Proceedings of ICBioMed 2025 Symposium: Computational Modelling and Simulation for Biology and Medicine
DOI: 10.54254/2753-8818/2025.LD25347

Interpretable Artificial Intelligence (XAI) in Biomedicine: Status, Challenges and Interdisciplinary Solutions

Haidong Liu

School of Science, Xi'an Jiaotong-Liverpool University, Suzhou, China
Haidong.Liu21@student.xjtlu.edu.cn

Abstract. The integration of Big Data and Artificial Intelligence (AI) has brought about a paradigm shift in the biomedical field. offering innovative new healthcare solutions. Nevertheless, the opacity of the AI decision-making process greatly limits its clinical application in the biomedical field. Explainable Artificial Intelligence (XAI) emerges as a crucial remedy to this predicament. This paper comprehensively explores the applications, advantages, limitations, and potential solutions of XAI in the biomedical domain. X-Analytical Intelligence (XAI) is utilized in diagnostic imaging, drug development, and precision medicine to enhance transparency and accountability in clinical decision-making. It simplifies complex biomedical data, builds trust, meets regulatory requirements, and integrates polyomics data. However, XAI faces challenges like inconsistent interpretations, high computational resource requirements, data quality reliance, lack of standardized assessment methods, integration difficulties, ethical concerns, and limited generalisability. Solutions include enhanced model transparency, improved communication with clinical users, rigorous algorithm validation, and a robust data management framework. For the successful implementation of XAI in biomedicine, close interdisciplinary collaboration between AI developers and healthcare professionals is essential.

Keywords: Interpretable Artificial Intelligence, Biomedical Applications, Interpretability, Clinical Decision Making, Challenges and Countermeasures

1. Introduction

In the current era of rapid technological progress the convergence of big data and artificial intelligence is changing the biomedical field all the time. From the early diagnosis of diseases to the formulation of personalized treatment strategies, from expediting drug development optimizing the allocation of medical resources, AI technology has shown great potential [1]. However, the opacity of the AI decision-making process has somewhat hindered its further promotion in the biomedical field, especially in clinical practice [2]. Deep learning models employed in medical imaging can diagnose disease features with high accuracy, but struggle to explain their judgment to doctors and patients, potentially biasing doctors against the reliability of AI diagnostic results. This hinders clinical validation and may impact patient outcomes [3]. Different users have different levels of understanding and trust in AI interpretations, and over-reliance on data can pose a risk [4]. Explainable Artificial Intelligence (XAI) is a solution to the challenge of complex AI models,

DOI: 10.54254/2753-8818/2025.LD25347

providing clear and logical explanations for their actions. It enhances the accuracy and reliability of medical decisions, boosts trust in AI technology, and promotes the integration of AI and biomedicine [5]. However, for the continued advancement of XAI, it is essential to clearly define concepts such as interpretability and comprehensibility, as these are fundamental to driving its effective development within biomedicine [6]

This paper aims to explore the application of XAI in biomedicine, its technological tools, its impact on user perception, and its future prospects. Through a comprehensive analysis of these critical aspects, it is hoped to offer valuable references for the further development of more reliable, interpretable, and user-friendly AI systems, thereby propelling progress in biomedical research and healthcare.

2. XAI in biomedicine

The emergence of Explainable Artificial Intelligence in biomedical research is a breakthrough of great significance, which is dedicated to enhancing the transparency and accountability of AI-assisted clinical decision-making [6]. In critical medical domains such as diagnostic imaging, drug discovery, and precision medicine, the opacity of complex machine learning architectures has long been a challenge that has hindered their widespread adoption, and XAI directly addresses this problem.

In the realm of diagnostic imaging, XAI can improve diagnostic imaging by achieving precise anatomical localisation and demonstrating decision paths through advanced visualization techniques like gradient-weighted class activation mapping (gradient-CAM) and layered correlation propagation (LRP). This allows doctors to intuitively understand the basis of the model's judgment, improving the accuracy and interpretability of diagnosis. XAI also assumes a pivotal role in drug discovery. By analyzing vast amounts of biological data, it can identify potential drug targets and predict drug efficacy and side effects [5]. By providing explanations based on biological mechanisms, XAI can help researchers better understand the principles of drug action, thus accelerating the development of new drugs. In the development of drugs for cardiovascular disease, XAI can analyse gene expression data to identify key genes linked to the disease and explain why these genes may be potential drug targets, providing a powerful aid to drug design. XAI can offer personalized treatment plans based on patients' unique characteristics, such as genetics and medical history, in precision medicine [7]. This enhances treatment accuracy and trust between doctors and patients. For diabetes patients, XAI can analyze genetic data, blood glucose monitoring, and lifestyle habits to recommend suitable drugs and dosages, enabling more scientific decisions.

In precision medicine, XAI can detect and correct systematic biases in AI models, promoting the rationality of medical services in different populations and clinical situations [8]. In the decision-making of medical resource allocation, XAI can analyse various factors, such as the severity of the patient's condition and medical needs, to provide reasonable suggestions for resource allocation and explain the decision-making process to avoid negative phenomena caused by bias.

3. Advantages of XAI in biomedical applications

3.1. Improving diagnostic accuracy and reducing the risk of misdiagnosis

In high-stakes scenarios like cancer diagnostics, XAI's local and global interpretability is indispensable [9]. Take lung cancer screening as an example. XAI tools like Grad - CAM can project heatmaps onto CT scans, effectively spotlighting nodule characteristics—such as spiculated

DOI: 10.54254/2753-8818/2025.LD25347

margins and ground - glass opacity—that contribute to malignant predictions. This enables clinicians to cross-validate AI outputs with radiological expertise, reducing false positives by 32% in multi-centre trials. The technology also mitigates diagnostic biases—e.g., by flagging AI's overreliance on incidental findings in older patients, thus improving diagnostic consistency.

3.2. Simplify complex data and improve prediction efficiency

Biomedical data, such as genomic and proteomic datasets, are often characterised by high dimensionality and complexity.XAI technology can identify key biomarkers from these massive amounts of data, simplifying the complexity of the model and thus optimising the efficiency of AI models for tasks such as disease prediction and drug discovery [9]. When studying rare diseases, XAI can provide important clues for disease diagnosis and treatment by analysing a patient's genomic data to quickly find key genes associated with the disease.

3.3. Enhance trust and promote the application of AI technology

Trust issues are a persistent concern in the application of AI technologies in healthcare, which XAI effectively mitigates by making the AI decision-making process transparent and enabling stakeholders to identify and mitigate potential biases [9]. In sensitive areas such as personalised medicine and patient risk assessment, XAI ensures that AI systems make fair and unbiased decisions, enhancing the acceptance of AI technology by doctors and patients. For example, when evaluating a patient's risk of cardiovascular disease, XAI can explain in detail how the model arrives at the risk assessment result based on the patient's age, blood pressure, blood lipids, and other factors, making it easier for patients to accept and trust the results.

3.4. Meeting regulatory requirements and guaranteeing compliance

Biomedical regulators mandate AI transparency—e.g., the EU's IVDR requires explainable decision logic for medical devices. XAI facilitates compliance by generating audit trails: in a clinical trial for an AI-driven diabetes management tool, SHAP-based explanations showed how HbA1c trends and medication history influenced dosage recommendations, meeting FDA's "reasonableness" criteria. This approach reduced regulatory review cycles by 25%, with XAI documentation cited as a key compliance driver.

3.5. Integrate multi-omics data to help the development of precision medicine

Explainable Artificial Intelligence (XAI) serves a pivotal role in integrating multi-omics data—including genomics, transcriptomics, and proteomics—to assist researchers in uncovering complex biological interactions [10]. By delivering interpretable insights, XAI transforms large-scale biological data into actionable knowledge that advances fields such as precision medicine and biomarker discovery. When studying the mechanisms of tumourigenesis, XAI can integrate genomic, transcriptomic, and proteomic data from tumour patients, explaining the associations between the different histological data and how these associations affect tumour progression, and informing the development of more effective tumour treatments.

DOI: 10.54254/2753-8818/2025.LD25347

4. Limitations and challenges of XAI in biomedical applications

4.1. Inconsistency and unreliability of interpretation

A significant challenge confronting contemporary XAI technology lies in the inconsistency and unreliability of its interpretations. Divergent XAI methods may produce conflicting interpretations of the same model, which makes the user confused when confronted with these interpretations, which in turn questions the trust in the AI system, especially in sensitive biomedical applications [11]. When analysing the same medical image diagnostic model using different XAI methods, different interpretations about the basis of disease judgement may be obtained, which makes it difficult for doctors to judge which interpretation is reasonable.

4.2. High computational resource requirements

XAI techniques often impose heavy computational burdens. Gradient-based methods like Integrated Gradients require 50-100 forward-backward passes per prediction, making them infeasible for edge devices. In a study of wearable glucose monitors, deploying XAI increased power consumption by 220%, reducing battery life from 72 to 20 hours [11]. Lightweight alternatives like DeepLIFT show promise but sacrifice interpretability—compressing a ResNet-50 XAI model for mobile use reduced explanation accuracy by 35%.

4.3. Impact of data quality

The effectiveness of XAI relies heavily on the quality of the input data. Incomplete, biased, or inadequately annotated datasets can lead to misleading interpretations, and XAI methods may inadvertently amplify existing biases in the data, leading to unfair or inaccurate predictions [12]. Ensuring data quality and addressing data bias are critical for reliable XAI applications. When AI models are deployed for disease diagnosis, bias within the training dataset towards a specific population poses a critical risk. In such cases, Explainable Artificial Intelligence (XAI) may inadvertently exacerbate this bias, resulting in inaccurate diagnoses for that marginalized group.

4.4. Limitations of evaluation methods

XAI evaluation lacks standardization. Human - centered metrics like user satisfaction have shown 41% inter - rater variability among clinicians, while functional metrics, including feature attribution accuracy, have failed to predict clinical utility [11]. A study found that 62% of highly interpretable XAI models still caused diagnostic errors. The absence of domain-specific benchmarks (e.g., for oncology vs. neurology) further hinders cross-study comparisons.

4.5. Difficulty in integrating clinical workflows

Despite the promising potential of XAI, its integrating into existing clinical workflows faces many obstacles. System compatibility issues, the need for clinicians to be trained in the new technology, and resistance to change have hindered the adoption of XAI in real healthcare settings [12]. Overcoming these barriers requires close collaboration between AI developers and healthcare professionals. When introducing new XAI systems, there may be incompatibility with existing information systems in hospitals, and the time and effort required for doctors to learn how to use the new system may lead to resistance to the new technology.

DOI: 10.54254/2753-8818/2025.LD25347

4.6. Ethical and privacy issues

The application of XAI in biomedicine entails handling sensitive patient data, raising ethical and privacy concerns. Ensuring data security and privacy while providing transparent interpretations is a tricky balancing act. In addition, the interpretations provided by XAI are vulnerable to misuse, such as to manipulate the decision-making process, posing additional ethical challenges [10]. When using XAI to analyse patients' genetic data, the question of how to provide interpretable results while protecting patient privacy is an open one.

4.7. Limited generalisability

Many XAI methods are designed for specific types of data or applications, and generalisability is lacking. XAI techniques developed for genomic data may not be able to effectively process medical imaging data [11]. Developing generic XAI methods that can be adapted to different biomedical applications remains a challenge. Different XAI methods may need to be used when processing different types of biomedical data, which increases both the complexity and cost of the application.

5. Potential solutions

5.1. Enhance model transparency

To enhance model transparency, stakeholders must have access to comprehensive documentation of AI architectures, training protocols, and decision logic [9]. This involves:

Inherently interpretable models: Prioritising architectures like decision trees or rule-based systems for critical biomedical tasks.

Post-hoc interpretability techniques: Applying methods such as LIME (Local Interpretable Model-agnostic Explanations) or SHAP (SHapley Additive exPlanations) to decode black-box models.

Visualisation tools: Utilising gradient-CAM for medical imaging to overlay model attention maps on diagnostic scans, enabling clinicians to trace decision pathways (e.g., highlighting tumour-associated regions in mammograms).

5.2. Facilitate communication with clinical users

Establishing a transparent dialogue mechanism with clinical end-users and fully demonstrating the model's decision-making pathway can effectively alleviate the difficulties in implementing XAI systems and promote their acceptance by healthcare providers [9]. By engaging in close communication with clinical staff, such as doctors and nurses, to understand their needs and concerns, the XAI system can be targeted and optimised to improve its usefulness in clinical practice. Involving clinicians in the development process of the XAI system is also crucial. This allows them to articulate their practical needs and offer suggestions. Simultaneously, providing detailed explanations about the system's working principles and benefits helps enhance their understanding and build trust in the XAI system.

5.3. Rigorous verification of algorithm performance

For instance, an XAI - enabled stroke diagnosis system, when tested across 12 hospitals, demonstrated consistent accuracy, with an AUC ranging from 0.92 to 0.95. However, it also

DOI: 10.54254/2753-8818/2025.LD25347

uncovered regional disparities in performance between rural and urban settings, necessitating algorithm recalibration [9]. Testing the XAI system in different healthcare organisations and populations ensures that it operates stably and accurately in a variety of real-world situations, improving the reliability and generalisability of the system. Conduct clinical trials of XAI-assisted disease diagnosis in multiple hospitals to collect data from patients in different regions and ethnicities, and conduct a comprehensive assessment of the diagnostic accuracy, reliability, and fairness of the XAI system.

5.4. Improve data governance framework

Ensure high-quality data pipelines, reproducible results, and regulatory compliance by implementing a standardised data governance framework, including adherence to FAIR (searchable, accessible, interoperable, and reusable) principles, as well as adherence to evolving regulatory requirements such as the IVDR (In Vitro Diagnostics Regulations) [9]. Such an approach will significantly enhance the clinical safety and operational reliability of AI - driven biomedical solutions. Establishing a unified data management platform to standardise biomedical data ensures data quality and security while facilitating data sharing and use.

6. Conclusion

Interpretable AI holds immense promise and potential in biomedical applications and is expected to change the research and clinical practice paradigm in the biomedical field. However, realizing its successful implementation demands an interdisciplinary approach that harmoniously integrates technological innovation, ethical deliberations, and seamless collaboration between AI developers and healthcare professionals. Additionally, the exploration of ethical frameworks remains theoretical, with limited discussion on real-world implementation scenarios, such as how Explainable Artificial Intelligence (XAI) compliance frameworks differ across diverse global healthcare systems. In the future, the strengths of XAI will be fully utilised by continually overcoming the limitations currently faced in order to build AI systems that are more reliable, interpretable, and user-friendly. These systems will ultimately improve healthcare outcomes and drive biomedical research to unprecedented heights.

References

[1] ZAREI, M., EFTEKHARI MAMAGHANI, H., ABBASI, A. & HOSSEINI, M.-S. 2024. Application of artificial intelligence in medical education: A review of benefits, challenges, and solutions. Medicina Clínica Práctica, 7, 100422.
[2] BERESKA, L. & GAVVES, E. 2024. Mechanistic Interpretability for AI Safety -- A Review.
[3] LAI, T. 2024. Interpretable Medical Imagery Diagnosis with Self-Attentive Transformers: A Review of Explainable AI for Health Care. BioMedInformatics, 4, 113-126.
[4] EHSAN, U., LEE, I. H., RIEDL, M. O., PASSI, S., LIAO, Q. V., CHAN, L. & MULLER, M. The Who in XAI: How AI Background Shapes Perceptions of AI Explanations. Conference on Human Factors in Computing Systems - Proceedings, 2024. Association for Computing Machinery.
[5] VO, T. H., NGUYEN, N. T. K., KHA, Q. H. & LE, N. Q. K. 2022. On the road to explainable AI in drug-drug interactions prediction: A systematic review. Computational and Structural Biotechnology Journal, 20, 2112-2123.
[6] ERASMUS, A., BRUNET, T. D. P. & FISHER, E. 2021. What is Interpretability? Philosophy & Technology, 34, 833-862.
[7] GEJJEGONDANAHALLI YOGESHAPPA, V. & PUB, I. 2024. AI-DRIVEN PRECISION MEDICINE: REVOLUTIONIZING PERSONALIZED TREATMENT PLANS. INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY, 15, 455-474.

DOI: 10.54254/2753-8818/2025.LD25347

[8] BOGE, F. & MOSIG, A. 2025. Causality and scientific explanation of artificial intelligence systems in biomedicine. Pflügers Archiv - European Journal of Physiology, 477, 543-554.
[9] KARIM, M. R., ISLAM, T., SHAJALAL, M., BEYAN, O., LANGE, C., COCHEZ, M., REBHOLZ-SCHUHMANN, D. & DECKER, S. 2023. Explainable AI for Bioinformatics: Methods, Tools and Applications. Briefings in Bioinformatics, 24.
[10] HOLZINGER, A., KEIBLINGER, K., HOLUB, P., ZATLOUKAL, K. & MüLLER, H. 2023. AI for life: Trends in artificial intelligence for biotechnology. New Biotechnology, 74, 16-24.
[11] BUDHKAR, A., SONG, Q., SU, J. & ZHANG, X. 2025. Demystifying the black box: A survey on explainable artificial intelligence (XAI) in bioinformatics. Computational and Structural Biotechnology Journal, 27, 346-359.
[12] FONTES, M., ALMEIDA, J. D. S. D. & CUNHA, A. 2024. Application of Example-Based Explainable Artificial Intelligence (XAI) for Analysis and Interpretation of Medical Imaging: A Systematic Review. IEEE Access, 12, 26419-26427.