Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 Dec; 6 (10s): 2444 -2460

Explainable AI and ML Models for Transparent Clinical Decision Support

Sasi Kumar Kolla*

*Independent Researcher, sasikkolla@gmail.com, ORCID: 0009-0004-9397-9533

Abstract

Artificial intelligence has the potential to augment clinical decision making. By learning patterns of risk and disease directly from empirical data, AI methods offer one solution to the difficulty health care professionals face in considering ever-increasing amounts of information. Clinicians making a medical decision for a patient want not only an accurate estimate of the risks associated with their patient's disease or treatment options but also an understanding of the reasoning behind these risks. This desire for explanation drives the growing interest in explainability in AI, particularly in AI for health care. Explainable Artificial Intelligence (XAI) is defined as methods that generate new AI models for which the behaviour can be understood, directly or indirectly, by humans. The concept of human understanding encompasses three different levels – transparency, interpretability and accountability. The heart of the concern for transparency in AI is the incomprehensibility of the learned representations, the "black box" nature of the complex function learned from the training data.

Keywords: Artificial Intelligence in Healthcare, Clinical Decision Support Systems, Explainable Artificial Intelligence (XAI), Model Transparency in Medicine, Interpretability of Machine Learning Models, Accountability in Clinical AI, Risk Prediction and Prognostic Modeling, Black-Box Model Limitations, Human-Centered AI Design, Trustworthy Medical AI, Ethical AI in Healthcare, Model Explainability Frameworks, Data-Driven Clinical Risk Assessment, AI Governance in Health Systems.

1. INTRODUCTION

Over the past decade, the performance of machine-learning methods in supervised classification tasks has improved significantly, achieving human-comparable performance on specific benchmarks. Subsequently, both academic researchers and industry apply these techniques to diverse application areas. One major area for application is health care, supported by advances in cheap data storage and processing, the digitization of patient histories, and the collection of large groups of clinical data.

However, these machine-learning methods do not provide any insight into how predictions are made and using them as a decision-making tool seems to be inconsistent with the principles of evidence-based medicine. CLINICIAN-ASSISTANT DECISION-SUPPORT TOOL Clinical decision support models support the medical staff during the examination and diagnosis of patients. Diagnostic support applies to disease detection based on patient data. Prognostic support predicts the chance of morbidity and mortality. Decision support determines the preferred therapy for the patient with certain diseases such as acute coronary syndrome. For simple and small medical datasets, traditional statistical methods (e.g., logistic regression, Cox regression) are still appropriate. However, the datasets are becoming larger and more complex, so that data-driven methods based on the machine-learning paradigm, with no or few a priori assumptions about the relationship between input and output, are replacing traditional statistical methods. However, these black-box models for clinical decision support may be inconsistent with the principles of evidence-based medicine, which rely on transparent reasoning by expert clinicians. Therefore, validation of ML models needs to estimate clinical relevance.

1.1. Overview of Explainable AI in Healthcare Context

Decision support systems are increasingly common across many healthcare domains, yet many use complicated machine learning models that are treated as a black box. For clinicians and patients to rely on these decision-support systems, there must be assurance that the developed models are indeed trustworthy. The subfield of Explainable Artificial Intelligence (XAI) investigates methods of increasing the transparency of these black boxes. Recent literature presents an overview of Explainable AI in the context of healthcare, along with a set of definitions specifically tailored for application in medical settings.

Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 Dec; 6 (10s): 2444 -2460

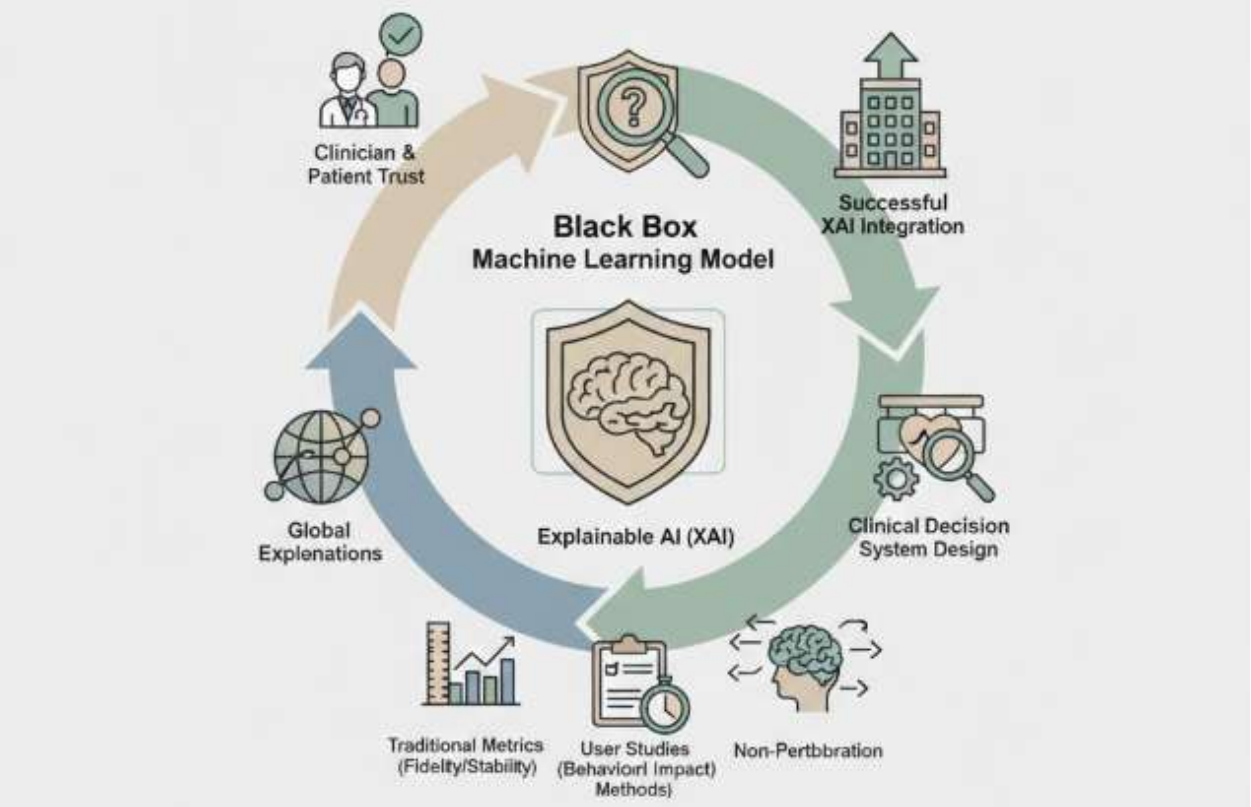
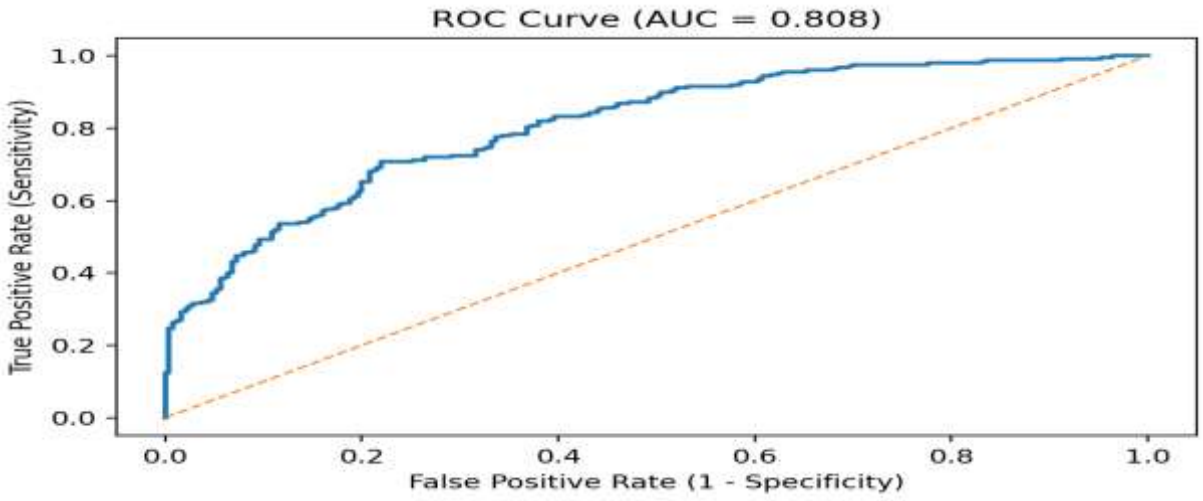


Fig 1: Beyond the Black Box: A Multi-Dimensional Framework for Evaluating Explainable AI (XAI) in Clinical Decision Support Systems

XAI researchers focus on model-agnostic perturbation-based approaches that are capable of providing either global or local explanations for a multitude of machine-learning model types. Evaluation of a given explanation can occur through traditional metrics (e.g., fidelity and stability), through user studies measuring the effects of explanations on clinician or patient behavior, or through examining the relationship between such explanations and non-perturbation methods of clinical prediction. Although user studies exploring the influence of XAI methods on real clinical decision-making are still largely absent, there are numerous studies examining the evaluation of explanations in clinical use cases. Such assessments are crucial to ensuring the future success of XAI and also demonstrate the importance of incorporating explanations into the design of clinical decision-support systems.



2. FOUNDATIONS OF EXPLAINABLE ARTIFICIAL INTELLIGENCE IN HEALTHCARE

Two complementary and interrelated subjects are foundational to Explainable AI in general, and its application in

healthcare in particular: definitions of explainability presented in the natural language processing context and a more comprehensive definition of Explainable AI that incorporates them in a larger perspective from the multi-dimensional field of machine learning. Choudhury's definitions are based on a qualitative analysis of 100 natural language processing papers that explicitly use the term explainability. The latter is built on existing definitions of transparency, interpretability, and accountability in the context of machine learning and provides a foundation for Explainable AI in other domains, including clinical-decision support systems based on machine learning and deep-learning. A discussion of their application in healthcare follows, focusing on transparency, interpretability, and accountability.
The term explainability is seldom explicitly defined in AI. One of its few definitions is provided by Choudhury, who uses the term explicitly as a synonym for the term justify. A second definition states that an explanation is a mapping from a complex function that approximates the function's behavior in a local region, Actual explanations generally do not achieve complete or perfect explanations. A general property of deep-learning models is that they perform well but are "black boxes." They can handle all types of data to make highly accurate predictions but do not provide any insights into how they make decisions. They can be viewed as complex pieces of software that convert an input into an output without revealing the logic of how the input is transformed into the output.

Equation 1) Logistic regression (risk/probability prediction)

1.1 Model equation

Let $x \in \mathbb{R}^d$ be features (labs, vitals, etc.), $y \in \{0,1\}$ be outcome.

Linear score

$$z = \beta_0 + \beta^\top x$$

Sigmoid (logistic) mapping to probability

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Predicted risk

$$p(x) = \Pr(y = 1 \mid x) = \sigma(\beta_0 + \beta^\top x)$$

1.2 Why the logit is linear (step-by-step)

Start with odds:

$$\text{odds}(x) = \frac{p(x)}{1 - p(x)}$$

Take log:

$$\log\left(\frac{p(x)}{1 - p(x)}\right) = \beta_0 + \beta^\top x$$

Exponentiate both sides:

$$\frac{p(x)}{1 - p(x)} = e^{\beta_0 + \beta^\top x}$$

Solve for p(x):

$$p(x) = \frac{e^{\beta_0 + \beta^\top x}}{1 + e^{\beta_0 + \beta^\top x}} = \frac{1}{1 + e^{-(\beta_0 + \beta^\top x)}}$$

1.3 Estimation via maximum likelihood (step-by-step)

Assume independent observations $\{(x_i, y_i)\}_{i=1}^n$

Bernoulli likelihood for each i:

$$\Pr(y_i \mid x_i) = p_i^{y_i}(1 - p_i)^{1 - y_i}, \quad p_i = \sigma(\beta_0 + \beta^\top x_i)$$

Total likelihood:

Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 Dec; 6 (10s): 2444 -2460

$$L(\beta_0, \beta) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1-y_i}$$

Log-likelihood:

$$\ell(\beta_0, \beta) = \sum_{i=1}^{n} [y_i \log p_i + (1 - y_i) \log(1 - p_i)]$$

Negative log-likelihood (a common loss to minimize):

$$\mathcal{J}(\beta_0, \beta) = -\ell(\beta_0, \beta)$$

2.1. Definitions and Scope

Diverse forms of transparency and explainability are required for successful AI systems functioning in different domains. Grounded in human-centered design principles for artificial intelligence, Explainable AI (XAI) is framed in this context as Artificial Intelligence Designed for Human Explanatory Needs, where users of AI-supported appliances and agents are imbued with a variety of human-centered explanatory needs by considerations of practical action: justification, understanding, learning, and reliability. AI is regarded as fundamentally different from other computer systems, such as graphical user interfaces with no rational, goal-oriented functionality. Accordingly, the comprehension and justification requirements of humans relying on AIs not only differ in focal contents but also involve the added dimension of human-centered design of the AI accounts for explaining AI-supported decisions or action recommendations, in a way that satisfies the explanatory needs of the AI users. Given the historical lack of distance of AI constructions and AI-supported decisions from such human-oriented explanatory points of view, AI models have intuitively been understood as opaque and largely interpreted as “black boxes” making unreliable decisions.

In particular, whereas natural language processing considering the explanatory needs of human subjects has produced much novel research on AI-based systems and their various underlying explanatory accounts or communicating vehicles, their success for core AI areas, such as vision and reasoning, has languished. Thus, the need for deep rational, functional understanding of explainability for core AI functionalities and systems couples an applied area—illustrating the significance of its requisite breadth of explanatory considerations and sources—with an underpinning conceptual framework for AI’s distinctive nature as an agent in need of a structural explanatory rationale. Further, the caution that XAI is a very broad area and the distinction between supporting machines with explainable models for human users and XAI systems with deep explanatory, advisory function for supporting users also resonate.

Concept	Core equation (symbolic)	In clinical decision support
Cox PH	h(t|x)=h0(t)exp(β^{T}x)	Dynamic survival/risk with covariates
SHAP	φi = Σ_{S⊆F\{i}} |S|!(M−|S|−1)!/M! [f(S∪{i})−f(S)]	Local feature contributions that add up to prediction
Surrogate fidelity	Fid = 1 − MSE(f,g) / Var(f)	How close surrogate g is to black-box f
Stability	Stab = E[||φ(x) − φ(x+δ)||]	Sensitivity of explanation to small perturbations

2.2. Core Concepts: Transparency, Interpretability, and Accountability

Many terms are associated with Explainable AI: explainable, interpretable, transparent, understandable, justifiable, trustworthy, and accountable. It is important to clarify definitions and avoid using terms ,interchangeably, particularly as the community works to define standards for AI developed for diagnoses in radiology, cardiology, pathology, dermatology, and other medical specialties without any question-able or biased model decisions.

Transparency refers to the understanding of the inner workings of a model. For example, statistical methods like logistic regression, support vector machines, and classification trees are more transparent than ensemble methods like random forests and gradient-boosting. The understanding of risk factors or paths to clinical events is less obvious for models developed with deep learning. Transparency can be improved with model-agnostic tools for clinical decision support and with methods like SHAP for risk prediction models. Transparency, however, does not imply that errors, false results, and unexpected results are less likely. Being a black box remains a characteristic of AI in general, and of explainable methods

of adapted and original neural-net models in particular.
Interpretation refers to the comprehension of model decisions by end users. This is pivotal in building trust and support for the adoption of AI in clinical practice. Model decisions can be justified and made understandable, but the models cannot be understood and explained by the end user, even when presented with visual aids. This aspect is being increasingly addressed by the development of novel interactive visual tools exploring embedding, quality assessment, uncertainty-aware and user-centered explanations with a focus on improving ML-augmented Davinci Test accuracy and understanding.
Accountability implies that model outcomes can be justified and are reliable. The cognitive appraisal of a model decision by the end user does not guarantee that such a model is reliable and accurate.

Equation 2) ROC curve and AUC (discrimination evaluation)

2.1 Confusion matrix rates (step-by-step)
Choose a threshold t. Predict $\hat{y} = 1$ if $p(x) \geq t$, else 0.

- True Positives: *TP(t)*
- False Positives: *FP(t)*
- True Negatives: *TN(t)*
- False Negatives: *FN(t)*

Then:

$$\text{TPR}(t) = \frac{TP(t)}{TP(t)+FN(t)} \quad \text{(Sensitivity)}\ \text{FPR}(t) = \frac{FP(t)}{FP(t)+TN(t)} = 1 - \text{Specificity}(t)$$

2.2 ROC curve
ROC is the parametric curve:

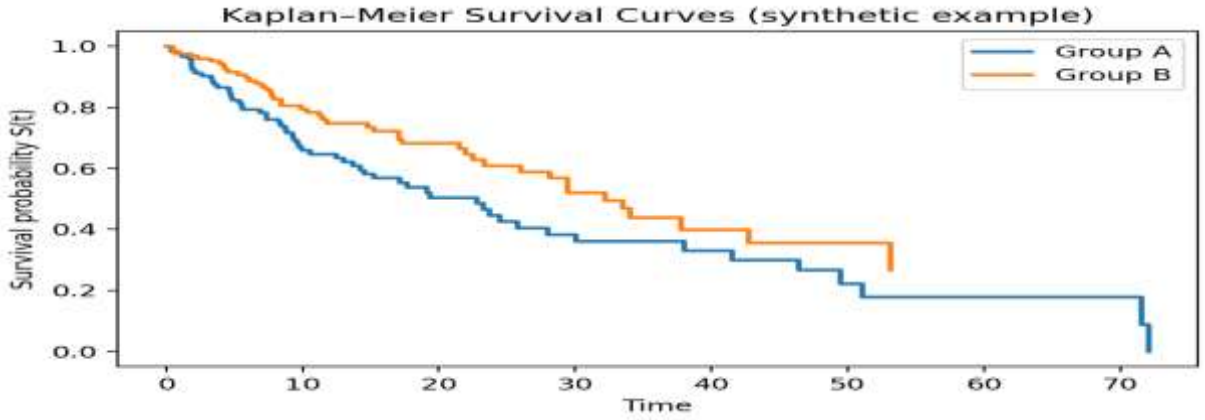
$$\big(\text{FPR}(t),\ \text{TPR}(t)\big) \quad \text{as } t \text{ varies from } 1 \to 0$$

2.3 AUC (area under ROC)
Continuous form:

$$\text{AUC} = \int_0^1 \text{TPR}\ \ d(\text{FPR})$$

Discrete approximation (trapezoids) for points (f_k, r_k) :

$$\text{AUC} \approx \sum_k \frac{(f_{k+1} - f_k)(r_{k+1} + r_k)}{2}$$



3. MACHINE LEARNING MODELS FOR CLINICAL DECISION SUPPORT

Decisions regarding a patient's diagnosis, prognosis, or treatment require the integration of clinical data with healthcare professionals' vast medical knowledge and heuristics, honed by education, training, and experience. These decisions are multifaceted and complex in nature. Clinicians often face numerous possible alternatives for a single decision, needing to

determine which option is correct, although usually based on a routine approach. In the face of increased patient complexity, time constraints and familiarity with technology, clinicians may utilize algorithms that predict clinical events and help guide their decisions. The CONCEPT system, which predicts the onset of complications in heart surgery, and ARTEMIS, which predicts acute renal failure, are two well-known examples of decision-support systems that use statistical modeling techniques to provide corrective guidelines for clinicians.

Different machine-learning techniques can be applied depending on the health event being predicted. In general, when there are few patients but many variables (as with various types of cancer) statistical modeling techniques are indicated. If there are many patients but few variables (as when predicting hypoglycemic events in diabetes), machine-learning methods such as decision trees or neural networks may perform better The most suitable algorithm should be chosen according to the clinical situation, considering not only accuracy but also other metrics such as the time needed for design and execution of the algorithm.

Statistical models rely on conventional clinical data input. However, in some clinical situations, these variables alone may not capture all the clinical complexity. Involvement of authors with both clinical and computational expertise enables more advanced data mining, leading not only to conventional categorical variables but also to the synthesis of algorithms that can receive both clinical and procedural data.

3.1. Traditional Statistical Methods in Clinical Decision Support

For decades, methods from statistical modeling have been widely used for developing models that help clinicians in their day-to-day work. In clinical medicine, a classic question is whether worsening function—with respect to a measured marker—or increasing levels of risk factors for a disease are predictive of incident disease during a specified period. Assessing the validation of such predictions is critical, particularly whether a specified level of a marker (a cutoff point) can be used in a logistic regression context in which the probability of reaching a disease threshold at the end of the prediction window is substantially different from the average probability of disease incidence in the population. The conventional approach to assessing this is by means of ROC curves.

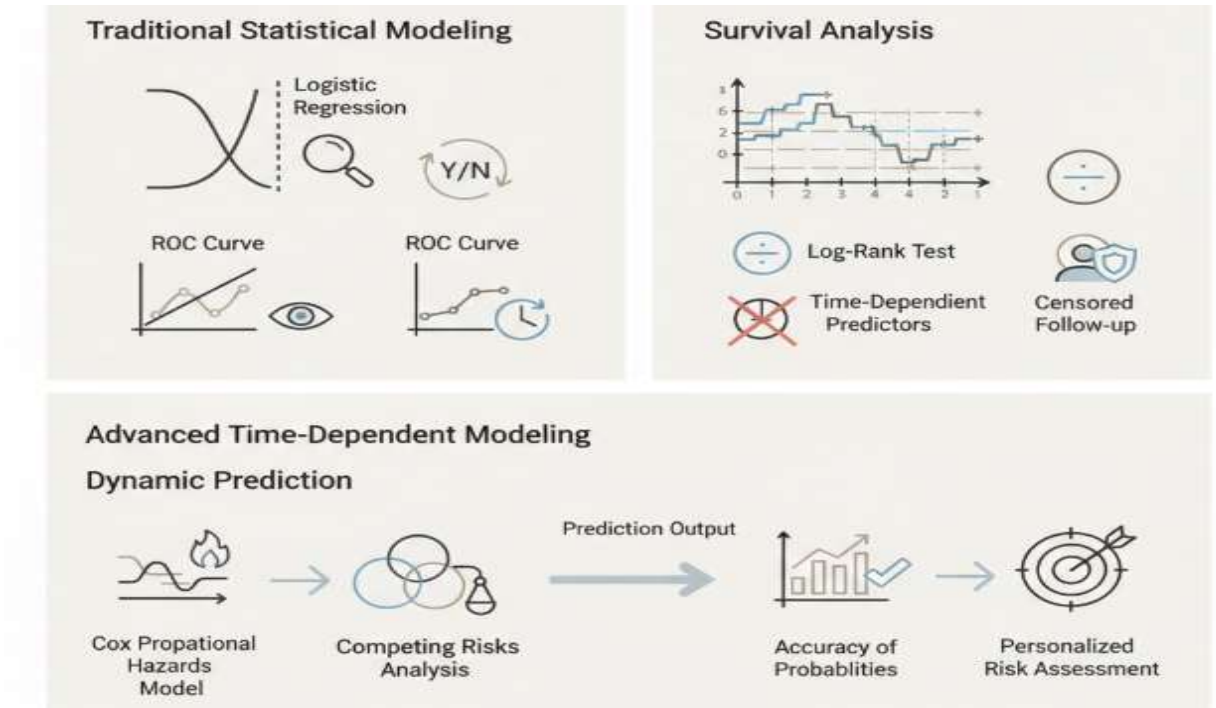


Fig 2: Advancing Clinical Prognostics: Evaluating Dynamic Prediction Models and Time-Dependent ROC Analysis in Survival Functions

Another standard technique for such analyses is the Kaplan-Meier plot with the log-rank test for k-fold stratification of the survival function with respect to a categorical description of the time-dependent variable. However, background risk stratification and covariate adjustment are not able to incorporate time-dependent predictors. Dynamic predictions of time-dependent survival probabilities can be obtained based on the Cox proportional hazards model or by using the cause-specific hazards in presence of competing risks. Validation can be accomplished with time-dependent ROC curves or through an assessment of the accuracy of the predicted probabilities. However, proposed stratifications and predictions using traditional statistical modeling remain limited because cessation of follow-up is due to reasons unrelated to the disease at hand.

3.2. Modern Machine Learning Approaches

Modern machine learning models, like their predecessors, receive training from data sets containing combinations of input features and corresponding outcomes. They then learn the relationships between input and output that are evident in the training data and apply these learned relationships on new data for which the outcome is unknown. However, while traditional statistical methods have often been based on parametric assumptions, and more recent non-parametric methods such as classification trees automatically reduced the number of considered input features through pruning, current machine learning models are more complex. They include neural networks, support vector machines, ensemble methods that combine many base learners such as random forests and gradient-boosted trees, and more. Some newer methods have also introduced concomitant novel approaches to speed up training by approximating a more complex model with a simpler one, more rapidly test a model's predictive ability using a limited portion of the data, or identify sub-groups of patients on whom a complex model can provide transparent predictions.

At times, a model that provides more accurate predictions on a test set is preferred without concern for whether it is considered interpretable. The rationale behind transparent machine learning has also diverged somewhat, differing from classical statistical reasoning focused on the value of understanding the key factors influencing predictions and of ensuring that predictions are reliable. A wider range of factors may contribute to the predictive accuracy of machine learning models than survival data alone and applying an accurate black-box model with faith should be sufficient. However, the desire for an accurate black-box model also leads to efforts supporting interpretability of machine learning models—these in addition being justified from a regulatory compliance perspective, the important need for users to develop appropriate trust in predictions or decisions made based on them, and the value of explanatory tools for debugging and testing prediction models.

Equation 3) Kaplan–Meier survival function and log-rank test

3.1 Kaplan–Meier estimator (step-by-step)

Let event times be ordered $t_{(1)} < t_{(2)} < \cdots$ At each event time $t_{(j)}$:

n_j: number at risk just before $t_{(j)}$

d_j: number of events at $t_{(j)}$

Conditional survival past $t_{(j)}$, given survival to just before it:

$$\Pr\left(T > t_{(j)} \mid T \geq t_{(j)}\right) = 1 - \frac{d_j}{n_j}$$

Multiply conditionals up to time t:

$$\hat{S}(t) = \prod_{t_{(j)} \leq t} \left(1 - \frac{d_j}{n_j}\right)$$

3.2 Log-rank test (step-by-step)

Compare two groups $g \in \{1,2\}$.. At each event time $t_{(j)}$:

n_{1j}, n_{2j}: at-risk counts

d_{1j}, d_{2j}:event counts

$n_j = n_{1j} + n_{2j}, d_j = d_{1j} + d_{2j}$

Expected events in group 1 under "same hazard" null:

$$E_{1j} = d_j \frac{n_{1j}}{n_j}$$

Aggregate:

$$O_1 = \sum_j d_{1j}\,, \quad E_1 = \sum_j E_{1j}$$

Variance (common form):

$$V_1 = \sum_j \frac{n_{1j} n_{2j} d_j \left(n_j - d_j\right)}{n_j^2 \left(n_j - 1\right)}$$

Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 Dec; 6 (10s): 2444 -2460

Test statistic:

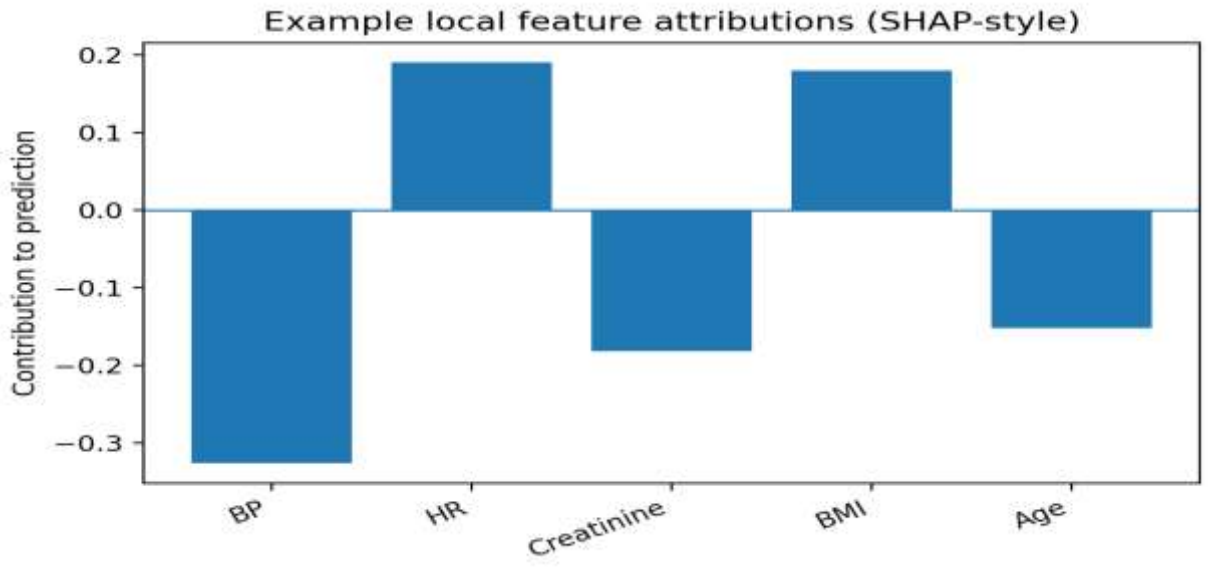
$$\chi^2 = \frac{(O_1 - E_1)^2}{V_1}$$

Under H_0, $\chi^2 \sim \chi^2_{(1)}$approximately.

4. TECHNIQUES FOR EXPLAINABILITY IN CLINICAL MODELS

The data-driven nature of clinical prediction models using machine learning techniques raises the question of whether these complex, high-dimensional models could be faithfully used to support clinical decision-making in a transparent manner. A distinction can be made between global and local predictions. Global approaches provide an understanding of the relationships between inputs and predictions, while local approaches highlight the main drivers for individual predictions. Furthermore, global or local interpretations can be embedded in or added on to any machine learning model using model-agnostic methods.

In many application areas, formulating a prediction model as a black box having only input–output pairs without any explanation of the inner mechanics of the model is accepted. In health care, however, a prediction from a model should be based on well-established and validated biological or physiological theories. Therefore, the efforts of the medical community to analyze the behavior of machine learning models, seeking to uncover the reasons behind predictions and ensuring that the predictions follow the underlining medical knowledge, are of particular importance. In particular, methods allowing clinicians to understand the reasons behind a single prediction, that is the local behavior of the models, have recently drawn interest.



4.1. Global Versus Local Interpretability

In an ideal scenario, clinicians could trust clinical-machine learning models as they do traditional statistical methods. Models would ideally provide global representations revealing their structure, overall behaviour, and reasoning processes. However, from experience, clinicians know that ML models can demonstrate behaviours that lead to unexpected or counterintuitive results, such as exploiting spurious correlations that might reflect noise in the training dataset. This limitation has pushed many researchers in the field to propose methods that improve local interpretability—the understanding of how features contribute to specific predictions. The most commonly used methods are based on surrogate models (e.g., locally weighted regression), Shapley values, and saliency maps.

Surrogate models express the prediction function learned by an ML model using a more interpretable one (e.g., a decision tree) trained on the predictions of the complex model. The use of such methods, however, is problematic. First, the surrogate model can misrepresent the internal logic of the complex model if the global behaviour of the latter—reflected in the loss function minimization—differs substantially from the loss function of the interpretable model (e.g., MSE for a decision tree). Also, if the complex model fails to learn a suitable interpretation, as in the case of associated-rule classifiers, the explanation would not be valid. Secondly, generalization is at the centre of building model-agnostic locality-aware interpretable methods. The most popular methods for interpreting ML models are based on Shapley values and saliency maps. The first approach explicitly incorporates the notion of fairness in interpretation. Saliency maps, on

the other hand, provide information about how input signals affect predictions without explicitly decomposing the prediction function.

4.2. Model-Agnostic Methods

For the majority of techniques described in this review, the mechanisms used to provide explanations operate independently of the clinical model at the heart of the CDSS. These model-agnostic methods are able to explain any underlying model and their use is therefore attractive, especially when developing complex predictive models that use state-of-the-art machine learning approaches such as deep learning or ensembles of decision trees. In such cases, the explanation techniques themselves are simpler than the core model being explained, and provide insights into the predictions made by a smoothed or a lower-dimensional version of the core model.

The Shapley Additive exPlanations (SHAP) framework is one such model-agnostic approach. Based on game theory, SHAP leverages a consideration of all possible subsets of features in a given model to create a global reference point for explanation that can then be decomposed into local attributions. A key advantage is that SHAP returns local importance scores that sum to the associated prediction, and communicates feature contributions in an intuitive and human-normalised manner. One variant of SHAP uses Monte Carlo methods for estimation, rendering it scalable to large data sets.

Model	Interpretability (0-1)	Performance (AUC-like)
Logistic Regression	0.9	0.78
Decision Tree	0.75	0.8
Random Forest	0.35	0.86
Deep NN	0.2	0.9

5. VALIDATION AND EVALUATION OF EXPLAINABLE CLINICAL MODELS

Techniques and methods for answer validation and evaluation, both technical and clinical, of interpretable models are essential. The common strategy to interpret models consists of providing explanations that resemble the models without giving away their full structure and parametrization. The contrast between explainability methods on interpretable models and on black boxes is the separation of the task into technical evaluation and clinical relevance and usability studies.

Technical evaluation metrics aim at examining how close the explanations generated by those techniques are to the actual predictors of the models—an inverse setting of common interpretability metrics of black-box models. These metrics usually rely on modeled explanations of the form of mappings, probabilistic or discriminative predictions, score functions, decision functions, or anything allowing to infer the classifier's final prediction in the whole input space. Assessing how well these models can make predictions in such settings could be defined as the oracular evaluation. As the black-box approach consists of approximating a complex model with a simpler one, methods based on the difference be-tween the original predictions and the approximated ones also represent a well-known way to evaluate the quality of the explanation on black-box classifiers, being popularly used for neural networks.

Equation 4) Cox proportional hazards model (dynamic prediction)

4.1 Cox model equation (step-by-step)

Hazard is instantaneous event rate:

$$h(t \mid x) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T < t + \Delta t \mid T \ge t, x)}{\Delta t}$$

Cox assumption:

Baseline hazard h_0 (t) (unknown function of time)

Covariates act multiplicatively:

$$h(t \mid x) = h_0(t)\exp(\beta^\top x)$$

4.2 Survival from hazard (key derivation)

Cumulative hazard:

$$H(t \mid x) = \int_0^t h\,(u \mid x)\,du = \int_0^t h_0\,(u)\exp(\beta^\top x)\,du = \exp(\beta^\top x)\underbrace{\int_0^t h_0\,(u)\,du}_{H_0(t)}$$

So:

$$H(t \mid x) = \exp(\beta^\top x)H_0(t)$$

Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 Dec; 6 (10s): 2444 -2460

Survival is:

$$S(t \mid x) = \exp\big(-H(t \mid x)\big) = \exp\big(-\exp(\beta^{\top} x) H_0(t)\big)$$

5.1. Technical Evaluation Metrics

To assess the technical aspects of explainable models, consider the model's fidelity, stability, and complexity. Such attributes are well-defined criteria for any predictive algorithm, but their importance is accentuated in the context of model interpretability. Increasing a model's interpretability often lowers its predictive performance and the cost/benefit assessment should factor in its influence on fidelity.

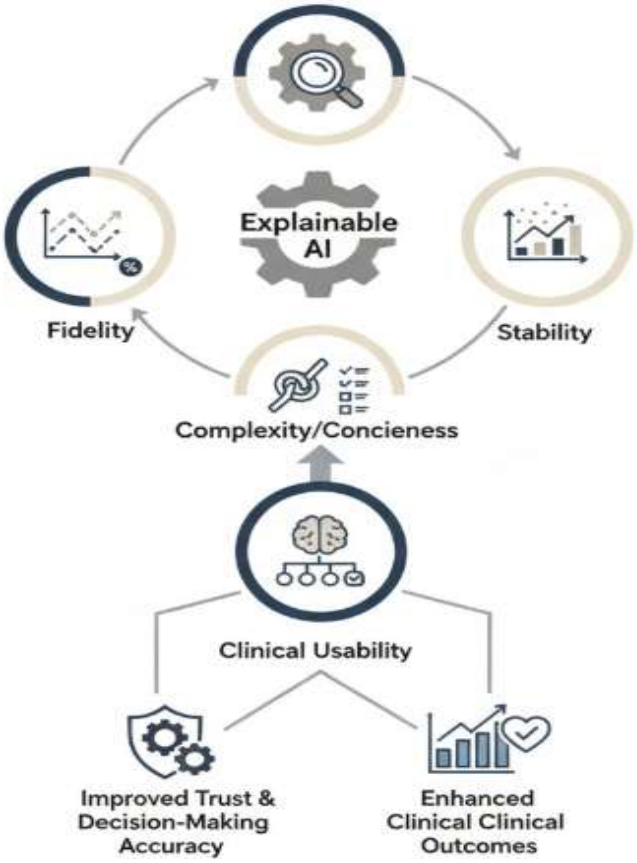


Fig 3: Evaluating Interpretability: A Multi-Dimensional Framework for Fidelity, Stability, and Clinical Usability in Explainable AI

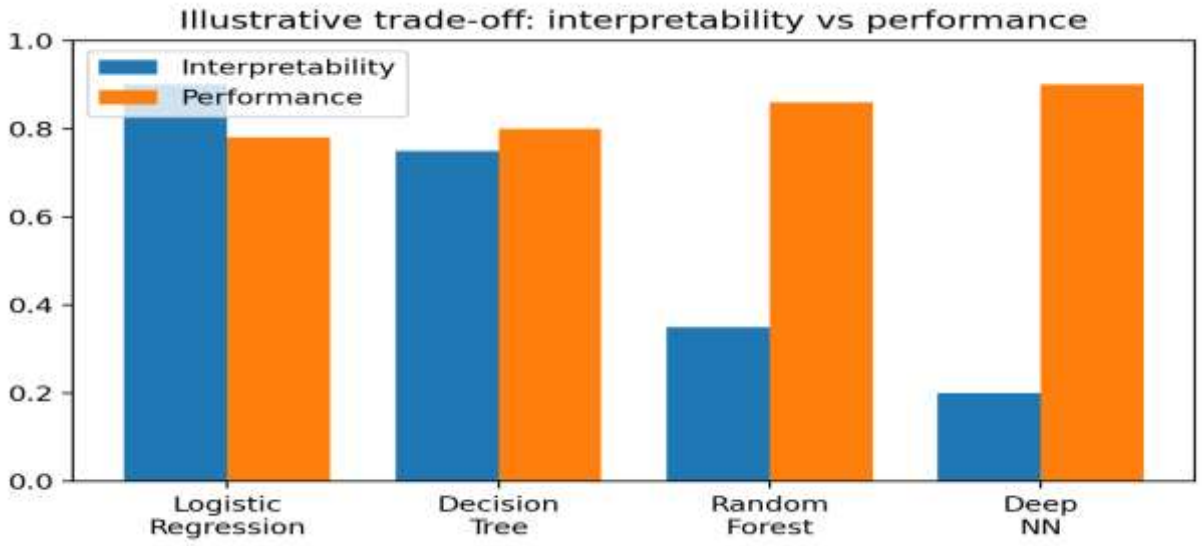
Model fidelity signifies how well the approximating model aligns with the target model's predictions. Effectively, it can be regarded as a metric suited for approaches that learn an approximation to a black-box model. Stability pertains to how much changes in the input data affect the provided explanation. High stability is required if the interpretation is intended to guide decisions, allowing the user to draw accurate insights and understand model outcomes better. Lastly, a simpler explanation is preferred over a more intricate one when all other factors are held constant, since simple explanations are generally easier to comprehend.

These three facets should ideally be minimized when evaluating a local surrogate model, and conciseness should be highlighted in the study of other forms of local explanations.Beyond these technical aspects, other validation methods are desirable in the context of clinical decision support. The ultimate test of a local explanation technique is through clinical usability studies aimed at determining whether the addition of an explanation positively influences clinical decision making.

5.2. Clinical Relevance and Usability Studies

Demonstrating the technological adequacy of explanations is insufficient for thorough validation. Explanations must also produce clinical impact when tested on real users. However, such evidence remains scarce. The justification of clinical prediction models is inconclusive and appears as a secondary task relative to prediction performance. Similarly, the authors of a usability study found that while narrative explanations increased user perception of transparency, they failed to reveal the inner workings of the model and to enhance judgement accuracy in clinical scenarios. These findings highlight the need for a systematic evaluation of human factors in clinical decision support tools.

Several user studies corroborate the importance of model-agnostic local explanations. Transparency is positively related to the perception of trustworthiness, a property that influences the intention to use the model's predictions in clinical settings. Yet, interpretability is not sufficient for trust. Although users prefer globally interpretable models, local explanations increase the transparency of otherwise opaque models and steer users' judgements toward the predictions. The lack of accountability for uninterpretable decisions, however, may produce the opposite effect. In conclusion, transparent and interpretable explanations should be regarded not as sufficient conditions for trust but as valuable facilitators.



6. IMPLEMENTATION IN CLINICAL PRACTICE

Integrating explainable clinical models with existing clinical decision support systems is key to facilitating their adoption in clinical practice. The primary applications of clinical decision support tools are the assessment of risk or prognosis, the recommendation or selection of treatment choices, and the prediction of adverse events after a surgical intervention. Applying these functionalities within the context of an EHR system naturally makes the model learnings available for use by patient care units or departments that routinely perform the same tasks, therefore increasing transparency and reliability while facilitating adoption. The required EHR influence is considerable, since changing surgical protocols or making major changes in how patient cohorts are defined and treated in clinical daily practice is exceedingly difficult.
Though the algorithms are designed for use by specifically trained clinicians, models also are being developed for use by non-specialists. These models address more common conditions and are based on data that are either trivially or not difficult to obtain. Despite the need for rigorous model testing, extensive user-centered human-factor studies are needed to ensure that the proposed systems will enhance effective decision making. Clinicians cannot afford to become data miners; hence, these recommendations are meant to assist, augment, and point out possible solutions to existing problems, enabling clinicians to focus on delivering the best possible patient treatment and care.

Equation 5) SHAP / Shapley values (local feature attributions)

5.1 Additive explanation form

For an instance *x*, SHAP explains:

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i$$

ϕ_0: base value (average prediction)
ϕ_i: contribution of feature *i*

5.2 Shapley value formula (step-by-step meaning)

Let *F* be the full feature set, $|F| = M$. For feature i, consider all subsets $S \subseteq F\backslash\{i\}$

f(S): model prediction when only features in S are "present" (others marginalized/perturbed)
Marginal contribution of adding *i* to subset *S:*

$$\Delta_i(S) = f(S \cup \{i\}) - f(S)$$

Weight each subset fairly by number of orderings:

$$w(S) = \frac{|S|!\ (M - |S| - 1)!}{M!}$$

Then:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} w\,(S)\,\Delta_i(S)$$

6.1. Integration with Electronic Health Records and Clinical Workflows

The primary goal behind the integration of clinical decision support tools with electronic medical records is to replicate classic regression-based decision-support tools such as Acute Physiology and Chronic Health Evaluation (APACHE) and Sequential Organ Failure Assessment (SOFA) scales. However, a straightforward "plug-and-play" of these tools inside electronic health records exposes serious risks, as they were primarily evaluated using "expert predictions" and not intended for direct use at the point of care. Automated tools for electronic health records populated with clinical data offer an opportunity to better validate those models and provide direct clinical value through clinical alerts and recommendations tailored to specific patient contexts. However, proactively and continuously monitoring risk and estimating physiological values in real time for an individual patient triggers the question of explainability and interpretability by design.

User-centered design of clinical decision-support tools requires the clinical state and the predicted behavior of the learning algorithm to be explicitly available to the user. These requirements are crucial for guiding additional data collection, making informed treatment decisions, and ultimately achieving a satisfactory recovery. While these algorithms can learn highly specialized rules based on historical data that cannot be easily interpreted by the algorithm itself, these explanations could be crucial for their continuous "training" to improve their reliability. A counter-argument against interpretability suggests that a system can be used as a black box as long as the risk of failure is low and its use leads to benefits greater than costs. However, this position raises serious ethical issues when considering real-time risk monitoring, as the reasons for the system's predictions impact decisions on the care of high-risk patients.

6.2. User-Centered Design for Clinicians and Patients

User-centered design principles should guide two key aspects of explainable AI-enabled decision-support systems: user interface design for clinicians and the provision of accessible explanations for patients. User-friendly graphical interfaces facilitate the integration of predictive and prescriptive models into the electronic health-record workflow and help clinicians make better use of generated predictions. Implementation studies for enterprise-class systems with millions of users worldwide report improved risk communication through a natural-language, evidence-based risk-assessment tool with a surgeon-facing interface. Optimizing the transparency of explanation and prediction generation also improves clinical acceptance. Providing patients with explanations tailored to their background knowledge and it enables process transparency, allowing for informed consent and enhancing patient-physician trust.

To be effective and ethically sound, an AI-enabled system must also address patient-centered and health equity needs. OpenXAI, a recently proposed design framework, provides a roadmap for centering equity in detailed explanations. Empirical studies demonstrate that personalized explanations inducing excitement and lowering anxiety promote healthy behavior. Following the precautionary principle, however, explanations must not portray AI as a supremely capable entity. Explanations that accurately reflect AI's capabilities and appropriately express reasonable expectations positively affect attitude and behavioral intention toward AI. Patient perception of AI capability is also critical for its acceptance, though the depth of explanation needed varies by patient type: expert patients prefer deeper details to satisfy their curiosity, while non-expert patients prefer high-level explanations that aid understanding without psychological burden.

7. CONCLUSION

Explainable AI in healthcare is an important and expanding area of research, whose long-term goal is to build transparent predictive models for clinical decision support. Such models must be interpretable to end users—including healthcare professionals as well as patients—to facilitate agent-in-the-loop decision making and increase clinical trustworthiness. Interpretability must therefore be aligned with, and reflect, the goals of the end users of clinical decision-support systems. User-centered studies have shown that not only clinicians but also patients value explanations of predictions made by AI and ML systems. Valid explanations can thus improve trust toward these systems, which supports their integration in the healthcare workflow. Nevertheless, it remains a research challenge to link the needs of end users to intuitive medical explanations.

Explainable AI in healthcare is also defined, in a narrower sense, as a rigorous probabilistic framework to justify any AI algorithm from first principles. It formulates the explainability of clinical prediction models using three basic concepts: transparency, interpretability, and the ability to provide proof of any result obtained. Transparency refers to the built-in capability of a model to be questioned and understood, which can be achieved by using simple and global models combined with fair-validation concepts. Interpretability is the extent to which concerned individuals can understand the reasons behind a decision made by a AI system. The ability to provide proof of the predictions made by such models is a formal way to build responsibility and trust toward the final results. The goal is to build clinical prediction-support systems in the medical-healthcare sector that comply with the laws of both Mother Nature and society.

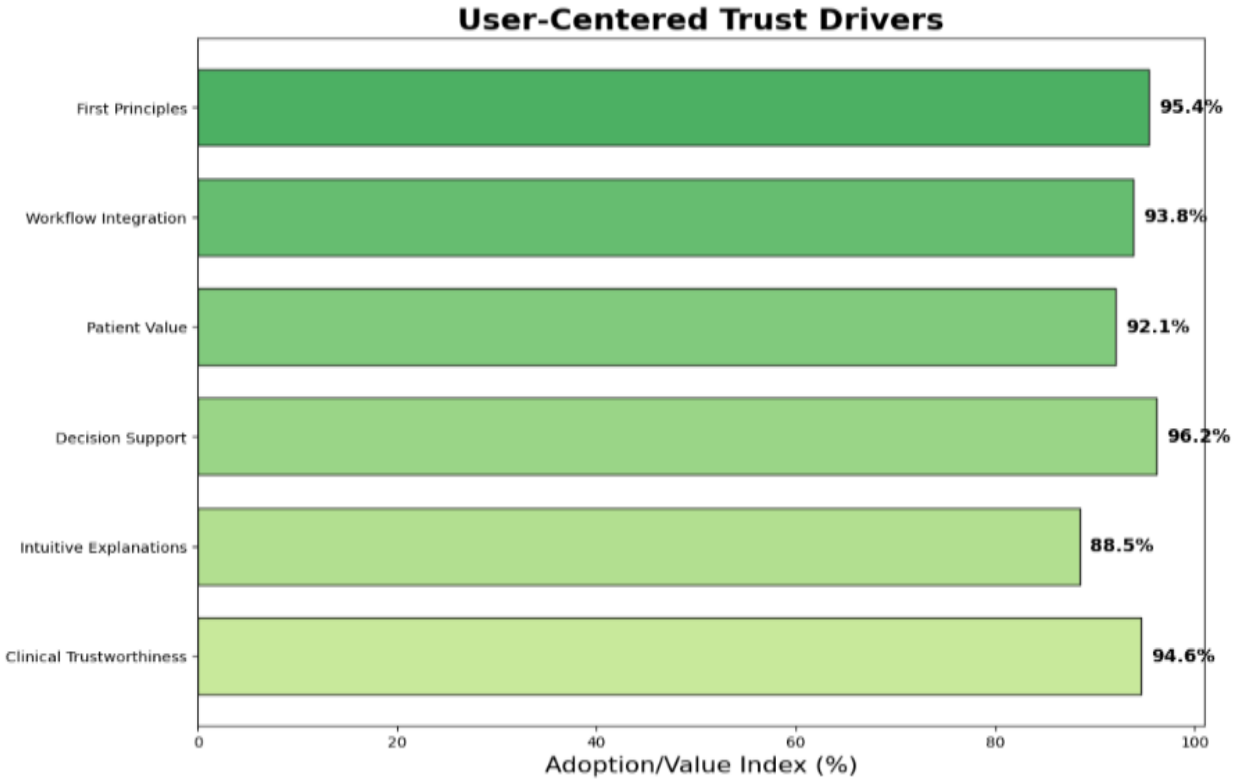


Fig 4: User-Centered Trust Drivers

7.1. Summary and Future Directions in Explainable AI for Healthcare

Explainable AI and ML are rapidly gaining traction within the healthcare domain as an essential aspect of AI research because of their ability to provide understanding of how decisions are made by algorithms when determining a diagnosis or predicting clinical outcomes. Identification of these factors can have an impact on the clinical workflow, as insights can support clinical decision-making in an explainable, comprehensible manner that aids in testing and validating diagnoses. The rise of the Explainable AI concept can partly be attributed to the growth of Deep Learning models, which, by virtue of their size and complexity, are severely criticized for their "black-box" nature. Automated decisions made without explanation are difficult for end users to accept. When a decision must be called, the user wants to know why that decision was made.

Nonetheless, this concern is not exclusive to Deep Learning. Although traditional statistical models are relatively simple, they are often treated as black-boxes without adequate attention for understanding the decision process, and assessing algorithm performance is usually reduced to a single number, for instance, an AUC value. X AI is concerned with characterizing an algorithm's behavior and decisions, identifying important input variables, and knowing how the model responds to them. By these means, the goal is to apply Explainable X AI techniques to produce machine learning models for clinical decision support that are comprehensible for users and where the reasoning behind individual or class predictions can be automatically inferred.

8. REFERENCES

1. Adadi, A., & Berrada, M. (2020). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 8, 52138–52160.
2. Goutham Kumar Sheelam, Hara Krishna Reddy Koppolu. (2022). Data Engineering And Analytics For 5G-Driven Customer Experience In Telecom, Media, And Healthcare. Migration Letters, 19(S2), 1920–1944. Retrieved from

https://migrationletters.com/index.php/ml/article/view/11938
3. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges. Information Fusion, 58, 82–115.
4. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318.
5. Meda, R. (2023). Intelligent Infrastructure for Real-Time Inventory and Logistics in Retail Supply Chains. Educational Administration: Theory and Practice.
6. Biecek, P., & Burzykowski, T. (2021). Explanatory Model Analysis. CRC Press.
7. Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability. Electronics, 8(8), 832.
8. Kushvanth Chowdary Nagabhyru. (2023). Accelerating Digital Transformation with AI Driven Data Engineering: Industry Case Studies from Cloud and IoT Domains. Educational Administration: Theory and Practice, 29(4), 5898–5910. https://doi.org/10.53555/kuey.v29i4.10932
9. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. Journal of the Royal Society Interface, 15(141), 20170387.
10. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint.
11. Koppolu, H. K. R., Sheelam, G. K., & Komaragiri, V. B. (2023). Autonomous Telecommunication Networks: The Convergence of Agentic AI and AI-Optimized Hardware. International Journal of Science and Research (IJSR), 12(12), 2253-2270.
12. Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable AI in health care. The Lancet Digital Health, 3(11), e745–e750.
13. Davuluri, P. N. Integrating Artificial Intelligence into Event-Driven Financial Crime Compliance Platforms.
14. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1312.
15. Guntupalli, R. (2023). Optimizing Cloud Infrastructure Performance Using AI: Intelligent Resource Allocation and Predictive Maintenance. Available at SSRN 5329154.
16. Jiang, F., Jiang, Y., Zhi, H., et al. (2017). Artificial intelligence in healthcare. Stroke and Vascular Neurology, 2(4), 230–243.
17. Avinash Reddy Aitha. (2022). Deep Neural Networks for Property Risk Prediction Leveraging Aerial and Satellite Imaging. International Journal of Communication Networks and Information Security (IJCNIS), 14(3), 1308–1318. Retrieved from https://www.ijcnis.org/index.php/ijcnis/article/view/8609
18. Johnson, A. E. W., Stone, D. J., Celi, L. A., & Pollard, T. J. (2021). MIMIC-IV. Scientific Data, 8, 257.
19. Gottimukkala, V. R. R. (2021). Digital Signal Processing Challenges in Financial Messaging Systems: Case Studies in High-Volume SWIFT Flows.
20. Kundu, S. (2021). AI in medicine must be explainable. Nature Medicine, 27, 1328.
21. Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36–43.
22. Varri, D. B. S. (2022). A Framework for Cloud-Integrated Database Hardening in Hybrid AWS-Azure Environments: Security Posture Automation Through Wiz-Driven Insights. International Journal of Scientific Research and Modern Technology, 1(12), 216-226.
23. [23] Molnar, C. (2022). Interpretable machine learning (2nd ed.). Lulu.
24. [24] Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. Digital Signal Processing, 73, 1–15.
25. [25] AI Powered Fraud Detection Systems: Enhancing Risk Assessment in the Insurance Sector. (2023). American Journal of Analytics and Artificial Intelligence (ajaai) With ISSN 3067-283X, 1(1). https://ajaai.com/index.php/ajaai/article/view/14
26. [26] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm. Science, 366(6464), 447–453.
27. [27] Nagubandi, A. R. (2023). Advanced Multi-Agent AI Systems for Autonomous Reconciliation Across Enterprise Multi-Counterparty Derivatives, Collateral, and Accounting Platforms. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 653-674.
28. [28] Rudin, C. (2019). Stop explaining black box machine learning models. Nature Machine Intelligence, 1, 206–215.
29. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (2019). Explainable AI: Interpreting, explaining and visualizing deep learning. Springer.
30. Amistapuram, K. (2022). Fraud Detection and Risk Modeling in Insurance: Early Adoption of Machine Learning in Claims Processing. Available at SSRN 5741982.
31. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR. IEEE Journal of Biomedical and Health Informatics, 22(5), 1589–1604.
32. Garapati, R. S. (2023). Optimizing Energy Consumption in Smart Build-ings Through Web-Integrated AI and Cloud-Driven Control Systems.
33. Sittig, D. F., & Singh, H. (2016). A socio-technical approach. Journal of the American Medical Informatics

Association, 23(4), 641–647.
34. Meda, R. (2023). Developing AI-Powered Virtual Color Consultation Tools for Retail and Professional Customers. Journal for ReAttach Therapy and Developmental Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2), 3577.
35. Topol, E. (2019). Deep medicine. Basic Books.
36. Van der Schaar, M., Alaa, A. M., Floto, A., et al. (2021). How machine learning can help healthcare systems. Machine Learning, 110(1), 1–20.
37. Wiens, J., Saria, S., Sendak, M., et al. (2019). Do no harm. Nature Medicine, 25(9), 1337–1340.
38. Aitha, A. R. (2023). CloudBased Microservices Architecture for Seamless Insurance Policy Administration. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 607-632.
39. Wehbe, R. M., et al. (2021). Deep learning in clinical NLP. Journal of the American Medical Informatics Association, 28(2), 1–15.
40. Varri, D. B. S. (2023). Advanced Threat Intelligence Modeling for Proactive Cyber Defense Systems. Available at SSRN 5774926.
41. Zhou, S., et al. (2023). A survey of explainable artificial intelligence in healthcare. Artificial Intelligence in Medicine, 138, 102473.
42. Unifying Data Engineering and Machine Learning Pipelines: An Enterprise Roadmap to Automated Model Deployment. (2023). American Online Journal of Science and Engineering (AOJSE) (ISSN: 3067-1140) , 1(1). https://aojse.com/index.php/aojse/article/view/19
43. Choudhury, A., & Naumann, F. (2022). Interpretable ML in healthcare. IEEE Access, 10, 104541–104557.
44. McCradden, M. D., Joshi, S., Anderson, J. A., et al. (2020). Patient safety and quality. npj Digital Medicine, 3, 1–5.
45. Gottimukkala, V. R. R. (2023). Privacy-Preserving Machine Learning Models for Transaction Monitoring in Global Banking Networks. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 633-652.
46. Björck, J., et al. (2021). Neural networks with monotonicity constraints. Proceedings of ICML.
47. Caruana, R., et al. (2015). Intelligible models for healthcare. Proceedings of KDD, 1721–1730.
48. Koh, P. W., & Liang, P. (2017). Influence functions. Proceedings of ICML, 1885–1894.
49. Meda, R. (2023). Data Engineering Architectures for Scalable AI in Paint Manufacturing Operations. European Data Science Journal (EDSJ) p-ISSN 3050-9572 en e-ISSN 3050-9580, 1(1).
50. Chen, J. H., & Asch, S. M. (2017). Machine learning and prediction in medicine. Annals of Internal Medicine, 167(3), 219–220.
51. Kummari, D. N., & Burugulla, J. K. R. (2023). Decision Support Systems for Government Auditing: The Role of AI in Ensuring Transparency and Compliance. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 493-532.
52. Rajkomar, A., et al. (2018). Scalable and accurate deep learning with EHRs. npj Digital Medicine, 1, 18.
53. Ramesh Inala. (2023). Big Data Architectures for Modernizing Customer Master Systems in Group Insurance and Retirement Planning. Educational Administration: Theory and Practice, 29(4), 5493–5505. https://doi.org/10.53555/kuey.v29i4.10424
54. Garapati, R. S. (2022). AI-Augmented Virtual Health Assistant: A Web-Based Solution for Personalized Medication Management and Patient Engagement. Available at SSRN 5639650.
55. Kumar Bandi, V. D. V. (2023). MLOps Frameworks for Reliable Model Deployment in Cloud Data Platforms. Journal of Artificial Intelligence and Big Data, 3(1), 81–101. Retrieved from https://www.scipublications.com/journal/index.php/jaibd/article/view/1368
56. Avinash Reddy Segireddy. (2022). Terraform and Ansible in Building Resilient Cloud-Native Payment Architectures. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 444–455. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/7905.
57. Guidotti, R., et al. (2019). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 1–42.
58. Inala, R. AI-Powered Investment Decision Support Systems: Building Smart Data Products with Embedded Governance Controls.
59. Raji, I. D., et al. (2020). Closing the AI accountability gap. Proceedings of FAT*, 33–44.
60. Keerthi Amistapuram. (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems. Educational Administration: Theory and Practice, 29(4), 5950–5958. https://doi.org/10.53555/kuey.v29i4.10965
61. Buolamwini, J., & Gebru, T. (2018). Gender shades. Proceedings of FAT*, 77–91.
62. European Commission. (2021). Ethics guidelines for trustworthy AI.
63. Rongali, S. K. (2023). Explainable Artificial Intelligence (XAI) Framework for Transparent Clinical Decision Support Systems. International Journal of Medical Toxicology and Legal Medicine, 26(3), 22-31..
64. Divya, V., & Bandi, V. K. (2023). Cloud-Native Model Lifecycle Management for Enterprise AI Systems. International Journal of Scientific Research and Modern Technology, 78. https://doi.org/10.38124/ijsrmt.v2i12.1236
65. Siva Hemanth Kolla. (2023). Deep Learning–Driven Retrieval-Augmented Generation for Enterprise ITSM

Automation: A Governance-Aligned Large Language Model Architecture . Journal of Computational Analysis and Applications (JoCAAA), 31(4), 2489–2502. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/4774
66. U.S. Food and Drug Administration. (2021). Artificial intelligence/machine learning software as a medical device.
67. Uday Surendra Yandamuri. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706. https://doi.org/10.5281/zenodo.18095256
68. Kummari, D. N. (2023). Energy Consumption Optimization in Smart Factories Using AI-Based Analytics: Evidence from Automotive Plants. Journal for Reattach Therapy and Development Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2), 3572.
69. Davuluri, P. N. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems.
70. Zhang, Q., et al. (2021). Interpreting deep learning models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10), 3378–3395.
71. Tonekaboni, S., et al. (2021). Clinician-centered explainable AI. Nature Machine Intelligence, 3, 40–47.
72. Inala, R. Revolutionizing Customer Master Data in Insurance Technology Platforms: An AI and MDM Architecture Perspective.
73. Louizos, C., et al. (2018). Causal effect inference. NeurIPS.
74. Garapati, R. S. (2022). Web-Centric Cloud Framework for Real-Time Monitoring and Risk Prediction in Clinical Trials Using Machine Learning. Current Research in Public Health, 2, 1346.
75. Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of causal inference. MIT Press.
76. Kolla, S. H. (2021). Rule-Based Automation for IT Service Management Workflows. Online Journal of Engineering Sciences, 1(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/ojes/article/view/1360
77. Nagabhyru, K. C. (2023). From Data Silos to Knowledge Graphs: Architecting CrossEnterprise AI Solutions for Scalability and Trust. Available at SSRN 5697663.
78. Gottimukkala, V. R. R. (2022). Licensing Innovation in the Financial Messaging Ecosystem: Business Models and Global Compliance Impact. International Journal of Scientific Research and Modern Technology, 1(12), 177-186.
79. Ribeiro, M. T., et al. (2018). Anchors. Proceedings of AAAI.
80. Wang, C., et al. (2022). Explainable boosting machines for healthcare. Artificial Intelligence in Medicine, 126, 102187.
81. Segireddy, A. R. (2021). Containerization and Microservices in Payment Systems: A Study of Kubernetes and Docker in Financial Applications. Universal Journal of Business and Management, 1(1), 1–17. Retrieved from https://www.scipublications.com/journal/index.php/ujbm/article/view/1352
82. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
83. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273–297.
84. Rongali, S. K. (2022). AI-Driven Automation in Healthcare Claims and EHR Processing Using MuleSoft and Machine Learning Pipelines. Available at SSRN 5763022.
85. Guntupalli, R. (2023). AI-Driven Threat Detection and Mitigation in Cloud Infrastructure: Enhancing Security through Machine Learning and Anomaly Detection. Available at SSRN 5329158.
86. Vaswani, A., et al. (2017). Attention is all you need. NeurIPS.
87. Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. International Journal of Scientific Research and Modern Technology, 1(12), 227–237. https://doi.org/10.38124/ijsrmt.v1i12.1111
88. Lundberg, S. M., et al. (2020). From local explanations to global understanding. Nature Machine Intelligence, 2, 252–259.
89. Kummari, D. N. (2023). AI-Powered Demand Forecasting for Automotive Components: A Multi-Supplier Data Fusion Approach. European Advanced Journal for Emerging Technologies (EAJET)-p-ISSN 3050-9734 en e-ISSN 3050-9742, 1(1).
90. Siva Hemanth Kolla. (2022). Knowledge Retrieval Systems for Enterprise Service Environments. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 495–506. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/8037