



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Explainable AI in Healthcare applications

¹K. Santosh Babu, ²Dr. Narendra

¹ research scholar, BBA department, KLEF, Vaddeswaram, Andhra Pradesh, India.

²V. Narendra, Assistant professor, BBA Department, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.

ABSTRACT :

AI algorithms now outperform clinicians on a handful of diagnostic tasks, but most top-rated algorithms are 'black boxes' and cannot be probed for explanation by clinicians, patients and regulators who need, and are increasingly demanding, them. Explainable AI (XAI) is an effort to close this explanatory gap in healthcare and this paper is the first structured narrative review of healthcare XAI so far: (i) classifies XAI techniques (LIME, SHAP, saliency and Grad-CAM, integrated gradients, attention, counterfactuals, and inherently interpretable models), (ii) maps techniques to clinical tasks (medical imaging, risk prediction, intensive care monitoring, genomics, drug discovery), (iii) summarizes criteria to evaluate explanations, (iv) identifies major issues faced in practice (fidelity and faithfulness, shortcut learning, bias, regulation) and (v) provides recommendations for XAI researchers, doctors, healthcare managers and regulators, with special reference to India. The review finds that explanation techniques are most developed for tabular patient data and medical images; unfaithfulness in post-hoc explanations could lead to false reassurance; and clinical validation of explanations with real users is infrequent. As a conclusion, XAI should be viewed as one element in a broader assurance framework that should encompass validation, continuous monitoring and governance.

Key Terms: Explainable AI, Interpretable Machine Learning, Healthcare, Medical Imaging, Clinical Decision Support, SHAP, LIME, Grad-CAM, Trust, India

INTRODUCTION

1.1 Background of the Study

Medicine is likely to be one of the most fertile and sensitive areas for AI use. Deep learning algorithms have been used to identify diabetic retinopathy using retinal fundus images (Gulshan et al., 2016), classify skin lesions at dermatologist levels (Esteva et al., 2017) and forecast clinical decline using electronic health records. Topol (2019) maintains that these types of devices could enhance accuracy, quality and workflow in medicine. But medical decisions influence human lives, and a prediction without an explanation cannot be acted upon easily. The clinician presented with the output high risk/low risk of a patient would need to know why, so that it can be compared with clinical familiarity. Several of the models that perform at or near state-of-the-art levels are not explainable – deep neural networks and large ensembles are two examples. The lack of transparency is known as the black-box problem (Adadi & Berrada, 2018; Arrieta et al., 2020).

like the General Data Protection Regulation and the EU AI Act transparency obligations, and the Indian Council of Medical Research guidelines on AI in biomedical research and healthcare. It is a series of techniques which make the behaviours of (or the decisions made by) AI systems comprehensible to humans. In healthcare, the topic of explainability has been linked to clinical trust, patient autonomy and informed consent, error analysis, accountability and legal compliance (Amann et al., 2020; Holzinger et al., 2019). The topic has gained additional prominence through regulatory initiatives

1.2 Statement of the Problem

The literature on XAI is relatively new and expands rapidly, but it suffers from two limitations: (1) it is often written from a computer-science perspective, where explanations are judged solely by their technical merit, with little focus on whether clinicians find them meaningful or whether they influence decision-making and (2) reviews of XAI in medicine tend to ignore how explanation techniques relate to different clinical contexts and mostly do not mention the specific challenges posed by the Indian environment, in which heterogeneity of data sources, resource limitations, and the scale and speed of growth of the digital health infrastructure all raise different issues. Researchers have also argued that the current methods of explanation may only provide a false sense of comprehension (Rudin, 2019; Ghassemi et al., 2021), so clinicians need to be aware of both their capabilities and limitations.

1.3 Research Objectives

- ♣ To categorize the primary XAI techniques employed in the healthcare domain and to explain their operational mechanisms.
- To see where these techniques fit into clinical application areas and what their representative use cases are.
- To list the criteria and method used to evaluate the quality of explanations.
- Uncover the technical, clinical, legal and compliance issues that have hampered the adoption of XAI.
- Aims To give practical suggestions to developers, clinicians, hospitals and regulators, with special emphasis on Indian scenario.

1.4 Research Questions

- What are the most popular XAI methods for healthcare and what are their advantages and limitations?
- How do we show XAI's efficacy in practice and for what kinds of data is it the most effective?
- In a clinical context, how can we measure the validity of an explanation?
- What are the barriers to deployment of explainable models in routine care?
- Should complex models be explained in post-hoc or have simpler and more interpretable models?

1.5 Scope and Importance of the Study

This paper is a narrative review driven by secondary sources. It addresses explanation methods for explaining the operation of machine-learning models used for clinical and biomedical purposes and does not describe new experiments, and does not attempt to be an extensive systematic review. This study provides help to clinicians where there is a necessity to interpret what an AI explanation presents and doesn't present; knowledge scientists engaged on healthcare-related datasets who're creating clinical algorithms; hospital executives, financial backers, or thought leader showing for an AI vendor; and regulators together with the Central Drugs Standard Control Organisation (CDSCO), ICMR and the National Health Authority defining governance standards for AI in health.

CHAPTER 2: LITERATURE REVIEW

This chapter covers the theoretical and empirical contributions that constitute the core of the field and defines the research gap.

2.1 Doshi-Velez & Kim (2017): Interpretability explained

Doshi-Velez and Kim (2017) claimed that interpretability has no formal definition and suggested a taxonomy of evaluation: application-grounded (testing with real users on real tasks), human-grounded (simpler tasks with lay users) and functionality-grounded (using proxy measures with no humans). This assessment framework is still the go-to for XAI evaluation, and illustrates how poorly this works in healthcare.

2.2 Ribeiro, Singh & Guestrin (2016); Lundberg & Lee (2017): Model-Agnostic Explanations

Ribeiro et al. (2016) developed LIME, which explains a single prediction by approximating it locally with a simple surrogate model. Lundberg and Lee (2017) came up with SHAP, which explains a prediction by attributing a contribution to each feature, taking the average of values they would obtain in a set of possible models, like calculating the Shapley values in cooperative game theory. These are the two most popular approaches in healthcare with tabular data, since they can generate explanations for nearly all models and they give a feature-level explanation that clinicians can interpret.

2.3 Selvaraju et al. (2017) and Sundararajan et al. (2017): Explanations for Images

Grad-CAM (Selvaraju et al., 2017) generates a heat map indicating the areas of an image that contributed to a convolutional net's decision, and is increasingly used for transparency in radiology, pathology, and dermatology model. Integrated gradients (Sundararajan et al., 2017) can decompose a model prediction to individual input features by accumulating a path integral of gradients from a baseline to an input, which satisfies desirable properties like sensitivity and implementation invariance.

2.4 Lundberg et al. (2018, 2020): XAI for Clinical Prediction

Lundberg et al. (2018) explained the same model that estimated the risk of hypoxaemia during the operation to anaesthetists, providing the explanation with SHAP-based methods. And finally, Lundberg et al. (2020) introduced TreeSHAP and used it for risk models trained on extensive health databases, demonstrating for the first time the suitability of a tree ensemble to be explained locally and globally. These are often referenced as the best examples of

explanations in a clinical workflow.

2.5 Caruana et al. (2015): A Cautionary Tale

Caruana et al. (2015), learned that those with asthma were at a reduced risk of death - a result driven by the fact that those with asthma received more aggressive treatment. The rule was identified using an interpretable generalised additive model and was then adjusted for. This example forms part of the classic set of justifications for interpretability in high-stakes settings.

2.6 Rudin (2019) and Ghassemi et al. (2021): Critiques of Post-hoc Explanation

Rudin (2019) suggested that for high-stakes decisions, it is better to use inherently interpretable models rather than explaining black box models ex-post, as post-hoc explanations can be unfaithful to the model. Ghassemi et al. (2021) studied prevalent practices in health care and found that they provide weak guarantees at the individual level and that model validation should take precedence over explanation.

2.7 DeGrave et al. (2021): Shortcut Learning

DeGrave et al. (2021) demonstrated that large, deep-learning based systems to detect COVID-19 from chest x-ray images depended on confounders like source image or marker rather than pathology - and that saliency-based explanations could reveal this. This article exemplifies diagnostic benefit of XAI for developers, as well as the danger of seemingly accurate models breaking down in different contexts.

2.8 Comments and Observations

Existing surveys by Adadi and Berrada (2018), Arrieta et al. (2020) and Tjoa and Guan (2021) categorize the various techniques of XAI and talk about how XAI can be utilized to medication. Holzinger et al.

3.3 Analytical Framework

Each technique and study was analysed using the following dimensions.

Table 3.1: Analytical framework used in the review

Dimension	Description	Question asked
Scope of explanation	Local (single prediction) or global (whole model)	Does the method explain one patient or the model overall?
Model dependence	Model-agnostic or model-specific	Can it be applied to any classifier?
Timing	Post-hoc or intrinsic (ante-hoc)	Is interpretability built in or added afterwards?
Data type	Tabular, image, signal, text, genomic	For which clinical data does it work?
Output form	Feature weights, heat map, rule, example, counterfactual	What does the clinician actually see?
Evaluation level	Functionality-, human- or application-grounded	Has it been tested with real users?

3.4 Limitations of the Methodology

- There was a narrative review and therefore may be subject to selection bias and this was not a meta-analysis.
- XAI is an evolving field and new publications may not be included.
- Performance of models reported in these studies are not directly comparable as datasets and tasks vary.
- This paper does not comment on any specific commercial product.

CHAPTER 4: ANALYSIS AND DISCUSSION

4.1 Overview of XAI Techniques in Healthcare

XAI approach into model-agnostic post hoc approaches, model specific post hoc approaches, and inherently interpretable models. We summarize in

Table 4.1 the techniques that are most frequently used in medicine.**Table 4.1: Major XAI techniques used in healthcare**

Technique	Type	Scope	Typical data	What the user sees	Main strength	Main limitation
LIME (Ribeiro et al., 2016)	Post-hoc, model-agnostic	Local	Tabular, text, images	Weights of features in a local surrogate	Simple, works with any model	Unstable; depends on sampling and neighbourhood
SHAP (Lundberg & Lee, 2017)	Post-hoc, model-agnostic	Local and global	Tabular, some images	Feature contributions per patient and overall	Theoretical grounding; fast version for trees	Costly for large models; assumes feature independence in some forms
Saliency maps (Simonyan et al., 2014)	Post-hoc, model-specific	Local	Images	Pixel-importance map	Cheap to compute	Noisy; sensitive to model changes
Grad-CAM (Selvaraju et al., 2017)	Post-hoc, model-specific	Local	Images (CNNs)	Heat map of important regions	Widely supported in radiology and pathology	Low resolution; may highlight regions that are not causal
Integrated gradients (Sundararajan et al., 2017)	Post-hoc, model-specific	Local	Images, signals, tabular	Attribution per input feature	Satisfies attribution axioms	Depends on baseline choice
Attention visualisation	Intrinsic/post-hoc	Local	Text, sequences, EHR	Attention weights over tokens or visits	Easy to display	Attention is not always a faithful explanation
Counterfactual explanations	Post-hoc	Local	Tabular, images	"What would need to change for a different outcome"	Actionable for patients and clinicians	May suggest clinically unrealistic changes
Decision trees, rule lists, GAMS	Intrinsic	Global and local	Tabular	Rules or additive feature curves	Transparent, easily audited (Caruana et al., 2015)	May be less accurate on complex data

4.2 Clinical Application Areas

Medical imaging. In radiology, dermatology, ophthalmology and pathology, convolutional networks rule, and Grad-CAM-type heat maps show the areas that led to a decision. These maps can be used to verify that a model is focusing on a lesion and not an artefact, as occurred in the shortcut-learning paper of DeGrave et al. (2021).

Clinical risk prediction: Patient models based on electronic health records forecast things like sepsis, readmission and mortality. SHAP is the go-to explanation that displays which vitals, labs and history elements have increased or lowered risk for an individual patient (Lundberg et al., 2018, 2020).

Critical care & monitoring. Streaming icons and vital signs from the ICU are explained with the origin traced back along the time axis to show the lead section that produced the alarm.

Genomics & Drug Discovery Feature attribution helps identify which genes, variants, or molecular substructures are associated with a prediction, thereby aiding hypothesis generation.

Clinical language processing. Use attention and token-level attribution to identify phrases in the clinical notes that contributed to the classification.

Table 4.2: Application areas, typical explanation methods and their purpose

Application area	Typical task	Commonly used XAI methods	Purpose of the explanation	Key concern
Medical imaging	Detection of diabetic retinopathy, skin lesions, pneumonia, tumours	Grad-CAM, saliency, integrated gradients	Check that the model attends to pathology, not artefacts	Heat maps may be unfaithful or low resolution
Clinical risk prediction	Sepsis, mortality, readmission, hypoxaemia	SHAP, LIME, GAMS	Show contributing risk factors per patient	Correlation presented as causation
ICU and wearable monitoring	Arrhythmia, seizure and deterioration alerts	Integrated gradients, attention	Locate the signal segment that triggered an alert	Alert fatigue; time-critical use
Genomics and drug	Variant effect, molecular	SHAP, attention,	Generate biological	Explanations need wet-

Application area	Typical task	Commonly used XAI methods	Purpose of the explanation	Key concern
discovery	property prediction	substructure attribution	hypotheses	lab validation
Clinical text	Note classification, coding, triage	Attention, token attribution, LIME	Highlight evidence in the note	Attention is not always explanation
Patient-facing tools	Self-assessment, treatment options	Counterfactuals, simple rules	Support understanding and consent	Health literacy and language

4.3 Stakeholder Needs for Explanation

Different users need different explanations, and a single method rarely satisfies all of them.

Table 4.3: Stakeholders and their explanation requirements

Stakeholder	What they need to know	Suitable form of explanation
Clinician	Why this patient was flagged; whether it fits clinical reasoning	Per-patient feature contributions, heat maps with the image
Patient	What the result means and what can be done	Plain-language reasons, counterfactual advice
Developer / data scientist	Whether the model learned the right signal; debugging	Global importance, error analysis, attribution audits
Hospital administrator	Whether the system is safe and fair to deploy	Subgroup performance, model documentation
Regulator / auditor	Whether the system meets legal and safety requirements	Documentation, validation evidence, audit trail

4.4 Evaluating Explanations

Following Doshi-Velez and Kim (2017), evaluations may be application-grounded, human-grounded or functionality-grounded. Table 4.4 lists commonly used criteria.

Table 4.4: Criteria for evaluating explanations

Criterion	Meaning	Example of assessment
Fidelity (faithfulness)	The explanation reflects what the model really does	Remove or perturb highlighted features and observe the change in prediction
Stability / robustness	Similar inputs give similar explanations	Compare explanations for near-identical patients
Comprehensibility	The user can understand the explanation	User studies; time to understand
Clinical plausibility	The explanation agrees with medical knowledge	Expert review of explanations
Usefulness	It improves decisions or detects errors	Clinician accuracy with versus without explanations
Sparsity / simplicity	Explanation is short enough to use	Number of features shown
Fairness insight	The explanation helps to reveal bias	Compare attributions across subgroups

A central observation is that most published work uses functionality-grounded proxies, while application-grounded studies with clinicians are rare. This matters because an explanation that looks convincing can raise confidence in a wrong prediction.

4.5 Comparing Interpretable Models and Post-hoc Explanations

Table 4.5: Inherently interpretable models versus post-hoc explanation of complex models

Aspect	Inherently interpretable model	Complex model with post-hoc explanation
Transparency	The model itself can be read and audited	Explanation approximates the model
Faithfulness	Exact by construction	May be unfaithful (Rudin, 2019)
Accuracy on tabular clinical data	Often comparable	Sometimes slightly higher
Accuracy on images and signals	Generally lower	Usually higher
Error detection	Easier (Caruana et al., 2015)	Possible, but explanations can mislead
Suitable when	Structured data and high-stakes decisions	Unstructured data where deep learning is needed

The evidence suggests that the choice should depend on the data type and the risk of the decision. For structured clinical data, an interpretable model should be tried first; for images and signals, post-hoc methods are usually unavoidable and must be validated carefully.

4.6 Challenges in Adopting XAI in Healthcare

Table 4.6: Principal challenges and possible mitigations

Challenge	Description	Possible mitigation
Unfaithful explanations	Post-hoc methods may not reflect model reasoning (Rudin, 2019)	Fidelity tests; prefer interpretable models where possible
False confidence	Plausible heat maps can increase trust in wrong outputs (Ghassemi et al., 2021)	Present uncertainty; train clinicians in limits of XAI
Shortcut learning	Models rely on artefacts or confounders (DeGrave et al., 2021)	External validation; multi-site data; attribution audits
Correlation versus causation	Attributions show association, not cause	Causal modelling; clinical review
Bias and fairness	Models may underperform for some groups (Obermeyer et al., 2019)	Subgroup evaluation; diverse training data
Lack of user studies	Few trials show benefit to clinicians	Application-grounded evaluations in real workflows
Workflow integration	Explanations add time and cognitive load	Co-design with clinicians; concise displays
Privacy and security	Explanations can leak information about training data	Access control; privacy-preserving methods
Regulation	Requirements for transparency and documentation are evolving	Documentation, audit trails, risk management
Data heterogeneity (India)	Varied equipment, languages and record quality	Local validation; standardised health data formats

4.7 The Indian Context

India's digital health landscape, including the Ayushman Bharat Digital Mission, is building the access to e-health records and may offer the data infrastructure needed for AI. Nevertheless, models trained on the data of other populations don't often translate well for Indian patients because of differences in disease prevalence, imaging equipment, record quality, etc., increasing the importance of local validation and clear explainability. The ICMR guidelines for AI in biomedical research and healthcare focus on principles like accountability, transparency and safety, and the Digital Personal Data Protection Act, 2023 legislates the use of personal data. AI-related medical software may also be covered by the regulations for medical devices and standards by the CDSCO. Explainability accommodates these requirements, allowing for audit, clinician oversight and informed consent, but it does not on its own ensure safety.

4.8 Key Findings

- XAI approaches have best applicability and maturity in tabular clinical data (SHAP, LIME, explainable additive models) and in image data (Grad-CAM, saliency, integrated gradients).
- Explanations are useful to developers for debugging and clinicians for sanity checking, but explanation does not provide any guarantee of the model being right or fair.
- Unfaithful post hoc explanations and visually deceptive explanations may generate false confidence.
- Technical proxies continue to dominate the evaluative landscape. Few clinical tests with users have been attempted.
- Of course interpretable models are the easier choice for very high-stakes decisions on structured data.
- Different stakeholder groups need different explanations, hence the design of XAI needs to be user-centred.
- In the Indian context, explanation is best supported with local validation, data governance and regulatory alignment.

CHAPTER 5: CONCLUSION AND RECOMMENDATION

5.1 Summary of the Study

The paper on explainable AI in healthcare presented: techniques were classified, mapped to clinical applications, summarised evaluation criteria and identified challenges and recommendations. The main point was: explainability is necessary but not sufficient for trustworthy clinical AI.

5.2 Conclusion

Four lessons are clear. First, the development of XAI techniques has matured enough to support debugging, validation and clinician support (including support for tabular data and medical images). Second, post-hoc explanations are necessarily approximation, and they need to be validated for fidelity and stability before they can be put to use. Third, the best evidence for the utility of explanations comes from model auditing (for example to uncover shortcut

learning and confounded rules) rather than the observed impact on patients. Fourth, people are the principal audience for explanations, not necessarily clinicians, but perhaps regulatory agencies, or perhaps patients or developers. These issues are discussed in a narrative review of a select subset of relevant papers, and should not be interpreted as a definitive assessment of any specific technique.

5.3 Recommendations

Table 5.1: Recommendations by stakeholder

Stakeholder	Recommendation
Developers and researchers	Try interpretable models first for structured data; test fidelity and stability of explanations; validate externally and across subgroups; document data and model limits
Clinicians	Treat explanations as aids for checking a prediction, not as evidence of correctness; seek training on what attribution methods show and do not show
Hospitals and administrators	Require local validation, post-deployment monitoring and clear escalation procedures before adopting AI tools; involve clinicians in design
Regulators and policymakers	Define documentation and validation expectations; encourage reporting of subgroup performance; support application-grounded evaluation standards
Patients and public	Provide plain-language information about AI use in care and the right to ask for a human review

5.4 Limitations

These restrictions of 3.4 hold. In particular, the review is selective, the evidence base is rapidly changing and comparison of results across studies is difficult.

5.5 Possibility for Further Studies

- C. Conduct application-grounded studies where clinicians use AI with and without explanations, and measure accuracy, time, and trust.
- "Develop benchmarks and measures for fidelity and clinical plausibility of explanations."
- Learn causality and counterfactual explanation approaches that are clinically plausible.
- Assess the XAI using Indian datasets and multilingual clinical text, and in public and private care environments.
- Explore descriptions of large language models and multimodal clinical systems.

REFERENCES

Foundational methods

1. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.
 2. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.
 3. 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. Proceedings of the IEEE International Conference on Computer Vision.
 4. Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. Proceedings of the 34th International Conference on Machine Learning, 3319-3328.
 5. Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Deep inside convolutional networks: Visualising image classification models and saliency maps. Workshop at the International Conference on Learning Representations.
 6. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. ArXiv:1702.08608.
- ### Healthcare XAI studies, reviews and critiques
7. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721-1730.
 8. Lundberg, S. M., Nair, B., Vavilala, M. S., et al. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomedical Engineering, 2, 749-760.

9. Lundberg, S. M., Erion, G., Chen, H., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2, 56-67.
10. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206-215.
11. 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), e745-e750.
12. DeGrave, A. J., Janizek, J. D., & Lee, S.-I. (2021). AI for radiographic COVID-19 detection takes a shortcut. *Nature Machine Intelligence*, 3, 610–619.
13. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*, 9(4), e1312.
14. 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, 310.
15. 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.
16. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138-52160.
17. Arrieta, A. B., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.
Clinical AI context
18. Gulshan, V., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402-2410.
19. A, et al. (2017). A dermatologist-level classification of skin cancer with deep neural networks *Nature* 542 115-118.
20. E. J. Topol, Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25, 44-56.
21. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.
22. Policy and regulatory documents
22. Ethical guidelines for application of artificial intelligence in biomedical research and healthcare (2023).
23. Government of India. Digital Personal Data Protection Act, 2023.
24. European Union. Regulation (EU) 2016/679 (General Data Protection Regulation) Regulation (EU) 2024/1689 (Artificial Intelligence Act).