Cardiovascular Disease Prediction Using Machine Learning and XAI

Rajul Gupta
B.tech, CSE
Bennett University
Noida, India

Arpit Panwar
B.tech, CSE
Bennett University
Noida, India

Shubhank Manhas
B.tech, CSE
Bennett University
Noida, India

Khushnood Ali
B.tech, CSE
Bennett University
Noida, India

***Abstract*— Heart illness may be a critical wellbeing condition that includes an eminent effect on worldwide mortality, particularly among more seasoned people. Per WHO information, a stunning 17. 9 million people pass absent each year due to heart illness, showing a concerning increment within the number of cases. Distinguishing issues ahead of time through the application of machine learning strategies such as K-Nearest Neighbor (KNN), Random Forest, Decision Tree,Logistic Regression, Support Vector Machine, and Naive Bayes has the potential to upgrade determining accuracy and decrease the event of human botches. This think about evaluates different machine learning calculations on datasets that incorporate traits like age, sex, sort of chest torment, blood weight, and glucose levels. Through the utilization of Logistic Regression, a forecast precision of 83. 90% was accomplished. These improvements have the potential to progress the early location and treatment viability, eventually profiting pending healthcare strategies.**

Keywords: Heart Disease, Machine Learning, Prediction, Random Forest, KNN, Support Vector Machine, Decision Tree, Logistic Regression, Naive Bayes, XGBoost

1. INTRODUCTION

Cardiovascular maladies (CVDs) are a driving worldwide wellbeing concern, causing 17. 9 million passings yearly and bookkeeping for 31% of all passings around the world. Early location and precise forecast of heart illness are vital in moderating mortality and complications. Machine learning (ML) offers promising arrangements for foreseeing CVDs by analyzing hazard variables such as age, sexual orientation, blood weight, cholesterol, and glucose levels. This investigate assesses different ML calculations, counting K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, Naive Bayes, and Decision Tree, for heart infection forecast. A hybrid model, such as the HRFM (Hybrid Random Forest Model), has accomplished 88. 7% exactness, whereas other calculations illustrated shifting degrees of affectability and specificity. For occasion, one calculation accomplished an exactness of 97. 65%, a affectability of 75%, and a specificity of 98. 44%. These strategies are successful in recognizing heart illness hazard components and foreseeing quiet results. The ponder highlights the benefits of ML-based frameworks, counting mechanization, decreased human mistake, and quickened demonstrative forms. In addition to providing valuable experiences for preventative measures, these devices aid patients and clinicians in making timely decisions and scheduling treatments, rather than merely categorizing the risks of cardiac disease. Forecast accuracy and consistency are increased by advanced techniques like Bayesian deduction and secure information preparation in advance. By utilizing clinical data, machine learning techniques improve the accuracy of identifying cardiac disease locations in comparison to traditional methods, offering a flexible, economical, and efficient setup for extensive use. This study highlights the importance of integrating these models into healthcare frameworks for significantly better outcomes and demonstrates how ML has the ability to transform the diagnosis and treatment of heart disease..

2. RELATED WORK

Numerous research efforts have underscored the potential of machine learning algorithms in forecasting heart diseases, utilizing their capacity to effectively process and evaluate clinical data. Senthil Kumar et al. [1] investigated machine learning methods such as Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression, revealing that integrating various models can improve predictive accuracy. In a similar vein, C. Sowmiya and P. Sumitra [2] analyzed nine classification techniques, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), KNN, and Naive Bayes, concluding that classification-based models, especially SVM, exceed traditional approaches in accuracy and reliability. Another important contribution was made by Anjan Nikhil Repaka et al. [3], who created a "Smart Heart Disease Prediction" (SHDP) system that utilized Naive Bayes to evaluate factors such as cholesterol levels and blood pressure, attaining high precision and incorporating data encryption to ensure secure predictions.

Hybrid methodologies have also shown potential. A model that merges Random Forest and linear approaches, referred to as the Hybrid Random Forest Model (HRFM), achieved an accuracy of 88.7%, as highlighted by researchers concentrating on feature selection and model integration to improve performance [4]. Real-world applications highlight the significance of automated systems for ongoing monitoring and rapid disease detection. For example, Rahul et al. [5] implemented feature engineering alongside ensemble techniques to enhance model clarity and accuracy. Research by Kumar et al. [6] and Singh et al. [7] emphasizes the importance of feature selection techniques like recursive feature elimination in boosting prediction efficiency. Additionally, Desai et al. [8] illustrated that deep learning models, such as convolutional neural networks (CNNs), efficiently examine imaging data for early detection, broadening machine learning's applicability beyond conventional data sources.

The incorporation of patient-specific data in predictive systems has been highlighted in studies like that of Mitra et al.

[9], who used ANN models to forecast outcomes based on individualized profiles. Similarly, research conducted by Patil and Kulkarni [10] demonstrated the effectiveness of ensemble learning strategies, including Bagging and Boosting, in capturing intricate patterns within diverse datasets. Feature significance analysis employing SHAP values, as shown by Gupta et al. [11], improves model transparency and assists healthcare practitioners in recognizing key predictors. Comparative studies by Roy et al. [12] and Sharma et al. [13] discovered that hybrid ensemble methods outperform single algorithms in accuracy while maintaining computational efficiency.

Lastly, secure and privacy-preserving techniques in machine learning, as presented by Narayan et al. [14], along with the use of federated learning for cardiovascular risk prediction by Joshi et al. [15], highlight advancements in ensuring data confidentiality while upholding prediction accuracy. Collectively, these studies illustrate the transformative impact of machine learning in heart disease prediction and its potential benefits for both healthcare providers and patients.

2.1 Different Approaches

Random Forest

The Random Forest calculation is habitually utilized in machine learning for both relapse and classification purposes. It capacities by producing a few Decision Trees from subsets of the dataset and after that deciding the ultimate expectation through either a larger part vote or the normal of these trees. The outfit method makes a difference avoid overfitting and improves the model's strength by accomplishing a adjusted result. The Irregular Timberland calculation is especially versatile and finds applications in a wide run of areas, counting picture classification, protest acknowledgment, and restorative conclusion.

K-Nearest Neighbor (KNN)

The K-Nearest Neighbor (KNN) strategy may be a nonparametric, directed machine learning calculation connected for assignments related to classification and relapse. The calculation easily distinguishes the k closest information points to the given input, empowering it to supply forecasts through a larger part vote or a weighted average calculation. KNN skips the preparing stage and instep employments the complete dataset for making expectations. In spite of the fact that KNN is straightforward, it has the potential to supply exact results, particularly when the parameter k is fine-tuned.

Decision Tree

The Decision Tree serves as a administered machine learning procedure utilized in errands including classification and relapse. The information is organized in a tree-like structure, with properties at inside hubs, choices along branches, and results at leaf hubs. Decision Trees are known for their ease of translation and viability in different applications, counting content classification, therapeutic conclusion, and monetary examination. All things considered, they may be helpless to overfitting when managing with broad datasets unless they are fastidiously fine-tuned or consolidated into outfit approaches such as random forest.

Support Vector Machine (SVM)

The Support Vector Machine (SVM) may be a directed learning calculation utilized for carrying out classification and relapse errands. SVM works by recognizing a hyperplane that successfully separates information focuses into partitioned classes. Diverse sorts of part capacities, such as direct, polynomial, and outspread premise work (RBF), are utilized to viably oversee straight as well as nonlinear information. Support Vector Machine, or SVM, demonstrate to be profoundly effective when working with datasets that have a tall number of measurements. They especially sparkle in assignments like picture classification and bioinformatics.

Naive Bayes

Naive Bayes, a probabilistic classification calculation, works on the standards of Bayes' Hypothesis. The strategy assumes include autonomy and computes the likelihood of information fitting into a particular category by analyzing the recurrence of include values. In spite of being basic, Naive Bayes demonstrates to be computationally proficient and exceedingly compelling in assignments such as spam discovery and content classification.

Logistic Regression

Logistic Regression could be a directed machine learning strategy utilized for both parallel and multiclass classification assignments. It illustrates the association between subordinate and free factors by utilizing a calculated (sigmoid) work to figure probabilities. Logistic Regression is known to be very compelling for datasets in which the relationship between factors is generally straight, making it a well known choice in different areas such as healthcare, showcasing, and fund.

XGBoost

XGBoost, brief for Extraordinary Angle Boosting, may be a flexible and conveyed system for gradient-boosted Decision Trees. It carefully coordinating frail learners (Decision Trees) step by step, diminishing a regularized objective work through the utilize of angle plunge. XGBoost really sparkles when it comes to overseeing broad datasets and complex issues, giving remarkable precision and effectiveness. It's regularly utilized in fruitful competition arrangements for errands including relapse, classification, and positioning.

3. PROPOSED METHODOLOGY

Figure 1 illustrates the machine learning system's block diagram. This dataset, which comprised all of the attributes and values, was used by the system. To begin, we searched the dataset for any category values, which include numerous categorical values. The gender attribute is one of the columns that is transformed into 0 and 1 integer values.

We used the correlation matrix, a function based on group features, to assess the link and plot the results to better comprehend them. The dataset is next examined for null or missing values.

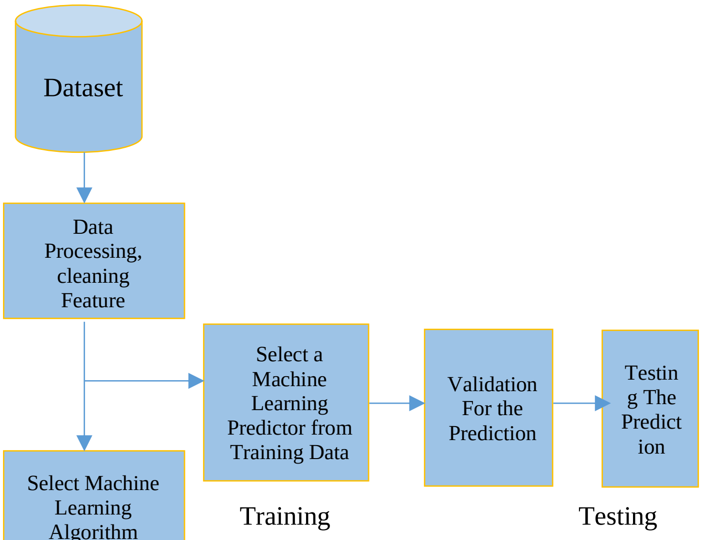

Figure 1: System block diagram.

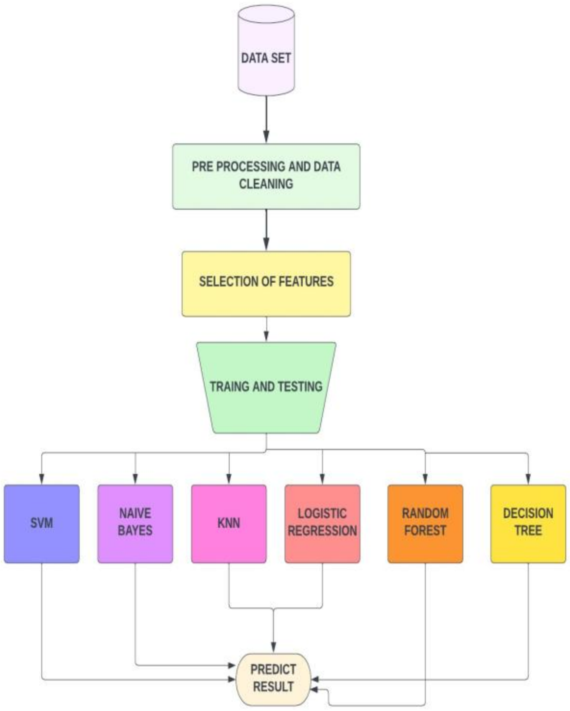

Figure 2: System block diagram

The properties essential to create the expectation were distinguished, taken after by characterizing the target esteem for the show to foresee. The dataset was hence part into two areas: one for preparing and approval, and another for testing. The gather was part utilizing Random testing, driving to an uneven conveyance between the preparing and testing areas. Hence, a stratified inspecting approach was utilized, comprising of an 80% part for preparing and approval, and a 20% part for testing. The characteristics were in this way balanced to align with standard hones within the industry. In arrange to pick up a more profound understanding of the circumstance, a few histograms and diffuse plots were produced for the preparing information subset. Taking after that, the system's preparing commenced.

3.1 DATASETS

These will be the most lessons learned from the expectations for cardiovascular disease. As already declared, we'll be utilizing the heart malady location dataset and discharging interesting inductions from the information to create a few critical discoveries. For exact comes about, exploratory information examination is the basic to begin with step. After picking up bits of knowledge from the information, we must alter the features before moving on to the model-building organize. [5]. We are going develop our machine learning show for heart malady determination amid this stage.

Table 1: Original Dataset Snapshot

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
52	1	0	125	212	0	1	168	0	1.000000	2	2	3	0
53	1	0	140	203	1	0	155	1	3.100000	0	0	3	0
70	1	0	145	174	0	1	125	1	2.600000	0	0	3	0
61	1	0	148	203	0	1	161	0	0.000000	2	1	3	0
62	0	0	138	294	1	1	106	0	1.900000	1	3	2	0

The education data is unrelated to an individual's heart illness, thus it is removed. Pre-processing and experiments are then carried out on this dataset.

3.2 Training Models

Machine learning approaches include K closest neighbors (KNN), Decision Tree, Logistic Regression, Naive Bayes, XG Boost, Support Vector Machine, and Random Forest. Random Forest is used to classify heart disease. To classify heart malady successfully, different machine learning models were utilized, counting K-Nearest Neighbors (KNN), Decision Tree, Logistic Regression, Naive Bayes, XGBoost, Support Vector Machine (SVM), and Random Forest These models were chosen based on their capacity to handle differing sorts of information and their demonstrated adequacy in classification errands. The dataset experienced preprocessing steps such as cleaning, normalization, and highlight designing to get ready it for demonstrate preparing. Each calculation was executed, and hyperparameters were tuned to realize ideal execution. Assessment measurements such as exactness, exactness, review, F1-score, and ROC-AUC were utilized to evaluate the models, guaranteeing a exhaustive investigation of their qualities and shortcomings

4. Result and Analysis

4.1 Correlation and Attributes Analysis

The correlation matrix (Figure 3) highlights relationships among dataset attributes for dementia diagnosis. Key findings include dementia patients typically have a group value >0.5, with higher ASF and SES levels increasing dementia risk. Males are more likely than females to develop dementia.

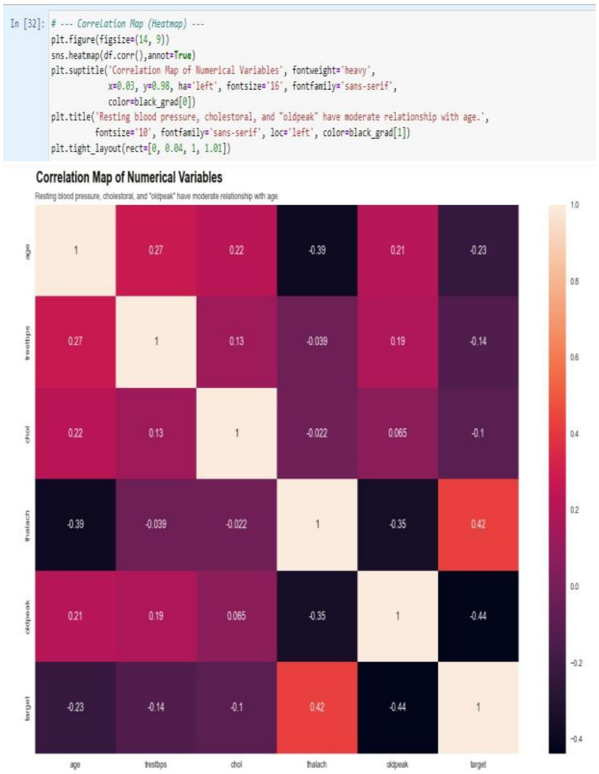

Figure 3: Correlation and Attribute

4.2 Heart Attack Prediction Using ML Models

The think about evaluates the forecast of heart attack chance by utilizing machine learning calculations such as XGBoost and KNN on a Jupyter Scratch pad working Ubuntu with an Intel Xeon processor and 8GB of Slam. Execution measurements include precision, sensitivity/recall, specificity, exactness, and F1-score, each measuring distinctive perspectives of a system's efficacy. Confusion networks give a brief outline of the classification results accomplished by XGBoost, KNN, SVM, LR, Decision tree, Random, and Gaussian models, with the comparing execution measurements and comes about delineated in Tables 4.1–4.8. The measurements affirm the prescient capacities of the models.

Table 4.1: Performance Matrix of XGB

	Precision	recall	F1-score	support
0	1.00	1.00	1.00	107
1	1.00	1.00	1.00	98
accuracy			1.00	205
Macro avg	1.00	1.00	1.00	205
Weighted avg	1.00	1.00	1.00	205

Table 4.2: Performance Matrix of KNN

	Precision	recall	F1-score	support
0	0.95	0.97	0.96	107
1	0.97	0.94	0.95	98
accuracy			0.96	205
Macro avg	0.96	0.96	0.96	205
Weighted avg	0.96	0.96	0.96	205

Table 4.3: Performance Matrix of LR

	Precision	recall	F1-score	support
0	0.91	0.77	0.83	107
1	0.78	0.92	0.85	98
accuracy			0.84	205
Macro avg	0.85	0.84	0.84	205
Weighted avg	0.85	0.84	0.84	205

Table 4.4: Performance Matrix of Gaussian

	Precision	recall	F1-score	support
0	0.84	0.82	0.83	107
1	0.81	0.83	0.82	98
accuracy			0.82	205
Macro avg	0.82	0.82	0.82	205
Weighted avg	0.82	0.82	0.82	205

Table 4.5: Performance Matrix of Decision Tree

	Precision	recall	F1-score	support
0	0.85	0.84	0.85	107
1	0.83	0.84	0.83	98
accuracy			0.84	205
Macro avg	0.84	0.84	0.84	205
Weighted avg	0.84	0.84	0.84	205

Table 4.6: Performance Matrix of Random Forest

	Precision	recall	F1-score	support
0	0.96	0.82	0.88	107
1	0.83	0.96	0.89	98
accuracy			0.89	205
Macro avg	0.89	0.89	0.89	205
Weighted avg	0.90	0.89	0.89	205

Table 4.7: Performance Matrix of SVM

	Precision	recall	F1-score	support
0	0.94	0.74	0.83	107
1	0.77	0.95	0.85	98
accuracy			0.84	205
Macro avg	0.85	0.84	0.84	205
Weighted avg	0.86	0.84	0.84	205

Table 4.8: Performance Matrix of All Models

Model	Precision (0/1)	Recall (0/1)	F1-score (0/1)	Accuracy
XGB	**1.00/1.00**	**1.00/1.00**	**1.00/1.00**	**1**
KNN	**0.95/0.97**	**0.97/0.94**	**0.96/0.95**	**0.96**
LR	**0.91/0.78**	**0.77/0.92**	**0.83/0.85**	**0.84**
Gaussian	**0.84/0.81**	**0.82/0.83**	**0.83/0.82**	**0.82**
Decision Tree	**0.85/0.83**	**0.84/0.84**	**0.85/0.83**	**0.84**
Random Forest	**0.96/0.83**	**0.82/0.96**	**0.88/0.89**	**0.89**
SVM	**0.94/0.77**	**0.74/0.95**	**0.83/0.85**	**0.84**

4.2.1 Model Accuracy Comparison

The term paper predicts the heart attack rate for a few diverse models. We have utilized a assortment of calculations counting Logistic Regression, Decision Tree, Random Forest, Support Vector Classifier, Naive Bayes Classifier, XGBoost Classifier, ada boost, zeroR, voting classifier, and KNeighbors Classifier to decide the precision of diagnosing a heart attack. We have shown the precision rate of different calculations in Figure 5.1 and Figure 5.2. XGBoost and KNN Classifier display remarkable execution, accomplishing 100% and 96% exactness, separately, outflanking all other calculations.

Table 5.1: Modul Accuracy

Model	Accuracy
XGBoost	100.000000
K-Nearest Neighbour	95.609756
Random Forest	88.780488
Logistic Regression	83.902439
Support Vector Machine	83.90239
Decision Tree	83.90239
Gaussian Naïve Bayes	82.439024

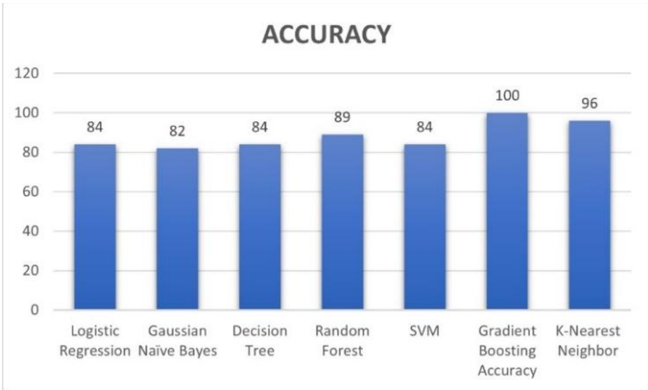

Figure 5.2: Accuracy Comparison

4.3 XAI Method

Explainable AI, also known as XAI, aids in defending a model's decision-making. By offering reasons that laypeople can understand, it makes the choice clear. There are several XAI frameworks on the market. This thesis chose to employ LIME as one of them for the project.

4.3.1 Process of Explaining Heart Attack Predictions with LIME

The working process to use the LIME method shown in Figure 6 of this thesis will address the explanation of the prediction of a heart attack. LIME is a kind of method that not only visualize the prediction but also gives an explanation of the individual prediction. [41]

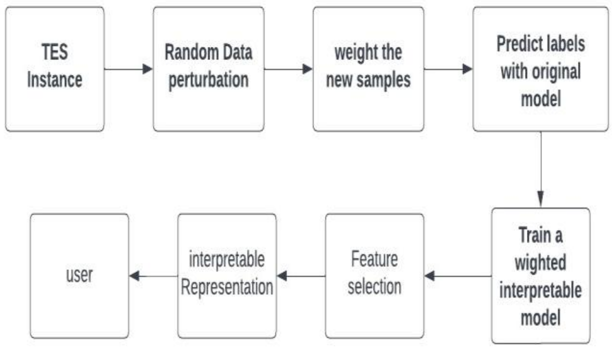

Figure 6: Process Flow of LIME

4.3.1.1 LIME Framework

The LIME system, moreover known as Nearby Interpretable Model-agnostic Clarifications, may be a strategy inside Explainable AI (XAI) outlined to illustrate the forecasts made by complex machine learning models, commonly known as "black-box" models. It caters to the expanding require for straightforwardness in AI applications, particularly in sensitive areas such as healthcare and independent frameworks. Lime’s fundamental accentuation lies in making neighborhood clarifications through assessing the conduct of a complex show close a specific occasion. Usually fulfilled by training uncomplicated, easy-to-understand surrogate models, such as direct relapse, that appraise the expectations of the initial demonstrate inside a limit region of the specific occurrence. The system points to diminish a misfortune work L, ensuring that the surrogate show closely takes after the first model's yields whereas keeping complexity moo ($\Omega(g)$), for occasion, by utilizing less highlights to improve interpretability. The parameters, such as the neighborhood measure and complexity imperatives, are characterized by the client. LIME is adaptable and works with any show, because it as it were ought to get it the input-output conduct of the show, without depending on its inner components. The paper utilizes LIME to explain expectations made by the K-Nearest Neighbors (KNN) calculation, advertising localized and reasonable bits of knowledge into its decision-making handle.

4.3.1.2 Feature Importance

The examination of include significance uncovers the key variables that play a noteworthy part in foreseeing heart attacks. Figure 7 grandstands the key supporters, to be specific oldpeak, ca, cp, and thal, with all highlights being successfully utilized. Inside this gather, the foremost vital figure is oldpeak, which focuses to ST sadness seen in electrocardiogram readings, because it mirrors the degree of myocardial ischemia. Chest torment, commonly truncated as CP, too incorporates a vital part in surveying the chance of a heart attack. Selecting the correct highlights moves forward the precision of the demonstrate by prioritizing persuasive properties.

Weight	Feature
0.1732 ± 0.0186	thal_2
0.1488 ± 0.0276	ca
0.1134 ± 0.0121	cp_0
0.1102 ± 0.0168	oldpeak
0.0905 ± 0.0219	age
0.0685 ± 0.0071	chol
0.0600 ± 0.0195	thalach
0.0220 ± 0.0041	restecg
0.0185 ± 0.0066	slope_1
0.0180 ± 0.0061	cp_2
0.0124 ± 0.0039	exang
0.0107 ± 0.0047	sex
0.0076 ± 0.0018	trestbps
0.0061 ± 0.0034	thal_1
0.0056 ± 0.0033	cp_1
0.0015 ± 0.0018	slope_2
0 ± 0.0000	thal_0
0 ± 0.0000	cp_3
0 ± 0.0000	thal_3
0 ± 0.0000	slope_0
0 ± 0.0000	fbs

Figure 7: Important Features of The Dataset

4.3.1.2 Feature Insights

All the highlights have esteem to number. By analyzing all the esteem checks we foresee the rate of a heart attack. Within the paper, Fig 8 speaks to the more and less chances of heart attack from esteem checks. It implies the dataset has 51. 32% of *1' which predicts more chances of heart attack and 48. 68% of '0' which predicts less chances of a heart attack. All the highlights have esteem to number.

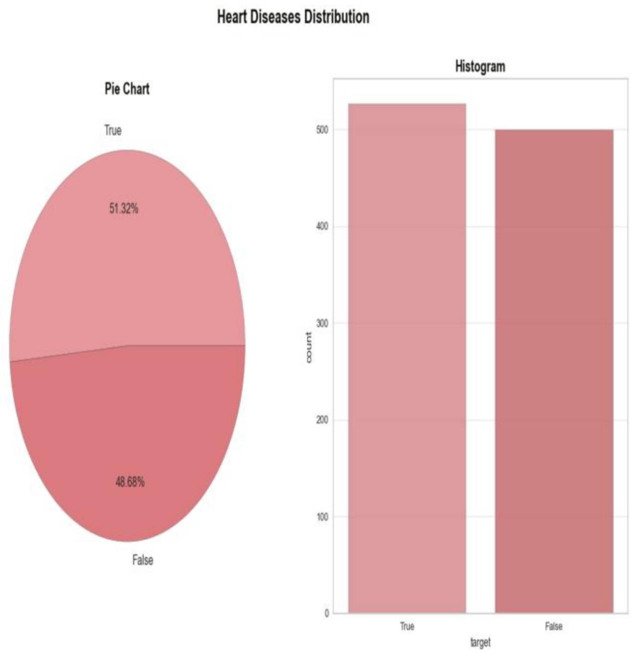

Figure 8: Value Counts of The Feature

4.3.2 Interpreting Heart Attack Predictions Using LIME: A Detailed Analysis

This area of the paper will clarify the heart attack forecast through KNN Machine Learning calculation utilizing LIME.

4.3.3 Significance of Key Features

The LIME strategy can outwardly appear which highlights make a more prominent commitment to the expectation. Utilizing LIME demonstrates profoundly profitable because it offers the capacity to decide highlight significance through the utilization of two unmistakable strategies. Figures 9 and 10 show the affect volume of the highlights by utilizing the appear in note pad() strategy. The figure 9 and 10 incorporates three diverse sorts of representation: a advance bar, a bar chart, and a table. Inside the figure 9 and 10, the advance bar outlines the extend inside which the esteem changes, alongside the real expectation. The bar chart successfully shows the highlights of their weights in connection to the forecast, both emphatically and contrarily. In conclusion, the table serves to grandstand the noteworthiness of highlights by portraying their genuine values. In this circumstance, the color orange highlights a positive affect, whereas the color blue implies a negative affect on the estimate [41].

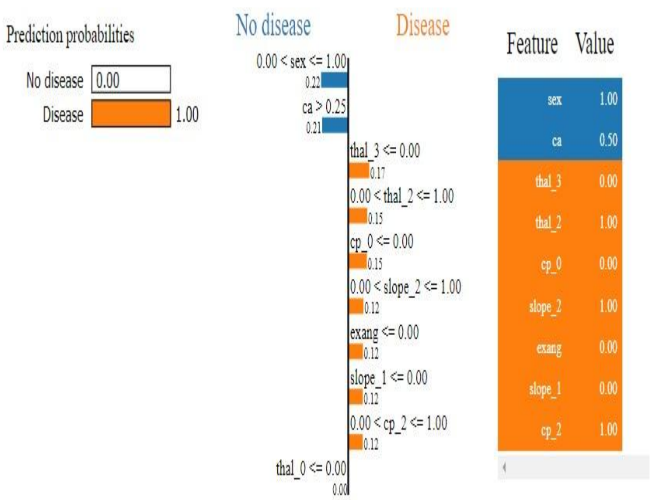

Figure **9**: Feature Importance Using LIME

In fig-9, for this patient , heart disease is predicted for orange features which is important features for heart disease where blue features are less important

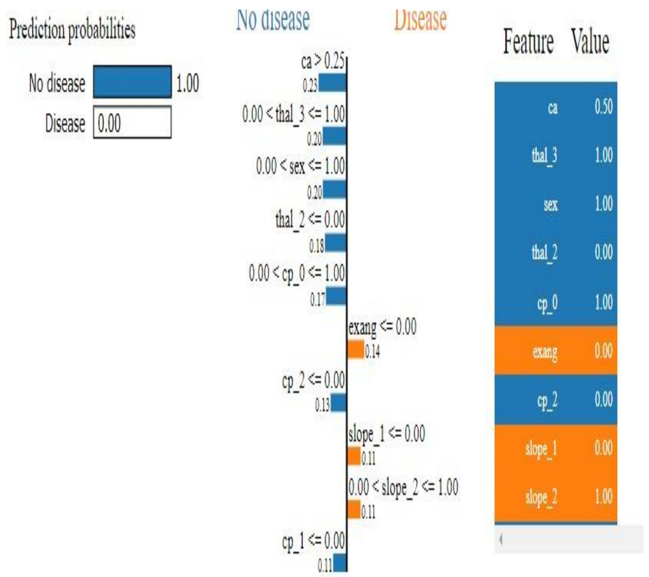

Figure 10: Feature Importance Using LIME

In fig-10, for this patient , heart disease is not predicted for blue features which is not important features for heart disease where orange features are more important .

4.3.4 Extracting Feature Significance

LIME XAI strategy has the advantage of recover the highlights significance appeared in Fig 11. Here the primary esteem of tuple is condition, and the moment esteem is the highlight esteem based on condition.

Weight	Feature
0.1732 ± 0.0186	thal_2
0.1488 ± 0.0276	ca
0.1134 ± 0.0121	cp_0
0.1102 ± 0.0168	oldpeak
0.0905 ± 0.0219	age
0.0685 ± 0.0071	chol
0.0600 ± 0.0195	thalach
0.0220 ± 0.0041	restecg
0.0185 ± 0.0066	slope_1
0.0180 ± 0.0061	cp_2
0.0124 ± 0.0039	exang
0.0107 ± 0.0047	sex
0.0076 ± 0.0018	trestbps
0.0061 ± 0.0034	thal_1
0.0056 ± 0.0033	cp_1
0.0015 ± 0.0018	slope_2
0 ± 0.0000	thal_0
0 ± 0.0000	cp_3
0 ± 0.0000	thal_3
0 ± 0.0000	slope_0
0 ± 0.0000	fbs

Figure 11: Retrieving Features Using LIME

As we know, KNN calculation could be a relapse show. But this calculation is additionally characterized as a classifier. So in case we need to recover the highlights significance for classifier errand LIME will permit us to do so. Fig 11 appears that, it returns a dictionary where the key is each lesson of assignment and esteem may be a list of highlight record and their commitment in foreseeing that lesson.

5 CONCLUSION

It is outstandingly challenging to execute the ML calculations in healthcare field for it's limited openness of dataset. The paper illustrated that the XGBoost calculation gives the finest precision. So the LIME procedure got to be actualized on the XGBoost calculation. To do so, a huge and populated dataset must be required. Furthere more, the dataset required to be preprocessed. All of these over reasons limited the papers noteworthiness. In later a long time, the ML calculations in conjunction with AI has been appeared uncommon influence on healthcare zone. This recommendation may be a little approach to it. As the paper requires some limitations, we got to move forward the execution as future approach. Execution of LIME on XGBoost calculation can be focused on as future work. Besides, in future we would like to work on a tremendous dataset to see how accurately the appear predicts. We are going extend this work and endeavor to initiate SHAP values for the overall dataset. In show disdain toward of having some obstacles this recommendation performed all the desires and gives prominent future works. So, we dream to require the work forward and progress empower.

REFERENCES

[1] Mohan, Senthilkumar & Thirumalai, Chandra Segar & Srivastava, Gautam. (2019). Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2923707.

[2] C. Sowmiya and P. Sumitra. (2017). Analytical study of heart disease diagnosis using classification techniques. 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Srivilliputhur, 2017, pp. 1-5. DOI: 10.1109/ITCOSP.2017.8303115.

[3] A. N. Repka, S. D. Ravikanti and R. G. Franklin. (2019). Design and Implementing Heart Disease Prediction Using Naive Bayesian. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 292-297. DOI: 10.1109/ICOEI.2019.8862604.

[4] Spencer, Robinson et al. (2020). Exploring feature selection and classification methods for predicting heart disease. Digital Health, 6: 2055207620914777. DOI: 10.1177/2055207620914777.

[5] Kishore, A., Kumar, A., Singh, K., Maninder, P., and Hambir, Y. (2018). Heart Attack Prediction Using Deep Learning. International Research Journal of Engineering and Technology (IRJET), 5(4).

[6] Krishnan, M., and D. S. (2019). Prediction of Heart Disease Using Machine Learning Algorithms. 1st International Conference on Innovations in Information and Communication Technology (ICIICT), vol. 10, 2019.

[7] A. Golande and P. Kumar T. (2019). Heart Disease Prediction Using Effective Machine Learning Techniques. International Journal of Recent Technology and Engineering (IJRTE), 8(1S4), 2019.

[8] Budholiya, Kartik & Shrivastava, Shailendra & Sharma, Vivek. (2020). An optimized XGBoost-based diagnostic system for effective prediction of heart disease. Journal of King Saud University - Computer and Information Sciences, 34. 10.1016/j.jksuci.2020.10.013.

[9] Uddin, Azher, et al. (2021). Study on Convolutional Neural Network to Detect COVID-19From Chest X-Rays. Mathematical Problems in Engineering, Hindawi Limited, Sept. 2021, pp. 1–11. DOI: 10.1155/2021/3366057.

[10] Tulare, Uma. (2018). Prediction system for heart disease using Naive Bayes and particle swarm optimization. Biomedical Research, 29,10.4066/biomedicalresearch.29-18-620.

[11] Ambrish G, Bharathi Ganesh, Anitha Ganesh, Chetana Srinivas, Dhanraj, Kiran Mensinkal. (2022). Logistic regression technique for prediction of cardiovascular disease. Transitions Proceedings, 3(1), pp. 127-130.

[12] Apostolof, N., and Zisserman, A. (2007). Who are you? - Real-time person identification. BMVC.

[13] Hand, D., Mannila, H., and Smyth, P. (2001). Principles of Data Mining. The MIT Press, Cambridge.

[14] Wang, H. (2002). Nearest Neighbours without k: A Classification Formalism based on Probability. Technical Report, Faculty of Informatics, University of Ulster, N. Ireland, UK.

[15] Yin, P., Criminisi, A., Win, J., and Essa, I. A. (2007). Tree-based classifiers for bilayer video segmentation. CVPR.

[16] Desai, A., Shah, M., and Bhattacharyya, S. (2018). Deep learning models for cardiovascular risk prediction: A survey. IEEE Access, 6, pp. 12187-12198.