Interpretable Artificial Intelligence in Cardiovascular Health: An In-depth Analysis of Heart Disease Data

”Ms Madhavilatha Working as Assistant Professor at Sri Venkateswara College Of Engineering,Tirupati.
Email: madhavilathavenkat@gmail.com”

”Ms G.T. Prasanna Kumari Working as Assosciate Professor at Sri Venkateswara College Of Engineering,Tirupati
Email: tabithaprasanna@gmail.com”February 2024

Abstract

With the escalating abundance of structured and unstructured data and the rapid advancements in analytical techniques, Artificial Intelligence (AI) is catalyzing a revolution in the healthcare industry. However, as AI becomes increasingly indispensable in healthcare, concerns are mounting regarding the lack of transparency, explainability, and potential bias in model predictions. Addressing these issues, Explainable Artificial Intelligence (XAI) emerges as a pivotal solution.
XAI plays a crucial role in fostering trust among medical practitioners and AI researchers, thereby paving the way for the broader integration of AI in healthcare. This paper aims to introduce diverse interpretability techniques, shedding light on the comprehensibility and interpretability of XAI systems. These techniques, when applied judiciously, offer significant advantages in the healthcare domain. Given that medical diagnosis models directly impact human life, it is imperative to instill confidence in treating patients based on instructions from seemingly opaque models.
The content of this paper includes illustrations grounded in the heart disease dataset, demonstrating how explainability techniques should be prioritized to establish trustworthiness when utilizing AI systems in healthcare.
Keywords: Explainable AI, Healthcare, Heart disease, Programming frame- works, LIME, SHAP, Example-based Techniques, Feature-based Techniques.

1 Introduction

For healthcare operations where an explanation of the essential sense is im- portant for people who form opinions, machine literacy’s lack of explainability restricts the wide-scale deployment of AI. However, also its threat of making a wrong decision may stamp its advantages of delicacy, speed, and decision-making efficacity If AI can not explain itself in the sphere of healthcare.
This would, in turn, oppressively limit its compass and mileage. thus, it’s very important to look at these issues closely. Standard tools must be erected before a model is stationed in the healthcare sphere. One similar tool is ex- plainability (or resolvable AI). The explanation behind the use of resolvable AI ways is to increase translucency, affect tracing, and model enhancement. For case, they explain why someone is distributed as ill or else.
This would increase the trust position of medical interpreters to calculate on AI. ultimately, XAI can be integrated into smart healthcare systems involving IoT, Cloud computing, and AI primarily used in the areas of cardiology, cancer, and neurology. These smart healthcare systems can also be used for diagnosing conditions and selecting the applicable treatment plan. In this paper, we look at some exemplifications of colorful XAI ways carried out on the Heart Disease Dataset from UCI 3 along with the use cases related to the fashion.

Objects: The idea of this paper is to study and use different explain- suit- able AI ways in the healthcare sector, as it gives translucency, thickness, fair- ness, and trust to the system. The particulars of the objects are – To study point-grounded and illustration-grounded resolvable AI ways using the heart complaint dataset.
– To draw out the consequences from the results of these ways and conclude the selection of one fashion over the other for a particular area of healthcare.
We've worked on colorful ways that give explanations of issues given by the black box models. The paper gives perceptivity on how these ways are profitable in different conditions. also, the different approaches followed by them are studied. figure. The following is the structure of the paper. Section 2 gives an overview of our approach, conforming to different phases of Machine literacy(similar to Model training and model deployment) and resolvable AI. A brief description of the dataset along with the explanation of features is presented in Section 3. Section 4 addresses the point grounded ways LIME and SHAP explaining the styles in detail with the support of exemplifications. We describe colorful partner-ample grounded ways in the coming Section 5. It gives perceptivity on the different ways available in the library justification and demonstrates their significance in resolvable AI. Incipiently, Section 6 discusses the findings of the entire paper and concludes the work.

2 Course of action

Course Of Action:
The idea of our exploration is to present our early exploration work on erect-

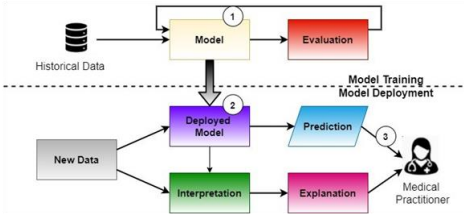

Figure 1: Machine Learning Journey Enhanced with Explainable AI (XAI)

ing a resolvable AI- AI-grounded approach for healthcare operations. Figure 1presents an ML life-cycle in confluence with XAI styles to gain lesser benefitfrom AI- AI-grounded black box models stationed in the healthcare sphere.Fig1**Model Training:** ML algorithms use the literal healthcare data collected
for training where they essay to learn idle patterns and relationships from data without hardwiring the fixed rules. We use the Heart Disease Dataset from the UCI ML Repository. This dataset contains 70 features. Its ideal is to describe the actuality of heart complaints in cases. The training leads to the generationof models, which are stationed in the product surroundings.
We use the XGBoost(Collaborative Extreme Grade Boosting) algorithm for training, which is a perpetuation of the grade-boosted decision trees developed for performance as well as speed. XGBoost(eXtreme Gradient Boosting) is a decision-tree-grounded machine literacy algorithm grounded on a boosting frame. Outliers don't have a significant impact on the performance of XGBoost. There's no demand for carrying out point engineering in XG- Boost moreover. This is one of the only algorithms that can decide point importance, which is an integral part of resolvable AI. XGBoost is largely suitable for any kind of bracket problem and thus, applicable to our paper and design.
Model Deployment and Interpretability: The ideal of enforcing partner-plainable AI ways along with the stationed model is to interpret how the vati-cination results are deduced. The data is supplied to the trained model and resolvable AI module. Enforcing resolvable AI ways, it allows us to give expla-nations along with vaticination results. The explanation can be consumed by medical interpreters to validate the predictions made by the AI models. Clinical records along with explanations can be used to induce deeper perceptivity and recommendations. Section 4 and Section 5 present the explanation generated by colorful resolvable AI ways.Before we present colorful resolvable AI ways and

their results, the coming section(Section 3) presents our heart complaint case study and describes the dataset.

3 Data Synopsis

Data Synopsis:

Heart Complaints are among the biggest causes of death in the entire world, the mortality rate will indeed increase in the post-COVID period as numerous heart problems arise due to it. vaticination of a heart complaint is one of the most important areas in the healthcare sector. Heart complaint refers to block- age or narrowing down of blood vessels, which can lead to chest pain or heart attack. The Heart Disease Cleveland UC Irvine dataset is grounded on vatici- nation if a person has a heart complaint or is not grounded on 13 attributes4. It's reprocessed from the original dataset having 76 features. In the following, we describe 13 features compactly

- 1. age An integer value signifying the age of an existent.
- 2. coitus 0 for womanish, 1 for manly.
- 3. CP stands for the casket Pain Type and ranges from 0- 3 depend- ing upon the symptoms endured by a case. They're classified using the symptoms endured by a case. The three main symptoms of angina are

– Substernal casket discomfort
– Provoked by physical exertion or emotional stress
– Rest and/ or nitroglycerine relieves

According to the symptoms endured, casket pain types are classified in thefollowing ways
(a) 0-Typical Angina passing all three symptoms
b) 1-Atypical Angina passing any two symptoms
c) 2-Non-Anginal Pain passing any one symptom
d) 3- Asymptomatic Pain Experiencing
none of the symptoms mentioned above still, lower exercise should be performed and the sugar and cholesterol position of the body should be maintained If cpis high.

4. trestbps (Resting Blood Pressure) shows the resting blood pressure calcu- lated in units of millimeters in mercury(mmHg). The ideal blood pressure is 90/ 60mmHg to 120/ 80mmHg. High blood pressure is considered to be anything above 140/ 90mmHg. still, lower exercise should be performed and if trestbpsis less, a separate drug should be taken, If trestbps is high.

5. chol(Serum Cholesterol) represents the cholesterol situations of an existent. A high cholesterol position can lead to blockage of heart vessels. immaculately, cholesterol situations should be below 170mg/ Dl for healthy adults. However, normal exercise should be performed, and lower unctuous food should be eaten,If cholestrol is high.

6. fbs (Fasting Blood Sugar) represents Fasting Blood Sugar situations of a case, by gauging the quantum of glucose present in the blood. A blood sugar position below 100 mmol/ L is considered to be normal. a) 1 signifies that the case has a blood sugar position in excess of 120mmol/ L

b) 0 signifies that the case has a blood sugar position lower than 120mmol/ L and If fbs is high, gusto diet should be taken, and frequent input of food in lower quantum should be taken

7 restecg(resting electrocardiogram) depicts the electrocardiograph results of a caseranges between 0- 2.

a) 0-Normal results in ECG

b) 1- The ECG Results have a ST- T surge abnormality for Heart complaint dataset https//www.kaggle.com/cherngs/ heart- complaint- cleveland- uci

c) 2- The ECG Results show a probable or definite left ventricular hyperactive- jewel by Estes criteria still, moderate exercise should be performed, gusto diet should be taken, If rest ecg is high

8 Thalach shows the maximum heart rate of an existent using a Thallium Test. A Thallium test is an unconventional system for checking heart complaint. It's carried out by edging in a small quantum of radioactive substance(Thallium in this case) into the bloodstream of an individual, while he she is exercising. Using a special camera, the blood inflow and the pumping of the heart can be determined. thalach denotes the maximum heart rate achieved during this Thallium test.However, proper exercise should be performed if thalach is low.

9 exang is a point that reveals whether a case has exercise convinced angina(signified by 1) or not(signified by 0). Exercise convinced angina is a kind of angina that's touched off by physical exertion and exertion due to an increasein the demand of oxygen.

10 Old peak is the quantum of depression of the ST Wave in the case of a case having a value of 1 in rest ecg (ST- T Wave abnormality set up in ECG Results). This peak is convinced by exercise and is measured relative to the re-sults at rest.However, lower exercise should be performed, and gusto diet shouldbe taken if old peak is high.

11 pitch pitch is also concerned with the ST Wave in ECG Results.

a) 0 signifies an upward pitch in the ST Wave

b) 1 signifies that the ST Wave is flat

c) 2 signifies a downcast pitch in the ST Wave

still, separate drugs should be taken, and gusto diet should be taken, If thepitch is high.

12 ca The heart has 3 main vessels responsible for blood inflow. An angiog- raphy is carried out and because of the color, the unblocked vessels show up on a X-Ray. immaculately, three vessels should be visible on the X-Ray as this would mean that none of the vessels are blocked.However, angioplasty should beperformed for the treatment of blocked vessels, at the after stage stent should be put if needed, If ca is high.

13 Thal denotes the results of Thallium test of an individual.that denotes the results of Thallium test of an existent.

a) 0 denotes that the results were normal.

b) 1 denotes a fixed disfigurement.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	condition
0	69	1	0	160	234	1	2	131	0	0.1	1	1	0	0
1	69	0	0	140	239	0	0	151	0	1.8	0	2	0	0
2	66	0	0	150	226	0	0	114	0	2.6	2	0	0	0
3	65	1	0	138	282	1	2	174	0	1.4	1	1	0	1
4	64	1	0	110	211	0	2	144	1	1.8	1	0	0	0

Figure 2: Instances of Dataset

c) 2 denotes a reversible disfigurement.

Then, disfigurement symbolises an inhibition in optimum blood inflow.Thalliumtest is done in a physically wielded state. Fixed disfigurement conveys a dis- figurement that stays indeed after the body is at rest. On the other hand, reversible disfigurement is a disfigurement that passes down as the body relaxes.

Figure 2 depicts illustration cases of the heart complaint dataset. Using this dataset, we present the explanation generated by colorful resolvable AI tech- niques.

4 Parameterized Techniques

Parameterized Techniques

This section introduces techniques for explaining feature-based models, which assess the impact of input features on a model's output. Various feature-based methods are available, such as Permutation Feature Importance, Partial De- pendence Plots (PDPs), Individual Conditional Expectation (ICE) plots, Ac- cumulated Local Effects (ALE) Plot, Global surrogate models, Local Inter- pretable Model-agnostic Explanations (LIME), and Shapley Additive Expla- nations (SHAP).

Local Interpretable Model-Agnostic Explanations (LIME):

LIME focuses on understanding a model by perturbing input data samples to observe changes in predictions. It enables local model interpretability, mod- ifying individual feature values to observe the resulting output impact. An example using a heart disease dataset illustrates the concept, displaying predic- tion probabilities and feature weights.

Shapley Additive Explanations (SHAP):

SHAP enhances model transparency using game theory concepts. It explains predictions by treating each feature value as a "player" in a game, where the prediction serves as the "pay-out." Shapley values determine fair distribution

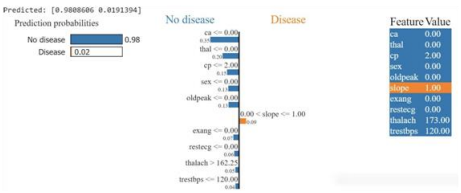

Figure 3: Generated by LIME

based on individual contributions. SHAP provides both local and global expla- nations, showcasing feature contributions in relation to a base value. Examples of local explanations using SHAP force plots demonstrate its behavior on specific instances, revealing the impact of individual features on predictions. Additionally, SHAP offers global explanations, showing how each feature contributes positively or negatively to the final prediction. Visualizations such as force plots and summary plots provide insights into feature importance and relationships between features. Scatter plots illustrate the effects of features on predictions at different values, aiding in understanding the model's behavior.

Territorial justification

To demonstrate how the SHAP values operate as a local explanation method, we shall run multiple instances. The three specific examples of randomly chosen cases that will be used to illustrate the approach are provided below. Every instance yields a unique collection of SHAP values, which allow us to understand the predictors' contributions as well as the reasons behind each case's specific prediction.
Using SHAP, we've observed two extreme scenarios similarly to LIME. Now, we examine a local explanation that is less extreme. In the provided figure, we observe a balanced distribution of Red and Blue feature values. Evaluating each case separately:

The key feature values impacting the model's decision regarding the presence of heart disease are thal=2 (indicating potential defects in blood supply and heart cell quality), cp=3 (suggesting asymptomatic chest pain, which, counter intuitively, is the most severe type among the four and leads the model to predict heart disease), and chol=307 (elevated cholesterol levels contributing to blood vessel blockage and reduced overall blood flow in and around the heart).
The primary feature value influencing the model's decision that the patient does not have heart disease is ca=0. This underscores the significance of this feature, indicating that none of the patient's vessels are obstructed.

Global explanation

Global explanation refers to an overview of how various features collectively contribute to a model's predictions. It entails analyzing the cumulative SHAP (SHapley Additive exPlanations) values to understand the extent and direction of each feature's influence on the final prediction. Various types of plots can effectively illustrate this global explanation, showcasing the overall impact and

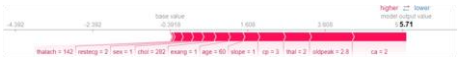

Figure 4: generated by SHAP force plot

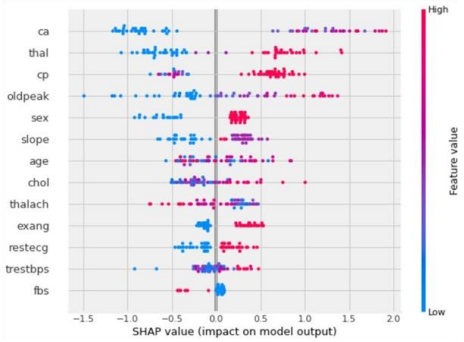

Figure 5: Enter Caption directionality of features on the model's outcomes

The global explanation of the model's predictions is based on the SHAP Val- ues, where -0.3918 represents the base value. If the total value exceeds -0.3918, it indicates the presence of the disease, while if it falls below, it suggests the absence of the disease. In the graph, the blue segment contributes to lowering the prediction, while the red segment increases it. Consequently, instances with predominantly red-colored features tend to be classified as 1 (indicating the presence of the disease), whereas those with mostly blue features are classified as 0 (indicating the absence of the disease).
The scatter plot graph depicted in Figure 5 offers a visual representation of how features influence predictions across different values. Features positioned towards the top of the plot hold greater significance for the model compared to those at the bottom, indicating higher feature importance. The color gradient in the plot corresponds to the value of each feature, with blue denoting low values, purple indicating median values, and red representing high values.
For instance, when examining the feature "ca," it becomes evident that when the dots appear blue, the SHAP values are predominantly negative. Conversely, when the dots shift towards red and purple, the SHAP values tend to be positive. This observation suggests that when no vessels are blocked (as indicated by blue dots), the likelihood of disease occurrence is low. However, as the number of blocked vessels increases, so does the likelihood of disease onset.

Counterfactual Theories of Causation

When applying a machine learning model to real-world data, understanding the rationale behind its decisions is crucial. In addition to explaining the reasons for a decision outcome, it's equally important to know what changes in features would lead to a different prediction. Counterfactual explanations, a model-agnostic technique within explainable AI (XAI), address this need by identifying the smallest adjustments to feature values required to switch the model's outputto a predefined outcome.

Counterfactual explainer methods are designed to work with black box mod- els, providing insights into how altering feature values can influence predictions. While they are most effective with binary datasets, they can also be applied to classification datasets with more than three target values, although their performance may not be as optimal in such cases. In essence, counterfactual explanations illustrate the impact of small changes in independent variables (e.g., X) on the dependent variable (e.g., Y), thus facilitating the calculation ofnecessary modifications needed to achieve desired outcomes.

References:

1. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual explanations without opening the black box: Automated decisions and the gdpr (2018) 2.Thampi, A.: Interpretable AI, Building explainable machine learning systems. Manning Publications, USA (2020)
3.Jiang,F.,Jiang,Y.,Zhi,H.,Dong,Y.,Li,H.,Ma,S.,Wang,Y.,Dong, Q.,Shen, H., Wang, Y.: Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology2(4), 230–243 (2017).https://doi.org/10.1136/svn-2017- 000101,https://svn.bmj.com/content/2/4/230
4. Looveren, A.V., Klaise, J.: Interpretable counterfactual explanations guidedby prototypes (2020)
5. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predic- tions.In: Advances in neural information processing systems. pp. 4765–4774 (2017)
6. Pawar,U.,O'shea,D.,Rea,S.,O'Reilly,R.:Explainable ai in health-care. In: 'Reasonable Explainability' for Regulating AI in Health (06 2020). https://doi.org/10.1109/CyberSA49311.2020.9139655
7. Ribeiro, M.T., Singh, S., Guestrin, C.: Anchors: High-precision model- agnostic explanations. In: AAAI Conference on Artificial Intelligence (AAAI) (2018)
8. Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep net- works.arXiv preprint arXiv:1703.01365 (2017)
9. Garc´ıa, M.V., Aznarte, J.L.: Shapley additive explanations for no2 fore- casting. Ecological Informatics56, 101039 (2020)
10.Dhurandhar,A.,Chen,P.Y.,Luss,R.,Tu,C.C.,Ting,P.,Shanmugam,K.,Das,P Ex- planations based on the missing: Towards contrastive explanations with pernent negatives. In:AdvancesinNeuralInformationProcessingSystems.pp.592– 603 (2018)