Heart Diases Classification using Machine Learning Techniques ans Explainable AI

Abdelrahman Osheba
Artificial Intillegence Science
Galala University

Ahmed Abou-el-ela
Artificial Intillegence Science
Galala University

Ahmed Alaa-eldin
Artificial Intillegence Science
Galala University

Youssef Hussieny
Artificial Intillegence Science
Galala University

Adham Mohamed
Artificial Intillegence Science
Galala University

Osama Adel
Artificial Intillegence Science
Galala University

Aser Mohammed
Artificial Intillegence Science
Galala University

Abstract

Heart disease encompasses a variety of conditions affecting the heart's structure and function and is the leading cause of death in the United States. Coronary artery disease (CAD), where arteries fail to deliver sufficient oxygen-rich blood to the heart, is the most prevalent type, affecting 20.5 million U.S. adults and causing approximately 660,000 deaths annually. The economic burden of heart disease is significant, with an annual cost of $219 billion in the United States. By 2030, the cost of heart failure alone is projected to reach $70 billion per year.

Early detection and diagnosis are crucial in improving patient outcomes. Machine learning (ML) and explainable AI (XAI) offer promising solutions for early heart disease detection. This study integrates various ML techniques, including Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbors, XGBoost, and Neural Networks, to predict heart disease using a public health dataset from Kaggle. This dataset includes 1,025 entries and 14 variables, such as age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate, exercise-induced angina, ST depression, slope of ST segment, number of major vessels colored by fluoroscopy, and thalassemia.

Ensemble learning with a Voting Classifier was employed to enhance performance and accuracy. Explainable AI techniques, particularly and Local Interpretable Model-agnostic Explanations (LIME), were used to provide transparency in model predictions. These methods help build trust among healthcare providers and patients by explaining how models arrive at diagnoses.

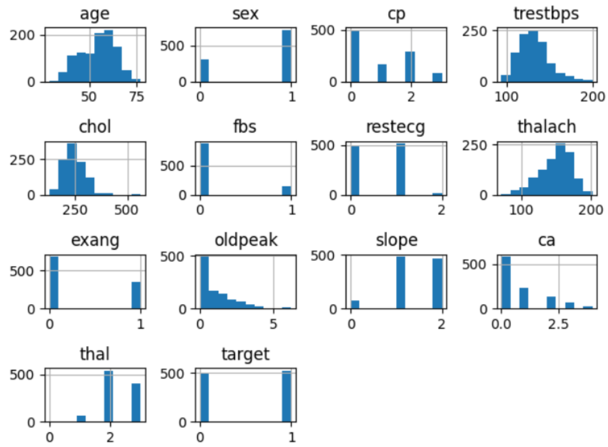
The dataset's analysis reveals critical indicators of heart disease, including age, sex, chest pain type, blood pressure, cholesterol, blood sugar levels, ECG results, maximum heart rate, exercise-induced angina, ST depression, slope, number of colored vessels, and thalassemia. These features are instrumental in developing accurate predictive models for early diagnosis and treatment planning of heart disease.

I. Introduction

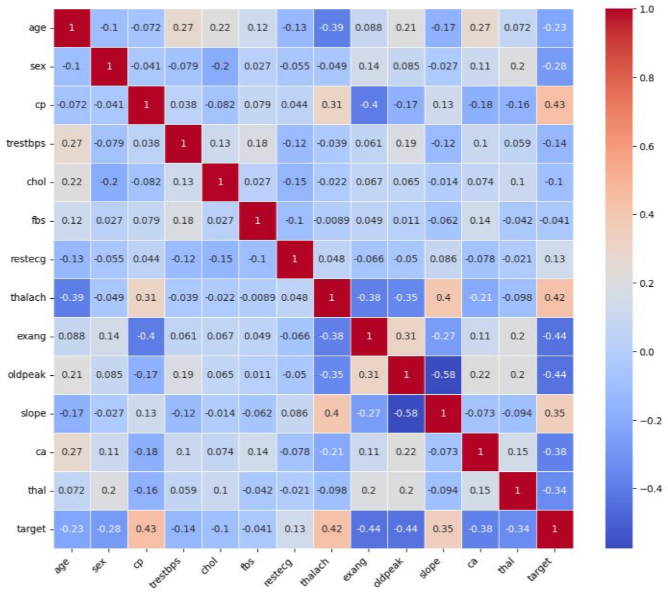
Heart disease is a catch-all phrase for a variety of conditions that affect the heart's structure and how it works also known as a range of conditions that affect the heart process. Heart disease is the leading cause of death in the United States. As example coronary heart disease is a type of heart disease where the arteries of the heart cannot deliver enough oxygen-rich blood to the heart. It is also sometimes called coronary artery disease or ischemic heart disease. About 20.5 million U.S. adults have coronary artery disease, making it the most common type of heart disease in the United States, according to the Centers for Disease Control and Prevention ,also in United states there is about 660,000 people die from heart disease every year that's 1 in every deaths and on average every 40 seconds , the heart diseases costs the united states about 219 $ billions each year this total includes all the services , According to **cardiovascular business** The annual cost of caring for a patient with Heart Failure as example is almost $30,000 in the United States. **Citing American Heart Association data**, the authors wrote that by 2030, Heart Failure costs in the United States are expected to be at least $70 billion per year by 2030, with the total cost of caring for HF patients approaching $160 billion.The largest economic burden linked to Heart Failure is from hospitalizations and rehospitalizations. According to the authors, 75% to 80% of the direct costs for Heart Failure are attributable to inpatient hospital stays. prompting the need for effective early detection methods. Early diagnosis can significantly improve patient outcomes by enabling timely intervention and management. Machine learning (ML) and explainable AI (XAI) are emerging as powerful tools in the healthcare sector, offering promising solutions for the early detection of heart disease. This introduction explores the integration of ML and XAI in this critical area. In machine learning techniques we worked with **Logistic Regression** a statistical model that predicts the probability of a binary outcome, such as the presence or absence of heart disease., **Random Forest** which an ensemble of decision trees that improves prediction accuracy by reducing overfitting ,**Support Vector Machine** which is a classification method that finds the hyperplane that best separates different classes ,**KNN** uses proximity to make classifications or predictions about the grouping of an individual data point , **XGBoost** that is an advanced gradient boosting algorithm that efficiently handles large datasets and complex models by iteratively improving predictions through ensemble learning, and tasks to enhance performance and accuracy and **Neural Networks** that are Models inspired by the human brain, capable of capturing complex relationships in data Finally from ensemble we used (Majority Voting) Voting Classifier to choose the best classifier among our models . Our Data is from Kaggle under the category of Public Health Dataset it consist of 1025 row entries and 14 columns This dataset is instrumental in the study and prediction of heart disease, encompassing a comprehensive range of patient demographics, clinical measurements, and diagnostic information. Then The Role of Explainable AI in our project is providing clear explanations of how a model arrives at a particular diagnosis, XAI helps build trust among healthcare providers and patients We mainly used one of the techniques of (LIME : Local Interpretable Model-agnostic Explanations) focuses on explaining the model’s prediction for individual instances

II. DATA COLLECTION

Our data source is a public health dataset from Kaggle, comprising 1,025 entries and 14 columns. The dataset includes the following variables:
Age: Age of the patient in years. Increasing age is generally associated with a higher risk of heart disease.
Sex: Sex of the patient, where 1 indicates male and 0 indicates female. Males are generally at a higher risk of heart disease compared to females.
Chest Pain Type (cp):
0: Typical angina
1: Atypical angina
2: Non-anginal pain
3: Asymptomatic
Different types of chest pain are indicators of heart disease, with typical angina being a stronger indicator.
Resting Blood Pressure (trestbps): Blood pressure measured in mm Hg on hospital admission. Higher blood pressure is a known risk factor for heart disease.
Serum Cholesterol (chol): Cholesterol level in mg/dl. High cholesterol levels can lead to the development of plaques in arteries, contributing to heart disease.
Fasting Blood Sugar (fbs): Indicates if fasting blood sugar is >120 mg/dl, where 1 is true and 0 is false. Elevated fasting blood sugar levels can indicate diabetes, which is a risk factor for heart disease.
Resting Electrocardiographic Results (restecg):
0: Normal
1: ST-T wave abnormality
2: Left ventricular hypertrophy by Estes' criteria
Abnormal ECG results can indicate heart problems.
Maximum Heart Rate Achieved (thalach): Maximum heart rate achieved during a stress test. Lower maximum heart rate might be an indicator of poorer heart function.
Exercise Induced Angina (exang): Indicates whether exercise induced angina (1 = yes, 0 = no). Angina induced by exercise is a strong indicator of heart disease.
ST Depression Induced by Exercise Relative to Rest (oldpeak): This value represents the difference in ST depression induced by exercise. Higher values can indicate significant heart problems.
Slope of the Peak Exercise ST Segment (slope):
0: Upsloping
1: Flat
2: Downsloping
The slope of the ST segment can indicate the severity of heart disease, with downsloping being the most severe.
Number of Major Vessels Colored by Fluoroscopy (ca): The number of major vessels (0-3) colored by fluoroscopy. Higher numbers indicate more significant heart disease.
Thalassemia (thal):
1: Normal
2: Fixed defect
3: Reversible defect
Different types of thalassemia defects can affect the heart differently, with reversible defects being a serious concern.
Target: Diagnosis of heart disease (0 = no heart disease, 1 = heart disease). This is the outcome variable we aim to predict using the features above.



Summary of Data Set:
Age and Sex: Older age and being male are associated with higher heart disease risk.
Chest Pain Type: Typical angina is strongly correlated with heart disease, while asymptomatic cases are less so.
Resting Blood Pressure and Cholesterol: High levels are significant risk factors.
Fasting Blood Sugar: Elevated levels suggest diabetes, which is a heart disease risk factor.
ECG Results: Abnormal results indicate potential heart issues.
Maximum Heart Rate: Lower rates during stress tests can indicate poor heart health.
Exercise Induced Angina: Indicates a strong likelihood of heart disease.
ST Depression and Slope: Higher ST depression and downsloping ST segment are significant indicators.
Fluoroscopy: Higher numbers of colored vessels suggest more severe heart disease.
Thalassemia: Certain types can impact heart health differently.



III. **O**VERALL, THIS DATASET CONTAINS CRITICAL INDICATORS THAT CAN BE USED TO PREDICT THE PRESENCE OF HEART DISEASE. ANALYZING THESE FEATURES USING MACHINE LEARNING MODELS CAN HELP DEVELOP ACCURATE PREDICTIVE MODELS TO AID IN EARLY DIAGNOSIS AND TREATMENT PLANNING

IV. METHEDOLOGY

Dataset description and preprocessing

The dataset from Kaggle, ‘heart.csv’, is a compilation of medical records, encapsulating metrics like Age, Cholesterol levels, and trestbps Acquired from a prominent medical research database, it offers vital indicators hinting at potential heart disease. For optimal predictive modelling, thorough data preprocessing is paramount. An initial inspection ascertained the absence of missing values. The entire heatmap is purple (or without yellow lines), which indicates there are no missing values in the dataset, as shown in the following FIG 1. Variables, namely ‘Sex’ and ‘chest pain type (4 values)’,they were transformed numerically from (0 – 3 label) without doing label encoding. Metrics ‘RestingBP’ and ‘Cholesterol’ underwent normalization using MinMaxScaler() to foster consistency, The data set contains 1225 sample instances as shown in Table 1. The dataset contains 13 clinical features as shown in Table 1. Different types of python libraries such as pandas, Sklearn, NumPy, matplotlib are used for processing the algorithms. Using explorative data analysis technique data was analyzed in VS Jupyter notebook. fold cross validation technique is used for spitting the data set into training and testing data.

Model training.

At the heart of any predictive analysis lies the model training phase. By splitting the dataset into training and testing subsets, this ensures that our model is not just memorizing the data (overfitting) but is genuinely deriving patterns and relationships within the data. In essence, this module metamorphoses raw data into discernible patterns and trends, which, when interpreted correctly, become invaluable insights that can drive proactive health interventions.

Models and Theoretical Framework

A. K-Nearest Neighbors (KNN)

Theory: KNN classifies a data point based on how its neighbors are classified.

Mathematics: Given a query point xxx, the algorithm computes the distances to all n points in the dataset and selects the k points with the smallest distances.

Assumptions: Assumes that similar points are close in the feature space.

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2}$$

where $d(x,y)$ is the Euclidean distance between points x and y.

B. Logistic Regression

Logistic Regression is a statistical model used for binary classification problems.

- Theory: Logistic regression estimates the probability that a given input point belongs to a certain class.
- Mathematics: The logistic function (sigmoid function) is used to model the probability:

Assumptions: Assumes a linear relationship between the

$$P(Y=1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n)}}$$

where β_i are the parameters to be estimated.

input features and the log-odds of the outcome.

C. Support Vector Machine (SVM)

SVM is a supervised learning model used for classification and regression tasks.

- Theory: SVM finds the hyperplane that best separates the classes in the feature space.
- Mathematics: The decision boundary is defined as:

Assumptions: Assumes that the data is linearly separable in the feature space or can be made linearly separable by a kernel function.

$$w \cdot x - b = 0$$

The optimization problem for SVM is:

$$\min \frac{1}{2}\|w\|^2 \quad \text{subject to} \quad y_i(w \cdot x_i - b) \geq 1 \, \forall i$$

D. Random Forest

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification or mean prediction for regression.

- Theory: Random Forest creates multiple decision trees using random subsets of the data and features, then aggregates their predictions.
- Mathematics: Each tree is built from a bootstrap sample of the data:
- Assumptions: Assumes that aggregating the predictions of multiple overfitted decision trees will reduce variance.

$$\text{Prediction} = \frac{1}{K}\sum_{k=1}^{K} T_k(x)$$

E. XGBoost

XGBoost is an ensemble learning method that uses gradient boosting for classification and regression.

- Theory: XGBoost builds an ensemble of trees sequentially, where each tree tries to correct the errors of the previous trees.
- Mathematics: The model minimizes the following regularized objective:

$$\text{Objective} = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k)$$

where l is a differentiable loss function and Ω is a regularization term.

Assumptions: Assumes that the weak learners (trees) can be improved by focusing on difficult cases.

F.Neural Network

Neural Networks are a set of algorithms designed to recognize patterns by learning from data.

- Theory: Neural networks consist of layers of interconnected nodes (neurons) where each connection has an associated weight.
- Mathematics: The output of each neuron is given by:

$$a^{(l)} = f\left(W^{(l)}a^{(l-1)} + b^{(l)}\right)$$

where $a^{(l)}$ is the activation of the l-th layer, $W^{(l)}$ are the weights, $b^{(l)}$ are the biases, and f is an activation function (e.g., ReLU, sigmoid).

- Assumptions: Assumes a large enough neural network can approximate any continuous function given sufficient data.

$$\hat{y} = \text{mode}(y_{(i)}) \quad \text{for} \quad i = 1, \ldots, k$$

2.**Methodological Approach**

2.1. Data Collection and Preprocessing

- Data Sources: Describe the origin of your data (e.g., public datasets, experiments, surveys).
- Preprocessing Steps: Include normalization, missing value handling, feature extraction, and selection.

$$X' = \frac{X - \mu}{\sigma}$$

where X' is the normalized data, μ is the mean, and σ is the standard deviation.

2.2. **Model Selection and Training**

- KNN: Select the number of neighbors kkk through cross-validation. Use distance metrics like Euclidean distance.

- Logistic Regression: Use maximum likelihood

$$\hat{\beta} = \arg\max_{\beta} \sum_{i=1}^{n} \left(y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)\right)$$

estimation to determine the parameters β\betaβ.

- SVM: Use a kernel trick if the data is not linearly separable. Select hyperparameters using grid search.

$$\hat{y} = \sum_{m=1}^{M} \gamma_m f_m(x) \quad \text{with} \quad \gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(m-1)} + \gamma f_m(x_i))$$

- Random Forest: Grow each tree using a bootstrap sample and select a random subset of features at each split.

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} T_t(x)$$

- Neural Network: Define the network architecture, initialize weights, and use backpropagation to minimize the loss function.

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i; \theta))$$

XAI

SHAP (SHapley Additive exPlanations)
SHAP values provide a unified measure of feature importance based on Shapley values from cooperative game theory. For a prediction
$f(x)$, the SHAP value for feature

$$\phi_i(f, x) = \sum_{S \subseteq \{1,\ldots,n\} \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} \left[f(S \cup \{i\}) - f(S)\right]$$

where S S is a subset of all features, and $f(S)$ f(S) is the model's prediction based on features in S S.
LIME (Local Interpretable Model-agnostic Explanations)
LIME explains individual predictions by fitting a local surrogate model around the prediction. The surrogate model is typically a simple, interpretable model like a linear model. The explanation is given by:

$$\min_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

where g g is the surrogate model from the family of interpretable models G G, L L is the loss function, π x is a proximity measure that weights the instances by their distance to x, and Ω is a regularization term to ensure simplicity of g

2.3. Model Evaluation

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

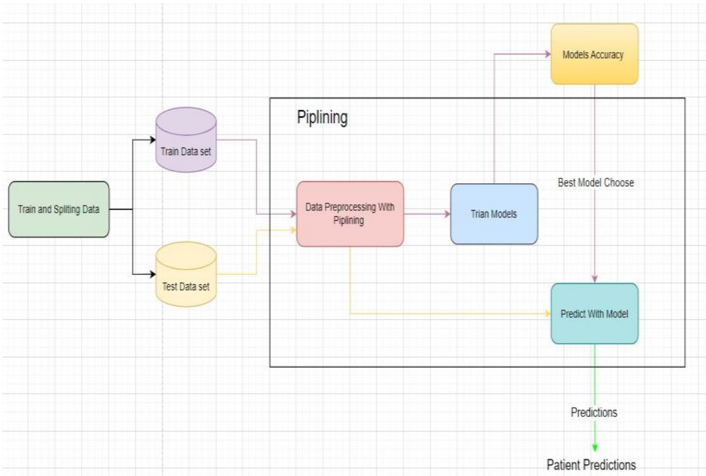
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

- Performance Metrics: Use accuracy, precision, recall, F1-score for classification, and RMSE for regression.

- Validation Methods: Apply k-fold cross-validation to assess model performance.

$$\text{CV Error} = \frac{1}{k} \sum_{i=1}^{k} \text{Error}_{\text{fold } i}$$

Model Flow



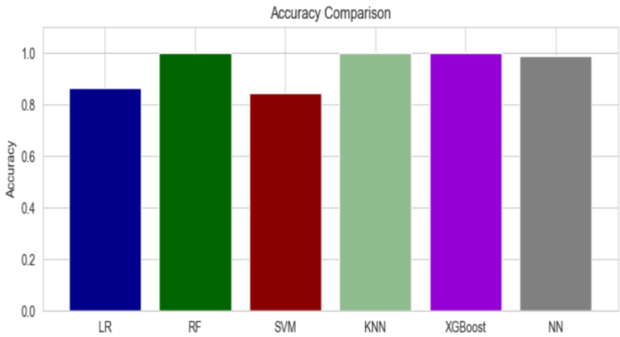
RESULTS

Accuracy

The percentage of occurrences that are successfully categorized as heart disease patients is known as accuracy. It is determined by dividing the total count of the presence of heart disease who were accurately predicted (true positive) by the total count of the absence of heart disease who were accurately predicted (true negative). It is determined as

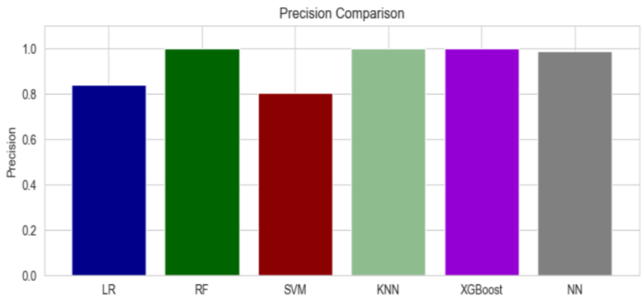
$$\frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Accuracy



Precession

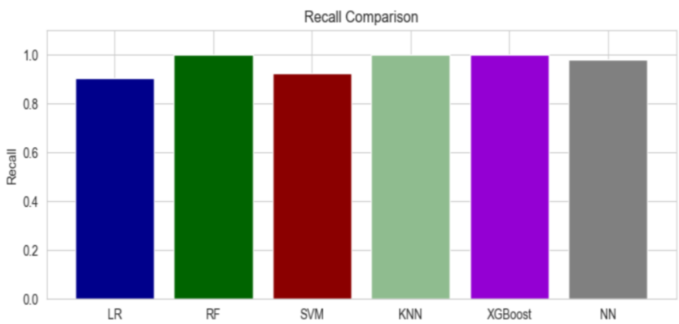
Precision is the measure to find the capacity of the heart disease prediction model to recognize only the relevant instances in the dataset. It is calculated as $\frac{TP}{TP+FP}$



Recall

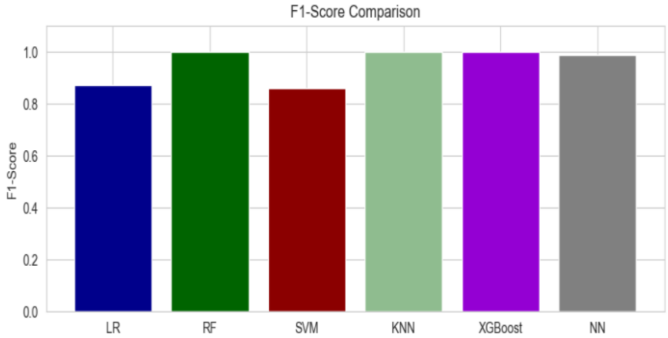
Recall can measure the heart disease prediction method's capacity to identify all the data instances of interest in a dataset. It is calculated as

$$\frac{TP}{TP + FN}$$



F1-Score

The F1-Score can measure the heart disease prediction method's capacity to balance precision and recall in a dataset. It is calculated as $2 \cdot \frac{Precesion * Recall}{Precesion + Recall}$



After Finding the models accuracy is too high to ensure choosing a model don't have over fitting we used SHAP in XAI to see the data and the impact of each features in model we found that the SVM model s the best describing data and not hallucinate so we have chosen it to be in the interface and then apply majority voting on them to choose the best and the chosen by majority voting used in our interface

Conclusion

This study demonstrates the potential of machine learning (ML) and explainable AI (XAI) techniques in the early detection and diagnosis of heart disease. By leveraging a public health dataset from Kaggle and applying various ML models, including Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbors, XGBoost, and Neural Networks, we achieved high accuracy in predicting heart disease. Ensemble learning, particularly using a Voting Classifier, further enhanced model performance and accuracy. Explainable AI techniques, such as Local Interpretable Model-agnostic Explanations (LIME), provided transparency in model predictions, fostering trust among healthcare providers and patients. The analysis identified critical indicators of heart disease, including age, sex, chest pain type, blood pressure, cholesterol, blood sugar levels, ECG results, maximum heart rate, exercise-induced angina, ST depression, slope, number of colored vessels, and thalassemia. These features are crucial in developing predictive models for early diagnosis and treatment planning. Overall, the integration of ML and XAI offers a powerful approach to improving patient outcomes through early detection and precise diagnosis of heart disease, potentially reducing the significant economic burden associated with this condition. The Support Vector Machine (SVM) model was selected for its balanced performance, avoiding overfitting while providing reliable predictions, making it a suitable choice for practical implementation in healthcare settings.

References

H. Niebling, "JetAdvice release Information | JetAdvice Updates," *JetAdvice*, Mar. 25, 2020. https://www.jetadvice.com/support/release-information/

J. W. Johnson, "Automatic Nucleus Segmentation with Mask-RCNN," *Advances in Intelligent Systems and Computing*, pp. 399–407, Apr. 2019, doi: 10.1007/978-3-030-17798-0_32.

N. Hou *et al.*, "Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost," *Journal of Translational Medicine*, vol. 18, no. 1, Dec. 2020, doi: 10.1186/s12967-020-02620-5.

Azizozmen, "Heart_Failure_Predict_8_Classification_Techniques," *Kaggle*, Oct. 20, 2021. https://www.kaggle.com/code/azizozmen/heart-failure-predict-8-classification-techniques

Arezoodahesh, "Heart failure prediction | With ensemble models," *Kaggle*, Jul. 15, 2022. https://www.kaggle.com/code/arezoodahesh/heart-failure-prediction-with-ensemble-models

Pythonafroz, "Heart disease prediction using 11 algorithms," *Kaggle*, Sep. 19, 2022. https://www.kaggle.com/code/pythonafroz/heart-disease-prediction-using-11-algorithms/log

E. Onose, "Explainability and Auditability in ML: Definitions, techniques, and tools," *neptune.ai*, Aug. 17, 2023. https://neptune.ai/blog/explainability-auditability-ml-definitions-techniques-tools

"Topics | IBM." https://www.ibm.com/topics/explainableai#:~:text=Explainable%20artificial%20intelligence%20(XAI)%20%20is,expected%20impact%20and%20potential%20biases

G. Visani, "Explainable machine learning - towards data science," *Medium*, Dec. 23, 2021. [Online]. Available: https://towardsdatascience.com/explainable-machine-learning-9d1ca0547ae0

M. Mittal, K. Kumar, and S. Behal, "Deep learning approaches for detecting DDoS attacks: a systematic review," *Soft Computing*, vol. 27, no. 18, pp. 13039–13075, Jan. 2022, doi: 10.1007/s00500-021-06608-1.

Diamond, "Predicting_Coronary_Heart_Disease_Using_an_Improved_LightGBM_Model_Performance_Analysis_and_Comparison," *Scribd*. https://www.scribd.com/document/639816395/Predicting-Coronary-Heart-Disease-Using-an-Improved%20LightGBM-Model-Performance-Analysis-and-Comparison

"What is coronary heart disease? | NHLBI, NIH," *NHLBI, NIH*, Dec. 20, 2023. https://www.nhlbi.nih.gov/health/coronary-heart-disease

"Annual heart failure costs in the US could surpass $70B by 2030," *Cardiovascular Business*, Sep. 01, 2023. [Online]. Available: https://cardiovascularbusiness.com/topics/clinical/heart-failure/annual-heart-failure-costs-us-could-surpass-70b-2030