin, N., Lingras, P., & Mago, V. (2024). Clinical Insights: A Comprehensive Review of Language Models in Medicine. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2408.11735

68. Noor, A., Manzoor, A., Qureshi, M. D. M., Qureshi, M. D. M., Qureshi, M. A., Qureshi, M. A., & Rashwan, W. (2025). Unveiling Explainable AI in Healthcare: Current Trends, Challenges, and Future Directions. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, *15*(2). https://doi.org/10.1002/widm.70018
69. Ossa, L. A., Starke, G., Lorenzini, G., Vogt, J. E., Shaw, D., & Elger, B. S. (2022). Re-focusing explainability in medicine. *Digital Health*, *8*. https://doi.org/10.1177/20552076221074488
70. Özdemir, Ö., & Fatunmbi, T. O. (2024). Explainable AI (XAI) in Healthcare: Bridging the Gap between Accuracy and Interpretability. *Journal of Science Technology and Engineering Research*, *2*(1), 32–32. https://doi.org/10.64206/0z78ev10
71. Pal, M., Saha, H. N., & Chakrabarti, A. (2026). The Trust-Aware XAI (TAXAI) framework: a quantitative model for interpretable and reliable clinical AI systems. *Scientific Reports*, *16*(1). https://doi.org/10.1038/s41598-026-44167-3
72. Palama, V., Kadiri, C., Babarinde, A. O., Nwanze, J., Adekoya, A. F., & Ejuone, O. (2026). Auditing and Monitoring Artificial Intelligence Systems in Healthcare: A Multilayer Framework for Bias Detection, Explainability, and Regulatory Compliance. *Cureus*, *18*(3). https://doi.org/10.7759/cureus.104547
73. Pappala, S., Malyadri, M., Naganjaneyulu, K. V., Prakashini, A., & Santosh, P. (2025). Explainable AI for healthcare professionals: Advancing risk assessment, diagnostic precision, and ethical clinical interventions. *World Journal of Biology Pharmacy and Health Sciences*, *22*(1), 446–453. https://doi.org/10.30574/wjbphs.2025.22.1.0425
74. Payrovnaziri, S. N., Chen, Z., Rengifo-Moreno, P., Miller, T., Bian, J., Chen, J. H., Liu, X., & He, Z. (2020). Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. *Journal of the American Medical Informatics Association*, *27*(7), 1173–1185. https://doi.org/10.1093/jamia/ocaa053
75. Phan, H. T., Pham, H. H., Nguyen, D. T., Nguyen, T. T., Le, H. N., Thai, D. Q., Tran, T. M., & Quan, T. (2026). Explainable multimodal knee osteoarthritis diagnosis from X-ray images using anomaly detection with clinical and structural biomarkers. *Scientific Reports*. https://doi.org/10.1038/s41598-026-51211-9
76. Qadri, Y. A., Shaikh, S., Ahmad, K., Choi, I., Kim, S. W., & Vasilakos, A. V. (2025). Explainable Artificial Intelligence: A Perspective on Drug Discovery. *Pharmaceutics*, *17*(9), 1119–1119. https://doi.org/10.3390/pharmaceutics17091119
77. Raees, M., Meijerink, I., Lykourentzou, I., Khan, V.-J., & Papagelis, K. (2024). From explainable to interactive AI: A literature review on current trends in human-AI interaction. *International Journal of Human-Computer Studies*, *189*, 103301–103301. https://doi.org/10.1016/j.ijhcs.2024.103301
78. RAFIQ, I., Jarir, Z., & Asri, H. (2025). A Comparative Review of AI, IoT, and Big Data in Healthcare: Towards a Data-Centric Approach for Enhanced Data Quality and Contextual Adaptability. *International Journal of Advanced Computer Science and Applications*, *16*(12). https://doi.org/10.14569/ijacsa.2025.0161245
79. Ramachandram, D., Joshi, H., Zhu, J. D., Gandhi, D., Hartman, L., & Raval, A. (2025). Transparent AI: The Case for Interpretability and Explainability. In *ArXiv.org*. https://doi.org/10.48550/arxiv.2507.23535
80. Rane, J., Kaya, Ö., Mallick, S. K., & Rane, N. L. (2024). *Enhancing black-box models: Advances in explainable artificial intelligence for ethical decision-making*. https://doi.org/10.70593/978-81-981271-0-5_4
81. Raz, A. E., Inbar, Y., Avnoon, N., Lifshitz-Milwidsky, L. B., Hofman, B., Honig, A. M., & Paz, Z. (2025). Who moved my scan? Early adopter experiences with pre- and post-market healthcare AI regulation challenges. *Frontiers in Medicine*, *12*, 1651934–1651934. https://doi.org/10.3389/fmed.2025.1651934
82. Rosenbacke, R., Melhus, Å., McKee, M., & Stückler, D. (2024). How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Health Care: Systematic Review. *JMIR AI*, *3*. https://doi.org/10.2196/53207
83. Sabet, C., Tamirisa, K., Bitterman, D. S., & Celi, L. A. (2025). Regulating medical AI before midnight strikes: Addressing bias, data fidelity, and implementation challenges. *PLOS Digital Health*, *4*(8). https://doi.org/10.1371/journal.pdig.0000986
84. Sadeghi, Z., Alizadehsani, R., Çifçi, M. A., Kausar, S., Rehman, R., Mahanta, P., Bora, P. K., Almasri, A., Alkhawaldeh, R. S., Hussain, S., Alataş, B., Shoeibi, A., Moosaei, H., Hladík, M., Nahavandi, S., & Pardalo, P. M. (2023). A Brief Review of Explainable Artificial Intelligence in Healthcare. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4600029
85. Sadeghi, Z., Alizadehsani, R., Çifçi, M. A., Kausar, S., Rehman, R., Mahanta, P., Bora, P. K., Almasri, A., Alkhawaldeh, R. S., Hussain, S., Alataş, B., Shoeibi, A., Moosaei, H., Hladík, M., Nahavandi, S.,

& Pardalos, P. M. (2024). A review of Explainable Artificial Intelligence in healthcare. *Computers & Electrical Engineering*, *118*, 109370–109370. https://doi.org/10.1016/j.compeleceng.2024.109370

86. Salih, A., Galazzo, I. B., Gkontra, P., Rauseo, E., Lee, A. M., Lekadir, K., Radeva, P., Petersen, S. E., & Menegaz, G. (2024). A review of evaluation approaches for explainable AI with applications in cardiology. *Artificial Intelligence Review*, *57*(9), 240–240. https://doi.org/10.1007/s10462-024-10852-w
87. Scarpato, N., Nourbakhsh, A., Ferroni, P., Riondino, S., Roselli, M., Fallucchi, F., Barbanti, P., Guadagni, F., & Zanzotto, F. M. (2024). Evaluating Explainable Machine Learning Models for Clinicians. *Cognitive Computation*, *16*(4), 1436–1446. https://doi.org/10.1007/s12559-024-10297-x
88. Sewada, R., Jangid, A., Kumar, P., & Mishra, N. (2023). Explainable Artificial Intelligence (XAI). *Journal of Nonlinear Analysis and Optimization*, *13*(1), 41–47. https://doi.org/10.36893/jnao.2022.v13i02.041-047
89. Shafik, W., Hidayatullah, A. F., Kalinaki, K., Gul, H., Zakari, R. Y., & Tufail, A. (2024). A Systematic Literature Review on Transparencyand Interpretability of AI models in Healthcare:Taxonomies, Tools, Techniques, Datasets, OpenResearch Challenges, and Future Trends. In *Research Square*. https://doi.org/10.21203/rs.3.rs-4419881/v1
90. Sharma, C., Sharma, S., Sharma, K., Sethi, G. K., & Chen, Y.-H. (2024). Exploring explainable AI: a bibliometric analysis. *Discover Applied Sciences*, *6*(11). https://doi.org/10.1007/s42452-024-06324-z
91. Sharma, K., Sethi, G. K., Sharma, S., & Kumar, R. (2025). *Explainable AI in Medical Imaging, Personalized Medicine, and Bias Reduction* (pp. 467–492). https://doi.org/10.1002/9781394249312.ch21
92. Shiddik, Md. A. B. (2026). Explainable Artificial Intelligence in Healthcare: Current Landscape, Challenges, and Future Directions. *Health Science Reports*, *9*(3). https://doi.org/10.1002/hsr2.72172
93. Siebra, C., Kurpicz-Briki, M., & Wac, K. (2024). Transformers in health: a systematic review on architectures for longitudinal data analysis. *Artificial Intelligence Review*, *57*(2). https://doi.org/10.1007/s10462-023-10677-z
94. Solanki, P., Grundy, J., & Hussain, W. (2022). Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. *AI and Ethics*, *3*(1), 223–240. https://doi.org/10.1007/s43681-022-00195-z
95. Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., & Cilar, L. (2020). Interpretability of machine learning-based prediction models in healthcare. *arXiv (Cornell University)*, *10*(5). https://doi.org/10.1002/widm.1379
96. Sun, Q., Akman, A., & Schuller, B. W. (2024). Explainable Artificial Intelligence for Medical Applications: A Review. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2412.01829
97. Talukder, Md. A., Talaat, A. S., Kazi, M., & Khraisat, A. (2025). XAI-HD: an explainable artificial intelligence framework for heart disease detection. *Artificial Intelligence Review*, *58*(12). https://doi.org/10.1007/s10462-025-11385-6
98. Thakur, Y., Dash, A. S., Patel, A., & Sayyad, Prof. A. (2025). A Comprehensive Review of Explainable AI (XAI) Methods in Deep Learning. *International Journal of Scientific Research in Science and Technology*, *12*(5), 333–352. https://doi.org/10.32628/ijsrst25126244
99. Tjoa, E., & Guan, C. (2020). 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
100. Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. In *arXiv (Cornell University)*. Cornell University. https://doi.org/10.48550/arxiv.1905.05134
101. Ueda, D., Kakinuma, T., Fujita, S., Kamagata, K., Fushimi, Y., Ito, R., Matsui, Y., Nozaki, T., Nakaura, T., Fujima, N., Tatsugami, F., Yanagawa, M., Hirata, K., Yamada, A., Tsuboyama, T., Kawamura, M., Fujioka, T., & Naganawa, S. (2023). Fairness of artificial intelligence in healthcare: review and recommendations. *Japanese Journal of Radiology*, *42*(1), 3–15. https://doi.org/10.1007/s11604-023-01474-3
102. Vasanthageethan, S. G. D. A. (2025). *Making Sense of Explainable AI in Healthcare and Exploring Its Current Impact and Future Possibilities*. https://doi.org/10.58445/rars.2143
103. Verhoeven, R., Bouisaghouane, W., & Hulscher, J. B. F. (2025). Explainable AI: Ethical Frameworks, Bias, and the Necessity for Benchmarks. *European Journal of Pediatric Surgery*. https://doi.org/10.1055/a-2702-1843
104. Viswan, V., Shaffi, N., Mahmud, M., Karthikeyan, S., & Hajamohideen, F. (2023). Explainable Artificial Intelligence in Alzheimer's Disease Classification: A

Systematic Review. *Cognitive Computation*, *16*(1), 1–44. https://doi.org/10.1007/s12559-023-10192-x

105. Watson, D., Krutzinna, J., Bruce, I. N., Griffiths, C. E. M., McInnes, I. B., Barnes, M. R., & Floridi, L. (2019). Clinical applications of machine learning algorithms: beyond the black box. *BMJ*, *364*. https://doi.org/10.1136/bmj.l886
106. Yang, W., Wei, Y.-C., Wei, H., Chen, Y., Huang, G., Li, X., Li, R., Yao, N., Wang, X., Gu, X., Amin, M. B., & Kang, B. H. (2023). Survey on Explainable AI: From Approaches, Limitations and Applications Aspects. *Human-Centric Intelligent Systems*, *3*(3), 161–188. https://doi.org/10.1007/s44230-023-00038-y
107. Zając, H. D., Ribeiro, J., Ingala, S., Gentile, S., Wanjohi, R., Gitau, S., Carlsen, J. F., Nielsen, M. B., & Andersen, T. O. (2024). "It depends": Configuring AI to Improve Clinical Usefulness Across Contexts. *arXiv (Cornell University)*, 874–889. https://doi.org/10.1145/3643834.3660707
108. Zhang, K., Wang, D., Lin, F., Xie, J., & Zhou, W. (2026). A comprehensive review of explainable artificial intelligence in healthcare methods, evaluation, and clinical integration. *iScience*, *29*(3), 115026–115026. https://doi.org/10.1016/j.isci.2026.115026
109. Ziethmann, P., Hummel, A., Althammer, A., Schlögl-Flierl, K., Heller, A. R., Brunner, J. O., & Bartenschlager, C. C. (2025). From Embedded Ethics to Explainable AI: Advancing Interdisciplinary Collaboration in Medical AI Development. In *Research Square*. https://doi.org/10.21203/rs.3.rs-7941605/v1