Demystifying Disease Prediction with Explainable Supervised Learning

Neel Modi[1] , Astha Soni[1] , Gokul Yenduri[2,*] , Rutvij H. Jhaveri[3,4], Stella Bvuma[4,*]

[1]Department of Information and Communication Technology, School of Technology, Pandit Deendayal Energy University, Gandhinagar, India

[2]School of Computer Science and Engineering, VIT-AP University, Amaravati, 522237, Andhra Pradesh, India

[3]Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar, India

[4]Department of Applied Information Systems School of Consumer Intelligence and Information Systems, College of Business & Economics, University of Johannesburg, Johannesburg, South Africa

Emails: Modi.Neel@gamil.com; ASoni.stha@gamil.com; Yenduri.Gokul@gamil.com; Rutvij.Jhaveri@sot.pdpu.ac.in; stellab@uj.ac.za

Abstract

The ever-worsening mortality rates due to various diseases such as heart disease, breast cancer, and kidney disease are of great concern. Early diagnosis of the disease can be of great help. This process can be automated with the help of Artificial intelligence (AI). But, the main worry of using AI in healthcare is its black-box behaviour. The majority of the models characterized by high accuracy are often black-box in nature. This can be overcome by the use of eXplainable Artificial Intelligence (XAI), which is capable of explaining the predictions made by these black box models. We have exploited 3 different XAI frameworks: SHAP, LIME, and DALEX, to understand the working and the facilities provided by the three frameworks and compare them. We have used 5 disease datasets (3 heart disease, 1 cancer and 1 kidney disease) to carry out our work. Each dataset was trained with 3 machine learning models, namely Support Vector Machine (SVM), Logistic regression (LR), and K-Nearest neighbours (KNN), and the best model was used to feed to the XAI framework. LR performed best for one of the heart disease datasets with 72.31%accuracy, while SVM outperformed in all the other datasets, thus proving the efficacy of such approaches for early disease prediction.

Keywords: Disease Forecasting; Explainable AI; Responsible Learning; Supervised Machine Learning; Healthcare

DOI: https://doi.org/10.54216/FPA.200213
Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

1 Introduction

The current state of healthcare involves a combination of challenges and advancements [1]. In recent years, a growing emphasis has been on preventive measures, early detection, and personalized treatment approaches. Medical researchers and experts are progressively directing their attention towards the management of risk factors, modifications in lifestyle, and public health initiatives aimed at enlightening individuals about the value of preserving a healthy heart and blood vessels. With almost 18 million deaths per year due to cardiovascular diseases (CVD), it is one of the leading causes of death globally. According to data from the Global Burden of Disease (GBD) study and the World Health Organization (WHO), cardiovascular disease is the leading cause of death worldwide each year. [1, 2]. A person's risk of developing cardiovascular disease (CVD) is typically estimated using a combination of clinical and non-clinical factors, including age, gender, blood pressure, cholesterol levels, diabetes, smoking, body mass index (BMI), physical activity, and socioeconomic status. Similarly, cancer is a complex and deadly disease affecting millions of people worldwide. Breast cancer prediction entails determining a person's risk based on factors such as personal and family medical history, genetic predisposition, lifestyle choices, and demographics. Kidney disorders pertain to a collection of ailments that impact the operation of the renal organs, which are crucial bodily components accountable for eliminating waste substances from the bloodstream and upholding the body's equilibrium of fluid and electrolytes. These illnesses demonstrate a range of severity, varying from moderate to intense, and can arise from a variety of sources, including genetic factors, infections, autoimmune diseases, specific medications, and underlying medical conditions such as diabetes or hypertension. A progressive deterioration in renal function is the hallmark of chronic renal disease (CKD). It is critical that diagnostic techniques for these conditions advance as quickly as possible. Healthcare and individualized therapy support are greatly impacted by emerging technologies like artificial intelligence (AI), deep learning (DL), and machine learning (ML). AI has the power to revolutionize healthcare by lowering costs, boosting efficiency, and improving patient outcomes. Artificial Intelligence can be applied to electronic health records, personalised medicine, medical imaging, diagnosis, and treatment, as well as predictive analytics and virtual health assistants. [3]. In order to predict the future, the analytical model will use the training data. When fresh data is presented to the model, it extracts valuable information based on prior learning. Through the analysis of vast amounts of patient data, including genetic data, clinical records, and medical images, AI algorithms are able to identify patterns and indicators of breast cancer. These algorithms can become more accurate and efficient over time by continuously learning from large amounts of data, which can help detect breast cancer early and lower the likelihood of late-stage diagnoses. Furthermore, they are capable of identifying risk factors that could accelerate the onset or worsening of kidney disease as well as subtle changes in kidney function. This tailored strategy improves treatment results, minimizes side effects, and maximizes the use of available resources. The diagnosis of kidney and heart conditions can also be aided by computer vision. Data scientists can use images from medical scans, such as computed tomography (CT) or magnetic resonance imaging (MRI) equipment, to train a convolutional neural network (CNN) to identify coronary artery disease.

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

Nevertheless, there exist possible disadvantages as well such as bias and lack of transparency, that should be considered. Perhaps there are a lot of advantages to using AI responsibly in healthcare, but it's important to make sure that patient privacy and ethical issues are taken into account. Artificial intelligence (AI) models that can be explained are essential in the healthcare industry due to the ethical dilemma surrounding the degree of transparency associated with AI and the lack of trust in the opaque operation of AI systems[4]. Explainable AI (XAI) methods are AI techniques that are used to explain AI models and their predictions[5]. The main objectives of XAI are to make AI more understandable and trustworthy, to create tactics and methodologies that assist humans in understanding how AI models arrive at their conclusions or predictions, and to make AI models interpretable by nature.

This need for explainability to establish trust in the black-box AI models has motivated us to use different XAI techniques on different datasets to determine which XAI technique is robust and to understand the drawbacks that we may face while working with these techniques. Black-box models are very powerful and can predict with very high accuracy, so it's important to use them in healthcare to predict diseases.

Even if we employ existing XAI frameworks and traditional ML algorithms, the novelty of our contribution lies in the comprehensive evaluation and comparison of these methods across multiple disease datasets. Our primary aim is to assess the utility of XAI frameworks in explaining the predictions of ML models for early disease detection across multiple diseases. The key contributions of our work are:

- We implemented three XAI frameworks: SHAP, LIME, and DALEX, on five different disease datasets.
- We provided insight into the manner in which the three different frameworks are distinct from one another.
- Five distinct datasets, each having unique properties, have been used. Asserting explainability and interpretability through the use of three distinct frameworks allows us to understand how XAI interacts with different datasets thanks to these differences in characteristics like size, number of features, and completeness.

The rest of the paper is structured as follows: The second section of the study discusses the literature review, while the third section introduces the methods we employed in our analysis. Results and discussion are illustrated in section V, just before the conclusion of our work, which is depicted in section 6, highlighting potential threats and future directions of this study.

2 Literature review

Artificial intelligence has shown enormous potential in the healthcare arena for the diagnosis of diseases, treatment planning, and patient monitoring. However, healthcare professionals find it challenging to accept AI models due to their black-box nature. Caruana et al. [6] underlined the relevance of explainability in healthcare AI systems, stating that interpretable models are more likely to be trusted and accepted by physicians. Explainable AI can increase transparency and empower medical professionals to make wise decisions by offering insights into the decision-making process. Holzinger

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

Table 1: Previous works of AI in healthcare.

Reference	Used ML model	Contributions	Limitations
[9]	Multilayer perceptron (MLP) models	Examines federated learning strategies to increase hospitalized patients' mortality prediction accuracy using electronic health record (EHR) information from various organizations.	Limited model generalizability due to limited data
[10]	COVNet (neural network)	Developed a completely automated system to use chest CT to identify COVID-19.	Lack of transparency and interpretability
[11]	Gradient boosting, Extreme gradient boosting, Bagging, Extra tree classifier	A classification system based on feature selection was developed to help detect Parkinson's disease early.	Limited dataset
[12]	Ensemble technique	feature selection model for detection of erythemato-squamous disease with improved speed and accuracy.	Limited dataset, Lack of transparency
[12]	Ensemble technique	feature selection model for detection of erythemato-squamous disease with improved speed and accuracy.	Limited dataset, Lack of transparency

Table 2: Previous works of XAI in healthcare.

Reference	Used ML model	Contributions	Limitations
[13]	LIME as XAI and Resnet50, VGG16 and Inception v3 deep learning	Utilized three deep learning models to predict the onset of Alzheimer's disease, and LIME was used to interpret the results.	Higher loss on prediction and limited data
[14]	SHAP and LIME as XAI, XGBOOST, Logistic Regression, Stochastic Gradient Descent, Support Vector Classifier, Kneighbors classifier, and Naive Bayes	Evaluated the predictability of the model's heart disease prediction by comparing various machine learning methods.	Limited dataset, XAI was not implemented on the best model.
[15]	Survival Gradient Boosting model, random forest, SHAP	Provided an analysis of explainability and a comparison of two models for predicting HF survival.	Substantial imbalance in target features that biases the model's performance
[16]	Ensemble Trees Classifier, SHAP	Developed an explainable CKD prediction model that can be used to explain how various patient clinical characteristics affect CKD early diagnosis.	Limitations on dataset
[17]	Logistic Regression, RF, GBM, SVM, Multilayer Perceptron Classifier, SHAP, LIME, DALEX.	Developed an (XAI) archetype to offer a qualitative understanding of the relationship between the model and the model's parameters for renal failure.	Limited dataset

et al. [7] discussed the importance of usability and explainability in AI, particularly in the context of medicine. They emphasize that transparency is essential for users to trust and use AI technologies, including healthcare professionals and patients. XAI approaches enable users to understand how and why AI models make specific predictions or judgments by offering interpretable explanations. Table 1 summarises the characteristics of previous works about the use of AI in healthcare.

Explainability is crucial not only for increasing user trust but also for addressing regulatory and ethical concerns. The European Union's General Data Protection Regulation (GDPR) demands that people have the right to know why decisions made by AI systems that impact them are made. Mittelstadt et al. [8] explored the ethical concerns of black-box AI systems, arguing that explainability is critical for maintaining fairness, avoiding discrimination, and facilitating accountability. This means that explainable artificial intelligence (XAI) is needed to replace the extremely complicated black box models. A list of prior XAI projects in the healthcare industry can be found in Table 2.

Three distinct post-hoc explainable techniques—SHAP, LIME, and DALEX—are compared in the proposed study. A unified framework called SHAP can be used to

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

Table 3: Previous works on same datasets

Ref	Disease	Dataset	Accuracy	Best model
[18]	Heart disease	UCI (270, 22)	87.69%	DT and SVM
[19]	Heart disease	UCI (303, 14)	87%	LOGREG
[20]	Heart disease	UCI (303, 13)	88.7%	HRFLM
[21]	Heart disease	UCI (303, 13)	85.60%	RF
[22]	Kidney disease	UCI (400, 26)	92%	RF
[23]	Kidney disease	UCI (400, 26)	98.46%	LSVM
[24]	Cancer disease	WDBC (569, 30)	98%	KNN

explain any ML model's output. It offers a global and local explanation of model predictions and is based on the idea of Shapley values from cooperative game theory. By quantifying each feature's contribution to the prediction outcome, SHAP values help to clarify the model's decision-making process [25]. One popular approach to deciphering individual predictions from black-box machine learning models is LIME. It creates interpretable "surrogate" models that approximately resemble the complex model's behavior in the immediate area of the instance being described. LIME assists users in understanding why a given prediction was produced by generating locally faithful explanations and providing insights into the key features and their contributions [26]. DALEX has the ability to generate a wide range of explanations for model behavior. It offers methods for partial dependence plots, accumulated local effects, computing variable importance, and more. These explanations offer valuable insights into how individual input features contribute to model predictions, allowing users to understand the factors driving the model's decisions [27]. Table 3 displays studies that used the same datasets as ours and the accuracies they achieved.

One of the leading causes of death worldwide is cardiovascular disease. To develop disease prediction models, researchers used supervised learning approaches combined with XAI technologies. Pedro et al. [28] undertook an explainability evaluation of the suggested heart failure survival prediction model in their work. Mehrdad et al. [29] stated a completely novel Genetic Algorithm (GA)-Adaptive Neural Fuzzy Inference System (ANFIS) algorithm for heart attack prediction. One important area where XAI methods have been used to improve the interpretability of AI models is cancer diagnosis. Philipp et al. [30] underlined in their work the great potential of XAI in cancer research, and the prediction of sample-wise networks was applied to proteomic data. In [31], Amoroso et al. used XAI frameworks to implement treatments for breast cancer. One important area of healthcare where AI advancements have shown promise is the diagnosis of Alzheimer's disease (AD). Shaker et al.[32] used Shapley values in their research to develop an interpretable model for AD diagnostic and progression detection as well as a two-layer model with random forest (RF) as a classifier algorithm.

Medical professionals' confidence in today's black box models may be increased by using XAI to make them easier to understand and interpret. The first step in understanding XAI is to distinguish between explainability and interpretability. Waddah et al.[33], in their work, the authors attempted to differentiate between the following two terms: interpretability refers to the extent to which the insights provided can be

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

understood in light of the targeted audience's domain knowledge, whereas explainability offers insights to a targeted audience to meet a need. In order to gain some important insights, we will compare several post hoc explainability methods (SHAP, LIME, and DALEX) in this work and analyze the results.

3 Methodology

By leveraging the capabilities of supervised machine learning algorithms, the suggested methodology seeks to address the explainability, interpretability, and transparency of predictions of three important health conditions: cancer, kidney disease, and cardiovascular disease. Models are trained on labeled data in supervised machine learning, where input features are matched with corresponding labels or targets. This approach allows the algorithms to learn patterns and make predictions based on the provided inputs. Three distinct supervised machine learning algorithms have been selected as viable candidates in order to achieve this: logistic regression, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). These three algorithms have shown promising results in the prediction of cardiovascular disease, kidney diseases, and cancer [34–39].

A binary classification algorithm called logistic regression models the relationship between the input features (X) and the probability that the target variable (y) will fall into a particular class. To estimate the probabilities, it makes use of the logistic function, commonly referred to as the sigmoid function. The logistic regression equation can be represented as follows: Probability of y belonging to class 1:

$$P(y = 1|X) = 1/(1 + exp(-(w * X + b))) \tag{1}$$

In which b is the bias term, X is the input feature vector, and w is the weight vector. Using optimization strategies like gradient descent, the logistic regression model is trained by minimizing the logistic loss function, such as the binary cross-entropy loss. The objective is to determine the ideal values for w and b in order to minimize the discrepancy between the actual labels and the predicted probabilities. The following represents the mathematical formula that gradient descent uses to update its parameters:

$$\theta\_new = \theta\_old - learning\_rate * \nabla(loss\_function) \tag{2}$$

Here, the step size is denoted by *learning_rate*, the updated parameter values are represented by θ*_new*, the current parameter values are represented by θ*_old* , and the gradient of the loss function with respect to the parameters is represented by ∇*(loss_function)*. Gradient descent seeks to identify the set of parameter values that minimizes the loss function through iteratively updating the parameters using this formula, enabling the model to produce predictions that are more accurate.

SVM is a potent classification algorithm that divides the various classes in a high-dimensional space by locating the ideal hyperplane. The goal of SVM is to maximize the difference between the closest data points of various classes and the hyperplane. The classification decision function in the case of linear SVM can be stated as follows:

$$f(X) = sign(w * X + b) \tag{3}$$

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

The input feature vector is represented by X, the bias term is b, the weight vector is w, and the sign function is *sign()*. The SVM model solves the following optimization problem in order to determine the ideal values for w and b:

$$Minimize : (1/2) * ||w||^2 \tag{4}$$

Subject to: $y\_i * (w * X\_i + b) >= 1$, for all training samples *(X_i, y_i)* Here, for every training sample *(X_i)*, the target label is *y_i*. The goal is to maximize the margin while identifying the hyperplane that appropriately divides the training samples.

KNN is a non-parametric algorithm that uses the k-nearest neighbors' majority vote to classify new data points. Instead of learning explicit models, it uses the similarity between instances in the feature space. This is a summary of the KNN algorithm:

1. Determine the distance using either the Manhattan distance or the Euclidean distance between the new data point (x) and all training data points (x_i).
   - Euclidean distance: *dist(x, x_i)* =

$$sqrt((x\_j - x\_i, j)^2)$$

   - Manhattan distance:

$$dist(x, x\_i) = |x\_j - x\_i, j|$$

2. Using the calculated distances as a guide, choose the k closest neighbors. Let *N_k(x)* represent the set of k nearest neighbors.
3. Locate at which of the k closest neighbors has the majority class label. Assume that for every neighbor x_i in *N_k(x)*, the class labels are y_i.
4. Assign the majority class label to the new data point (x). When it comes to binary classification, the majority class label among the neighbors would be the predicted class label.

Mathematically, the prediction step can be represented as follows:

$$y\_pred = argmax((1[y\_i = c])) \tag{5}$$

for each class label c in the set of k nearest neighbours *N_k(x)*. In this case, the indicator function 1 [*condition*] returns 1 in the event that the *condition* is true and 0 in the absence of it. The *argmax* function returns the class label that maximizes the sum of indicators, indicating the majority class label among the neighbors. It is necessary to set the hyperparameter k, which stands for the number of closest neighbors before the model can be trained. It affects the balance between bias and variance in the model's predictions. A smaller value of k tends to capture local patterns, while a larger value of k considers a broader context but may introduce more noise.

These algorithms have shown effectiveness in various classification tasks and are widely used in the field of medical research. These algorithms were applied to the five different datasets (3 heart diseases, 1 cancer, and 1 kidney disease). The proposed methodology is as shown in Fig. 1

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

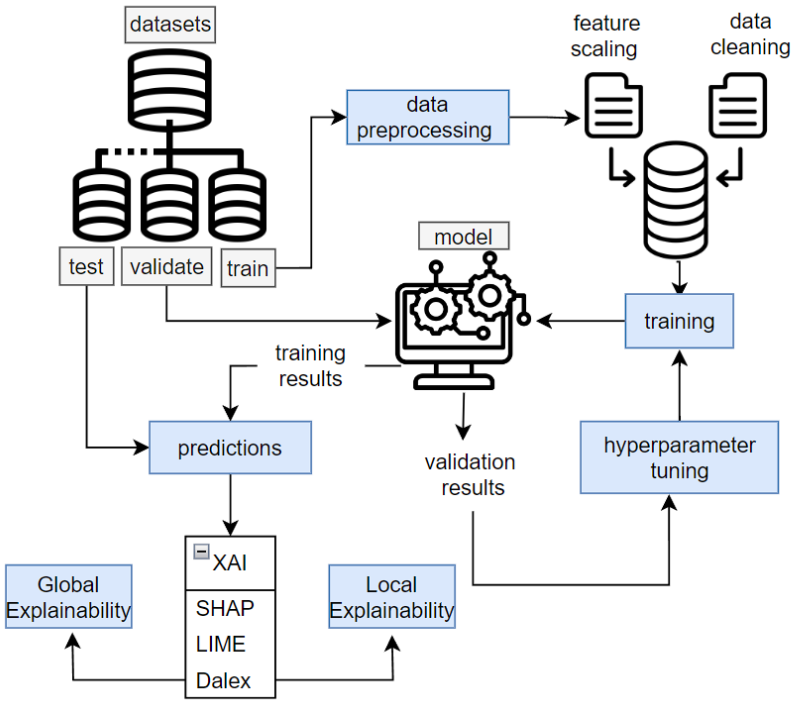


Fig. 1: Block diagram for methodology including each steps

Before training the models, the datasets undergo preprocessing steps to ensure data quality and suitability for analysis. This involves handling missing values by employing appropriate techniques such as dropping samples with missing values, replacing missing values with zero, mean, median, or mode, interpolation, or extrapolation of the missing values depending on the different datasets. Categorical variables are encoded into numerical representations to enable compatibility with the algorithms, and numerical features are scaled to comparable ranges by a min-max scaler to avoid dominance by any particular feature. Each value is substituted using the following formula after the minimum and maximum values from the data are obtained:

$$x\_Normalized = x_j - x\_min / x\_max - x\_min \tag{6}$$

where j in $1, ...n$ and n is the number of features.

After that, the datasets are divided into training and testing sets using the widely accepted 70:30 ratio. The training set is used to train the models, adjusting their parameters and hyperparameters to optimize their performance. This iterative process involves finding t he b est c ombination o f s ettings t hat y ields t he m ost accurate predictions. The trained models are subsequently evaluated using the testing set to assess their predictive performance. Various evaluation metrics, such as accuracy, precision, recall, and F1-score, are computed to measure the models' ability to correctly

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

predict the presence or absence of each health condition. To ensure the robustness of the results, additional techniques like cross-validation will be employed. The available data is divided into k non-overlapping folds (subsets) denoted as $D_1, D_2, ..., D_k$. For each iteration i ($1 \leq i \leq k$), the model is trained on the union of all folds except the i-th fold and evaluated on the i-th fold. The model's performance is measured using a chosen evaluation metric, such as accuracy, precision, recall, or F1 score. The performance measure for each iteration is denoted as P_i. The average performance, denoted as P_avg, is calculated by taking the mean of the performance measures from all iterations:

$$P\_avg = (P_1 + P_2 + ... + P_k)/k \tag{7}$$

The final model for predicting the corresponding health condition is chosen from among the models that show the best predictive performance across the various datasets and algorithms. Throughout the entire research process, strict adherence to ethical guidelines and regulations is paramount. Measures are taken to ensure the privacy and confidentiality of patient information.

Finally, the selected models are integrated into a framework that facilitates explainable AI (XAI) techniques. Three distinct XAI approaches, namely SHAP, DALEX, and LIME, are employed to provide interpretation at both the global and local levels. The contribution of the feature to the prediction is indicated by the SHAP value. Assigning a SHAP value (ϕ) to every feature (i) in the input vector is the aim of SHAP, signifying its input into the prediction. The local accuracy, missingness, and consistency properties are used to calculate the SHAP values. Conversely, LIME seeks to identify a sparse linear model that approximates the behavior of the complex model. Given a complex model f and an instance x, LIME defines a local linear model $g(x')$, where x' is a perturbed version of x, that approximates f near x:

$$g(x') = w \cdot x' + b \tag{8}$$

Here, w represents the weights assigned to each feature,x' is a perturbed version of x, and b is the bias term. DALEX measures the significance of each feature over the whole dataset to provide global explanations. Evaluating each feature's effect on the model's predictions is the aim. Computing the Shapley values is a popular method for estimating the significance of a feature. By breaking down the model's prediction for a particular instance into the contributions of individual features, DALEX additionally offers local explanations. This aids in comprehending the variables influencing the forecast in that specific case. One way to depict the local explanation is as follows:

$$f(x) = \phi_0 + \sum_{i=1}^{N} \phi_i \cdot x\_i \tag{9}$$

Here, the value of feature i in instance x is denoted by x_i, the contribution of feature i is represented by ϕ_i, the model's base value or intercept term is denoted by ϕ_0, and the model's prediction is represented by $f(x)$, for instance, x.

The objective of this study is to increase the predictability and understandability of ML models by utilizing LIME for local interpretations and SHAP and DALEX

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

for global interpretations. XAI techniques have been employed to enhance trust, decision-making, and health condition insights by improving the transparency and understandability of machine learning models' predictions.

4 Results and discussion

This work focuses more on improving the interpretability, justification, and explainability of black-box machine learning models because the study compares various XAI frameworks. The study uses supervised learning techniques like SVM, logistic regression, and KNN for prediction purposes. Weight optimization improves prediction accuracy. A confusion matrix is used to visualize classification algorithm performance, measuring recall, precision, specificity, accuracy, and AUC-ROC curves. True positive, false positive, true negative, and false negative values are used to evaluate performance, identifying whether patients are correctly classified as having a disease or not. While we acknowledge the value of measurements such as precision and recall, we want to underline that prioritizing accuracy enables clarity and comparability across models and datasets. Since accuracy gives a simple measure of prediction correctness, which is consistent with our goal of disease prediction, we believe that it is the necessary and sufficient metric in determining the effectiveness of our supervised approaches. The SVM model outperformed the logistic regression model with an accuracy of 90.16%, 98.05%, 99.12% and 98.33% on heart disease 1, heart disease 3, Breast cancer, and kidney disease datasets. Logistic regression proved to be a better model for the heart disease 2 datasets and gave us an accuracy of 72.13%. SVM uses a linear kernel because the data may be linearly segregated, or divided along a single line. Along with accuracy the four models of SVM also gave high F1-score, 0.86 (CVD-1), 0.98 (CVD-3), 0.97 (breast cancer), 0.99 (kidney disease). The F1 score for logistic regression that gave better accuracy for the heart disease-2 dataset was found to be 0.7. Table 4 summarises the results of our work.

Table 4: Comparison of different models on different datasets.

Disease Datasets	Best performing model	Accuracy	Best global interpretation	Best local interpretation	Dataset source
Heart disease 1(303, 14)	SVM	90.16%	SHAP	LIME	UCI
Heart disease 2(70000, 12)	Logistic regression	72.31%	SHAP	LIME	Kaggle
Heart disease 3(1025, 14)	SVM	98.05%	SHAP	LIME	Kaggle
Breast Cancer disease (569, 31)	SVM	99.12%	SHAP	LIME	WDBC
Kidney disease (400, 25)	Logistic Regression	98.75%	SHAP	LIME	UCI

There are three phases of an XAI framework they are pre-modelling, explainable modeling, and post-modeling. The proposed work follows all three phases so as to generate accurate explanations. Since SVM outperformed logistic regression, the pre-modelled SVM model was fed to the three XAI models: SHAP, LIME, and DALEX. Although SHAP and DALEX provide both global and local interpretations, LIME focuses on interpreting locally instead of providing a global model interpretation.

DOI: https://doi.org/10.54216/FPA.200213

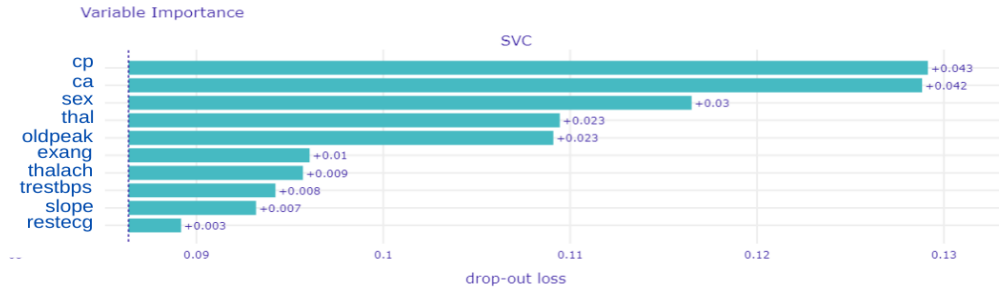
Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

4.1 Global interpretation

By taking into account the full dataset or a sizable portion of it, global interpretation enables us to comprehend the behavior and decision-making process of ML models holistically. The goal of global interpretation is to shed light on how the model functions generally throughout all cases. We have used SHAP and DALEX to understand global interpretations of the models.

4.1.1 Heart disease

We have used 3 different datasets of cardiovascular disease with 303, 70000, and 1025 instances and 14, 12, and 14 features in the respective datasets.



s

Fig. 2: DALEX CVD-1 summary plot.

A weak heart muscle is typically brought on by coronary artery disease or a heart attack, but other factors like malfunctioning heart valves, chronic high blood pressure, a high maximum heart rate, and chest pain may also be to blame. A weakened heart may occasionally be caused by a combination of conditions. The analysis of datasets revealed that features such as age, cholesterol level, thalach (maximum heart rate achieved), and chest pain played a significant r ole i n d etermining t he l ikelihood of heart disease.

Higher levels of cholesterol, maximum heart rate achieved (thalach), and systolic blood pressure were found to be associated with an increased risk of heart disease. Increasing thalach values increases the likelihood of heart disease, reflecting the heart rate-cardiovascular risk relationship. High blood pressure also increases the risk of developing heart disease, [40]. Hence, SHAP values revealed important information about the relationships between these factors and heart disease. (as shown in the SHAP summary plots in Fig. 3, Fig. 5, and Fig. 7).

The DALEX plots did provide us with the same SHAP features that are influencing the results for each individual dataset. DALEX explainer revealed that the most important feature for predicting heart disease was chest pain (CP) for dataset 1. Similarly, for dataset 2, the most important feature is systolic blood pressure (ap-hi), and for dataset 3, it is thalach (according to Fig. 2, Fig. 4, and Fig. 6).

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

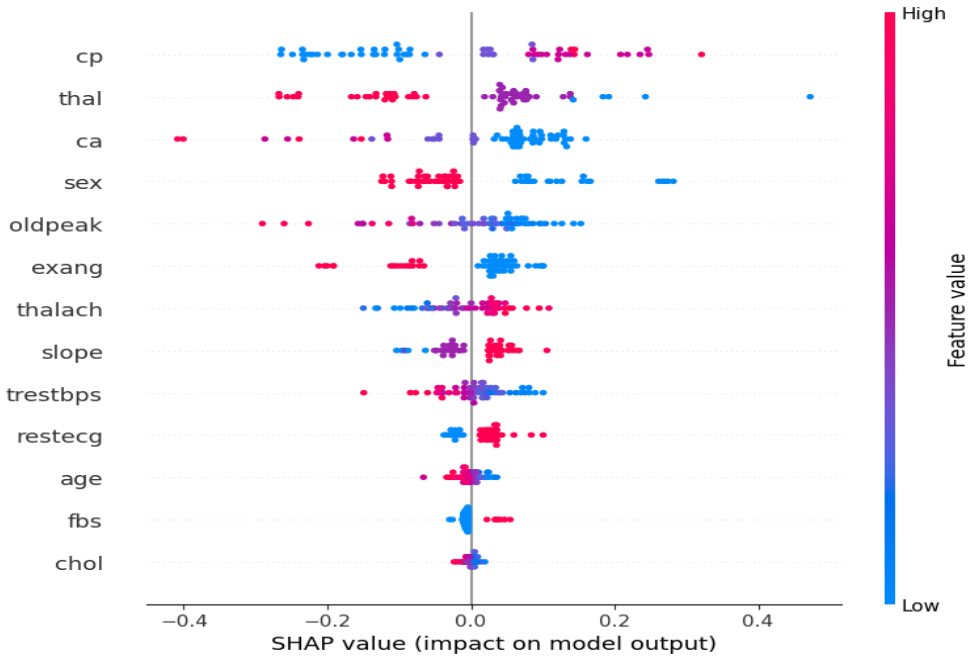
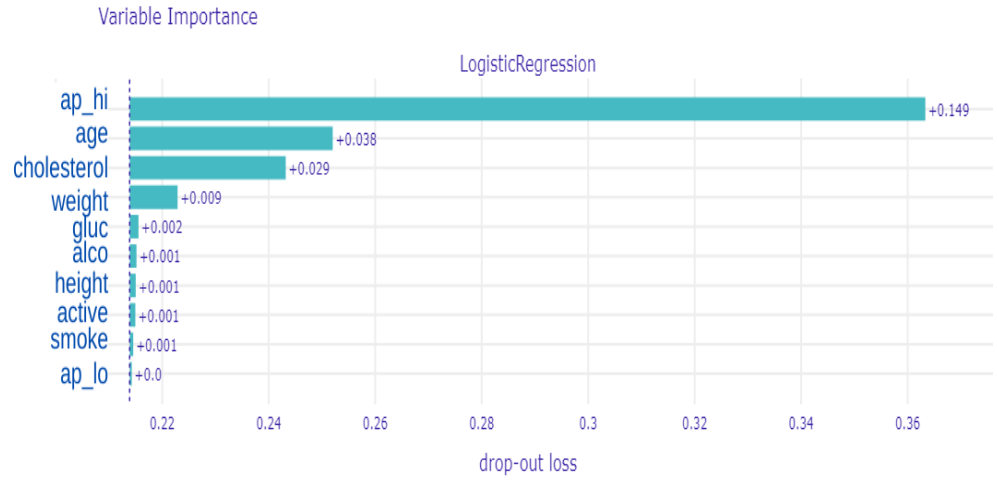


Fig. 3: SHAP CVD-1 summary plot



s

Fig. 4: DALEX CVD-2 summary plot

The SHAP explanations provided vital details about the specific relationships between these risk factors and heart disease. They could, for example, reveal that high

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

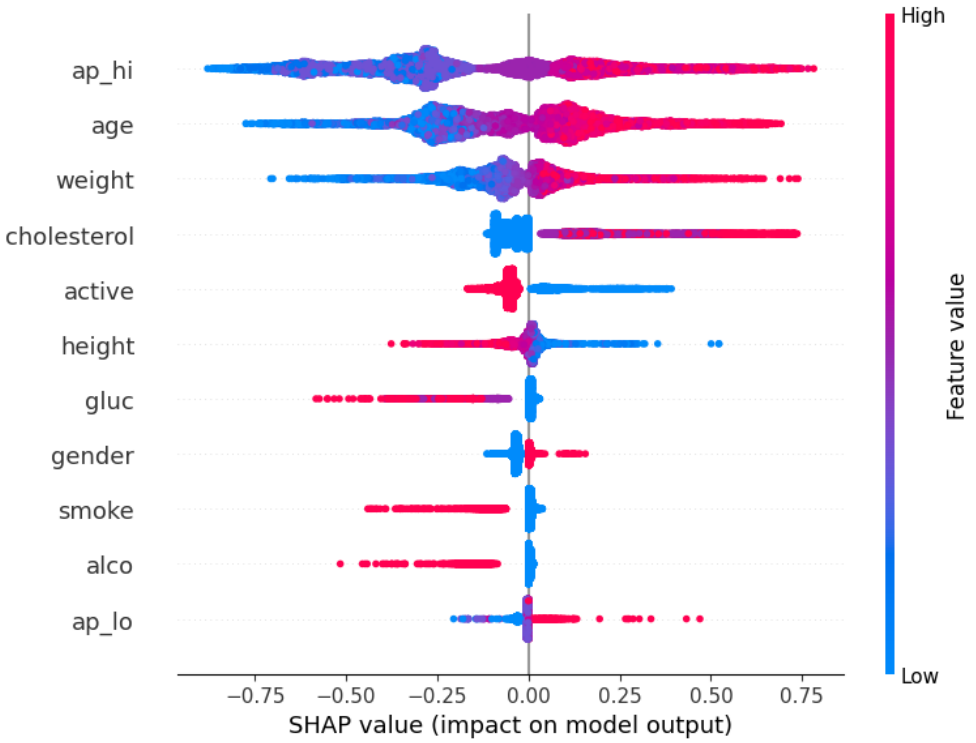
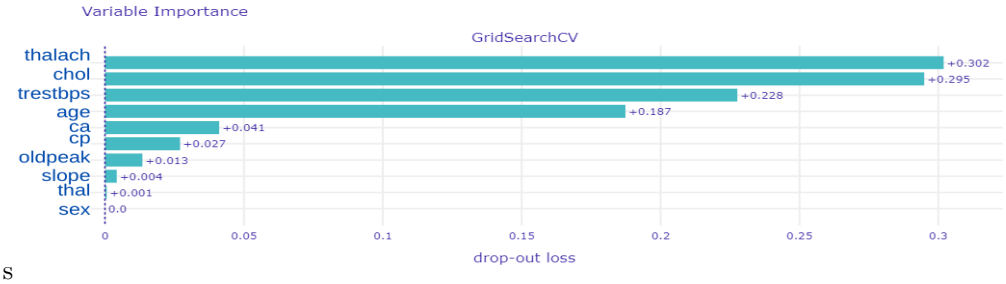


Fig. 5: SHAP CVD-2 summary plot



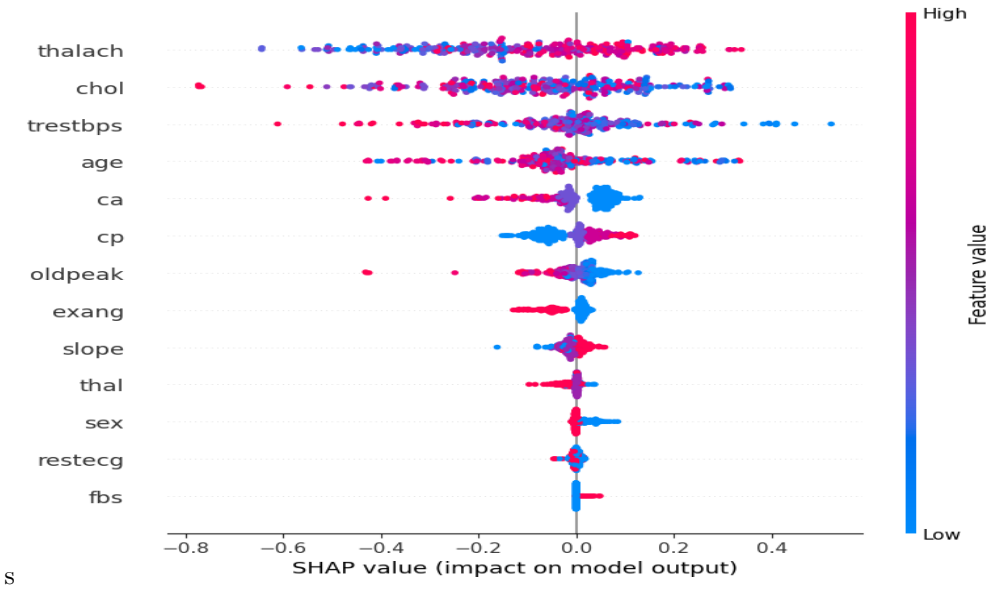
s

Fig. 6: DALEX CVD-3 summary plot

cholesterol levels combined with advanced age have a synergistic effect on the model's predictions, increasing the likelihood of heart disease even further. Such insights can be extremely useful in clinical practice, allowing doctors to more accurately identify high-risk patients and tailor their interventions accordingly.

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025



s

Fig. 7: SHAP CVD-3 summary plot

4.1.2 Breast Cancer disease

When applying XAI using SHAP and DALEX to the breast cancer dataset, the SHAP and DALEX summary plot's analysis identified several influential factors in predicting the likelihood of breast cancer. Features such as tumor size, tumor grade, and lymph node status were found to be particularly significant in determining the predictions, providing valuable insights for understanding and decision-making in breast cancer diagnosis.

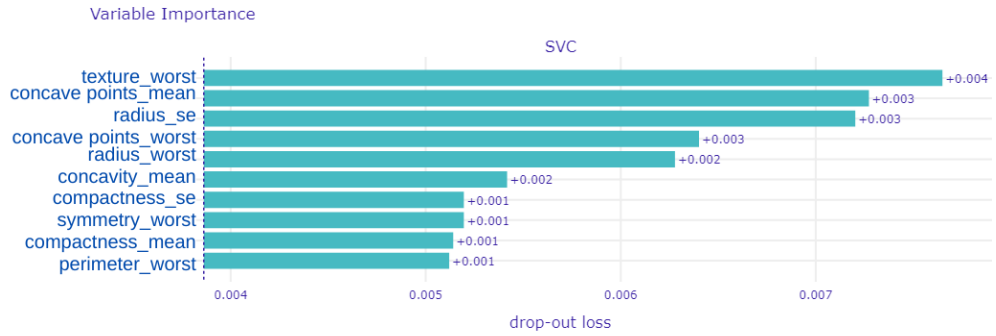
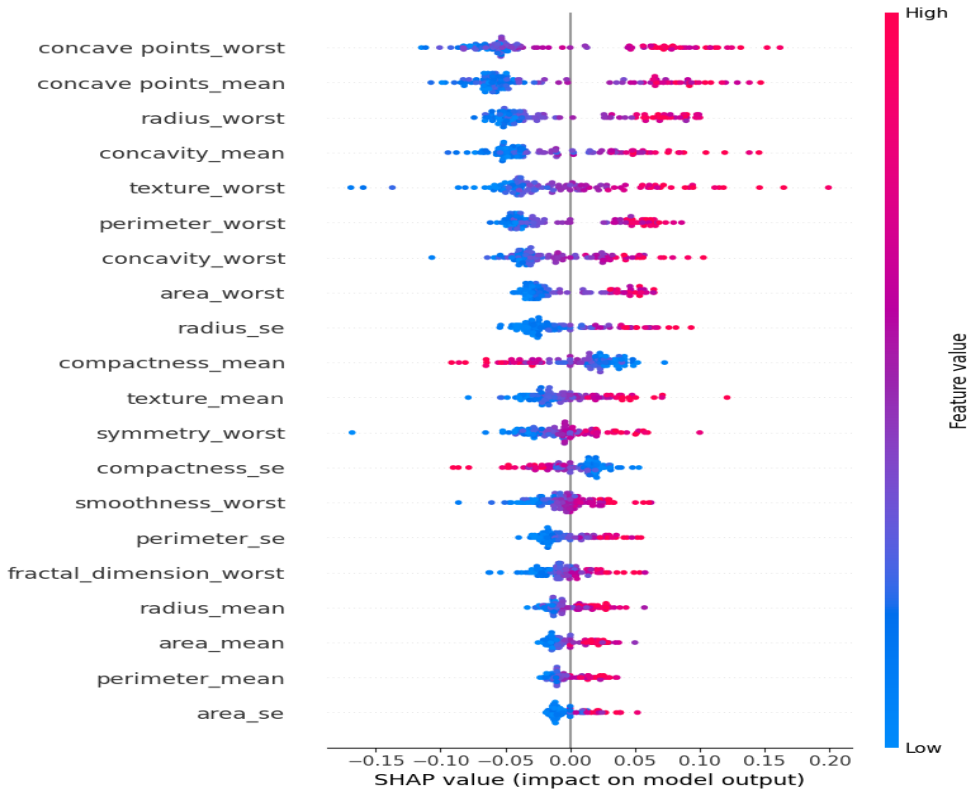


Fig. 8: Breast cancer DALEX summary plot.

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025



s

Fig. 9: Breast Cancer SHAP summary plot

Tumor size was identified as a crucial predictor, with larger tumor sizes exhibiting positive SHAP values, indicating a higher probability of breast cancer (as shown in Fig. 9). The DALEX (as depicted in Fig. 8), on the other hand, provides only the extent of influence of tumor size on the predictions of breast cancer and does not provide the relationship of the size of the tumor to breast cancer, unlike SHAP. The SHAP explanations emphasized the direct relationship between tumor size and the model's predictions, underscoring the importance of considering tumor size as a primary factor in diagnosing breast cancer. This aligns with the findings of Shah, Aamera, et al. [41] that most cases of breast cancer present in advanced stages are associated with a larger size of the tumor.

4.1.3 Kidney Disease

When applying XAI using SHAP and DALEX to the kidney disease dataset, the analysis of Fig. 11 and Fig. 10 unveiled the significant influence of variables such as diabetes mellitus, hypertension, and specific gravity on the predictions. Positive SHAP values indicate an increased likelihood of kidney disease for each of these factors. Diabetes mellitus emerged as a crucial factor, indicating that individuals with this

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

condition have a higher risk of kidney problems due to potential damage to the kidneys' blood vessels and nerves.

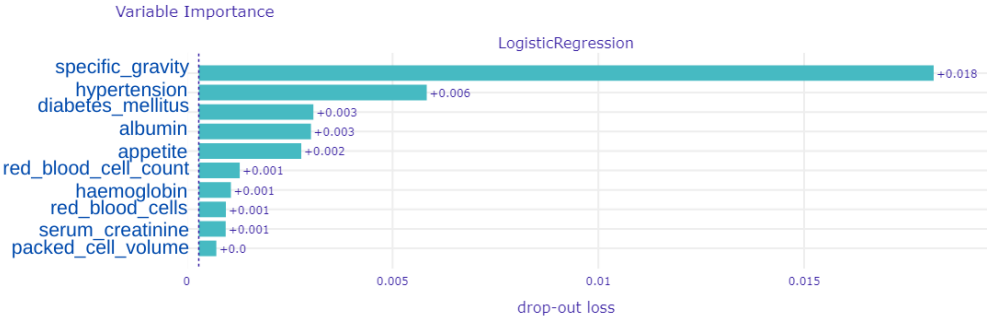
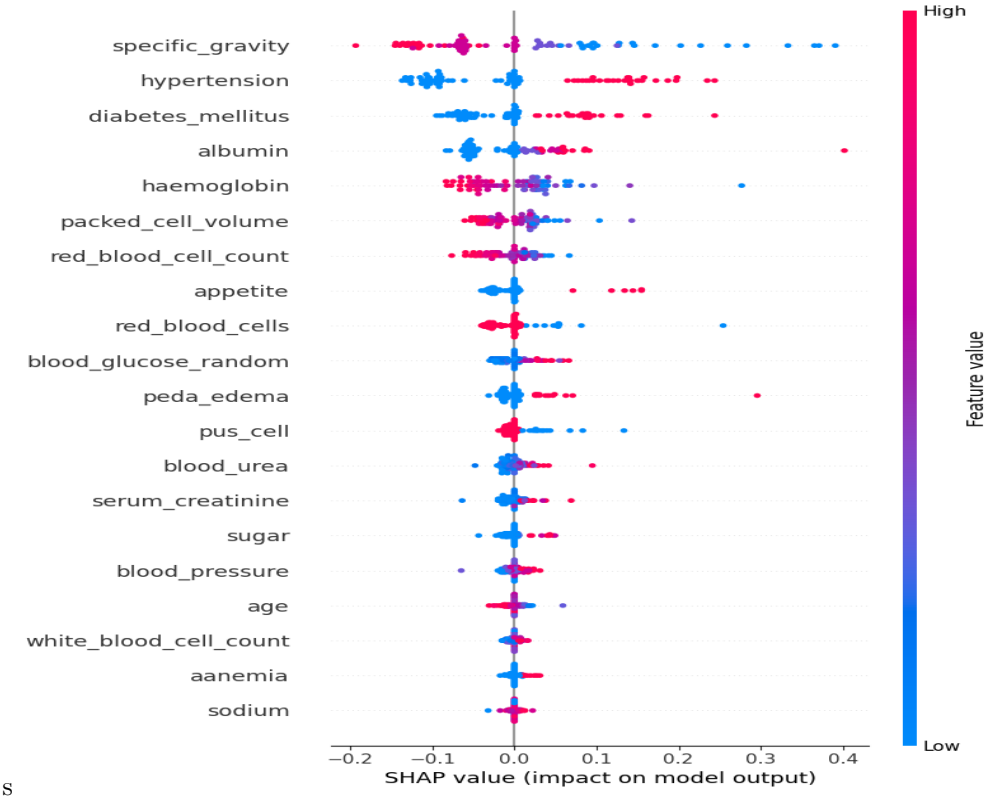


Fig. 10: Kidney disease DALEX summary plot.



s

Fig. 11: Kidney disease SHAP summary plot.

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

Hypertension, or high blood pressure, was also identified as a significant predictor of kidney disease, as it can lead to damage in the kidney's blood vessels, impairing their ability to filter waste and maintain fluid balance. Abnormal levels of specific gravity, a measure of urine concentration, were found to be associated with impaired kidney function, suggesting that this variable plays a crucial role in predicting kidney disease.

The SHAP explanations provide valuable insights for clinicians in diagnosing kidney disease and identifying potential underlying causes, particularly in patients with kidney injury. Individuals with a history of diabetes mellitus, hypertension, and abnormal specific gravity levels should be particularly vigilant about their kidney health. Regular monitoring, consultation with healthcare professionals, and appropriate preventive measures are essential to prevent or manage kidney disease effectively.

4.2 Local Interpretation

SHAP, LIME, and DALEX are techniques that provide local interpretation of data, explaining why a data instance was classified into its target class. LIME builds surrogate models to mimic the behavior of the original model, Dalex uses statistical importance, and SHAP uses cooperative game theory to capture feature importance. However, they are model-agnostic and do not significantly differ in interpretation and every technique is model-agnostic.

4.2.1 Heart diesease

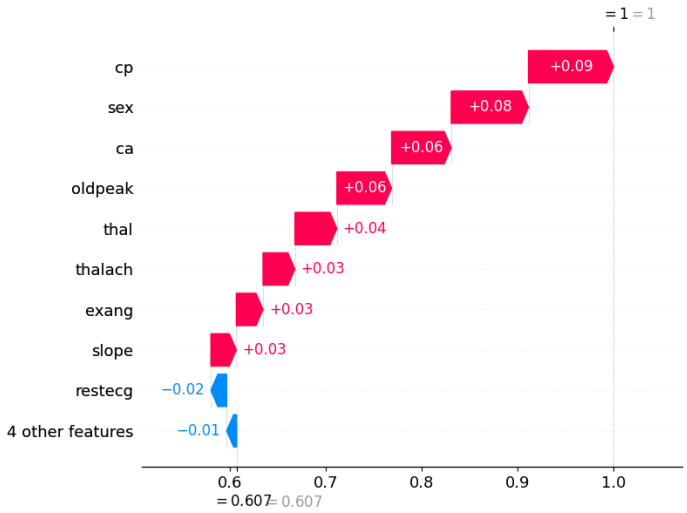


Fig. 12: CVD-1 SHAP instance explanation

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

Based on the interpretations provided by LIME, SHAP, and DALEX for a specific instance of the heart disease-1 dataset (illustrated in Fig. 12, Fig. 14 , Fig. 13), It is evident that the primary characteristics influencing the predictions in all three approaches are sex, thalassemia, and the number of major vessels (0–3) colored by fluoroscopy.

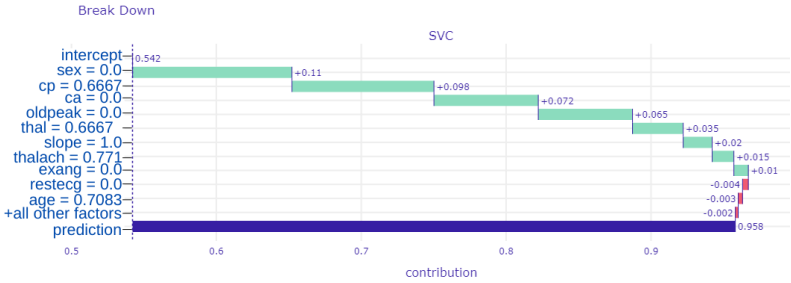


Fig. 13: CVD-1 DALEX instance explanation

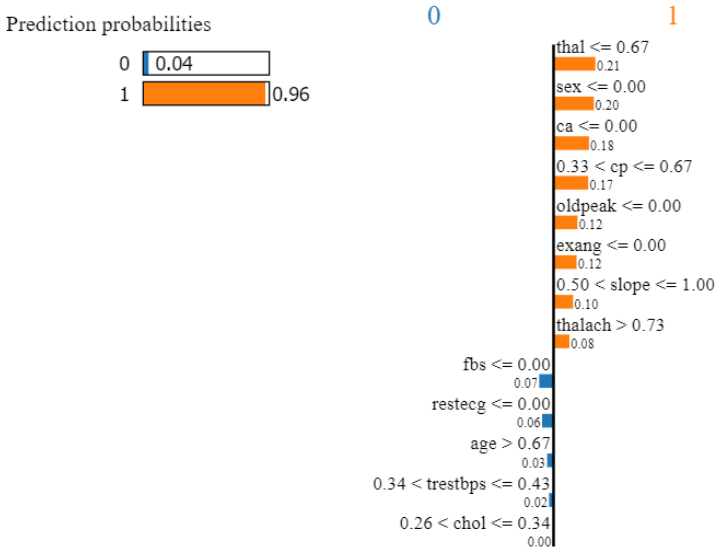


Fig. 14: CVD-1 LIME instance explanation

For the particular instance, thethal value, thalassemia, indicated "no blood flow in some part of the heart," and the ca value—the number of major vessels colored by fluoroscopy—was found to be two, indicating that two out of three major vessels were found by the fluoroscopy technique, further indicating blockage in one of the vessels. Both of these are obvious indications that the patient has heart disease.

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

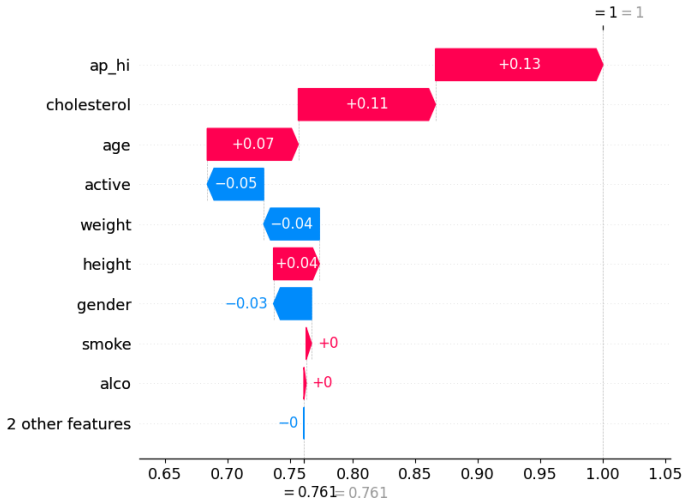


Fig. 15: CVD-2 SHAP instance explanation

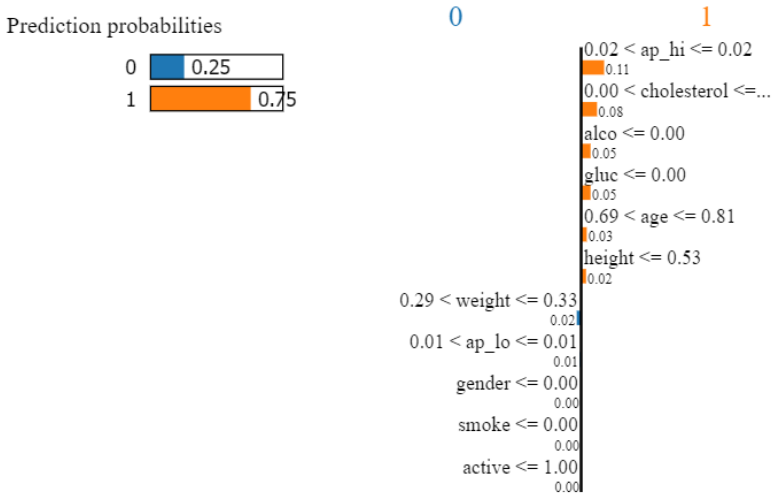


Fig. 16: CVD-2 LIME instance explanation

The XAI techniques, when applied to a single instance of the heart disease dataset 2 (as shown in Fig. 15, Fig. 16, Fig. 17), resulted in ap-hi (systolic blood pressure), cholesterol, and age as major factors in the prediction of the diagnosis. The systolic blood pressure for the patient was found to be 140 mm HG, which is higher than it should be (normally it should be less than 120 mm HG). Moreover, the age is 20,388 days (a little less than 56 years), which further indicates that age is one more influential factor in the prediction. The cholesterol of the patient was above normal, which strengthens our predictions.

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

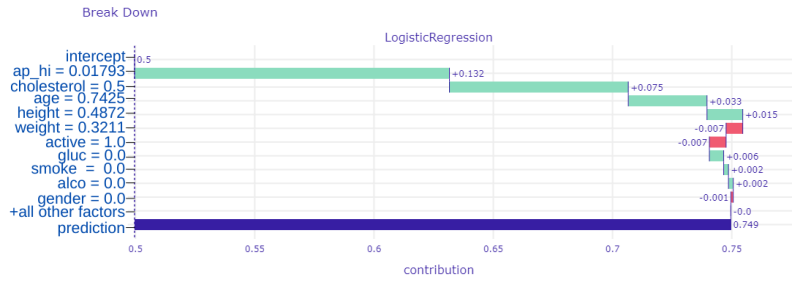


Fig. 17: CVD-2 DALEX instance explanation

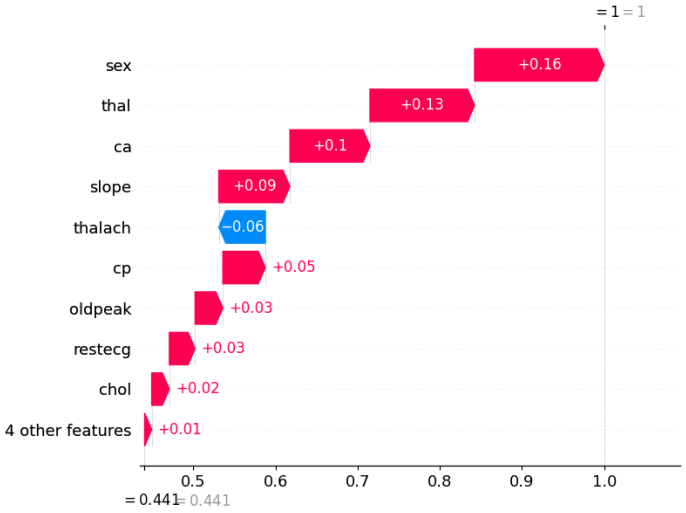


Fig. 18: CVD-3 SHAP instance explanation

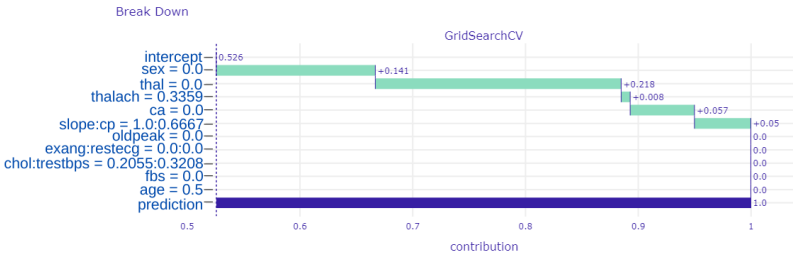


Fig. 19: CVD-3 DALEX instance explanation

For the heart disease-3 dataset again, the three models provide us with ca, sex, and slope as the top features that contribute towards predicting that the given patient

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

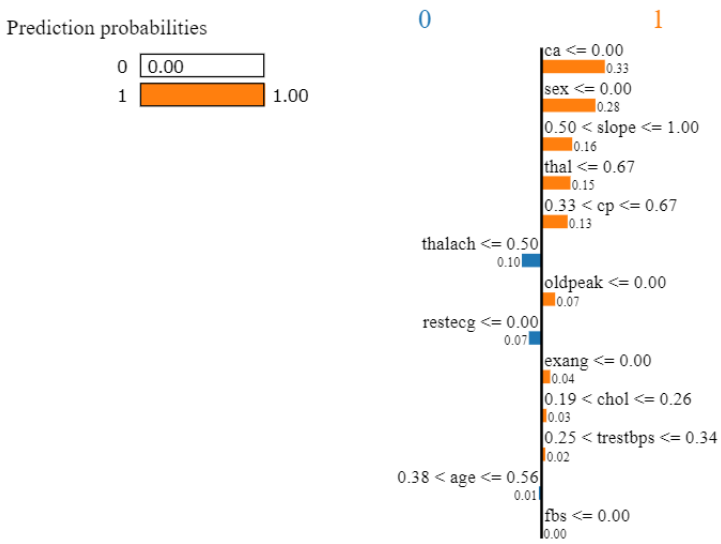


Fig. 20: CVD-3 LIME instance explanation

is diagnosed with cardiovascular disease (as seen in Fig. 18, Fig. 20, Fig. 19). For this particular patient, the ca value of 0 indicated that no vessels were found during the process of fluoroscopy, which indicates that there is a blockage of major blood vessels. Furthermore, the peak exercise ST segment's slope is negative, whereas a healthy patient's slope should be positive, according to the slope value of 1. Therefore, the models came to the conclusion of classifying the patient as having been diagnosed with cardiovascular disease.

4.2.2 Breast Cancer disease

When SHAP, LIME, and DALEX were applied to the breast cancer dataset for a local interpretation, concave points worst, concave points mean, and texture worst were the top 3 features that helped in determining that the cancer is malignant (as can be seen in Fig. 22, Fig. 23, and Fig. 21).

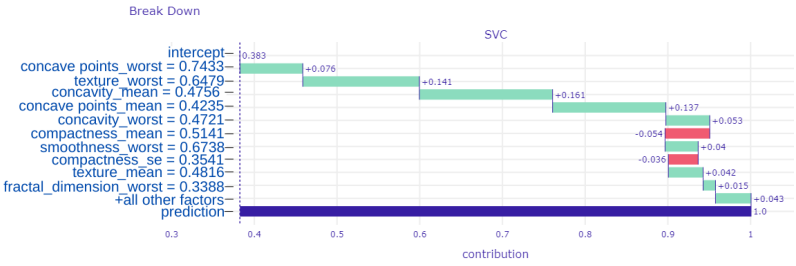


Fig. 21: Breast cancer DALEX instance explanation

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

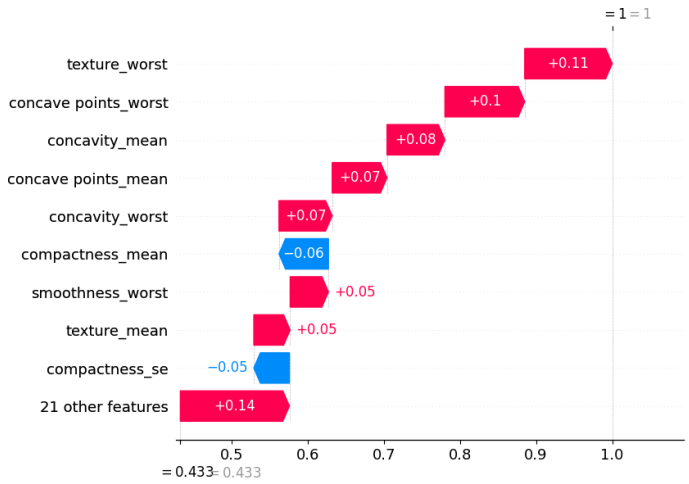


Fig. 22: Breast cancer SHAP instance explanation

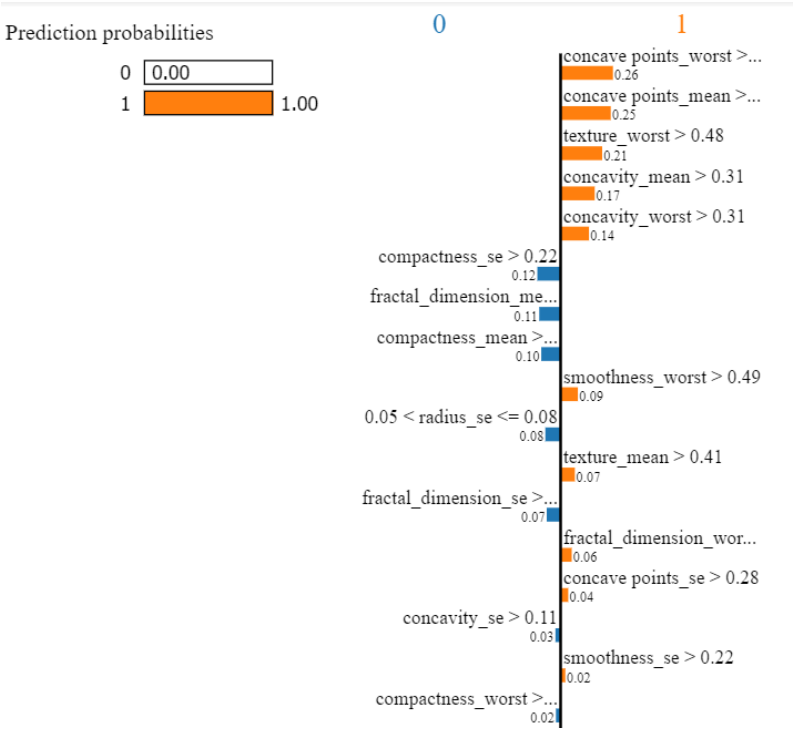


Fig. 23: Breast cancer LIME instance explanation

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

Concave points indicate how many concave areas there are in the contour, and the worst concave points represent the highest mean value. The worst texture is indicated by the highest mean value of the standard deviation for greyscale values. The standard deviation is necessary to determine the variation of the data and to explain how to spread out the numbers. The greyscale is frequently used to locate tumors. The value of concave points worst is 0.2163, the concave points mean is 0.0852, and the texture worst is 36.33, all three of which are far too large for the cancer to be benign. Hence, the XAI techniques classified the patient's cancer as malignant, providing these 3 features as the most important factors in determining so.

4.2.3 Kidney disease

In the local interpretation of the SHAP, LIME, and DALEX frameworks for a kidney disease dataset (as depicted Fig. 24, Fig. 26, and Fig. 25), The top three characteristics in SHAP that substantially influence the prediction are different from those in LIME and DALEX.

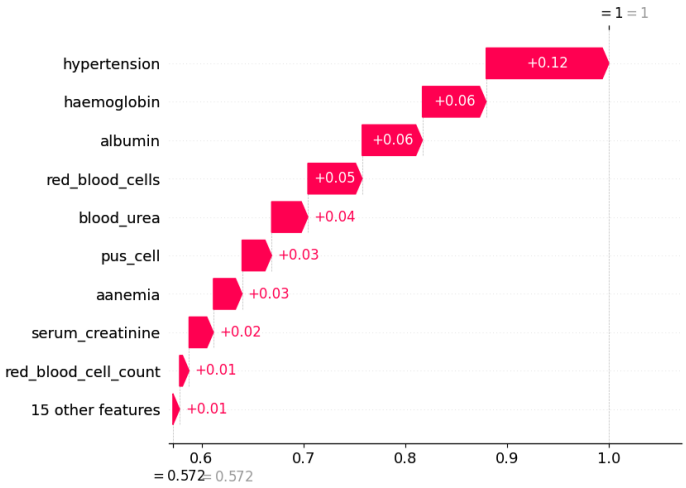


Fig. 24: Kidney disease SHAP instance explanation

It can be seen that while SHAP provided only the features that contributed towards diagnosing the patient as positive, LIME and DALEX also provided a few features that also worked against the predictions. Hypertension (value = 1) in all 3 frameworks was the most decisive factor in the prediction. The discrepancy for the other features can be due to a few factors, like the composition of the dataset and the quality of the data. From the plots, it can be seen that LIME and DALEX also provide the probability of predicting a class.

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

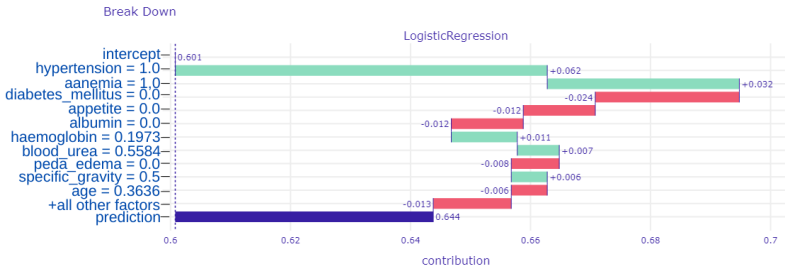


Fig. 25: Kidney disease DALEX instance explanation

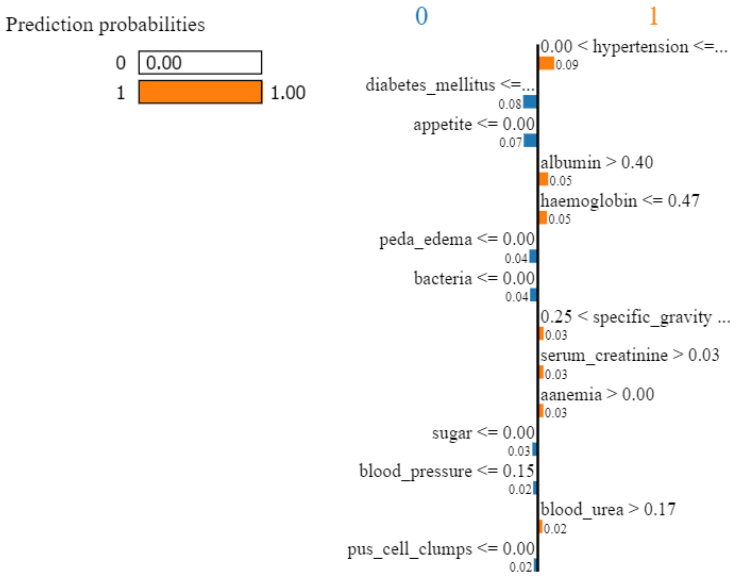


Fig. 26: Kidney disease LIME instance explanation

5 Threats to validity

Our work is still subject to various limitations, and there are threats to its validity. Upon properly assessing, we got to know that:

1. SHAP is one of the most reliable XAI techniques, but it was observed that, with an increase in the number of features and instances, the computation time for SHAP values also increased. SHAP is computationally very expensive. SHAP is computationally very expensive. For instance, while calculating SHAP values for second dataset of cardiovascular disease which constituted 14000 test instances it took SHAP 23 minutes and 18 seconds.

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

2. The results we got on our particular models may not be the same for other models. As every ML technique has its own approach to predicting, the results may vary with changing models.
3. Neural networks, one of the most complex models, were not part of our study.

6 Conclusion and future works

AI is very helpful in the healthcare sector because it not only cuts costs but also lessens the time taken to complete a task, be it the analysis of patient data or drug development. However, the question of its predictions' reliability remains a problem. Reliability cannot be achieved without establishing explainability in these black-box models. Before XAI can develop to a powerful enough level to handle interpretability, AI predictions will inevitably involve some risks and failures, just like any new technology, treatment, or medication we intend to offer in the healthcare industry [42]. XAI helps us achieve this through different interpretations of the predictions made by our models. While there are various XAI frameworks to use, we used three of the more commonly used frameworks: SHAP, LIME, and DALEX. To understand the interpretations produced by these frameworks, we used five different datasets with different qualities in each of them.

SHAP offers a very good visualization of both the global and local interpretations. It provides both trend and feature ranking in a single plot for the purpose of global interpretation, and it does so for all of the features. On the other hand, LIME provided great local interpretation. It provides the feature importance plot of all the features for a single instance, unlike SHAP and DALEX, who provide the feature importance plot of the top 10 features only. For the local interpretation of a model, LIME may be the most appropriate if the dataset contains a lot of features. Another advantage of using LIME is that users have the choice of determining how many of their features will be ranked. While considering DALEX, it gives a number of visualization tools that enable you to look into how the model responds to changes in the values of its input features. Although DALEX offers a greater number of visualization tools, the majority of these tools are not user-friendly. This means that it may be difficult for a non-specialist to interpret the plots that are provided because they are dependent on statistical information. Moreover, the computation time required by DALEX is an issue. This is due to the fact that DALEX might have to analyze the whole dataset in order to provide a local explanation. One important limitation of our work is that all five datasets contain linearly separable data.

AI is a dire need in the healthcare sector. Our work was limited only to tabular data, but there are other types of medical data that can be evaluated using XAI. Although the focus of our work has been on post-hoc explainability, establishability can also benefit greatly from ante-hoc explainability. While aiming for optimal accuracy or minimal error, ante-hoc techniques usually focus on examining a model's explainability from the beginning and during training to make it naturally explainable [43]. Trust in AI cannot be entrenched without integrating explainability into current black-box models. When it comes to time constraints, the current healthcare system still has a lot of issues. The majority of these issues can be resolved by using AI, but the first

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

and most important step in doing so is determining whether or not AI is explainable. Only then can reliability and trust be ensured in these black-box models.

Acknowledgements. The authors would like to gratefully acknowledge the grant GUJCOST/STI/2021-2022/3922 from the Government of Gujarat, India, and support from Pandit Deendayal Energy University (PDEU) to perform this work.

Conflict of Interest Declaration

The authors have no conflicts of interest to declare.

References

[1] Roth, G.A., Mensah, G.A., Johnson, C.O., Addolorato, G., Ammirati, E., Baddour, L.M., Barengo, N.C., Beaton, A.Z., Benjamin, E.J., Benziger, C.P., *et al.*: Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the gbd 2019 study. Journal of the American College of Cardiology **76**(25), 2982–3021 (2020)

[2] Chestnov, O.: World health organization global action plan for the prevention and control of noncommunicable diseases. Geneva, Switzerland (2013)

[3] Alowais, S.A., Alghamdi, S.S., Alsuhebany, N., Alqahtani, T., Alshaya, A.I., Almohareb, S.N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H.A., *et al.*: Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Medical Education **23**(1), 689 (2023)

[4] Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608 (2017)

[5] Khedkar, S., Subramanian, V., Shinde, G., Gandhi, P.: Explainable ai in healthcare. In: Healthcare (April 8, 2019). 2nd International Conference on Advances in Science & Technology (ICAST) (2019)

[6] Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N.: Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1721–1730 (2015)

[7] Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H.: Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **9**(4), 1312 (2019)

[8] Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L.: The ethics of algorithms: Mapping the debate. Big Data & Society **3**(2), 2053951716679679 (2016)

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

[9] Vaid, A., Jaladanki, S.K., Xu, J., Teng, S., Kumar, A., Lee, S., Somani, S., Paranjpe, I., De Freitas, J.K., Wanyan, T., *et al.*: Federated learning of electronic health records to improve mortality prediction in hospitalized patients with covid-19: machine learning approach. JMIR medical informatics **9**(1), 24207 (2021)

[10] Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., *et al.*: Using artificial intelligence to detect covid-19 and community-acquired pneumonia based on pulmonary ct: evaluation of the diagnostic accuracy. Radiology **296**(2), 65–71 (2020)

[11] Verma, A.K., Pal, S., Kumar, S.: Comparison of skin disease prediction by feature selection using ensemble data mining techniques. Informatics in Medicine Unlocked **16**, 100202 (2019)

[12] Nahar, N., Ara, F., Neloy, M.A.I., Biswas, A., Hossain, M.S., Andersson, K.: Feature selection based machine learning to improve prediction of parkinson disease. In: Brain Informatics: 14th International Conference, BI 2021, Virtual Event, September 17–19, 2021, Proceedings 14, pp. 496–508 (2021). Springer

[13] Shad, H.A., Rahman, Q.A., Asad, N.B., Bakshi, A.Z., Mursalin, S.F., Reza, M.T., Parvez, M.Z.: Exploring alzheimer's disease prediction with xai in various neural network models. In: TENCON 2021-2021 IEEE Region 10 Conference (TENCON), pp. 720–725 (2021). IEEE

[14] Ahsan, M.: Heart attack prediction using machine learning and xai. PhD thesis, Brac University (2022)

[15] Moreno-Sanchez, P.A.: Improvement of a prediction model for heart failure survival through explainable artificial intelligence. Frontiers in Cardiovascular Medicine **10** (2023)

[16] Moreno-Sánchez, P.A.: Data-driven early diagnosis of chronic kidney disease: Development and evaluation of an explainable ai model. IEEE Access **11**, 38359–38369 (2023)

[17] Gaur, L., Biswas, M., Bakshi, S., Gupta, P., Si, T., Mallik, S., Maulik, U.: An integrated model to evaluate the transparency in predicting chronic kidney disease using a trio-embedded explainable model. Available at SSRN 4129888

[18] Pires, I.M., Marques, G., Garcia, N.M., Ponciano, V.: Machine learning for the evaluation of the presence of heart disease. Procedia Computer Science **177**, 432–437 (2020)

[19] Muppalaneni, N.B., Ma, M., Gurumoorthy, S., Kannan, R., Vasanthi, V.: Machine learning algorithms with roc curve for predicting and diagnosing the heart disease. Soft computing and medical bioinformatics, 63–72 (2019)

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

[20] Mohan, S., Thirumalai, C., Srivastava, G.: Effective heart disease prediction using hybrid machine learning techniques. IEEE access **7**, 81542–81554 (2019)

[21] Magesh, G., Swarnalatha, P.: Retracted article: Optimal feature selection through a cluster-based dt learning (cdtl) in heart disease prediction. Evolutionary intelligence **14**(2), 583–593 (2021)

[22] Pal, S.: Prediction for chronic kidney disease by categorical and non_categorical attributes using different machine learning algorithms. Multimedia Tools and Applications, 1–14 (2023)

[23] Chittora, P., Chaurasia, S., Chakrabarti, P., Kumawat, G., Chakrabarti, T., Leonowicz, Z., Jasiński, M., Jasiński, Ł., Gono, R., Jasińska, E., *et al.*: Prediction of chronic kidney disease-a machine learning perspective. IEEE Access **9**, 17312–17334 (2021)

[24] Al-Azzam, N., Shatnawi, I.: Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer. Annals of Medicine and Surgery **62**, 53–64 (2021)

[25] Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. Advances in neural information processing systems **30** (2017)

[26] Ribeiro, M.T., Singh, S., Guestrin, C.: " why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144 (2016)

[27] Baniecki, H., Kretowicz, W., Piatyszek, P., Wisniewski, J., Biecek, P.: Dalex: Responsible machine learning with interactive explainability and fairness in python. The Journal of Machine Learning Research **22**(1), 9759–9765 (2021)

[28] Moreno-Sanchez, P.A.: Improvement of a prediction model for heart failure survival through explainable artificial intelligence. arXiv preprint arXiv:2108.10717 (2021)

[29] Aghamohammadi, M., Madan, M., Hong, J.K., Watson, I.: Predicting heart attack through explainable artificial intelligence. In: Computational Science–ICCS 2019: 19th International Conference, Faro, Portugal, June 12–14, 2019, Proceedings, Part II 19, pp. 633–645 (2019). Springer

[30] Keyl, P., Bockmayr, M., Heim, D., Dernbach, G., Montavon, G., Müller, K.-R., Klauschen, F.: Patient-level proteomic network prediction by explainable artificial intelligence. NPJ Precision Oncology **6**(1), 35 (2022)

[31] Amoroso, N., Pomarico, D., Fanizzi, A., Didonna, V., Giotta, F., La Forgia, D., Latorre, A., Monaco, A., Pantaleo, E., Petruzzellis, N., *et al.*: A roadmap towards

DOI: https://doi.org/10.54216/FPA.200213
Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025

breast cancer therapies supported by explainable artificial intelligence. Applied Sciences **11**(11), 4881 (2021)

[32] El-Sappagh, S., Alonso, J.M., Islam, S.R., Sultan, A.M., Kwak, K.S.: A multilayer multimodal detection and prediction model based on explainable artificial intelligence for alzheimer's disease. Scientific reports **11**(1), 2660 (2021)

[33] Saeed, W., Omlin, C.: Explainable ai (xai): A systematic meta-survey of current challenges and future opportunities. Knowledge-Based Systems **263**, 110273 (2023)

[34] Garg, A., Sharma, B., Khan, R.: Heart disease prediction using machine learning techniques. In: IOP Conference Series: Materials Science and Engineering, vol. 1022, p. 012046 (2021). IOP Publishing

[35] Boukhatem, C., Youssef, H.Y., Nassif, A.B.: Heart disease prediction using machine learning. In: 2022 Advances in Science and Engineering Technology International Conferences (ASET), pp. 1–6 (2022). IEEE

[36] Faieq, A.K., Mijwil, M.M.: Prediction of heart diseases utilising support vector machine and artificial neural network. Indonesian Journal of Electrical Engineering and Computer Science **26**(1), 374–380 (2022)

[37] Monirujjaman Khan, M., Islam, S., Sarkar, S., Ayaz, F.I., Kabir, M.M., Tazin, T., Albraikan, A.A., Almalki, F.A., et al.: Machine learning based comparative analysis for breast cancer prediction. Journal of Healthcare Engineering **2022** (2022)

[38] Rahman, M.M., Rahman, A., Akter, S., Pinky, S.A.: Hyperparameter tuning based machine learning classifier for breast cancer prediction. Journal of Computer and Communications **11**(4), 149–165 (2023)

[39] Swain, D., Mehta, U., Bhatt, A., Patel, H., Patel, K., Mehta, D., Acharya, B., Gerogiannis, V.C., Kanavos, A., Manika, S.: A robust chronic kidney disease classifier using machine learning. Electronics **12**(1), 212 (2023)

[40] Kim, J., Das, R.N., Lee, Y., Mukherjee, S., An, H., Medda, S.K.: Interrelationships of cholesterol with cardiac factors for heart patients. Journal of Cardiovascular Disease Research **10**(4) (2019)

[41] Shah, A., Haider, G., Abro, N., Hashmat, S., Chandio, S., Shaikh, A., Abbas, K.: Correlation between site and stage of breast cancer in women. Cureus **14**(2) (2022)

[42] Di Martino, F., Delmastro, F.: Explainable ai for clinical and remote health applications: a survey on tabular and time series data. Artificial Intelligence Review **56**(6), 5261–5315 (2023)

[43] Vilone, G., Longo, L.: Explainable artificial intelligence: a systematic review. arXiv preprint arXiv:2006.00093 (2020)

DOI: https://doi.org/10.54216/FPA.200213

Received: February 08, 2025 Revised: April 02, 2025 Accepted: June 06, 2025