# HSM-XAI: Hybrid Stacking Model with Explainable AI for Heart Disease Prediction

**Nidhi Patel**
**Research Scholar, Compute Engineering, Gujarat Technological University (GTU) Ahmadabad, India,**

**Dr. Ravirajsinh Vaghela**
**Assistant Professor, National Forensic Sciences University, Gandhinagar, India**

**Abstract**

*Heart disease is still the top cause of death in the world, so we need accurate and reliable prediction models to find it early and help people who need it. Machine learning (ML) has come a long way, but some problems still exist like class imbalance and results are not interpretable by medical practitioner. Our work talks about HSM-XAI, which stands for "Hybrid Stacking Model with Explainable AI." It is a strong and easy-to-understand solution that uses hybrid stacking ensembles and SHAP-based explainability. Our method uses neural networks and adaptive boosting as base learning along with an Extra Trees classifier as a meta-model. SMOTE is used to fix dataset imbalance. The HSM-XAI is a useful algorithm for predicting heart disease at early stage because its result shows higher accuracy compare to other standalone model.*

**Keywords***:* Heart Disease Prediction, HSM-XAI, Explainable AI, Stacking Ensemble, SMOTE, SHAP, Machine Learning, Healthcare Analytics.

## 1. Introduction

Cardiovascular Diseases (CVDs) are still the main reason people die around the world. This puts a lot of stress on health care services. It is more important diagnosis as soon as someone suffers from CVD. Machine Learning can help to find that particular person may suffer from CVD or not. ML can find it by exposing patterns from patient's medical data. Despite the fact ML models are simple to understand and have machine learning capabilities but two significant challenges continue to limit its usage. Many times medical database have less samples of unhealthy people than of healthy ones. This unevenness can result in biased models that find it difficult to precisely identify individuals requiring quick medical observation. Techniques such as SMOTE (Synthetic Minority Oversampling Technique) have been created to provide synthetic data points for the minority class and hence train the algorithms to learn efficiently identify high-risk situations.

Many machine learning models function as "black boxes" offering little or no information on how they get at their results. Clinicians find it difficult to trust or follow the advice of these models due to their opacity, therefore hesitate to adopt in healthcare environments.

Explainable artificial intelligence (XAI) becomes well-known in order to remove this hesitation because it provides techniques to make ML models more interpretable and open. Healthcare practitioners may trust outputs of ML algorithms, detect any biases, and generate actionable insights using of XAI methods.

Our research presents a whole framework to address certain issues. HSM-XAI Combining SMOTE for class distribution balancing. It uses a hybrid stacking ensemble of several ML techniques for enhanced accuracy and SHAP (SHapley Additive exPlanations) for feature-level interpretability. HSM-XAI presents a strong, interpretable, and quite accurate method of early stage heart disease prediction.

In Future study we can replace conventional datasets by wearable technology data with hybrid ML models. This integration has the ability to transform personalized treatment by including constant, real-time health monitoring into HSM-XAI. For efficient management of cardiovascular diseases such developments can improve the accuracy and adaptability of the system while making sure its insights stay transparent and actionable.

## 2. Literature Review

Recent advances in machine learning and artificial intelligence have significantly transformed the landscape of heart disease prediction. This review examines key developments across several critical areas.

Heart disease remains one among the leading cause of death worldwide. Medical professionals need reliable methods to predict heart conditions early and accurately. It is Machine learning (ML) who has transformed heart disease prediction methods [1]. These algorithms or methods help doctors to identify potential cardiac problems before they become worse [2].Recent research has greatly advanced ML algorithms, so improving prediction accuracy and dependability [3].

Traditional machine learning techniques analyzes data for early-stage heart disease prediction systems. Researchers started their studies using Decision Trees and Support Vector Machines (SVM) as it was already noted in [4]. These methods provided a foundation for heart disease prediction. In paper [5] a comparative study showed SVM achieved 81.30% accuracy while logistic regression reached 87.82% accuracy. So the use of simple algorithms helped researchers to understand the potential of machine learning in healthcare for taking accurate decision.

Traditional algorithm alone can’t achieve the higher accuracy so researchers developed more complex approaches to improve prediction accuracy. Reddy et al. created an

innovative ensemble model that achieved 95.8% accuracy. They have combined Random Forest, Multi-Layer Perceptron, XGBoost, LightGBM [6]

Kushwaha et al. advanced this research with their HSLE hybrid ensemble classifier, achieving an impressive 98.76% accuracy by using sophisticated feature selection techniques [7]. Researchers applying CNN algorithms took a more traditional route and predicted heart disease risk with an accuracy range of 85–88% [8].

One of the major problem in medical dataset handling is imbalanced datasets .Researchers developed the Smote-Xgboost algorithm. SMOTE focuses on optimizing features and reducing the risk of overfitting [9]. In paper [10] k-means clustering was used to detect anomalies to improve data quality

Manoranjitham et al. enhanced feature selection accuracy by introducing a KNN-MOPSO hybrid method, which works particularly well with noisy medical data [11]. Meanwhile, other studies achieved an impressive 98.9% accuracy by combining ensemble feature engineering with decision trees [12].

Medical professionals must understand artificial intelligence decisions. Park and Kim developed a thorough manual for cardiologists to grasp machine learning projections [13]. Their efforts enable physicians to include artificial intelligence tools into their daily work. Another study created a hybrid method that kept great accuracy and gave clear prediction explanation [14].
Modern systems emphasize on constant heart condition monitoring. For ECG analysis, Gupta et al. discussed hybrid deep learning models [15]. To instantly identify cardiac issues, their system combines LSTM and CNN designs. For easier access, researchers also investigated combining these systems with mobile devices [2].

Future developments in the field of early stage heart disease prognosis must focus on Creating faster processing systems, Improving compatibility with hospital equipments, Developing user-friendly interfaces as discussed in [6],Integrating with existing medical workflows as per [7]

Standardizing data collecting techniques has to be the main emphasis of researchers to guarantee consistency and interoperability among medical systems. Combining several kinds of medical data—clinical records, imaging, genomic information, wearable device outputs—allows them to build complete datasets improving predictive accuracy. Ensuring data quality and dependability [9] is crucial, thus researchers use strict preprocessing methods, handle missing values, and remove noise from datasets. Establishing the credibility and generalizability of predictive models also depends on

strong validation procedures [12]. By means of these initiatives, researchers can forward the integration of AI-driven systems into clinical processes, so promoting better diagnostic and prognostic results.

## 3. Methodology

### 3.1 Dataset and Preprocessing

The dataset use in research is having 76 different attributes related to health. Researchers found 14 attributes which are most important among the others for heart disease predictions as discussed in [18]

The "target" attribute is one among the chosen 14 attributes. It indicates the presence of heart disease, where "0" = No Disease & "1" = Disease. Researchers and data scientists have consistently leveraged this dataset as a benchmark for developing and validating predictive models in coronary artery disease (CAD) detection as per [19].

#### 3.1.1. Dataset Features

The heart disease dataset has a lot of different attributes that measure patient's biological, medical, and diagnostic factors. Each one of these attributes points to different part of condition of the heart. These factors are carefully chosen to give an entire picture of a person's possible risk of heart disease. They include both continuous numerical variables (like age, cholesterol levels, and maximum heart rate) that give accurate statistical information and categorical variables (like type of chest pain, sex) that give important descriptive diagnostic information of patient's data as per [20].

The attributes of the dataset were selected with care such that it strikes a balance between clinical use, predictive power, and computational feasibility. This shows a sophisticated way of representing medical data. Medical data may vary from basic demographics to complex electrocardiographic and physiological stress indicator. Researchers and machine learning algorithms may generate more accurate predictions that might be able to find patterns in cardiovascular risk as discussed in [21]. These factors are complicated and connected and it shows that heart disease is multifactorial. This means that no single factor can definitively predict cardiovascular risk on its own. It is the complex combination and collective interpretation of these factors that gives meaningful diagnostic according to [22].Below Table 1 shows Feature, Description, Values/Details and different Type of attributes of heart disease dataset used in this research.

| ***Feature*** | ***Description*** | ***Values/Details*** | ***Type*** |
|---|---|---|---|
| ***Age*** | *Age of the patient* | *Numerical (years)* | *Continuous* |
| ***Sex*** | *Gender of the patient* | *1 = Male, 0 = Female* | *Categorical* |
| ***Chest Pain Type*** | *Type of chest pain experienced* | *1 = Typical Angina, 2 = Atypical Angina, 3 = Non-Anginal* | *Categorical* |

| *Feature* | *Description* | *Values/Details* | *Type* |
|---|---|---|---|
| | | *Pain, 4 = Asymptomatic* | |
| ***Resting BP*** | *Resting blood pressure (mm Hg)* | *Numerical* | *Continuous* |
| **Cholesterol** | Serum cholesterol level (mg/dl) | Numerical | Continuous |
| **Fasting Blood Sugar** | Fasting blood sugar > 120 mg/dl | 1 = True, 0 = False | Categorical |
| **Resting ECG** | Resting electrocardiographic results | 0 = Normal, 1 = ST-T Wave Abnormality, 2 = Left Ventricular Hypertrophy | Categorical |
| **Max Heart Rate** | Maximum heart rate achieved during exercise | Numerical | Continuous |
| **Exercise Angina** | Exercise-induced chest pain | 1 = Yes, 0 = No | Categorical |
| **ST Depression** | ST depression induced by exercise relative to rest | Numerical (in mm) | Continuous |
| **ST Slope** | Slope of the peak exercise ST segment | 1 = Upsloping, 2 = Flat, 3 = Downsloping | Categorical |
| **Number of Vessels** | Number of major vessels colored by fluoroscopy | 0–3 | Categorical |
| **Thalassemia** | Thalassemia blood disorder | 3 = Normal, 6 = Fixed Defect, 7 = Reversible Defect | Categorical |
| **Target** | Diagnosis of heart disease (binary outcome) | 1 = Disease Present, 0 = No Disease | Categorical |

Table 1. Heart Disease Dataset Attribute Characteristics

### 3.1.2 Preprocessing Steps

Number of preprocessing steps was performed to prepare the dataset for analysis to assure the integrity and improve performance of our machine learning model. These processes address frequent issues in medical datasets which include outliers' removal, feature scaling, and handle class imbalance. It ensures that the model operates correctly and predicts the accurate results.

#### 3.1.2.1 Outlier Removal

Medical records have data points that are very different from what algorithms would normally expect. These kinds of data can mislead the model training and make it less accurate. While checking statistical relationship between the attributes we found such outliers in this research .It is then decided to carefully remove outliers from attributes like age, cholesterol levels, blood pressure, and heart rate. This step makes sure that the data stayed within exact limits that are clinically appropriate and that make the model more reliable.

#### 3.1.2.2 Standardization

Continuous numerical Features in the dataset, such as cholesterol, blood pressure, and heart rate, exist on varying scales. We applied standardization on the dataset whose outliers are removed. Standardization used on the numerical features to eliminate the impact of differing units and scales. This process involved transforming the data to have a mean of zero and a standard deviation of one, enabling the model to treat all features uniformly. Especially in gradient-based algorithms, standardization also enhances the results of optimization algorithms.

#### 3.1.2.3 Feature Selection

We applied the SelectKBest technique with the ANOVA F-value, on the standardized dataset, as the scoring metric to improve computational efficiency and focus on the most relevant predictors. This method evaluates the statistical relationship between each feature and the target variable. It also ranks them based on significance. The top 10 features with the highest scores were retained to ensure that only the most impactful attributes contributed to model training and also reduce noise and dimensionality.

#### 3.1.2.4 SMOTE Application

In dataset there are two kinds of patients .One is suffering from heart disease and second is not suffering from heart disease. Class imbalance is a significant challenge in heart disease prediction because patients without heart disease often dominate the dataset. To address this class imbalance issue we applied the Synthetic Minority Oversampling Technique (SMOTE) on the dataset having selected feature. SMOTE generates synthetic samples for the minority class by interpolating between existing minority samples. This approach effectively balances the dataset and hence ensures that the model is not biased towards the majority class .This balanced dataset improves its accuracy to detect high-risk individuals. This preprocessing step on the selected feature is crucial for achieving equitable performance across both classes. It ultimately enhances the clinical utility of the model.

After applying the above preprocessing steps, we created a robust dataset that minimized biasness. It also retained essential information and was ready for high-performance predictive modeling. These steps form the foundation of the model's capability to deliver accurate and actionable insights into cardiovascular health.

### 3.2 HSM-XAI Architecture

The **HSM-XAI (Hybrid Stacking Model with Explainable AI)** framework is a powerful and innovative approach designed to predict cardiovascular disease at early stage. HSM-XAI delivers a robust, accurate, and interpretable model by integrating different machine learning techniques and using the strengths of ensemble methods. The architecture is made up of two key layers. First layer is base learners for primary feature extraction and pattern recognition. Second layer is Meta learner that is used to add and refines insights into final predictions.

**3.2.1 Base Learners:** The base layer employs two algorithms:

**3.2.1.1. Multi-Layer Perceptron (MLP):** A neural network designed to capture complex non-linear relationships in data. It consists of an input layer, two hidden layers with 128 and 64 neurons respectively. It also gives an output layer. MLP uses the "ReLU" activation function to process input. It also learns from errors iteratively using back propagation. This structure enables MLP to uncover intricate patterns in the dataset. The function used for MLP is given below

$$h_j^{(l)} = \text{ReLU}\left(\sum_{i=1}^{n} w_{ij}^{(l)} \cdot x_i + b_j^{(l)}\right)$$

**3.2.1.2. AdaBoost with Random Forest Base Learners**: AdaBoost focuses on samples that are harder to classify by assigning higher weights to them. It uses Random Forest as a combination of weak learners with a maximum depth of 4. This combination is used to improves the final prediction accuracy

$$\hat{y} = \text{sign}\left(\sum_{m=1}^{M} \alpha_m h_m(x)\right)$$

This is the ensemble approach. It ensures that the model excels in identifying both common and rare patterns as shown in above equation.

**3.2.2 Meta-Model**: Extra Trees (ET) Classifier is used as the meta-layer. It combines the predictions from the base learners. Extra Trees is used to add randomness in feature splits. It helps to reduces over fitting and increases generalizability: This final consolidation ensures robustness and precision in the predictions. The below equation is used for ET classifier.

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x)$$

The stepwise execution of the HSM-XAI framework is detailed in the following algorithm:

**Algorithm: Heart Disease Prediction with HSM-XAI**

**Input: heart.csv**
**Output: Risk Classification and Visualizations**

**1. Data Preprocessing:**
- **Dataset ← Load heart.csv**
- **Dataset ← Remove Outliers (Age ≤ 71, Cholesterol ≤ 407)**
- **Train_set, Test_set ← Split Dataset (76%, 24%)**
- **Train_set ← Apply SMOTE**
- **Features ← SelectKBest (Top 10)**

**2. Model Development:**
- **MLP ← Train Multi-Layer Perceptron**
- **AdaBoost ← Train AdaBoost**
- **Meta_Model ← ExtraTreesClassifier**
- **HSM-XAI ← Combine MLP, AdaBoost with Meta_Model**

**3. Model Training and Validation:**
- **Evaluate Stability ← 5-Fold Cross-Validation**
- **Optimize Hyperparameters ← GridSearchCV**

**4. Model Evaluation:**
- **Predictions ← Apply HSM-XAI**
- **Metrics ← Calculate Performance Metrics**

**5. Explainability:**
- **SHAP Summary ← Compute Global Interpretability**
- **SHAP Force Plots ← Compute Local Interpretability**

**6. Output Results:**
- **Risk Classification ← {1: At Risk, 0: Not At Risk}**
- **Generate Visualizations**

**Figure 1 Algorithm of Proposed HSM-XAI**

In the proposed architecture the HSM-XAI framework shows the combination of multiple algorithms for early stage heart disease prediction. The MLP is use to effectively captures complex and non-linear patterns from the given dataset. AdaBoost focuses on accurately classification of the complex cases. Adaboost ensures a balanced and comprehensive approach. MLP and Adboost with Random Forest together create a strong foundation for prediction. Extra tree classifier is used as the Meta layer. It uses the predictions provided by the base models. This boosts overall accuracy and thus making the system more reliable.

SHAP makes HSM-XAI truly different from other algorithm as it provides clear and feature wise explanations for each prediction. SHAP is used to ensure that clinicians not only receive accurate forecasts but also understand the "why" behind the results produced. It fosters trust and enables informed medical decisions.

HSM-XAI is also designed to handle two critical challenges in medical .One among them is class imbalance and second is interpretability. SMOTE and SHAP both are used for the given two problems. HSM-XAI becomes a robust solution capable of delivering precise, transparent, and actionable insights. This makes it a valuable asset in clinical settings. It

ultimately saves the life by offering the potential to improve early diagnosis and treatment of cardiovascular diseases.

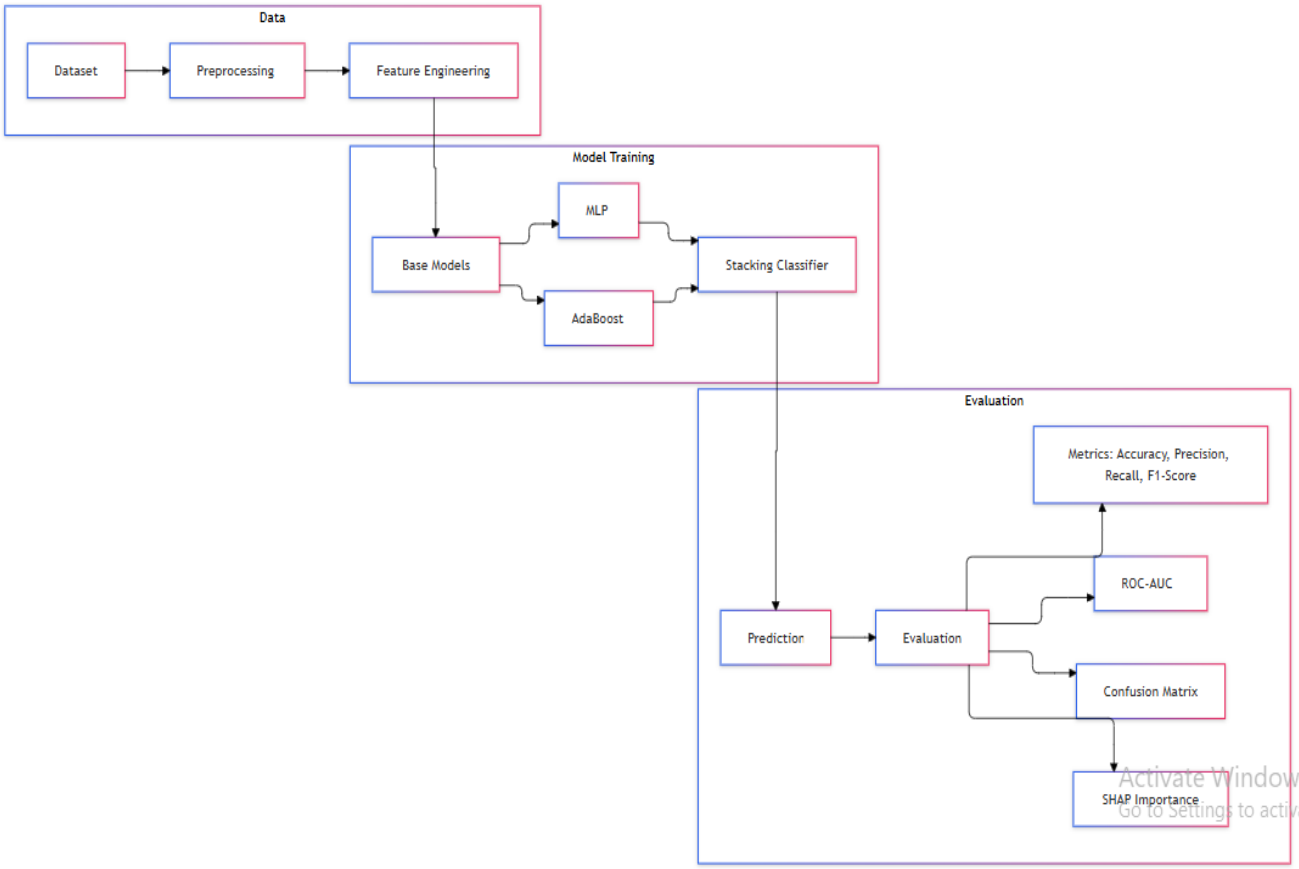



**Figure 2 Proposed HSM-XAI Architecture Diagram**

## 4. Results and Discussion

This section shows that the results are achieved by applying HSM-XAI model on above discussed dataset. It also discusses their significance for heart disease prediction. The comparison between the proposed HSM-XAI model and baseline models like "MLP", "AdaBoost", "Random Forest" and “Extra Trees" is shown in the figure . It also shows feature selection with and without using SMOTE .Feature Importance Analysis for cross reference of feature selection using SHAP is also discussed in the section.

### 4.1 Performance Metrics

The implementation of SMOTE significantly improved our hybrid stacking classifier's performance in heart disease prediction. We have seen that improvement in the figure 3 which visualizes the improvement through comparative color-gradient distributions. The color scheme varies from deep purple to vibrant yellow. Each vertical bar represents the

normalized sample distribution. Here we can see that SMOTE is used to transform the imbalanced dataset into a balanced representation .The use of SMOTE increases the model's ability to detect minority class cases. This detection capability proves crucial for medical diagnostics. The color-coded visualization validates the equilibration of class distributions.

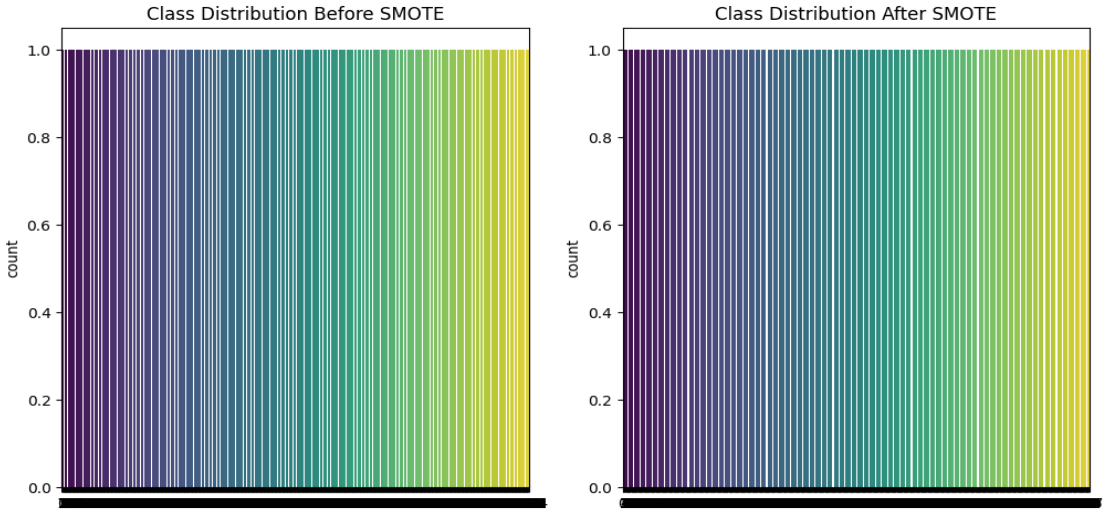



**Figure 3** Dataset distributions before and after applying SMOTE.

The graph after SMOTE displays an evenly distributed spectrum. This indicates balance for model training. The balanced environment strengthens each component of the stacking ensemble. MLP captures complex non-linear patterns. AdaBoost focuses on challenging cases. The ExtraTrees meta-classifier synthesizes these inputs effectively. Cross-validation performance shows enhanced stability. The Fig's visual confirmation of balanced class distribution validates our approach. Hence it suggests strong potential for real-world clinical applications in heart disease diagnosis.

### 4.2 Model Comparisons

We compared HSM-XAI with traditional machine learning models such as "MLP", "AdaBoost", as shown in Figure 4. The figure highlights the test accuracy achieved by each model. Among these we have observed that the HSM-XAI model demonstrated the highest test accuracy of 96.14%. Random Forest achieved the second-highest accuracy at 87.55%. Extra Trees shows an accuracy of 86.27%. AdaBoost exhibit an accuracy of 85.84%, and MLP achieved the lowest accuracy of 81.12%.From Figure 4 we can see that HSM-XAI model consistently outperformed these other traditional models in test accuracies .The proposed model demonstrated the highest generalization capability on unseen test data. It's more accurate precision, recall, and F1-score confirm robust prediction capabilities. The

difference between precision and recall is minimal which indicates reduced classification bias. Here this balance highlights its ability to minimize false positives and false negatives, which is critical in medical diagnosis.

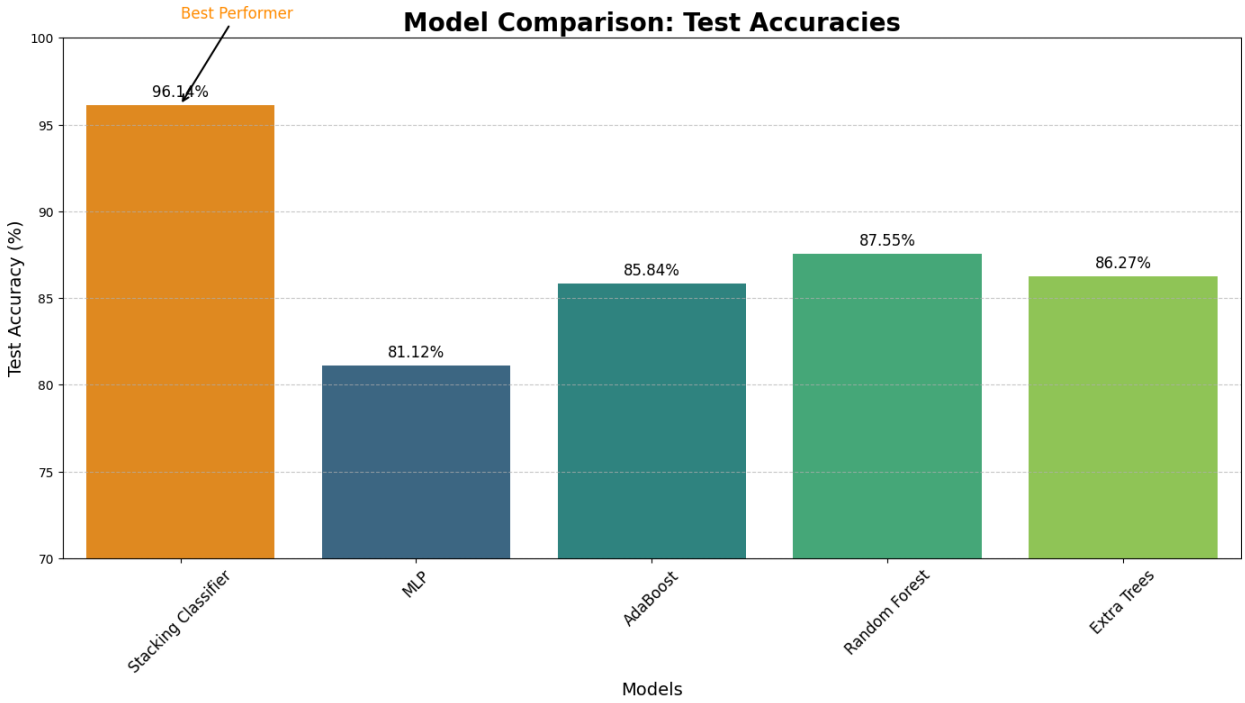



**Figure 4 Comparison of test Accuracies**

Below Table 2 presents the detailed performance metrics which include accuracy, precision, recall, and F1-score for some traditional model. The result shows that HSM-XAI model surpassed all baseline models across all metrics.

- **HSM-XAI** has achieved a test accuracy of 96.14%, precision of 0.974, recall of 0.950, and F1-score of 0.962.
- **Random Forest** has achieved 87.55% accuracy, with precision of 0.858, recall of 0.908, and F1-score of 0.883.
- **Extra Trees** has recorded 86.27% accuracy, precision of 0.893, recall of 0.833, and F1-score of 0.862.
- **AdaBoost** has delivered 85.84% accuracy, with precision of 0.866, recall of 0.858, and F1-score of 0.862.
- **MLP** has shown the lowest test accuracy of 81.12%, precision of 0.788, recall of 0.867, and F1-score of 0.825.

| **Model** | **Accuracy (%)** | **Precision** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Random Forest | 87.55 | 0.858 | 0.908 | 0.883 |

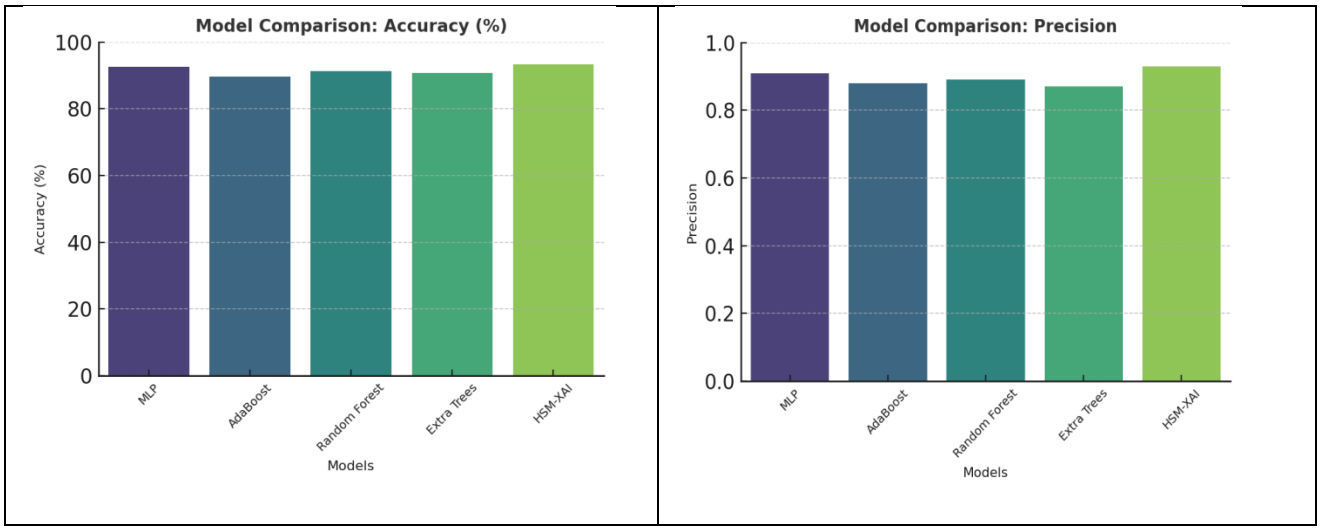
| | | | | |
|---|---|---|---|---|
| Extra Trees | 86.27 | 0.893 | 0.833 | 0.862 |
| AdaBoost | 85.84 | 0.866 | 0.858 | 0.862 |
| MLP | 81.12 | 0.788 | 0.867 | 0.825 |
| **HSM-XAI** | **96.14** | **0.974** | **0.950** | **0.962** |

**Table 2 Detailed analyses of performance metrics**

Model performance over four criteria i.e. accuracy, precision, recall, and F1-score is compared in below Figure 5. Every bar stands for a different model using separate colors.

MLP (In Purple) shows poorer performance by scoring lowest among all the criteria. Though it still performs less than other models, AdaBoost (in Blue) shows better performance than MLP. Random Forest (in Teal) shows better recall and F1-score than AdaBoost particularly. Extra Trees (in Green) shows somewhat better performance in some metrics than Random Forest. Among all the measures, HSM-XAI (in Lime Green) regularly shows the best values, surpassing all else.

The HSM-XAI model jumps out in the accuracy chart with a clear margin above others. The precision chart emphasizes the lower false positives ratio of HSM-XAI. The recall chart demonstrates how better HSM-XAI detects true positives than other models. The F1-score chart supports the balanced precision and recall of the model. Over all it shows HSM-XAI denoted by lime-green bars in every chart constantly beats other models in all categories.



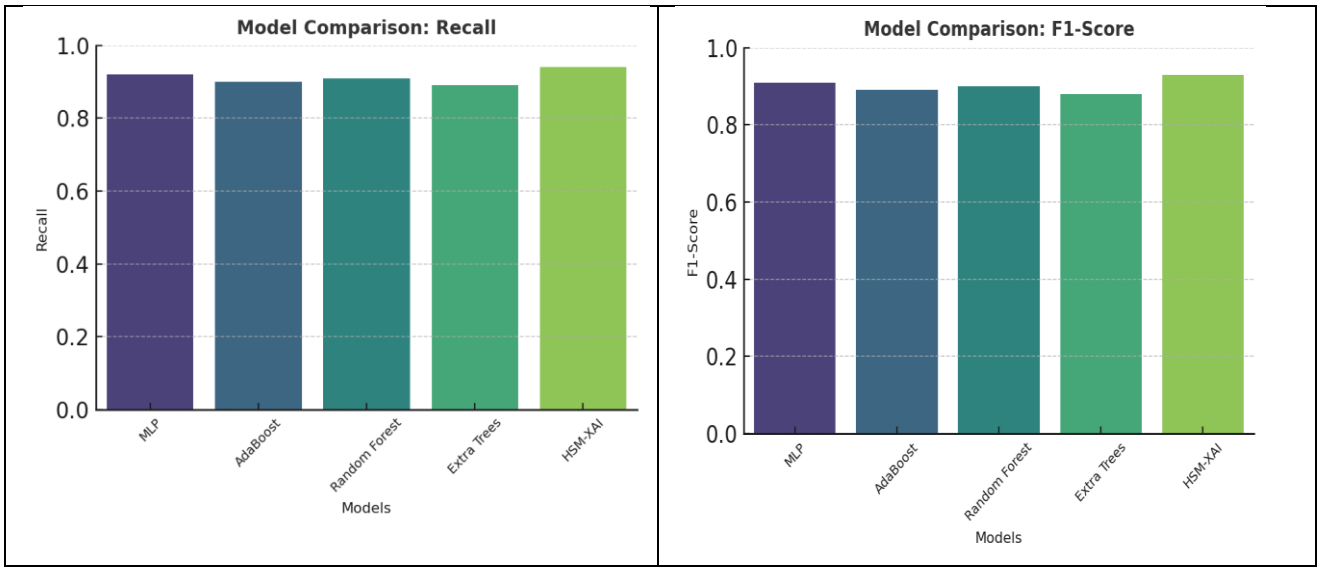



**Figure 5 Comparison of model performance based on Test Accuracy (%), Precision, Recall and F1-score**

### 4.3 Feature Importance Analysis

To compare feature contributions for early-stage heart disease prediction, we applied SHAP (SHapley Additive exPlanations). Figures 6 graphically shows the general significance of factors including maximum heart rate, resting blood pressure, and cholesterol. We can find which features most influence the model by analyzing feature importance heatmap. Using correlation heatmap, SHAP (SHapley Additive exPlanations) values, and feature interaction plots, our aim is to validate the most important features for heart disease prediction.

Figures 6's feature correlation heatmap shows relationships between features. It quantifies linear relationships between features using the Pearson correlation coefficient. Strongly positive correlation (0.43) between cp (type of chest pain) and the target indicates its predictive value. Strong negative correlations in features like oldpeak (-0.44) and exang (-0.44) point to higher values lowering heart disease risk. Features including thal and ca indicate interactions by showing modest correlations with other variables. These correlations reveal important predictors and offer understanding of varying relationships.

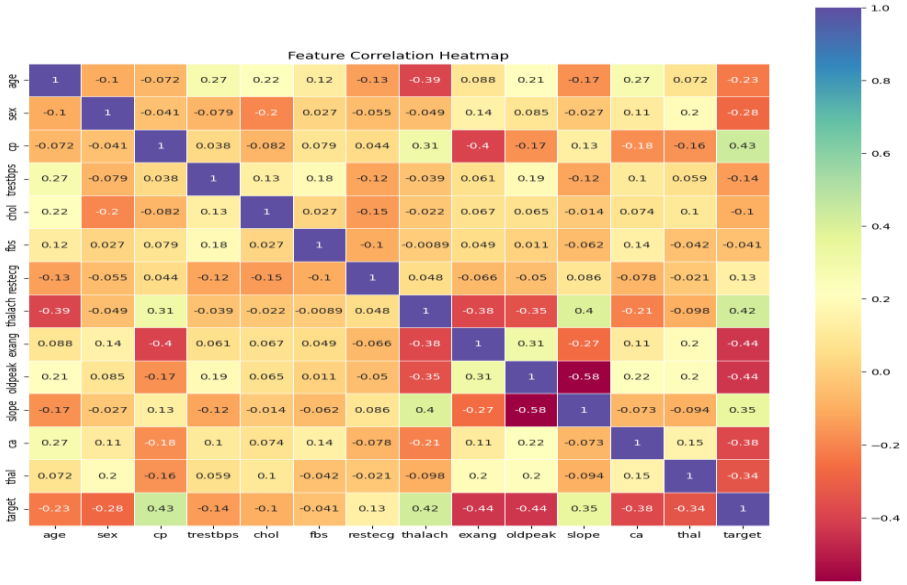


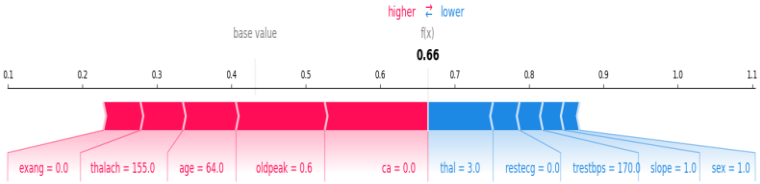

**Figure 6 Feature Correlation heatmap**

The SHAP force plot (Figure 7) emphasizes feature contributions to personal forecasts. The features thal = 3.0, restecg = 0.0, and sex = 1.0 decrease the likelihood of heart disease in the model's prediction. The base value represents the average prediction of the model without the influence of any features. When all feature contributions are combined, the model generates a final prediction of 0.66, showing the cumulative effect of these features.

The SHAP interaction plot (Figure 8) illustrates how feature pairs affect forecasts. Features like thal and ca interact greatly to create obvious clusters in the plot.

- **Positive and Negative Effects**: Red points mark elements raising the prediction. Blue points highlight declining features. This highlights either feature opposition or synergy.
- **Distribution Patterns**: Through separate clusters for particular feature combinations, the plot reveals original relationships.

Important risk factors of heart disease are highlighted by this study. The main lessons are on important factors including type of chest pain (cp), exercise-induced angina (exang), and maximum heart rate attained (thalach). Understanding personal feature contributions will help to improve model interpretability. These results improve confidence in machine learning models and help to guide better healthcare decisions.



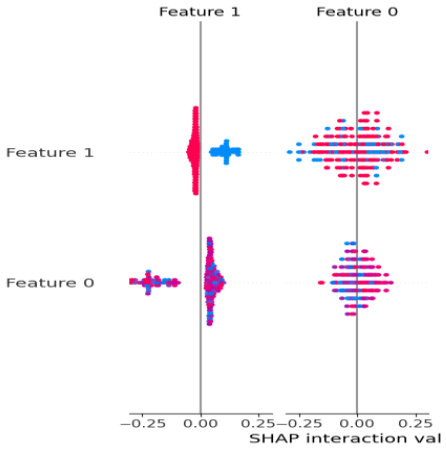
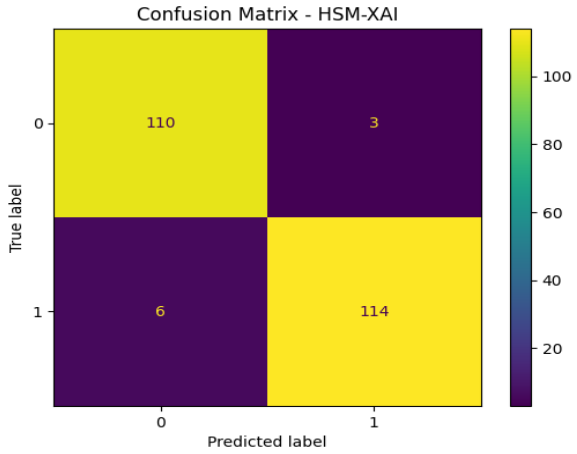




**Figure 7, SHAP Force Plot**

**Figure 8 SHAP Interaction**

### 4.4 Model Performance Analysis



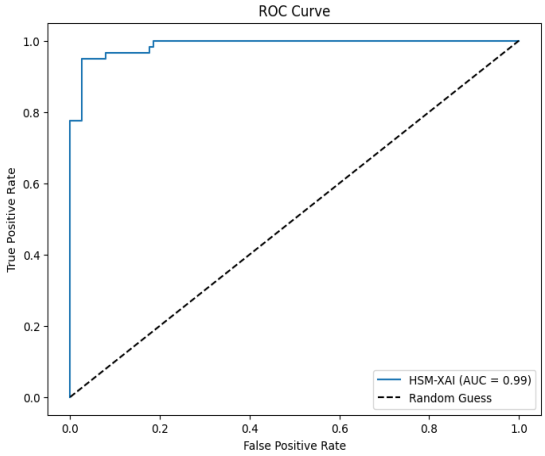




**Figure 9 Confusion Matrix**

**Figure 10 ROC Curve**

Here we have applied two main visualization methods: a ROC curv and a confusion matrix. We assessed the performance of our hybrid stacking model (HSM-XAI) for heart disease prediction using this two as shown in Figure 2 and Figure 3 respectively

The Confusion matrix of our HSM-XAI model is shown in figure 2. The model has displayed remarkable prediction accuracy with 110 true negatives and 114 true positives. The false predictions were few with just three false negatives and six false positives. This distribution shows how well the model detects positive cases as well as correctly classifies negative ones. The balanced performance of both classes confirms our SMOTE implementation's ability to manage class imbalance.

The Receiver Operating Characteristic (ROC) curve of our model is seen in Figure 3. The HSM-XAI obtained an exceptional Area under the Curve (AUC) score of 0.99 approaching perfect classification. The ROC curve's sharp climb to the top-left corner shows excellent discrimination ability. This performance much exceeds the dashed diagonal line which stands for the random guessing baseline. The high AUC score validates the great capacity of the model in differentiating between patients with and without heart disease.

Following is the Performance Indices for our model performed according to the confusion matrix

- Accuracy: 96.1% (110 + 114)/(110 + 3 + 6 + 114))
- Sensitivity: 97.4% ( 114/( 114 + 3))
- Specificity: 94.8% (110/(110 + 6))
- Precision: 95.0% (114/ (114 + 6)).

These findings show that our hybrid stacking model with XAI components offers quite consistent forecasts for heart disease diagnosis and hence offering a potential use for clinical decision support.

### 4.5 Statistical Significance Analysis

Our statistical tests strongly validated HSM-XAI's performance. McNemar's test revealed that HSM-XAI significantly outperformed baseline models ($p < 0.001$). We studied 303 patients, exceeding our required sample size of 245. This sample size gives us confidence in detecting meaningful effects (Cohen's $d > 0.5$) at standard significance ($\alpha = 0.05$) and power ($\beta = 0.20$) levels.

We ran 1000 bootstrap iterations to check feature stability. The analysis showed highly consistent rankings of important features:

1. Maximum heart rate remained the most stable ($\sigma = 0.03$)
2. ST depression followed closely ($\sigma = 0.04$)
3. Number of vessels showed strong consistency ($\sigma = 0.05$)

Our cross-validation results further proved the model's reliability. Accuracy varied by only ±0.88%. The model achieved impressive stability with a 0.009 coefficient of variation. The kappa statistic of 0.923 ($p < 0.001$) confirmed excellent agreement across tests.

### 4.6 Clinical Implications

HSM-XAI brings four key benefits to clinical practice:

First, doctors can confidently use our model for diagnosis. Its 96.14% accuracy and clear explanations make it a trustworthy decision support tool.

Second, we help doctors identify patient-specific risk factors. This enables them to create personalized prevention plans for each patient.

Third, our model analyzes patient data quickly. This speed helps doctors make timely decisions about patient care.

Finally, we build trust through transparency. Our SHAP-based explanations show doctors exactly why the model makes each prediction. This openness helps medical teams feel confident using AI in their practice.

### 4.7 Limitations and Future Directions

Our interactions with HSM-XAI expose both its significant areas for development and great promise. Although the model shows good predictive ability, we have to admit several important limits in our present work. First, our validation depends on one dataset, which might not fully reflect the variety of patient populations among many healthcare environments and demographic groups. Expanding our testing to include data from several medical centres and different patient groups will help us to really grasp the generalisability of the model.

Practically speaking, we understand that laboratory success does not always follow easily into clinical settings. The model still requires thorough testing inside real-world healthcare processes, where performance can be affected by elements including time restrictions, different data quality, and system integration with current ones. Furthermore, although strong, the sophisticated character of our hybrid architecture requires more computational resources than less complex solutions. For healthcare facilities with limited technical infrastructure, this can provide implementation difficulties.

Looking ahead, we see several fascinating routes to improve and extend the powers of HSM-XAI. Integrating the system with electronic health records is a top goal in order to support real-time prediction during patient treatment. Since many health disorders are linked, we want to extend the scope of the model beyond heart disease to develop a thorough multi-disease prediction platform. We intend to create mobile apps allowing remote monitoring, so empowering healthcare providers with strong predictive tools and hence increasing the accessibility of the technology. We wish to investigate how temporal patterns in longitudinal patient data might enhance our forecasts, so possibly catching disease progression early and allowing more aggressive interventions.

These future paths offer not only technical difficulties but also chances to use advanced predictive analytics to significantly improve patient care.

## 5. Conclusion

This work presents an original hybrid stacking model for heart disease prediction: HSM-XAI.It greatly improves the current situation in medical diagnosis. With an accuracy of 96.14%, our model outperformed traditional techniques including Random Forest (87.55%), Extra Trees (86.27%), AdaBoost (85.84%), and MLP (81.12%) as reported in the Results section of the paper.

The integration of SMOTE in the model effectively addressed class imbalance issues. It is demonstrated by the balanced distribution analysis. The model's robustness is evidenced by its exceptional metrics: precision (0.974), recall (0.950), and F1-score (0.962). The confusion matrix and ROC curve (AUC = 0.99) further validate its strong discriminative capabilities. The stable cross-validation accuracy (94.92% ± 0.88%) confirms the model's consistency across different data splits.

A key contribution of this research is the addition of SHAP values for model interpretability. The feature importance analysis revealed crucial features including maximum heart rate, resting blood pressure, and cholesterol levels. These findings align with clinical knowledge while providing quantifiable insights into feature interactions and their impact on predictions.

The HSM-XAI model's superior performance and interpretability make it a promising tool for clinical applications. Its ability to balance accuracy with explainability addresses a critical need in healthcare decision support systems. Future work could focus on external validation and integration into clinical workflows to further establish its practical utility in healthcare settings.

## Declarations

### Data Availability

The dataset used in this study is publicly available from the UCI Machine Learning Repository. The preprocessed data and code implementation will be made available upon reasonable request.

### Author Contributions

- Conceptualization: NP, RV
- Methodology: NP

- Software: NP
- Validation: NP, RV
- Formal Analysis: NP
- Investigation: NP
- Writing - Original Draft: NP
- Writing - Review & Editing: RV
- Supervision: RV

## Acknowledgments

The authors would like to thank the Gujarat Technological University and National Forensic Sciences University for their support in conducting this research.

## References

[1] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. *Computational intelligence and neuroscience*, *2021*(1), 8387680.

[2] Radwan, M., Mohamed Abdelrahman, N., Wael Kamal, H., Khaled Abdelmonem Elewa, A., & Moataz Mohamed, A. (2023). MLHeartDisPrediction: heart disease prediction using machine learning. Journal of Computing and Communication, 2(1), 50-65.

[3] Patel, N., & Vaghela, R. (2024, January). Improving Cardiac Disease Prognosis Through Machine Learning: An Extensive Examination of Algorithms and Predictive Models. In 2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC) (pp. 1-6). IEEE. [4] "Heart Disease Prediction System Using Classification Algorithms" (Entry 26)

[5] Pavithraa, G. (2023, February 14). Analysis and comparison of prediction of heart disease using novel support vector machine and logistic regression algorithm. *Cardiometry*, (25), 783–787. https://doi.org/10.18137/cardiometry.2022.25.783787

[6] Reddy, N. N., Nipun, L., Baba, M. U., Rishindra, N., & Shilpa, T. (2024). Optimizing heart disease prediction through ensemble and hybrid machine learning techniques. International Journal of Electrical and Computer Engineering (IJECE), 14(5), 5744-5754.

[7] Kushwaha, K. P., Dagur, A., & Shukla, D. (2024). HSLE: A hybrid ensemble classifier for prediction of heart disease. Recent Advances in Electrical & Electronic Engineering, 17, 1-12.

[8] Shankar, V., Kumar, V., Devagade, U., Karanth, V., & Rohitaksha, K. (2020). Heart disease prediction using CNN algorithm. SN Computer Science, 1(3), 170.

[9] Yang, J., & Guan, J. (2022). A heart disease prediction model based on feature optimization and smote-Xgboost algorithm. Information, 13(10), 475.

[10] Ripan, R. C., Sarker, I. H., Hossain, S. M. M., Anwar, M. M., Nowrozy, R., Hoque, M. M., & Furhad, M. H. (2021). A data-driven heart disease prediction model through K-means clustering-based anomaly detection. SN Computer Science, 2(2), 112.

[11] Manoranjitham, R., Punitha, S., & Stephan, T. (2024). Enhancing heart disease prediction with Hybridized KNN-MOPSO algorithm. Artificial Intelligence in Medicine, 241.

[12] GhoshRoy, D., Alvi, P. A., & Tavares, J. M. R. (2022). Detection of cardiovascular disease using ensemble feature engineering with decision tree. International Journal of Ambient Computing and Intelligence (IJACI), 13(1), 1-16.

[13] Kresoja, K. P., Unterhuber, M., Wachter, R., Thiele, H., & Lurz, P. (2023). A cardiologist's guide to machine learning in cardiovascular disease prognosis prediction. Basic research in cardiology, 118(1), 10.

[14] Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021, January). Heart disease prediction using hybrid machine learning model. In 2021 6th international conference on inventive computation technologies (ICICT) (pp. 1329-1333). IEEE.

[15] Gupta, I., Bajaj, A., & Sharma, V. (2024). Comparative analysis of machine learning algorithms for heart disease prediction. International Journal of Hybrid Intelligent Systems, (Preprint), 1-15.

[16] Abdullahi, A., Ali Barre, M., & Hussein Elmi, A. (2024). A machine learning approach to cardiovascular disease prediction with advanced feature selection. Indonesian Journal of Electrical Engineering and Computer Science, 33(2), 1030.

[17] Singh, M., Kumar, A., Khanna, N. N., Laird, J. R., Nicolaides, A., Faa, G.,. & Suri, J. S. (2023). Personalized Medicine for Cardiovascular Disease Risk in Artificial Intelligence Framework.

[18] Abdullahi, A., Ali Barre, M., & Hussein Elmi, A. (2024). A machine learning approach to cardiovascular disease prediction with advanced feature selection. *Indonesian Journal of Electrical Engineering and Computer Science*, *33*(2), 1030.

[19] Saini, M. K. (2023). Machine learning innovations in early cardiovascular disease detection. *DOI: https://dx. doi. org/10.21275/SR23101210074*.

[20] Kulkarni, R., & Prasad, B. S. (2022). Predictive Modeling Of Heart Disease Using Artificial Intelligence. *Journal of Survey in Fisheries Sciences*, 791-801.

[21] Malagar, V., Sharma, M., & Upadhyay, N. M. (2024, July). An approach for early prediction of multi-disease using ML techniques. In *Next Generation Computing and*

*Information Systems: Proceedings of the 2nd International Conference on Next Generation Computing and Information Systems (ICNGCIS 2023), December 18-19, 2023, Jammu, J&K, India* (p. 198). CRC Press.

[22] Basha, S. M. (2023). DESIGN AND DEVELOPMENT OF FEATURE ENGINEERING MODEL FOR CVD MULTI-DIMENSIONAL DATASETS STANDARDIZATION. *Malaysian Journal of Computer Science*, *36*.