

# Optimizing Prediction of Cardiac Conditions Using Hyper-Adaboost-Integrated Machine Learning Models

Abdulrahman Ahmed Jasim<sup>\*</sup>, Layth Rafea Hazim<sup>\*\*</sup>, Hajer Alwindawi<sup>\*\*\*</sup>, Oguz Ata<sup>\*\*\*\*</sup>

<sup>\*</sup>Dept. of Electrical and Computer Engineering, Altinbas University, Istanbul, Turkey  
Email: abdulrahman.alsalmany@gmail.com  
<https://orcid.org/0000-0002-1002-5792>

<sup>\*\*</sup>Dept. of Electrical and Computer Engineering, Altinbas University, Istanbul, Turkey  
Email: layth1985it@gmail.com  
<https://orcid.org/0000-0001-8066-2175>

<sup>\*\*\*</sup>Dept. of Artificial Intelligence Engineering, Bahçeşehir University, Istanbul, Turkey  
Email: hajir.alwindawi@bahcesehir.edu.tr  
<https://orcid.org/0009-0003-5393-3461>

<sup>\*\*\*\*</sup>Dept. of Electrical and Computer Engineering, Altinbas University, Istanbul, Turkey  
Email: oguz.ata@altinbas.edu.tr  
<https://orcid.org/0000-0003-4511-7694>

## Abstract

Recent advancements in machine learning have played a crucial role in the healthcare industry, particularly in predicting heart disease with assorted datasets. Despite the registration results indicating promising accuracy and resilience in heart disease prediction models, it is still low compared to what we expected. In this article, a new technique has been demonstrated to predict heart disease based on state-of-the-art machine learning with a focus mostly on Hyper Adaboost classifiers. We evaluate the performance of several key machine learning algorithms, including Random Forest, Extra Trees, LightGBM, Decision Tree, and Hyper Adaboost, on well-known heart disease datasets. Our results are mixed, but they suggest that the Hyper Adaboost classifier provides stronger accuracy, with performance metrics consistently exceeding 97%, much better than similar models. This research highlights the promise of newer packages under Hyper for detecting and screening individuals with heart disease early, which can offer insights into advancing methodologies in future studies as well as clinical use. Our technique exhibited strong predictive power, encouraging additional study with a larger and more comprehensive dataset for the validation of our model.

**Keywords-** Machine Learning techniques, Heart disease prediction, Predictive modelling, Hyper classifiers.

## I. INTRODUCTION

Heart disease is a specific kind of abnormality that involves problems with the heart and blood arteries. It may involve issues with heart rhythm and blood vessel health [1]. In addition to assessing the patient's medical history, the doctor checks blood pressure, cholesterol, and blood sugar measurements to help in the diagnosis of heart disease. Many of the most thorough medical exams are used to assess a person's heart condition. It includes exercise stress tests, X-rays, radioactivity testing, magnetic resonance imaging (MRI), coronary angiography (CAD), cardiac echocardiograms (CEGs), and CT scanning [2]. Appropriate treatment, when paired with heart disease diagnosis, is essential to lower the risk of heart disease and halt the rising death rate. The use of artificial intelligence in early disease detection is growing in popularity [3]. Finding the more subtle trends in the data allows one to make conclusions that are consistent with new knowledge. One example of this would be heart problems, which are the most severe and fatal illness globally, taking the lives of almost 17.9 million people each year [4]. Using this method, a large number of academics from many fields were studying the issue since early detection of cardiac disease can lower death rates.

The creation of computer algorithms that can access and alter current data in order to predict data in the future is the primary goal of the machine learning technique [5]. There are also other tried-and-true methods to improve model accuracy. These include handling outlier and missing values, selecting attributes, expanding the dataset, optimising methods, performing cross-validation, and data balancing.

In this article, the model's performance is assessed using five-fold cross-validation on both benchmark datasets. Moreover, it employs a Hyper Adaboost classifier to improve the accuracy of the ML model. This website features The Performance Rating Analysis Witch as one of its significant works. This study employs eight machine learning approaches to analyse performance rating elements using a dataset related to cardiac conditions. For the best prediction accuracy, five-fold cross-validation is used. Four ML models additionally use the Hyper Adaboost classifier to improve model accuracy.

With a special emphasis on the Hyper Adaboost classifier, this work enhances the prediction of cardiac illness by using state-of-the-art machine learning methods. The principal aim is to enhance the precision and resilience of heart disease prognostic models by use of optimised machine learning methodologies.

## II. RELATED WORKS

CAD data sets have recently seen the use of numerous ML and DM techniques to support a range of goals. Machine learning methods are becoming more and more popular for diagnosing computer-aided design (CAD) because of their low complexity, ease of deployment, and measurement possibilities. In a current evaluation report, Kolukisa and Gungor [6] indicated that 84 research papers on ML-based ensemble feature selection CAD diagnosis had been published as of 2023, and that comparisons of the CAD detection study measures' performance in a range of classifiers and FS approaches for three distinct datasets had shown results of over 91%, 85.55%, and 85.47%. In an additional study [7], The best 14 common features from several datasets were sorted using entropy-based feature engineering, which included the imputing of missing values (IMV) and outliers (OR) and evaluation procedures utilising various machine learning techniques. These features were then evaluated and eliminated, yielding an AUC of 96.8%. The results exceeded previous methods in terms of accuracy (92.7%), specificity (91.5%), accuracy (92.5%), and F1 score (0.931). Baviskar and associates. [8], used the feature selection technique, which was based on the results of the African buffalo algorithm, genetic algorithm, and particle swarm optimisation, to choose the relevant characteristics and eliminate the unnecessary and irrelevant features. They created (Deep Progressive Attention-RNN+LSTM) to improve the methods even more and raise the categorisation rate. Narayan and P in [9], were chosen for their work, and the cardiac datasets were cleaned and processes carried out using the Hammering Selection function. To precisely detect the existence of heart disease, the researchers used deep learning models, such as deep belief networks and a cuckoo search algorithm, using datasets from Cleveland, South Africa, Alizadeh Sani, Framingham, and the Heart Disease Statlog. The percentages were 89.2%, 89.5%, 89.7%, and 91.2% in that sequence. In this investigation [10] A classification and regression tree (CART) was the instrument used. Supervised machine learning is the technique employed by the algorithm. The characteristics that impact heart disease are prioritised in the findings. considering each and every performance standard. The reliability of the model is validated by its 87% forecast accuracy. Mohapatra together with others. [11], were reported in their study. A predictive ability model for the evaluation of heart disease based on stacks was proposed, with unique classifiers at both tiers. Working together with a variety of learners produces consistent model results. 92% prediction accuracy, 92.6% sensitivity, and 91% specificity were attained by the model. In an additional study [12] To extract pertinent characteristics from various high-resolution datasets (Statlog + Hungary + Cleveland + Switzerland + Long Beach VA datasets), the authors employed a deep learning CNN model. D-t-SNE (Distributed-t-Stochastic Neighbour Embedding) was then applied to solve the overlay and lower the dimensionality. Furthermore, the accuracy of the work classification using hyperparameter-tuned MLP (H-MLP) was 96.68%. Peddakrishna and Chandrasekhar [5], heart disease prognosis was enhanced. This work is distinctive in that it makes use of GridSearchCV Validation to refine hyperparameters, find ideal parameters, and assess efficacy through machine learning techniques and accuracy metrics.

## III. METHODOLOGY

This section explains how to use publicly available data to predict heart disease. This study is divided into two parts: the data gathering phase comes first, and the heart disease prediction segment comes last. In the first stage, data transformation and balancing techniques can be used to pre-process data. Several machine learning algorithms are employed in the next step to construct the recommended model. The following part of the method uses a Hyper Adaboost strategy to improve the model's accuracy. Forecasts made by the dominant class will largely come true, while characteristics of the minority class will be discounted as statistical noise. SMOTE was used to address the imbalance and keep the positive and negative classes in balance. To determine the most important and relevant variables for the classification task, the performance of each classification model was assessed using a variety of measures, such as accuracy, F1, precision, recall, and area under the curve (AUC). To validate the output of the classification model, the data is divided into training and test portions, with the training data making up 80% of the overall data set and the test portions making up 20%. A thorough description of the process architecture is given in Figure 1.

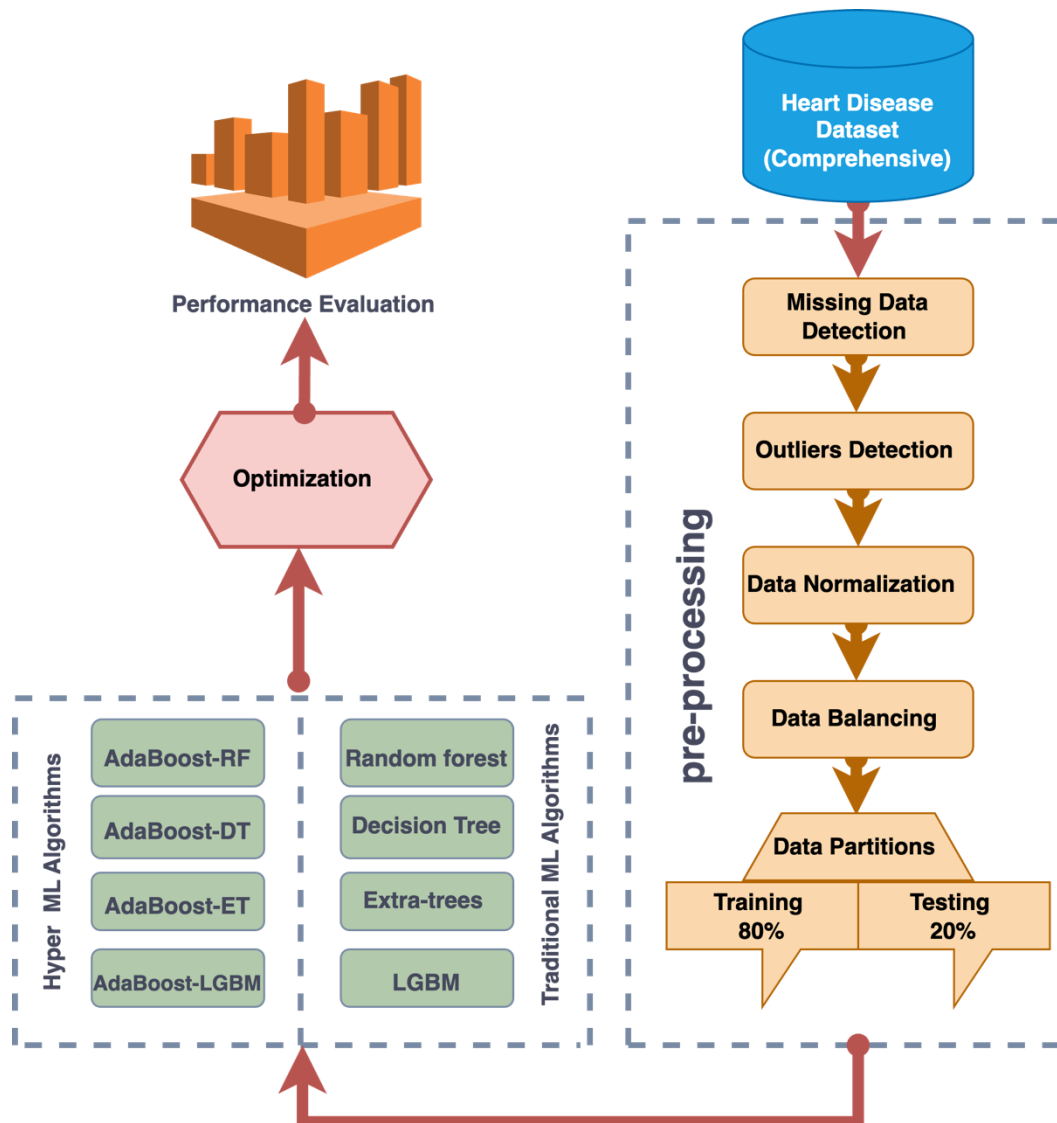


Figure 1, Workflow of the model.

### 1. Dataset

Five well-known data sets on heart disease were evaluated [13], comprising all non-medical and technological attributes, to create the data set used in this investigation. There were 1190 clinical files in this extensive data collection for the entire year. Maximum heart rate, exercise-induced heartburn, gender, age, kind of chest pain, blood pressure, resting cholesterol levels, rapid blood sugar levels, resting ECG, old peak, and ST sloping are the eleven variables in the data set. It also has a single objective statistic, which might be "average" or "heart disease." Figure 2 illustrates a typical ECG signal, highlighting key components such as Oldpeak Distribution and ST Slope Distribution. These components represent different phases of the cardiac cycle.

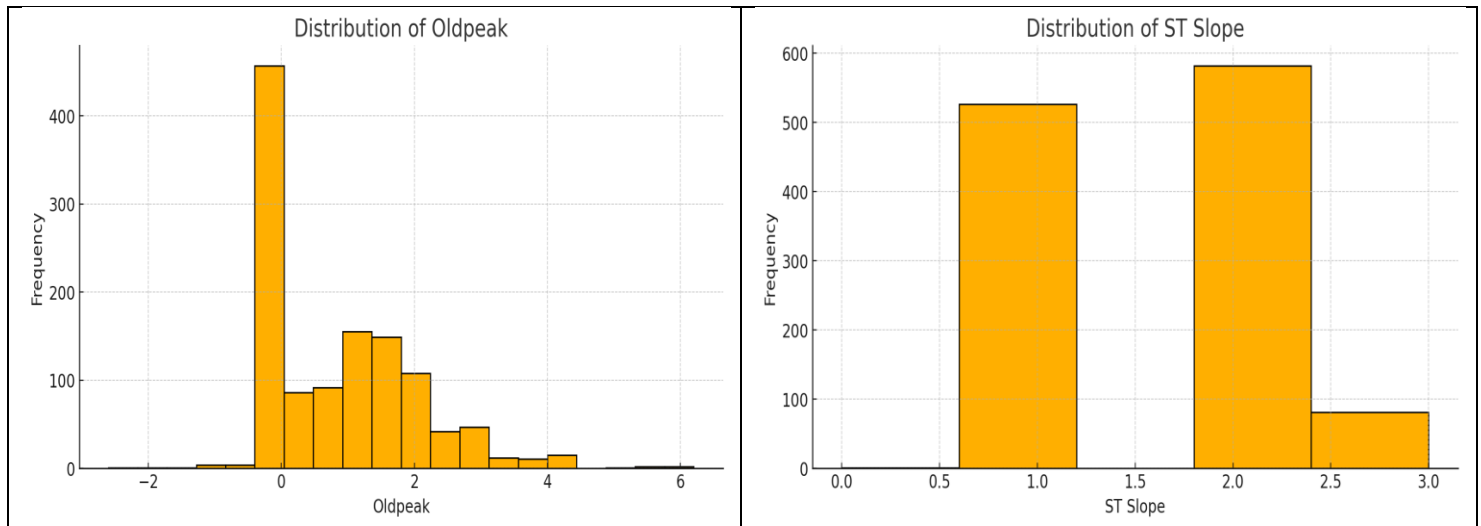


Figure 2, Oldpeak and ST Slope Distribution.

## 2. Resampling Technique

Unbalanced datasets are a common cause of problems for machine learning systems. These datasets can be divided into two categories: imbalanced multi-class data, which has more than two classes, and imbalanced binary data, which has two classes. [14]. This work focusses on imbalanced binary and multi-class datasets with a disproportionately low number of occurrences of a specific group, called minority groups. Figure 3 illustrates the binary imbalance issue with the dataset that was used, and how the random oversampling approach (ROST) was used to solve it. Unlike traditional sampling methods, ROST prioritises the processing of minority data above majority data in order to increase class effectiveness. Utilising extra random samples selected from comparable surrounding samples, ROST increases the precision of predictions for small groups. ROST is a powerful machine learning algorithm that generates new data from nearby minority group data in order to further enhance the minority group data share and meet class balance requirements.

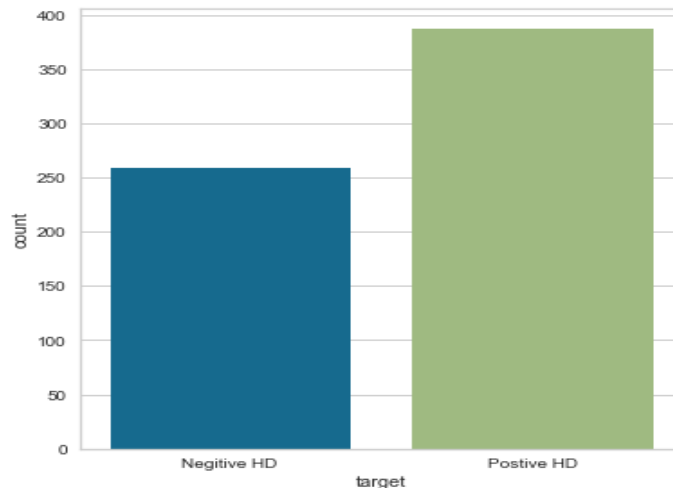


Figure 3, Sample distribution of heart dataset.

## 3. Machine Learning

In order to generate complex models and retrieve medical data, machine learning is utilised, giving specialists and physicians fresh perspectives [15], [16]. Better guidelines for inpatient care decision-making in clinical settings may be found by predictive machine learning algorithms. In this work, several machine learning models were used[17].

### A. Random Forest

A decision tree extension called random forest (RF) consists of many single decision trees, each of which offers a distinct category of prediction outcomes. The ultimate prediction outcome of the random forest classifier is influenced by the category that received the most votes in the forest. Figure 4 illustrates that of the six single-decision trees in the forest, two have predictive values of 0 and the

other four have prediction scores of 1. Consequently, the RF's predicted outcome is 1. The forest's trees are typically disconnected to one another, which ensures that the decisions they make collectively are superior to those made individually and accounts for the classifier's excellent performance [18].

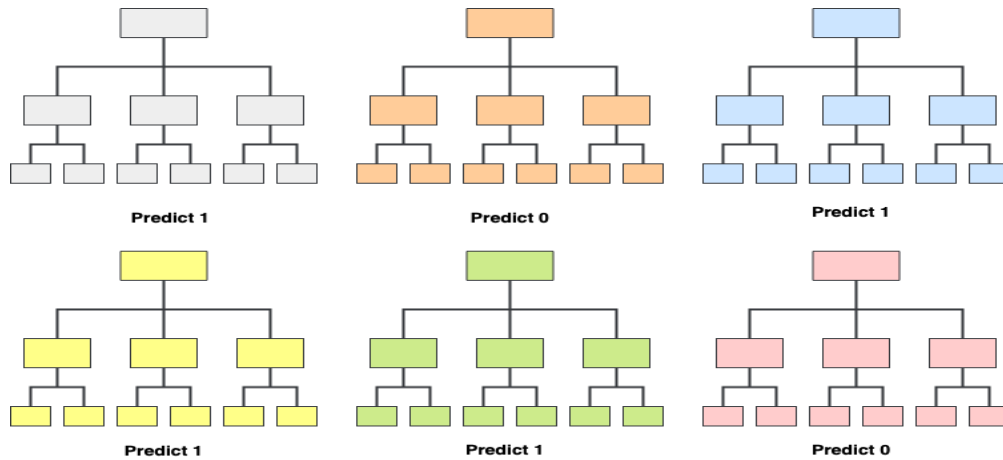


Figure 4, An Illustration of a Random Forest Model's Prediction Process.

### B. Decision Tree

A decision tree (DT) [19] is a crucial and widely applied learning technique that addresses problems with regression and classification. Figure 5 shows that a Decision Tree (DT) is made up of leaf nodes that predict the result, decision nodes that assess the value of an attribute, and edges that link to the next node. In the DT, each feature of a dataset is represented as a node, with the root node being a single, distinct node. We have to start at a single node and make our way down the decision tree to satisfy our demands and make a decision in order to illustrate how the decision tree functions.

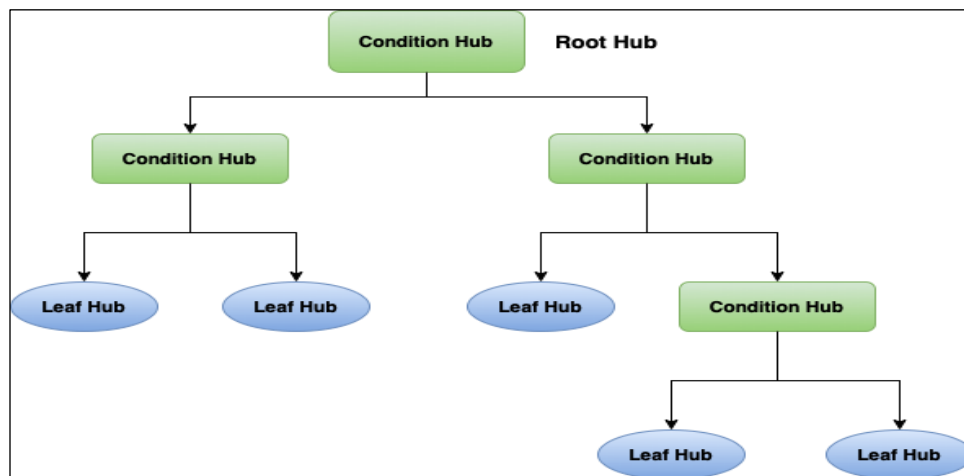


Figure 5, The Structure of The Decision Tree [19].

### C. Extra- Trees

Extra-trees [20] Using the entire training dataset, produce a great deal of unique decision trees. Based on a partially random cut point and a randomly selected subset of attributes (K), the approach selects a split rule for the root node. To split the parent node into two equally chosen child nodes, it chooses a random split. Up until you get to a leaf node, you repeat this process in each child node. Leaf nodes are those nodes that have no children. Each tree's forecast is pooled, and the result is decided by a majority vote.

### D. Light Gradient Boosting Machine (LGBM)

In order to increase classification performance, the LGBM technique adjusts the weight of the samples and gathers the classified results in a linear fashion by learning from numerous classifiers throughout the training phase [21]. In 2011, the gradient's functions were extended by the scientist Friedman, who introduced the gradient reinforcement machine. Reducing the model's loss function is

the goal. The tree is trained repeatedly to enhance the new tree's regression. For negative training, the model's loss value is estimated throughout each iterative training session. Lastly, the residual value is updated, and the current regression tree is inserted. Until the algorithm hits the user-specified limit, it goes through the training procedure repeatedly.

### E. Hyper AdaBoost Method

AdaBoost [20] permits the creation of a single classifier known as a "strong classifier" by merging multiple "weak classifiers". People call these trees "Decision Stumps." This method gives each data point the same weight in order to develop a model. Consequently, misclassified points are assigned a higher weight. In the model that follows, each point has a higher weight. The models will be trained until the error is as small as possible. The recall, precision, accuracy, and f1 score of the algorithms RF, DT, ET, and LGBM classifiers are used to evaluate how effective they are. Then, using the previously stated metrics, the Adaboost classifier is integrated with each of these individual classifiers to create a homogenous Adaboost classifier. The model that performs the best for the early identification of heart disease is then suggested for use[22].

### 4. Normalization

Noise in the classification performance is caused by the dataset's feature values' varying distribution [23]. The dataset is therefore normalised to lie inside the uniform [0,1] range. Normalisation is calculated using the formula below:

$$f_i^{new} = \frac{f_i^{old} - \min(F)}{\max(F) - \min(F)} \tag{Eq.(1)}$$

The "minimum and maximum" values of characteristic curves are min (F) and max (F), and the function's current value is fi.

### 5. Evolution Metrics

In this work, we evaluated each generated prediction model based on multiple supervised machine learning algorithms using different criteria. As can be seen in Table 1, the evaluation metrics employ various categorisation metrics and a confusion matrix to predict evaluated outcomes both well and poorly.

TABLE I. TABLE 1. CONFUSION MATRIX VALUES.

Metrics	Description
True Positive	The overall number of patients that receive accurate labels after testing positive.
False Positive	The overall number of patients who, although testing positive, are wrongly classified as negative.
True Negative	The overall number of patients who received a negative classification.
False Negative	The overall number of patients that received false negative labels.

In this investigation, recall, accuracy, F1 score, and precision are calculated using the following formulas: (2), (3), (4), and (5) [24], to evaluate the models' performance.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$F1 - Score = 2 * \frac{Recall*Precision}{Recall+Precision} \tag{5}$$

## IV. RESULTS

The dataset contained 1190 rows and 11 characteristics. Numerical features were converted from categorical features, and duplicate data was removed after cleaning and preprocessing, leaving 646 rows. Of the original data, 272 rows were found to be duplicates. When the data reached 720 after using the balancing process, four perfect features were found using feature selection. Outliers were

controlled by combining the data into a uniform range in order to improve model performance. Heart disease was diagnosed four hyper-classifiers (these single classifiers fused with the hyper-Adaboost model) and four single classifiers (RF, DT, ET, and LGBM) were employed in this work. We split the data 80-20 and used the test-train split method to assess our models' performance. Before any features were put to classifiers, they were all normalized. Numerous performance measures are used in this work, including as area under the ROC curve, F1 score, recall, accuracy, and precision. Two thirds of the dataset were used for testing, while the remaining 80% were used for training.

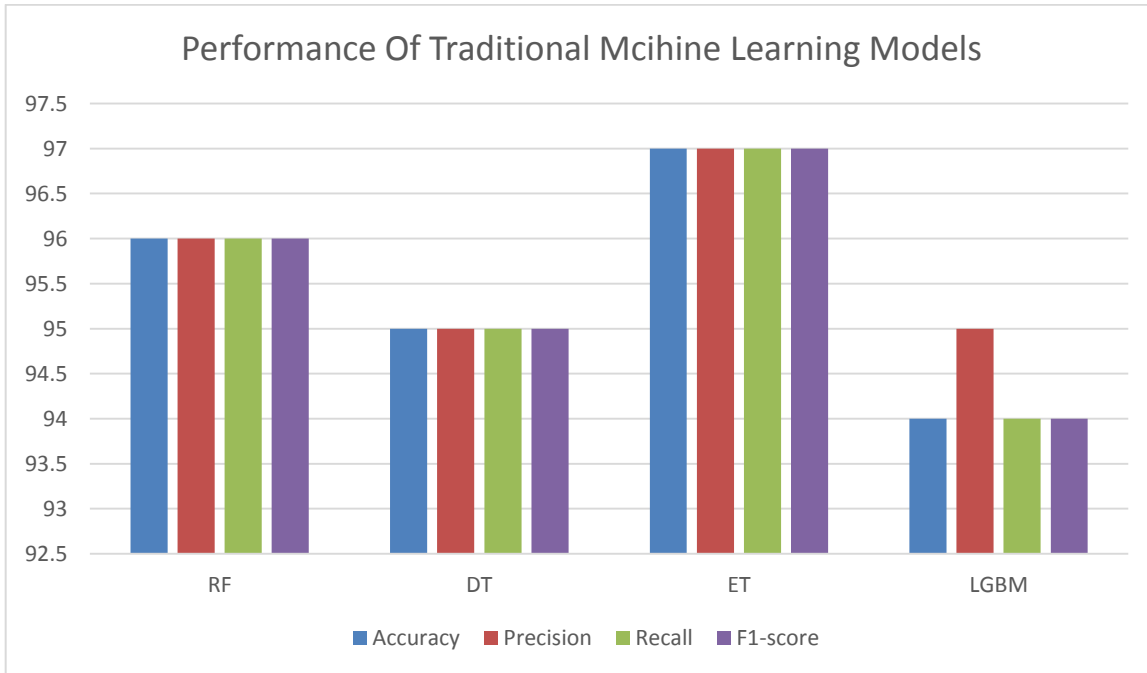


Figure 6, Classification performance on heart disease dataset FOR Traditional Machine Learning Models

The dataset's overall results for traditional machine learning are displayed in Figure 6. Recall, accuracy, precision, and f1 score were also used as evaluation criteria. 80% of the data utilized for each classifier's training and 20% for testing. The ET model, which used all features and had a 97% accuracy rate, generated the most favorable evaluation outcome using the original data.

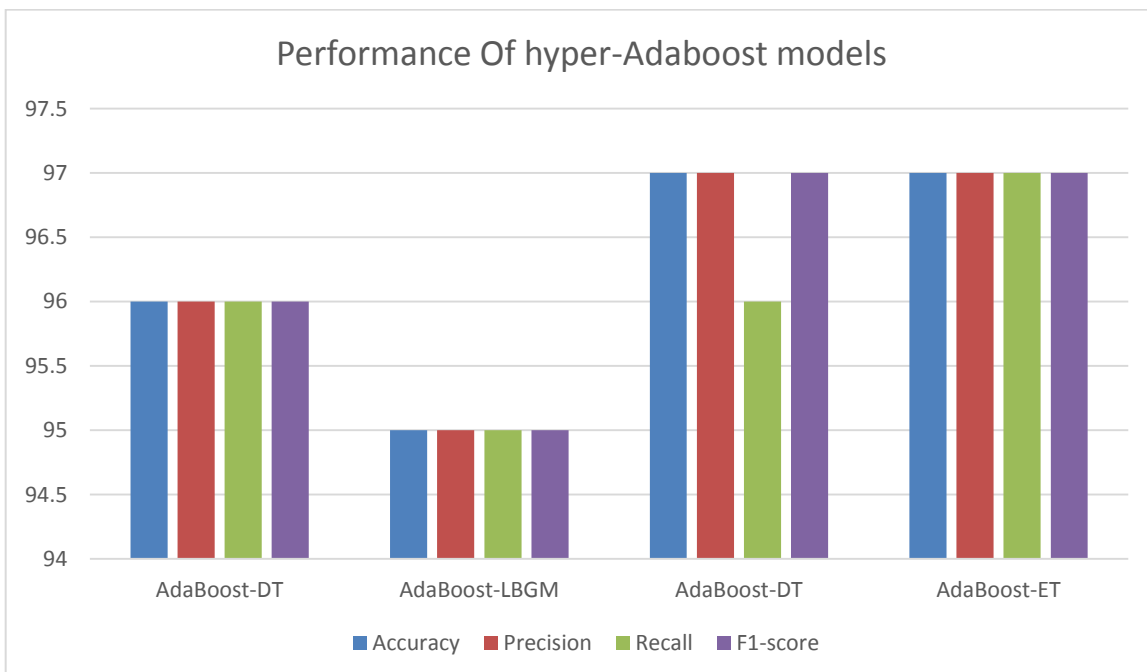


Figure 7, Classification performance on heart disease dataset for hyper-Adaboost model.

The dataset's overall results for hyper-Adaboost models are displayed in Figure 7. Recall, accuracy, precision, and f1 score were also used as evaluation criteria. 80% of the data utilised for each classifier's training and 20% for testing. The optimal evaluation result for the original data was obtained by the Hyper-ET model, which made use of all features and had a 97% accuracy rate.

A variety of machine learning classifiers were applied to the heart disease dataset, as shown in Figure 1, to determine the presence of heart illness: four single classifiers (RF, DT, ET, and LGBM) and four hyper-classifiers (where these single classifiers are fused with the hyper-Adaboost model) were used. The findings indicated that the hyper-Adaboost-ET model (97%), the RF and DT models (96% and 95%, respectively), and LGBM (94%), had the highest accuracy among the other models. The ET model had the highest accuracy (97%). The categorization was visualized using a confusion matrix, The prediction results of a classification task are stored in a confusion matrix. The total number of right and wrong predictions for each class is calculated using count values. The confusion matrix created by fitting the recommended model is shown in Figures 8 and 9. Whereas the y-axis shows expected problems, the x-axis shows actual defects. The diagonal of the matrix displays the forecasts that came true.

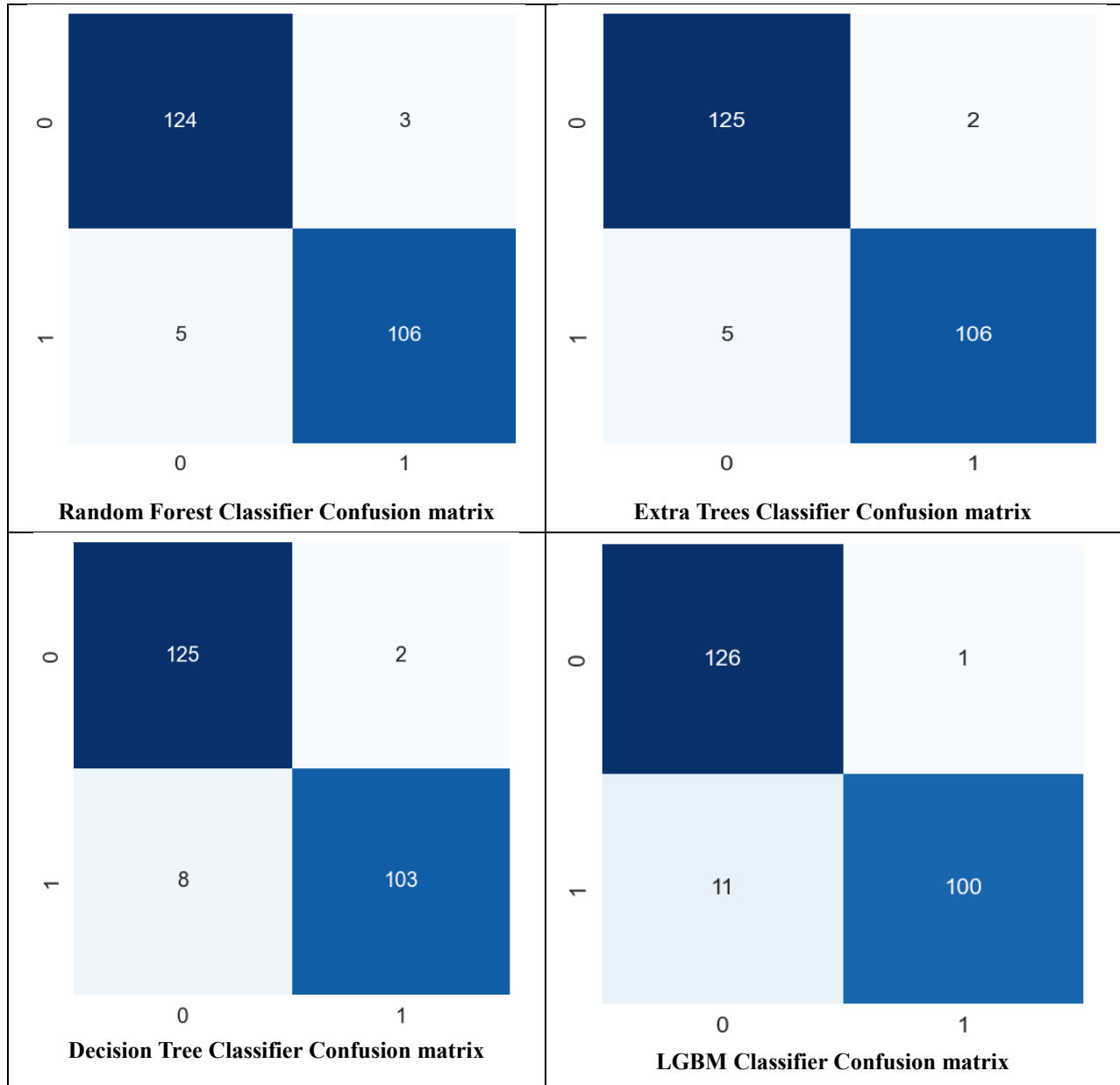


Figure 8, Confusion matrix For Traditional ML classifier

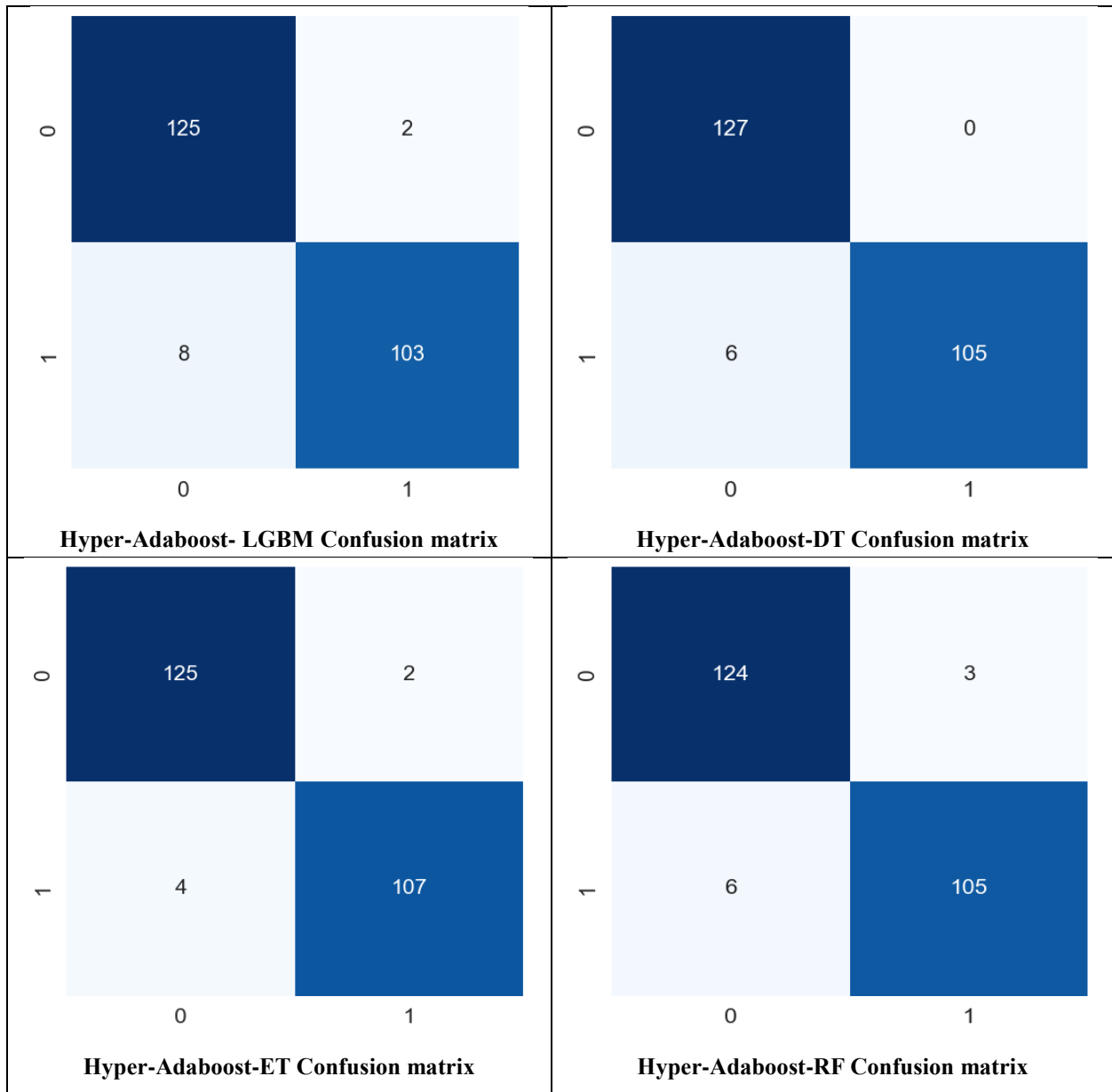


Figure 9, Confusion matrix for Hyper ML classifier

The receiver operating characteristic (ROC) curve provides a visual representation of the performance of a binary classifier. It compares the false positive rate (FPR) to the true positive rate (TPR) using a variety of classification criteria. The classifier's overall performance is determined by measuring not only its sensitivity and specificity but also its area under the ROC curve (AUC), a scalar statistic. As can be seen in Figure 10, all models have an AUC greater than 100%.

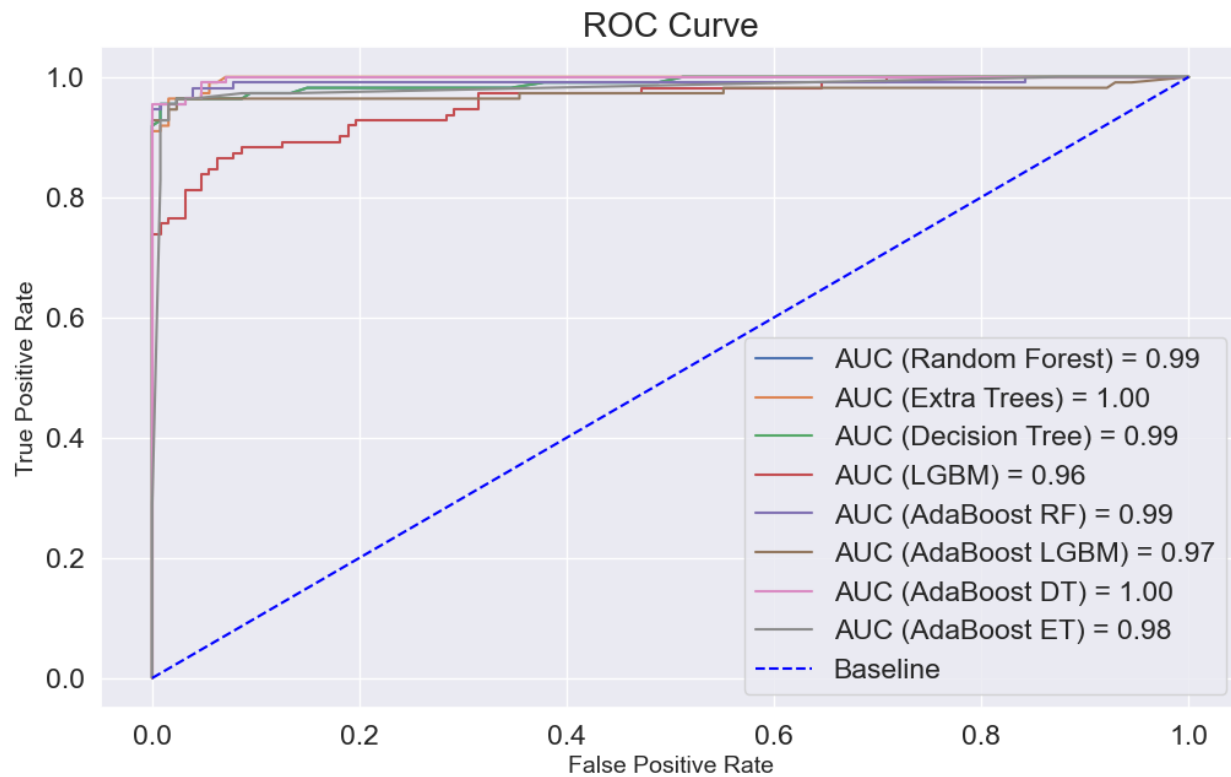


Figure 10, ROC Curve of all ML models

Table 2 shows that the suggested method outperforms earlier research. With an accuracy of 97%, the comparison clearly demonstrates the improved performance of the Hyper Adaboost-ET model above the accuracies found in previous studies.

TABLE II. TABLE 2. COMPARING PERFORMANCE WITH EARLIER RESEARCH

Author	Year	Classifier	Accuracy (%)
[12]	2023	MLP	96
[11]	2023	Stacking classifier	92
[10]	2023	CART	87.25
[5]	2023	Soft voting ensemble classifier	93.44
[7]	2022	Ensemble model	92.7
Proposed	2024	Hyper-Adaboost-ET	97

The findings show that the Hyper Adaboost classifier greatly improves the conventional ML models' ability to predict heart disease. In particular, the Hyper Adaboost-ET model stands out for its robust performance across a range of measures and great accuracy. These results highlight the promise of cutting-edge machine learning methods for medical diagnosis, especially for the early identification and treatment of cardiac disease.

## V. CONCLUSION

This study provides an example of how well-established Hyper Adaboost classifiers can be used to accurately predict heart disease. We improved these models' performance by using Hyper Adaboost on top of them, together with traditional machine learning approaches such as Random Forest, Decision Trees, and Extra Trees, besides Light Gradient Boosting Machine. In particular, the model Hyper Adaboost-ET showed a high accuracy of 97%, suggesting that this one is more capable of being resilient to avoid overfitting. Our results highlight that the nature of medical data is so complex and has non-linear relationships that this dataset can only be dealt with through advanced machine learning techniques. The proposed method not only enhances predictability but opens up early disease detection and prevention of heart diseases, which may result in better healthcare outcomes and economic savings globally. While our positive predictive value is high, this evaluation has its limitations. Although the dataset is extensive, it may not encompass all types of patients or clinical scenarios, and findings might therefore have only limited generalizability. Moreover, the computational cost and complexity of expanding Hyper classifiers make them unsuited to responsive online applications or large-scale deployments. In order to validate these models, more research with bigger, more varied data sets is necessary, and merging validations should be done carefully. Further research on other Hyper strategies and the development of computationally efficient models are also necessary to enhance predicting performance and facilitate its application in clinical settings. The field's ongoing advancements hold enormous promise for improving heart disease preventative and treatment approaches, among other serious health diseases. In the future, the incorporation of other physiological signals, such as real-time ECG monitoring, could be investigated to improve the predictive accuracy of the suggested model. Furthermore, increasing the dataset to encompass a broader demographic range and examining longitudinal research could provide further insights and increase the model's generalizability to diverse demographics and clinical situations.

## VI. References

- [1] A. V. Ferry *et al.*, "Presenting symptoms in men and women diagnosed with myocardial infarction using sex-specific criteria," *J Am Heart Assoc*, vol. 8, no. 17, Sep. 2019, doi: 10.1161/JAHA.119.012307.
- [2] "Coronary heart disease - Diagnosis - NHS." Accessed: Feb. 03, 2024. [Online]. Available: <https://www.nhs.uk/conditions/coronary-heart-disease/diagnosis/>
- [3] R. Williams, T. Shongwe, A. N. Hasan, and V. Rameshar, "Heart Disease Prediction using Machine Learning Techniques," *2021 International Conference on Data Analytics for Business and Industry, ICDABI 2021*, pp. 118–123, 2021, doi: 10.1109/ICDABI53623.2021.9655783.
- [4] "Cardiovascular diseases." Accessed: Feb. 03, 2024. [Online]. Available: [https://www.who.int/health-topics/cardiovascular-diseases#tab=tab\\_1](https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1)
- [5] K. K. Wong, D. N. Ghista, A. W. Ip, W. Zhang, N. Chandrasekhar, and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," *Processes 2023, Vol. 11, Page 1210*, vol. 11, no. 4, p. 1210, Apr. 2023, doi: 10.3390/PR11041210.
- [6] B. Kolukisa and B. Bakir-Gungor, "Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis," *Comput Stand Interfaces*, vol. 84, p. 103706, Mar. 2023, doi: 10.1016/J.CSI.2022.103706.
- [7] R. Rajendran and A. Karthi, "Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers," *Expert Syst Appl*, vol. 207, p. 117882, Nov. 2022, doi: 10.1016/J.ESWA.2022.117882.
- [8] V. Baviskar, M. Verma, P. Chatterjee, and G. Singal, "Efficient Heart Disease Prediction Using Hybrid Deep Learning Classification Models," *IRBM*, vol. 44, no. 5, p. 100786, Oct. 2023, doi: 10.1016/J.IRBM.2023.100786.
- [9] N. P and S. Narayan, "Cardiac disease detection using cuckoo search enabled deep belief network," *Intelligent Systems with Applications*, vol. 16, p. 200131, Nov. 2022, doi: 10.1016/J.ISWA.2022.200131.
- [10] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthcare Analytics*, vol. 3, p. 100130, Nov. 2023, doi: 10.1016/J.HEALTH.2022.100130.
- [11] S. Mohapatra *et al.*, "A stacking classifiers model for detecting heart irregularities and predicting Cardiovascular Disease," *Healthcare Analytics*, vol. 3, p. 100133, Nov. 2023, doi: 10.1016/J.HEALTH.2022.100133.
- [12] S. P. Barfungpa, L. Samantaray, H. Kumar Deva Sarma, R. Panda, and A. Abraham, "D-t-SNE: Predicting heart disease based on hyper parameter tuned MLP," *Biomed Signal Process Control*, vol. 86, p. 105129, Sep. 2023, doi: 10.1016/J.BSPC.2023.105129.
- [13] A. M. Al-Ssulami, R. S. AlSORori, A. M. Azmi, and H. Aboalsamh, "Improving Coronary Heart Disease Prediction Through Machine Learning and an Innovative Data Augmentation Technique," *Cognit Comput*, vol. 15, no. 5, pp. 1687–1702, Sep. 2023, doi: 10.1007/S12559-023-10151-6.
- [14] E. I. Alyasin, O. Ata, H. Mohammedqasim, and R. Mohammedqasem, "Enhancing Self-Care Prediction in Children with Impairments: A Novel Framework for Addressing Imbalance and High Dimensionality," *Applied Sciences 2024, Vol. 14, Page 356*, vol. 14, no. 1, p. 356, Dec. 2023, doi: 10.3390/APP14010356.

- [15] O. S. Tătaru *et al.*, “Artificial Intelligence and Machine Learning in Prostate Cancer Patient Management—Current Trends and Future Perspectives,” *Diagnostics 2021, Vol. 11, Page 354*, vol. 11, no. 2, p. 354, Feb. 2021, doi: 10.3390/DIAGNOSTICS11020354.
- [16] N. Maleki, Y. Zeinali, and S. T. A. Niaki, “A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection,” *Expert Syst Appl*, vol. 164, p. 113981, Feb. 2021, doi: 10.1016/J.ESWA.2020.113981.
- [17] L. H. H and B. M. A. M. A., “Comparative Study of Different Machine Learning Algorithms to Classify EEG Commands,” *Al-Iraqia Journal for Scientific Engineering Research*, vol. 3, no. 2, pp. 28–35, Jun. 2024, doi: 10.58564/IJSER.3.2.2024.176.
- [18] V. Chang, J. Bailey, Q. A. Xu, and Z. Sun, “Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms,” *Neural Comput Appl*, pp. 1–17, Mar. 2022, doi: 10.1007/S00521-022-07049-Z/FIGURES/17.
- [19] J. Singh Kushwah, A. Kumar, S. Patel, R. Soni, A. Gawande, and S. Gupta, “Comparative study of regressor and classifier with decision tree using modern tools,” *Mater Today Proc*, vol. 56, pp. 3571–3576, Jan. 2022, doi: 10.1016/J.MATPR.2021.11.635.
- [20] G. Alfian *et al.*, “Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method,” *Computers 2022, Vol. 11, Page 136*, vol. 11, no. 9, p. 136, Sep. 2022, doi: 10.3390/COMPUTERS11090136.
- [21] M. Gong, Y. Bai, J. Qin, J. Wang, P. Yang, and S. Wang, “Gradient boosting machine for predicting return temperature of district heating system: A case study for residential buildings in Tianjin,” *Journal of Building Engineering*, vol. 27, p. 100950, Jan. 2020, doi: 10.1016/J.JOBE.2019.100950.
- [22] E. Najjar and A. M. Breesam, “Supervised Machine Learning a Brief Survey of Approaches,” *Al-Iraqia Journal for Scientific Engineering Research*, vol. 2, no. 4, pp. 71–82, Dec. 2023, doi: 10.58564/IJSER.2.4.2023.121.
- [23] H. M. Qasim, O. Ata, M. A. Ansari, M. N. Alomary, S. Alghamdi, and M. Almeahmadi, “Hybrid Feature Selection Framework for the Parkinson Imbalanced Dataset Prediction Problem,” *Medicina 2021, Vol. 57, Page 1217*, vol. 57, no. 11, p. 1217, Nov. 2021, doi: 10.3390/MEDICINA57111217.
- [24] A. A. Jasim, L. R. Hazim, H. Mohammedqasim, R. Mohammedqasem, O. Ata, and O. H. Salman, “e-Diagnostic system for diabetes disease prediction on an IoMT environment-based hyper AdaBoost machine learning model,” *Journal of Supercomputing*, vol. 80, no. 11, pp. 15664–15689, Jul. 2024, doi: 10.1007/S11227-024-06082-0/TABLES/4.