CSIT (April 2023) 11(1):57–70
https://doi.org/10.1007/s40012-023-00381-2

ORIGINAL RESEARCH

EASE© framework in design and development of clinical artificial intelligence applications

Sujoy Kar[1] · **Triyanka Tiu**[1] · **Sangita Reddy**[1]

Received: 20 February 2023 / Accepted: 27 March 2023 / Published online: 12 April 2023

Abstract Building Clinical artificial intelligence (AI) applications requires a delicate balance between clinical need, technical knowhow and ethical considerations. Many Clinical AI models have issues like unproven value, model opacity, model drift, disutility and persistent resistance for adoption. At Apollo Hospitals, we design, develop, validate and deploy Clinical AI solutions with ethically sourced clinical data which feature facets from accountability and accuracy to fairness and inclusivity. Without proper oversight, Clinical AI applications can quickly become flawed, hindering their potential impact. To prevent such issues, the EASE framework was developed to guide the processes towards the right path. EASE encompasses Ethics, Adoption, Suitability and Explainability. As digital technology continues to permeate all aspects of society, the discussion and debate surrounding data integrity will only grow more pressing. Data ethics are relevant to everyone—government, businesses, and individuals alike—and should be approached with openness and transparency. The process of discussion and knowledge-sharing is vital for raising awareness and promoting ethical practices.

Keywords Ethics · Artificial intelligence · Research · Software · Intersectional framework

✉ Sujoy Kar
drsujoy_k@apollohospitals.com

Triyanka Tiu
drtriyanka_t@apollohospitals.com

Sangita Reddy
sangita_reddy@apollohospitals.com

[1] Apollo Hospitals, Jubilee Hills, India

1 Introduction

The advent of digital technologies, particularly AI and its subset, Machine Learning (ML), is transforming the fields of medicine, medical research, and public health. AI holds tremendous potential and has already led to significant advancements in areas such as drug development, genomics, radiology, pathology, and disease prevention. By reducing errors and allowing healthcare professionals to concentrate on providing care, AI can greatly benefit the healthcare industry. The benefits of AI in healthcare and its economic potential are driving an increased worldwide adoption of these technologies.

However, use of Clinical AI has four fundamental challenges that clinicians and AI practitioners should address during the process for design, development and deployment of Clinical AI. These are—

(a) Unproven Value—How to ensure AI—ML models to be more clinically relevant, particularly when they are used to address clinical uncertainties. How do we design Randomized Control Trials to unearth the value for clinical AI solutions as adjunct tools to improve decision making [1].
(b) Model Opacity—How inexplainable models hinder in adoption as clinicians cannot confidently trust ML Models and their outputs due to the lack of understanding [1].
(c) Model Drift—How AI—ML models overcome geographic, ethnic, gender, machine and individual bias in development, validation, deployment and prospective use. As medicine continues to evolve, how do these models prepare to learn from the feedback loops of new learnings and associated data [1].

(d) AI versus Clinician Conundrum—We have traditionally put AI against Clinicians which has put in an inevitable resistance to change and adopt new technologies [2].

Furthermore, there are lack of comprehensive international guidance on the ethical and human rights-compliant use of Clinical AI. Nations lack regulations specifically aimed at governing the use of AI in healthcare, and existing laws around the use of data and AI are evolving. The World Health Organization [3] and the European Commission [4] are among the organizations working towards creating regulations that prioritize ethics and human rights.

2 Our approach

At our organizations' network, the healthcare systems leverage technology to build integrated healthcare delivery models, which facilitate seamless electronic medical records. Artificial intelligence, machine learning and the deep-learning, in particular, are empowering the use of labelled clinical data from these electronic medical records—'big' in terms of volume, variability, velocity, or scalability—with significantly enhanced computing power and cloud storage. At our organization' clinical practice and population health perspective, this is making initial steps to have an impact at these fundamental levels:

- For patients, by enabling them with better decisions to access and to process their own data, federated—secured—meaningfully used to promote health,
- For clinicians, predominantly via supported, accurate interpretation of patient data including electronic health records and images;
- For hospitals, by improving throughput, enhancing patient safety and the potential for reducing healthcare cost;
- For technology providers—constructing digital health platforms and creating positive network effects where patients, providers, payers and physicians can derive intense clinical value, and lastly
- For Community—by creating solutions that have last mile accessibility.

Our organization has formulated a framework named EASE—Ethics, Adoption, Suitability and Explainability to propagate Ethical Practices in Clinical AI. They are enumerated below.

(a) Ethics

a. Safety and Non-Maleficence—The paramount objectives of Ethical Principles in AI and Healthcare Data Management include that no Clinical Algorithm or its implementation would lead to avoidable harm or unintended consequences to individuals (patients) using the product.
b. Accountability—Conclusions and recommendations about AI and Healthcare Data Management work must be documented and auditable.
c. Efficacy and Accuracy—The efficacy of the models or clinical AI algorithms developed shall be verified in compliance with accepted international standards. The accuracy of the model or the algorithms for diagnosis, risk prediction or prognosis and associated Clinical Decision Support shall be transparently stated and compared with the existing systems.
d. Integrity—Models and Algorithms derived from high quality and ethically sourced data, collected for a clearly defined purpose—with appropriate consent is reliable and evidenced based.
e. Fairness and Inclusivity—Models and Algorithms designed and developed under the aegis of the collaboration shall not discriminate with respect to ethnicity, geography, gender, age, religion and patient's pre-existing clinical condition (e.g. PLHIV). It shall not exclude vulnerable groups.

(b) Adoption

a. Interoperability and Universalization—Algorithm prepared on one organization's EMR need to be translatable across others and hence requires extensive validation in diverse population.
b. Calibration and Benchmarking—Our organization strongly advocates Clinicians + AI—they are mutually complimentary to the decision making
c. Clinical Relevant Outcomes—like Sensitivity, Predictive Values, Likelihood Ratios etc.
d. Patient Safety—Meeting accepted standards for Clinical Benefit.

(c) Suitability

a. Solve the right clinical or public health problem.
b. Address Bias—Bias generated from geographical, different clinical setting, investigators, different machines (Tests /Imaging / Graphs (ECG)) impact on the development and validation of an AI model.
c. Practical Accuracy—come with a solution that reflects—Knowledge Representation, Causal Structure, Missing Observation, Flow of Information.
d. Machine and Human Handovers—algorithms learn to handover/defer to clinical expert in decision making and vice versa.

(d) Explainability

a. Verifying before Deciding—Validating model's explanation against expert knowledge before taking an action.
b. Building Trust in System—View the reasons behind decisions and predictions. Verify that the performance is optimum.
c. Counterfactual Reasoning—Test through all the what-if scenarios which confirm the accuracy of its predictive power.
d. New Clinical Insights—Data on a scale have access to undiscovered medical insights, like drugs and interactions (Fig. 1).

3 Discussion

In this section, we discuss the important aspects of 4 EASE Components and some of the 18 EASE Subcomponents.

3.1 Ethical considerations

3.1.1 Safety and non-maleficence

Principles of patient safety are derived from the safety in design and deployment of AI Ethics Principles (EU) and Health Ethics Principles [5]. Notable areas of safety include—

(a) Patient Interest—The Clinical AI Solutions should always work in patients' interests and in doing so, it should encompass all stakeholders for buy in, be part of an eco-system that enables effective, affordable access to high quality health services and which is never used to discriminate, persecute or deny necessary care.
(b) Use of Clinical AI and associated decision tools should always be to improve a patient's outcomes and those of others, through appropriate consent from patients for information to be shared. Care providers should acknowledge their absolute duty of care with respect to the use of patient information.
(c) "Do no harm" should apply to the use, particularly in design and deployment of solution to ensure the that the risks of malicious use, unintended use and use in different clinical context other than what the model

Fig. 1 EASE Framework—Our organizations' concept

has been designed for are anticipated, safeguarded and these risks are mitigated.

(d) Evolving Genre of ML Systems—With AI ML algorithms anticipated to evolve and become more proactive and reward driven, it is imperative that they need to be monitored to ensure they are working as desired.

3.1.2 Accountability

Rising concern for the societal implications of AI systems has inspired a wave of academic and media literature [6]. Systems deployed were audited for unacceptable results, by investigators outside organizations deploying the algorithms. It is expected that it is the responsibility of any reputed organization, to explain or justify their AI ML Models. Auditing is strategically important to detect embedded biases in the Clinical AI ML and ensuring transparency and fairness.

Raji et al. introduced a framework for algorithmic auditing that supports an AI system development end-to-end. This is to be applied throughout the internal organization development life-cycle. Each stage of the audit yields a set of documents that together form an overall audit report, drawing on an organization's values or principles to assess the fit of decisions made throughout the process. The proposed auditing framework is intended to contribute to closing the accountability gap in the development and deployment of large-scale AI systems by embedding a robust process to ensure audit integrity [7].

3.1.3 Efficacy

The foundation of any AI Algorithm needs to follow a set of standards or principles. Challenges faced by High Efficacy Clinical Algorithms deployed in Clinical Settings include the following:

(a) Retrospective versus Prospective Study

Retrospective investigations are frequently criticized because the majority of errors are caused by confounding factors, and bias is more common. In retrospective studies, the odds ratio estimates relative risk. The desired outcome should be common; otherwise, the number of observed outcomes will be too small to be statistically significant (indistinguishable from those that may have arisen by chance). All efforts should be made to avoid sources of bias, such as the failure to follow up on individuals. Prospective studies have fewer sources of bias and confounding variables.

(b) Metrics and Clinical Reality

Clinical AI based Software as Medical Devices (SaMD) need to be scalable and measurable. Each diagnosis is different for each individual. To cater to this complex nature a high level of efficacy in the real world is required for these medical devices. Continuous testing and validations at various scales is important. Without constant evolution of the technology and numerous testing, the results might give an inaccurate picture.

(c) Issues related to the Model Design and Development

Randomized Control Trials using the TRIPOD Checklists (Appendix 1) [8] in Development and Validation for transparent reporting of multivariable prediction models for prognosis, diagnosis or decision support systems are highly recommended. It is also important that the Clinical AI ML models have clear and uniform metrics for evaluation. Difficulty comparing different algorithms in different setting are serious roadblocks for adoption as they lack consistency.

Another area which requires insights for Clinical AI is particularly in Deep Learning where there are accidental fitting confounders versus true signals.

3.1.4 Integrity

To foster a culture of integrity, AI should be held to human standards. A reasonable AI Model is reflective of its development and validation data sets. Ethical sourcing of data, deidentification and anonymization protocols and multiple rounds of data cleansing via multi layered deep neural networks—where the bad data, duplicates and incorrect data are filtered out, till it meets the set standard. To add to these aspects, organization building Clinical AI solutions should very clearly have sets of policies on use of synthetic data, imputation and associated techniques. We found it very challenging in meeting the standards in Large Clinical Language Datasets in their Named Entity Recognition and Relation Extraction steps of Natural Language Processing.

3.1.5 Fairness and inclusivity

AI is about inclusion and diversity in systems that impact millions from various backgrounds. The end users of this algorithm are humans, and they are affected by the models' decision. It is therefore essential that the algorithm is trained in diverse data sets and population. This in turn results in better inclusivity, preventing bias and giving more accurate results. The AI model needs to be explainable to the population so that an ethical analysis of the process is possible

[6]. This also helps in building trust in the AI models in the following ways—

1. A participatory process that involves key stakeholders, including frequently marginalized populations.
2. Considers distributive justice within specific clinical and organizational contexts.
3. Different technical approaches can configure the mathematical properties of machine-learning models to render predictions that are equitable in various ways.
4. The existence of mathematical levers must be supplemented with criteria for when and why they should be used—each tool comes with trade-offs that require ethical reasoning to decide what is best for a given application.
5. Incorporating fairness into the design, deployment, and evaluation of machine learning models.
6. Machine learning should not harm the protected groups by being inaccurate, diverting resources, or worsening outcomes, especially if the models are built without consideration for these patients.

3.1.6 How does bias sets in

During model development, differences in the distribution of features used to predict a label between the protected and non-protected groups, may bias a model to be less accurate for protected groups. Moreover, the data used to develop a model may not generalize to the data used during model deployment (training–serving skew). Biases in model design and data affect patient outcomes through the model's interaction with clinicians and patients [6].

1. Model Development—
    a. Label Bias—Inappropriate Labeling of Clinical Images or Signals, mostly unintentional due to pre-existing biases against certain diagnosis or artefacts.
    b. Missing Data Bias—Missing data and clinical values are highly common in Clinical AI ML Models. However, a lot of these models using machine learning, do not report adequate information on their presence and how would they handle the missing data.
    c. Cohort Bias—preselected cohort with cherry-picking clinical variables.
    d. Minority Bias and informativeness of data.

2. Model Deployment—
    a. Privilege Bias—Individuals left out and not being served by the model.
    b. Agency Bias—which generally represents informed mistrust.
    c. Skewness between Training and Serving.

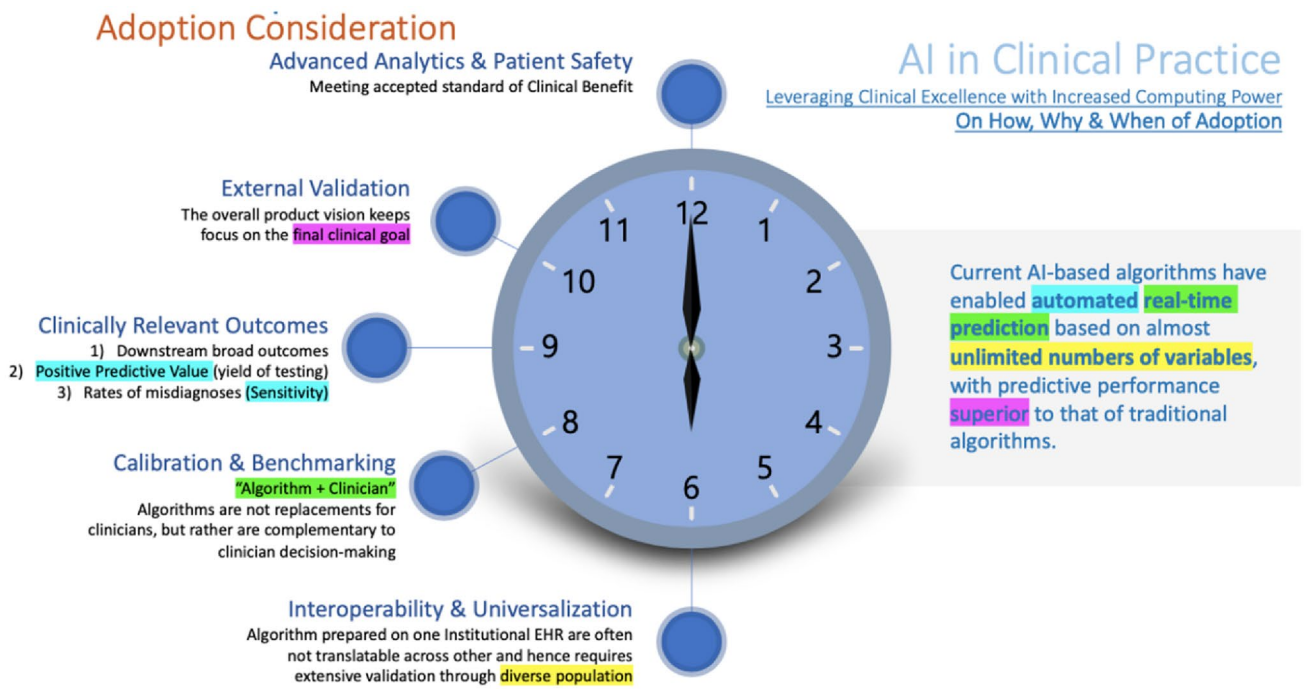

Fig. 2 Adoption Considerations

3.2 Adoption

Adoption of the Clinical AI ML is the acid test for success in research and its meaningful use. The Clinical AI algorithms require real time predictions, mostly made at the patient bedside or in the testing fields with almost unlimited variables providing a predictive performance (Positive Predictive Value and Likelihood Ratios) better than traditional algorithms. This requires technical skills for the users, secured datasets and very importantly, internet connectivity to take it to the last mile (Fig. 2).

3.2.1 Interoperability and universalization

An algorithm prepared on one organizations' EMR needs to be translatable across another organization's EMR. This requires extensive validation in a diverse population.

Interoperability is the ability of different information systems, devices and applications (systems) to access, exchange, integrate and cooperatively use data in a coordinated manner, within and across organizational, regional and national boundaries, to provide timely and seamless portability of information and optimize the health of individuals and population globally [9]. Functions of interoperable components include data access, data transmission and cross-organizational collaboration regardless of its developer or origin. Similar to compatibility, interoperability helps organizations achieve higher efficiency and a more holistic view of information [10]. Whereas, universalization lays emphasis on the need to develop a transcendent, collaborative model of human interaction that looks beyond the limited confines of current human relationships.

Data Validation is done internally as well as externally using the help of interested third parties, with the same goals. It is an important step in AI to ensure the quality of input data before it is used to develop models and insights.

3.2.2 External validation

Any form of technology, invention or discovery needs to be externally validated. The Clinical AI and its algorithmic models should be verifiable from a third party, outside of the organization. It should be tested and accuracy of results validated. This testing determines how well the models work against new and variable datasets. This technological peer review helps in determining areas of improvements and helps make the technology better. At our organization, we have been able to develop different Data Consortium which help in bringing the perspectives of running our Clinical AI APIs on resident data sets in other healthcare organizations across the world particularly using Federated Learning and Confidential Computing (Trusted Execution Environment).

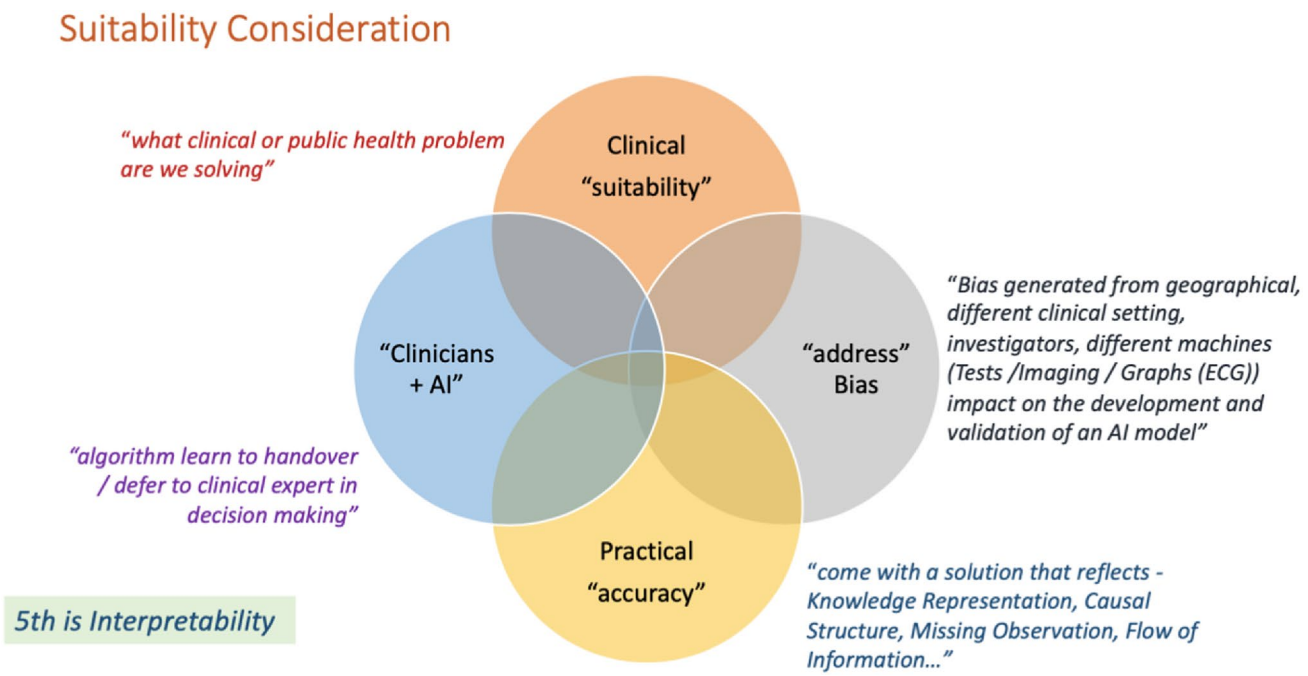

Fig. 3 Suitability Considerations

3.3 Suitability

Suitability is a measure of what that would work and what wouldn't in designing or framing a clinical or public health problem (Fig. 3).

3.3.1 Clinical need

Before starting any new project, a clinical need assessment is necessary, to save time and resources. It is a process by which information is gathered regarding the scope and potential impact of gaps or deficiencies in the current delivery and practice of health care [11]. The clinical or public health problems for which a solution is being developed should be clearly identified.

This involves critical review of information to identify gaps in health care. These include, identification of methods or technologies that may help address shortfalls, stakeholder requirements and consideration of potential barriers and facilitators for uptake of new technologies.

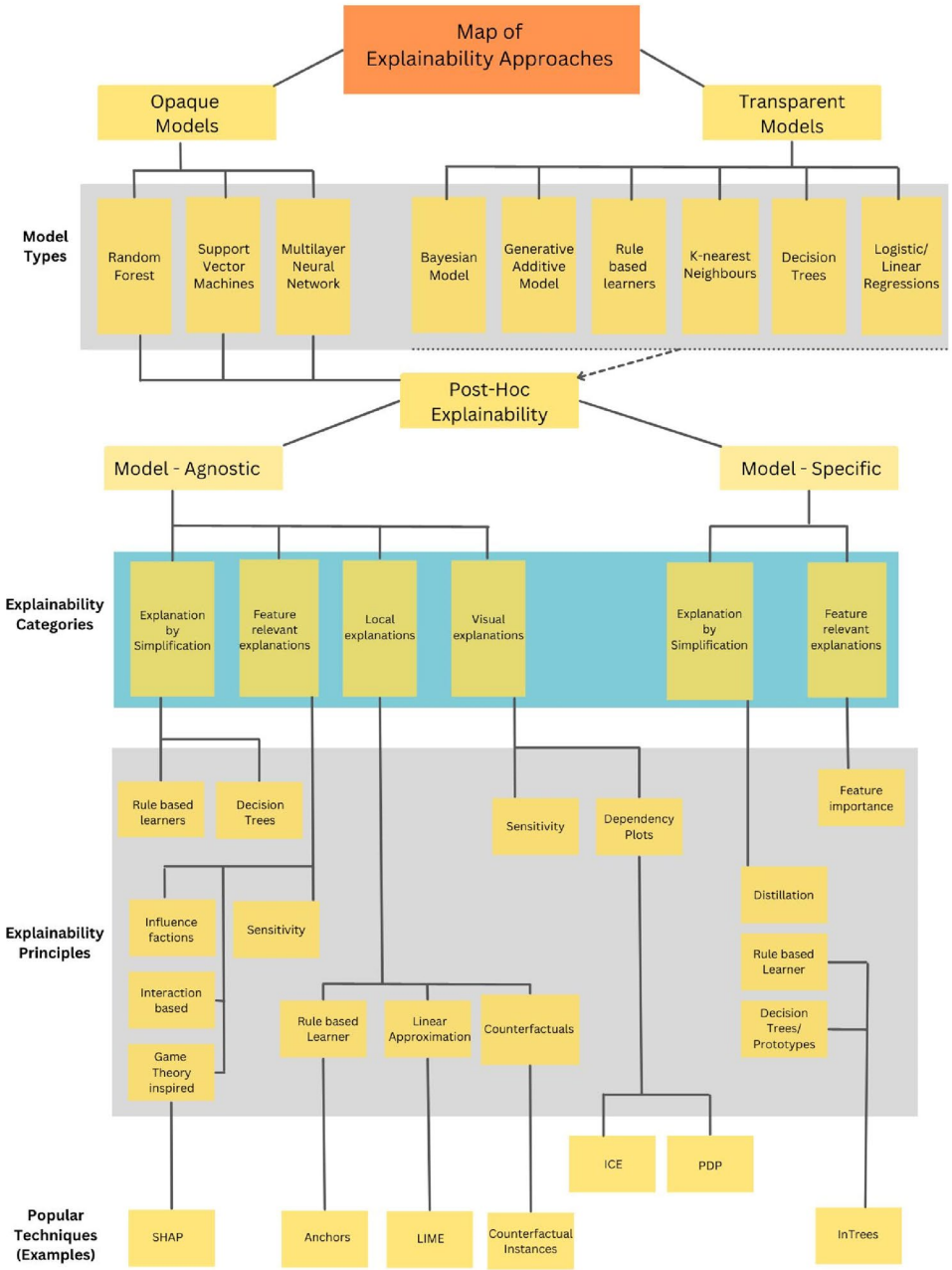

Fig. 4 Map of Explainability

3.3.2 Machine and human handover

It is also important to build trust among staff to report any safety concerns with the AI, part of effective human–AI teaming is handover from AI to the healthcare professional. To achieve this, AI needs to recognize the need for handover and then execute the handover effectively. Handover between healthcare professionals is a recognized safety–critical task that remains surprisingly challenging and error prone in practice. Consideration should be given to the development of comparable approaches for the structured handover between AI and healthcare professionals. Various researches have shown that general human–machine cooperation is achievable using a non-trivial, but ultimately simple, set of algorithmic mechanisms [12].

3.4 Explainability

Clinical AI Explainability is a set of tools which make the predictions of the models comprehensible and sensible to the Clinicians. These processes are designed to integrate into Clinical AI solutions developed at Our organization for their better acceptance and adoption. We strongly believe that a model that explains itself better to clinicians has a higher chance of adoption as clinician in turn can explain this well to patients leading to a flow of trust developing in the system.

Figure 4 gives an overview of Map of Explainability Approaches [10]. Its goal is to provide Clinicians and AI researchers a thorough taxonomy that can serve as reference material. This would stimulate future research advances, and encourage experts and professionals from other disciplines to embrace the benefits of AI.

3.4.1 Verifying before deciding

Verification activities need to ensure that data requirements are properly met. Validation activities ensure that the labelling and feature engineering activities are properly driven by the experts' input and user requirements.

At Our organization, we took the approach of validation of the AI models with the help of data consortium partners using Federated Learning. For the verification step, the approach was to go through comprehensive review by ISO 13485:2016, ISO 14971 and IEC 62304—certification process.

3.4.2 Building trust in the system

The use of more advanced and increasingly autonomous AI technologies presents novel challenges that require further study and research. AI technologies can augment what people do in ways that was not possible when machines simply replaced physical work, but in order to do this effectively, the AI needs to able to communicate and explain to the people its decision-making. This can be very challenging when using ML algorithms that produce complex and inscrutable models.

A direction that AI algorithms will need to consider is that of explainable AI. Several powerful AI algorithms are employing a so-called "black box" approach, where it is difficult to understand how the results have been obtained. In explainable AI, the features that a model is using to make a decision need to be traceable and understood by a human. As an example, transparent techniques like decision trees, relying on interpretable features, can provide a "white-box" approach for diagnosis [13].

3.4.3 Counterfactual reasoning

Counterfactual reasoning means thinking about alternative possibilities for past or future events. In interpretable ML, counterfactual explanations can be used to explain predictions of individual instances. It means testing through what-if scenarios which confirm the accuracy of its predictive power. An obvious first requirement is that a counterfactual instance produces the predefined prediction as closely as possible. Another quality criterion is that a counterfactual should be as similar as possible to the instance regarding feature values. The counterfactual should not only be close to the original instance, but should also change as few features as possible. It is often desirable to generate multiple diverse counterfactual explanations so that the decision-subject gets access to multiple viable ways of generating a different outcome [14].

3.4.4 New clinical insights

A software as medical device (Clinical AI) should always be open to change and should be able to incorporate future advancements without risking functionality and relevance of the core of the algorithm. The technology should always be open to evolving, as new ideas emerge.

In the last decade we have been facing a continuous and rapid exponential growth of usage and production of clinical data. In general, technological developments associated with healthcare (new powerful imaging machines) on one side have improved the general healthcare quality. Nevertheless, on the other side they have produced much more data than expected. Conversely, our developments in data mining techniques have been growing much slower than expected or at least not as fast as the production of data. This data represents then an almost unexplored source of potential information that can be used to develop clinical prediction models,

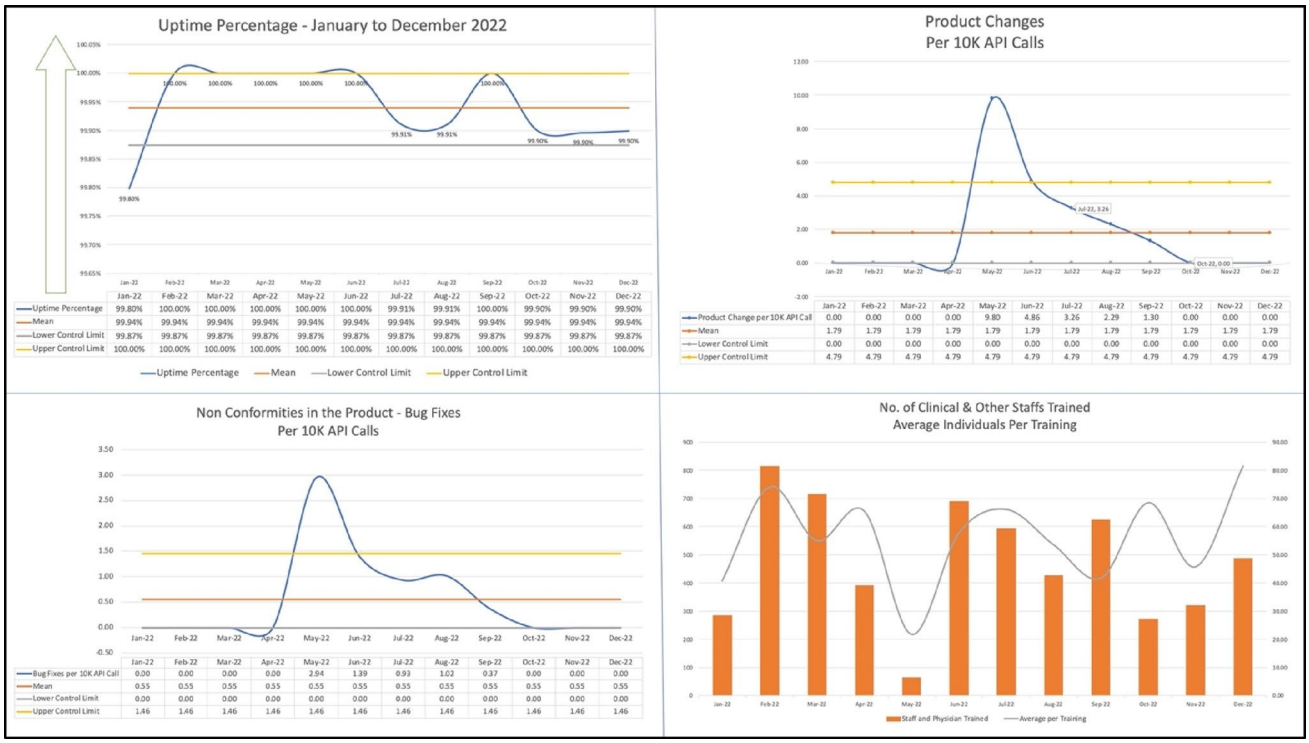

Fig. 5 EASE implementation indicators

using all the information (e.g. imaging, genetics banks, and electronic reports) available in medical institutions [15].

3.5 Effects from implementation of the program

The organization implemented the EASE Framework (Checklist in Appendix 2) in Early 2021 and it has helped in ISO 13485 certification process. Organization was successfully certified and re certified for 5 Clinical APIs using these guidelines in February 2022 and January 2023. From the process perspective, it was beneficial by having tracked following monthly trackers (Fig. 5) of deployment—

- 99.4% Uptime for over 300 thousand API calls in India and abroad,
- Product Changes in 1.79 instances per 10,000 API Calls,
- 0.55 bug fixes and non-conformity rate per 10,000 API Calls,
- No issues registered with Regulatory or Certifying Authority on the performance and inferences of the Clinical AI API models, and most importantly,
- Over 3000 Physicians and Health Care Workers trained in ethical use of Clinical AI API in just last one year.

4 Conclusion

Adoption of EASE Framework at Our organization have provided impetus to Clinical AI researchers and clinicians to design, develop and deploy these solutions within and beyond the institution.

5 Appendix 1

TRIPOD Checklist

Section/topic	Item		Checklist item
Title and abstract			
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted

Section/topic	Item		Checklist item
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions
Introduction			
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both
Methods			
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres
	5b	D;V	Describe eligibility criteria for participants
	5c	D;V	Give details of treatments received, if relevant
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted
Predictors	7a	D;V	Clearly define all predictors used in developing the multivariable prediction model, including how and when they were measured
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors
Sample size	8	D;V	Explain how the study size was arrived at
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation
	10c	V	For validation, describe how the predictions were calculated
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done

Section/topic	Item		Checklist item
Risk groups	11	D;V	Provide details on how risk groups were created, if done
Development versus validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors
Results			
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome)
Model development	14a	D	Specify the number of participants and outcome events in each analysis
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point)
	15b	D	Explain how to use the prediction model
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance)
Discussion			
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data)
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research
Other information			
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets
Funding	22	D;V	Give the source of funding and the role of the funders for the present study

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

6 Appendix 2

EASE Checklist

Objective		Clinical area		
Framework	Definition	Justification	Score	Remarks
Ethics				
Safety and non-maleficence	The paramount objectives of Ethical Principles in AI and Healthcare Data Management include that no Clinical Algorithm or its implementation would lead to avoidable harm or unintended consequences to individuals (patients) using the product	Objectives Met (10) Partially Met (05) Not Met (05)		
Accountability	Conclusions and recommendations about AI and Healthcare Data Management work must be documented and auditable	Objectives Met (10) Partially Met (05) Not Met (05)		
Efficacy	The efficacy of the models or clinical AI algorithms developed shall be verified in compliance with accepted international standards	Objectives Met (10) Partially Met (05) Not Met (05)		
Accuracy	The accuracy of the model or the algorithms for diagnosis, risk prediction or prognosis and associated Clinical Decision Support shall be transparently stated and compared with existing systems	Objectives Met (10) Partially Met (05) Not Met (05)		
Integrity	Models and Algorithms derived from high quality and ethically sourced data, collected for a clearly defined purpose—with appropriate consent shall be reliable and evidenced based	Objectives Met (10) Partially Met (05) Not Met (05)		
Fairness and Inclusivity	Models and Algorithms designed and developed under the aegis of the collaborative shall not discriminate in respect to ethnicity, geography, gender, age, religion and patient's pre-existing clinical condition (e.g. PLHIV). It shall not exclude vulnerable groups	Objectives Met (10) Partially Met (05) Not Met (05)		
Adoption				
Interoperability and Universalization	Algorithm prepared on one organization's EMR need to be translatable across other and hence requires extensive validation in diverse population	Objectives Met (10) Partially Met (05) Not Met (05)		
Calibration and Benchmarking	Our organization strongly advocates Clinicians + AI—they are mutually complimentary to decision making	Objectives Met (10) Partially Met (05) Not Met (05)		
Clinical Relevant Outcomes	Like Sensitivity, Predictive Values, Likelihood Ratios etc	Objectives Met (10) Partially Met (05) Not Met (05)		
Patient Safety	Meeting accepted standards for Clinical Benefit	Objectives Met (10) Partially Met (05) Not Met (05)		
Suitability				

Objective		Clinical area		
Framework	Definition	Justification	Score	Remarks
Clinical Need	Solve the right clinical or public health problem	Objectives Met (10) Partially Met (05) Not Met (05)		
Address Bias -	Bias generated from geographical, different clinical setting, investigators, different machines (Tests/Imaging/Graphs (ECG)) impact on the development and validation of an AI model	Objectives Met (10) Partially Met (05) Not Met (05)		
Practical Accuracy	Come with a solution that reflects—Knowledge Representation, Causal Structure, Missing Observation, Flow of Information	Objectives Met (10) Partially Met (05) Not Met (05)		
Machine and Human Handovers	Algorithm learn to handover/defer to clinical expert in decision making and vice versa	Objectives Met (10) Partially Met (05) Not Met (05)		
Explainability				
Verifying before Deciding	Validating model's explanation against expert knowledge before taking an action	Objectives Met (10) Partially Met (05) Not Met (05)		
Building Trust in System	View the reasons behind decisions and predictions. Verify that the performance is optimum	Objectives Met (10) Partially Met (05) Not Met (05)		
Counterfactual Reasoning	Test through what-if scenarios which confirm the accuracy of its predictive power	Objectives Met (10) Partially Met (05) Not Met (05)		
New Clinical Insights	Data on a scale have access to undiscovered medical insights, like drugs and interactions	Objectives Met (10) Partially Met (05) Not Met (05)		
Overall	Combining Every aspect	Total out of 180		

Programs above 125 are automatically selected for Step 2—Data and ETL.

Acknowledgements The authors want to acknowledge the employing organization of the authors: Our organization Group. In particular, we want to thank Dr Sangita Reddy that supported the initial discussions of this paper. The authors would also like to thank Dr. Ganapathy for his very helpful comments.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Wixom B, Someh I, Beath C (2022) Building AI explanation capability for the AI—powered organization. MIT CISR research briefing, vol XXII. MIT, Boston
2. Shelmerdine SC, Martin H, Shirodkar K, Shamshuddin S, Weir-McCall JR (2022) Can artificial intelligence pass the Fellowship of the Royal College of Radiologists examination? Multi-reader diagnostic accuracy study. BMJ. https://doi.org/10.1136/bmj-2022-072826
3. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization, Geneva (2021). Licence: CC BY-NC-SA 3.0 IGO. https://www.who.int/publications/i/item/9789240029200. Accessed 28 June 2021
4. Ethics guidelines for trustworthy AI | Shaping Europe's digital future (europa.eu) (2022)
5. Schröder-Bäck P, Duncan P, Sherlaw W, Brall C, Czabanowska K (2014) Teaching seven principles for public health ethics: towards a curriculum for a short course on ethics in public health programmes. BMC Med Ethics 15:73. https://doi.org/10.1186/1472-6939-15-73
6. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH (2018) Ensuring fairness in machine learning to advance health equity. Ann Intern Med 169(12):866–872. https://doi.org/10.7326/M18-1990
7. Raji ID, Smart A, White RN, Mitchell M, Gebru T, Hutchinson B, Smith-Loud J, Theron D, Barnes P (2020) Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of the 2020 conference on fairness, accountability, and transparency, pp 33–44. https://arxiv.org/abs/2001.00973
8. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 162(1):55–63

9. Interoperability in Healthcare | HIMSS
10. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115
11. Weigl BH, Gaydos CA, Kost G, Beyette FR Jr, Sabourin S, Rompalo A, de Los ST, McMullan JT, Haller J (2012) The value of clinical needs assessments for point-of-care diagnostics. Point Care 11(2):108–113. https://doi.org/10.1097/POC.0b013e31825a241e
12. Crandall JW, Oudah M et al (2018) Cooperating with machines. Nat Commun 9:233. https://doi.org/10.1038/s41467-017-02597-8
13. Linardatos P, Papastefanopoulos V, Kotsiantis S (2021) Explainable AI: a review of machine learning interpretability methods. Entropy 23(1):18. https://doi.org/10.3390/e23010018
14. Counterfactual Reasoning—Interpretable Machine Learning (2022). https://christophm.github.io/interpretable-ml-book/counterfactual.html
15. Bohr A, Memarzadeh K (2020) The rise of artificial intelligence in healthcare applications. In: Artificial intelligence in healthcare, pp 25–60. https://doi.org/10.1016/B978-0-12-818438-7.00002-2