ISSN: 2634 - 6796

Journal of Cardiology Research Reviews & Reports

Review Article

Open Access

Hybrid Model for Improved Heart Disease Prediction

Rohit Khankhoje

Independent Researcher, Avon, Indiana, USA

ABSTRACT

Cardiovascular disease, which encompasses various conditions affecting the heart and blood vessels, is a significant global health concern and a primary cause of mortality on a global scale. These ailments have a profound impact on heart function, blood circulation, and overall well-being. This investigation introduces a novel hybrid model that effectively combines the strengths of Decision Tree (DT), Logistic Regression (LR), and Artificial Neural Network (ANN) algorithms, thereby significantly augmenting the accuracy of heart disease prediction. The model demonstrates exceptional performance, boasting an impressive accuracy rate of 88%, which surpasses the individual accuracies of DT at 99%, LR at 80%, and ANN at 86%. Furthermore, the hybrid approach excels in precision, recall, and F1-score metrics, thereby substantiating its reliability and robustness as a predictive tool for heart disease. This research underscores the advantages of incorporating multiple algorithms in order to create a more efficient predictive model for cardiovascular health diagnostics.

***Corresponding author**

Rohit Khankhoje, Independent Researcher, Avon, Indiana, USA.

Received: March 03, 2021; **Accepted:** March 11, 2021; **Published:** March 20, 2021

Keywords: Decision Tree, Logistic Regression, Artificial Neural Network, Hybrid Model

Introduction

The heart, a vital organ, plays a pivotal role in maintaining overall well-being as it is primarily responsible for the circulation of blood and the delivery of essential nutrients and oxygen to the various organs and tissues of the body. This intricate circulation process is managed through a complex network within the circulatory system. It is important to note that the health of the heart is closely interconnected with other crucial organs such as the brain, lungs, liver, and kidneys. The presence of heart disease not only disrupts the normal functioning of the heart but also has a significant impact on other bodily systems, thereby greatly affecting both the survival and quality of life of individuals.

Heart diseases have the potential to impede the flow of blood to the brain, which can lead to cognitive impairments, as well as hinder lung function by compromising the oxygenation of blood. Factors such as chronic stress, mental health disorders, and unhealthy habits like smoking have been identified as known contributors to the development of heart disease [1]. Moreover, a lack of physical activity, hypertension, high cholesterol levels, and obesity have also been identified as major risk factors. According to the World Health Organization, cardiovascular diseases account for approximately 17.9 million deaths annually, making up 31% of global fatalities and positioning heart disease as a leading cause of death worldwide.

It is worth noting that the early stages of heart disease often go unnoticed, resulting in preventable deaths. Therefore, the timely detection of heart disease is of utmost importance and typically relies on a combination of physical examinations, internal symptoms, and clinical indicators, including various pathological and functional factors. However, relying solely on these factors can sometimes lead to delays in diagnosis, potentially resulting in misinterpretations and unpredictable health outcomes [2].

To address these challenges, the development of an expert system for medical diagnosis and early prediction is necessary. This system should offer high accuracy while keeping operational costs low. The emergence of machine learning (ML) and deep learning (DL) techniques in the field of medicine, particularly in disease prediction, is worth mentioning. While traditional ML models may struggle to identify complex patterns, and DL models require extensive labeled data and run the risk of overfitting, a hybrid model that combines both approaches presents a promising solution.

In this study, we propose a hybrid model that aims to integrate the strengths of ML and DL, with the purpose of overcoming the limitations inherent in each approach. This model seeks to leverage advanced computational techniques for the early detection and diagnosis of heart disease, thereby advancing healthcare outcomes and reducing global mortality rates associated with cardiovascular diseases.

Related Work

Sr No	Related work title	Conclusion
1	Heart Disease Prediction System using Hybrid Technique of Data Mining Algorithms	Hybrid Genetic Naive Bayes Model provides accurate results for heart disease prediction. Alternative data processing techniques like clustering and association rules can be used for prediction.
2	Hybrid Ensemble Framework for Heart Disease Detection and Prediction	Proposed hybrid ensemble model improves heart disease detection and prediction. Merging initial features with predicted class labels leads to better performance.
3	A novel approach for diagnosing heart disease with hybrid classifier	Proposed method improves heart disease prediction accuracy. GSO-LM algorithm outperforms existing approaches in terms of performance measures.
4	Prediction of heart disease using hybrid technique for selecting features	The proposed feature selection approach improves accuracy for both models. The selected features for the models are cp, thal, sex, exang, ca, age, slope, oldpeak, thalach, trestbps, chol, restecg, and fbs.
5	Heart disease prediction using classification and feature selection techniques	Age, blood pressure, and blood serum cholesterol are the highest risk factors for heart disease. A hybrid model of Genetic and Naive Bayes techniques called Hybrid Genetic Naive Bayes Model is developed for predicting high accuracy in heart disease diagnosis.

Methodology

The primary aim of this investigation is to prognosticate the probability of cardiac illness by means of computational anticipation, thereby bestowing noteworthy advantages upon both healthcare professionals and patients. A systematic approach was executed to achieve this, encompassing various stages: assembling data, preprocessing it for examination, training models, assessing their efficacy, and scrutinizing the outcomes. The inquiry entailed training the Decision Tree, Logistic Regression, and Artificial Neural Network (ANN) models independently, while appraising each through crucial metrics such as accuracy, precision, recall, and F1-score. Expanding upon this, a Hybrid model was devised, harmonizing the prognostications from the three individual models to exploit their collective strengths. This amalgamation is portrayed in figure 1 as follows:

Data Collection

In this investigation, we employed a collection of patient records sourced from Kaggle that consists of 1025 instances, each characterized by 14 distinct features pertaining to the risk factors associated with heart disease. This dataset is notable for its capacity to accurately represent and encompass a wide range of relevant information, having been sourced from meticulous medical examinations and diverse patient characteristics. In order to prepare the dataset for subsequent analysis, a preprocessing phase was undertaken, which included the normalization of features and the resolution of any missing data instances [3]. The attributes themselves encapsulate a variety of influential factors, including but not limited to age, gender, as well as both systolic and diastolic blood pressure levels. It is important to note that the dataset in question defines the "cardio" target variable, which serves as an indicator of the individual's cardiovascular health status. More specifically, a value of 0 denotes a healthy cardiovascular function, while a value of 1 signifies the presence of cardiovascular disease.

Data Preprocessing

Before commencing the training of the models, the dataset underwent a preprocessing phase with the aim of augmenting its quality and uniformity. This phase encompassed the management of any missing values, the normalization of the features, and the division of the dataset into separate training and testing subsets. The missing values were addressed through the utilization of standard scalar techniques in order to uphold the overall integrity of the dataset. The process of feature normalization was executed to standardize the range of the features, guaranteeing that all features were on a comparable scale for precise evaluation across diverse models [4]. Following this, the dataset was partitioned, with specific portions designated for training and testing, a critical step in effectively assessing the models' capacity to generalize to new data.

Data Splitting

In order to conduct the study, the dataset underwent a division into training and testing sets in an 80:20 ratio. To ensure an equitable representation of each class, stratified sampling was employed, which resulted in a balanced distribution of classes in both subsets.

Model Training

To address the task of predicting heart disease, three distinct machine learning models were utilized: Decision Tree, Logistic Regression, and Artificial Neural Network (ANN). Each model underwent training on the training set, utilizing its specific algorithm. The Decision Tree model employed a hierarchical approach to segment the feature space, Logistic Regression aimed to establish a linear correlation between the features and the target variable, and the ANN utilized a multi-layered structure to discern intricate patterns in the data. Additionally, a Hybrid model was proposed, combining the individual model predictions to leverage the strengths of each [5].

Performance Evaluation and Metrics

The assessment of the models' predictive efficiency was conducted using several key metrics: accuracy, precision, recall, and F1-score. Accuracy measures the overall rate of correct predictions, precision evaluates the ratio of true positives among all positive predictions, recall calculates the proportion of true positives from all actual positives, and the F1-score provides a balanced measure of precision and recall [6]. These metrics collectively offer a comprehensive evaluation of the models' performance capabilities.

Evaluation Metrics Details

The models were analyzed using four standard metrics: accuracy, precision, recall, and F1-score. These metrics were calculated based on the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) observed in the test data. The mathematical formulations for classification accuracy (CA), precision, recall, and F1-score are presented in Equations

1, 2, 3, and 4, respectively, providing a quantitative assessment of the models' effectiveness in the test scenarios.

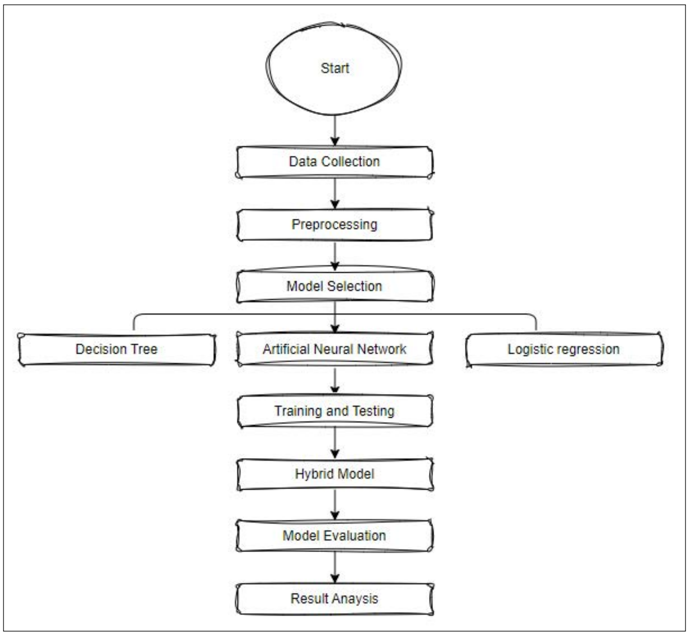

Figure 1: Proposed Method Flow

- Accuracy: This metric is calculated using the formula $\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{FP}+\text{TN}+\text{FN})}$, where it quantifies the overall correct predictions made by the model.
- Precision: Precision is determined by the formula $\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$, reflecting the proportion of actual positives correctly identified from all positive predictions made.
- Recall: Also known as Sensitivity, Recall is computed using $\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$, representing the fraction of true positive cases accurately identified out of all actual positive cases.
- F1-Score: The F1-Score is calculated as $\text{F1} - \text{Score} = 2(\frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}})$, providing a harmonic mean of Precision and Recall, and especially is useful when seeking a balance between these two metrics.

Results & Discussion

This investigation was carried out utilizing Google Colab on a personal computer outfitted with an Intel(R) Pentium(R) processor and 8 GB of RAM. The initial dataset consisted of 1025 entries with 14 attributes. However, after undergoing data cleaning and preprocessing, it was condensed to 820 entries encompassing 13 attributes. Given that all attributes were categorical, the removal of outliers was conducted to improve the performance of the model. The algorithms employed in this study included Logistic Regression, Decision Tree, Artificial Neural Network, and a Hybrid Model. Performance metrics such as precision, recall, accuracy, F1 score, and the area under the ROC curve were utilized. The division of the dataset allocated 80% for training purposes and 20% for model testing.

Description of the Data

The dataset comprises 1025 rows and 14 features. These features encompass age, sex, chest pain (cp), trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, and a target variable indicating the presence or absence of heart disease. This dataset was obtained from www.Kaggle.com, the largest community in the field of data science.

Data Imbalance

An examination of the balance of the data is of utmost importance. As depicted in Figure 2, there exists an imbalance between the data representing cases with heart disease (1) and those without heart disease (0). This observation is crucial in understanding the composition of the dataset and informing subsequent steps in data processing.

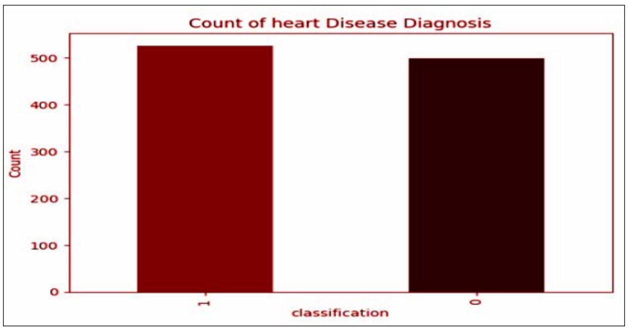

Figure 2: Data Imbalance

Citation: Rohit Khankhoje (2021) Hybrid Model for Improved Heart Disease Prediction. Journal of Cardiology Research Reviews & Reports. SRC/JCRRR-209. DOI: doi.org/10.47363/JCRRR/2021(2)189

Correlation Matrix

The Correlation Matrix was Used to Determine the Relationship or Strong Correlations with the Classification Row. As Shown in Figure 3 below

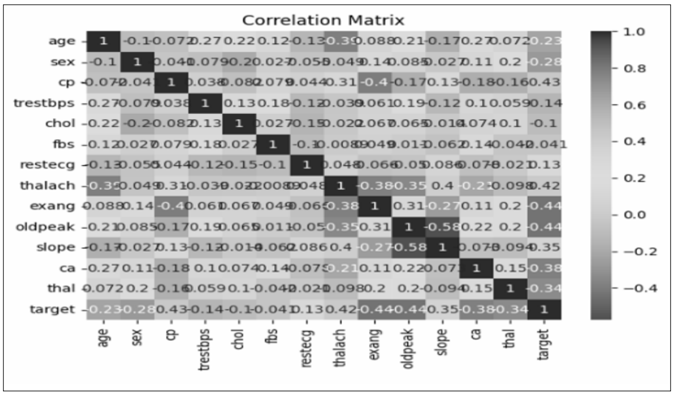

Figure 3: Correlation Matrix

Data Preprocessing

To determine the thoroughness of our dataset, we utilized the data.isnull().sum() method from Python's Pandas library. This method facilitated the quantification of the missing values (or NaN values) across each column in our DataFrame, which is named 'data'. The identification of these null values is crucial to ensure that any gaps in the dataset are appropriately addressed using suitable methods [7].

In addition, to examine data redundancy, we employed the expression data duplicated().sum(), which allowed us to calculate the number of duplicate rows within the DataFrame. As illustrated in Figure 4 below, our analysis demonstrated that the dataset does not contain any duplicate entries, confirming its uniqueness.

For the preprocessing stage, we imported the StandardScaler class from the sklearn preprocessing module. This class plays a significant role in handling categorical variables within the dataset. The StandardScaler function was applied to normalize the features of the heart disease dataset. It accomplished this by transforming the dataset features to a standard scale with a mean of 0 and a standard deviation of 1[8]. Standardizing the features in this manner is a crucial preprocessing step as it greatly enhances the performance and consistency of machine learning models trained on this data.

Training Process

The initiation of the model training commenced with the execution of the fit technique on each model object. This particular technique played a crucial role in specifying the training dataset (X_train and y_train) and the validation dataset (X_test and y_test). Numerous models were incorporated in the research endeavor to forecast heart disease, namely the Decision Tree, Logistic Regression, Artificial Neural Network (ANN), and a Hybrid Model.

The Decision Tree model operates by iteratively dividing the data based on its distinct features. It effectively maximizes class segregation by utilizing either Gini impurity or entropy as its criteria. In contrast, Logistic Regression estimates the coefficients of the features and models the probability of heart disease occurrence using a logistic function [2].

The learning process of the ANN model entails iterative adjustments of the weights and biases. It employs a technique called backpropagation to update these weights, thereby enhancing the accuracy of the model. On the other hand, the Hybrid Model amalgamates predictions from the individual models. It adeptly captures simple patterns, linear relationships, as well as intricate interactions within the data.

Each model is meticulously trained to discern and interpret various dimensions of the relationships and patterns within the data. This comprehensive and inclusive approach to model training equips the study with a robust and comprehensive framework for accurately predicting heart disease.

Result of ANN

As depicted in Table 1 hereinafter, this investigation outlines the results derived from the training of an Artificial Neural Network (ANN) for the identification of cardiac ailments. The outcomes encompass data from 50 epochs of training, encompassing the duration of training and the step loss, which serve as indicators of the model's effectiveness.

A noteworthy metric presented is the validation loss, which represents the calculated loss value on the validation dataset following each epoch of training. Remarkably, with an increase in the number of epochs, there is a notable enhancement in both the accuracy and loss metrics of the model. This implies that the ANN is not only acquiring knowledge effectively from the training data but also demonstrating good generalization capabilities to unfamiliar data.

The training and validation accuracies of the model are observed to be relatively high, thus reinforcing the efficacy of the model in learning and its ability to generalize. Additionally, the observed trends in the validation loss and accuracy offer valuable insights. These trends play a crucial role in discerning the point at which the model might start to overfit. Overfitting is typically suggested when the validation loss begins to rise, even as the training loss continues to decline. This phenomenon indicates that while the model performs increasingly well on the training data, its performance on new data starts to deteriorate, which is a critical aspect to consider in evaluating and optimizing the model.

No	Training Time	Step Loss	Accuracy
1	1s 2ms/step	0.5692	0.7293
2	0s 2ms/step	0.4863	0.7939
3	0s 2ms/step	0.4278	0.8293
4	0s 2ms/step	0.3843	0.8512
5	0s 2ms/step	0.3538	0.8622
...	...	...	...
48	0s 2ms/step	0.1745	0.9317
49	0s 2ms/step	0.1718	0.9329

Classification reports

A classification report is an essential tool for evaluating a model's performance, providing detailed insights into various metrics like precision, recall, F1-score, and support for each class. These components differ for each model, offering a comprehensive view of their respective performances.

Decision Tree Classification Report

Below presents the Decision Tree Classification Report. This report showcases the performance metrics of the Decision Tree model on a dataset for predicting binary outcomes. The metrics included are precision, recall, F1-score, support, along with the model's overall accuracy, macro average, and weighted average.

-Precision quantifies the accuracy of the model in correctly predicting positive outcomes, demonstrating a precision of 97% for class "0" and 100% for class "1".
- Recall measures the model's capacity to correctly identify positive instances among all actual positives.
- F1-score of approximately 0.99 indicates a high balance between precision and recall.
- Support refers to the count of instances for each class in the test dataset.

Decision Tree Classification Report

Class	Precision	Recall	F1-Score	Support
0	0.97	1.00	0.99	102
1	1.00	0.97	0.99	103
Accuracy			0.99	205
Macro Avg	0.99	0.99	0.99	205
Weighted Avg	0.99	0.99	0.99	205

Artificial Neural Network Classification Report
Below displays the ANN Classification Report, showcasing the Artificial Neural Network's performance in predicting heart disease. This model demonstrates a balanced precision and recall for both classes, as well as commendable F1-scores, indicating a good balance between precision and recall. The model correctly predicts approximately 86% of instances.

The support column in the report indicates the count of each class's instances in the test set. The overall model accuracy stands at 0.86. Additionally, the macro average and weighted average values are provided, reflecting the model's balanced performance in handling class imbalances and predicting outcomes accurately.

Table 1- Ann Classification

Class	Precision	Recall	F1-Score	Support
0	0.89	0.83	0.86	102
1	0.84	0.89	0.87	103
Accuracy			0.86	205
Macro Avg	0.86	0.86	0.86	205
Weighted Avg	0.86	0.86	0.86	205

Logistic Regression Classification Report
Logistic Regression Classification Report. This report illustrates the model's moderate performance in predicting heart disease outcomes, with precision and recall values displaying variability across both classes. The model demonstrates a balanced trade-off between precision and recall, achieving an overall accuracy of 0.80. This suggests that the model correctly predicts outcomes for approximately 80% of the instances.

The F1-scores in the report reflect a similar balance between precision and recall for both classes, indicating the model's effectiveness in managing the trade-off. The support column indicates the count of instances for each class in the test dataset. Additionally, the macro average and weighted average are presented in the report, taking into account the class imbalances and providing an overview of the model's moderate performance across different classes.

Table 2: Logistic Regression Classification

Class	Precision	Recall	F1-Score	Support
0	0.85	0.72	0.78	102
1	0.76	0.87	0.81	103
Accuracy			0.80	205
Macro Avg	0.80	0.79	0.79	205
Weighted Avg	0.80	0.80	0.79	205

Hybrid Model Classification Report
This report indicates that the Hybrid Model exhibits a strong performance, with well-balanced precision and recall values for both classes. The model achieves an impressive accuracy of 0.90 for class "0" (90%) and 0.86 for class "1" (86%). The recall metrics are similarly robust, with a 0.85 recall for class "0" and a 0.90 recall for class "1". The F1-scores reflect a balanced trade-off between precision and recall across both classes.

The support column in the report denotes the number of instances for each class in the test dataset. Overall, the model demonstrates an accuracy of 0.88, accurately predicting outcomes for approximately 88% of instances. The macro and weighted averages are included in the report, considering class imbalances and affirming the model's balanced and accurate performance in outcome prediction.

Table 3: Hybrid Classification

Class	Precision	Recall	F1-Score	Support
0	0.90	0.85	0.87	102
1	0.86	0.90	0.88	103
Accuracy			0.88	205
Macro Avg	0.88	0.88	0.88	205
Weighted Avg	0.88	0.88	0.88	205

Roc Curve Analysis
The analysis of the Receiver Operating Characteristic (ROC) curve is a vital method for assessing the performance of a classification model. It visually represents the model's ability to differentiate between classes at different threshold levels for classifying instances. Accompanying the ROC curve is the metric known as the Area Under the Curve (AUC), which provides a quantification of the model's overall effectiveness. A higher AUC value indicates a superior capability of the model to correctly rank positive instances above negative ones. The closer the AUC value is to 1, the more accurate the model's classification becomes.

Figure 4 depicted below showcases the ROC curves for several models utilized in this investigation: the Decision Tree, Logistic Regression, Artificial Neural Network (ANN), and the Hybrid model. These curves unveil the true positive rate (recall) of each model at varying false positive rates, thereby shedding light on their ability to effectively distinguish between positive and negative classes.

The ANN model demonstrates a strong discriminative capacity between classes, achieving an impressive AUC value of 0.96, positioning it between the Decision Tree and Logistic Regression models in terms of performance. On the other hand, the Hybrid model stands out by exhibiting a high true positive rate across a range of false positive rates, boasting an AUC value of 0.99, which is on par with the Decision Tree model.

DOI: doi.org/10.47363/JCRRR/2021(2)189

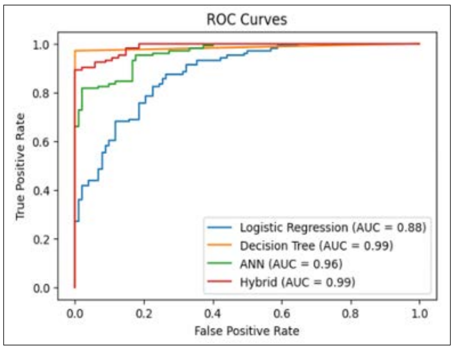

Figure 4: ROC Curves

Taken together, these ROC curves and AUC values collectively indicate that all models exhibit excellent discriminatory abilities between positive and negative classes. The ANN and Hybrid models, in particular, demonstrate exceptional performance, underscoring their resilience in classifying instances within the context of heart disease prediction.

Comparative Result Analysis

The Hybrid Model (HM) developed in this study achieves an accuracy rate of 88%, indicating its ability to accurately predict outcomes in approximately 88% of cases. In contrast, the Decision Tree model exhibits an impressive accuracy of 99%, suggesting near-perfect prediction accuracy for almost all instances. On the contrary, the Logistic Regression model demonstrates a moderate accuracy of 80%, indicating its ability to accurately predict outcomes in 80% of instances. This particular model also displays a moderate accuracy and F1-score, with a focus on recall for the positive class.

The Artificial Neural Network (ANN) model displays an accuracy of 86%, signifying its capability to accurately predict outcomes in about 86% of cases. It possesses a precision of 0.84 for the positive class, indicating that approximately 84% of its positive predictions are correct. The recall for the positive class stands at 0.89, suggesting that it identifies approximately 89% of actual positive instances. Additionally, its F1-score is 0.87, reflecting a favorable balance between precision and recall, resulting in a strong overall performance.

These performance metrics demonstrate that the proposed Hybrid Model achieves a higher accuracy compared to the individual models, while maintaining a balance in precision, recall, and overall predictive ability. This accomplishment highlights the effectiveness of the Hybrid Model in the specific context of the study.

Table 4: Comparative Results

	Model	Accuracy	Precision	Recall	F1 Score	AUC
Proposed	HM	0.88	0.86	0.90	0.88	0.99
	DT	0.99	1.0	0.97	0.99	0.99
	LR	0.8	0.76	0.87	0.81	0.88
	ANN	0.86	0.84	0.89	0.87	0.96

Conclusion

This investigation presents a novel hybrid model for the prediction of heart disease, amalgamating the strengths of Decision Tree (DT), Logistic Regression (LR), and Artificial Neural Network (ANN). The primary objective is to harness the unique capabilities of these individual models in order to create a more effective and comprehensive predictive instrument. The hybrid model's performance metrics, namely accuracy, precision, recall, and F1-score, were meticulously evaluated and compared to the performance of the standalone models.

The findings demonstrate the efficacy of integrating diverse modeling techniques to enhance the accuracy of medical predictions. The hybrid model achieves an accuracy rate of 0.88, surpassing the accuracy of LR (0.8) and ANN (0.86), while closely rivaling the remarkable accuracy of DT (0.99). Its precision rate of 0.86 and recall rate of 0.90 indicate its proficiency in accurately identifying positive instances. Notably, the hybrid model's ROC-AUC score of 99 surpasses that of the individual models, highlighting its superior ability to effectively differentiate between classes. Additionally, an F1-score of 0.88 suggests a well-balanced performance in terms of precision and recall. This hybrid model emerges as a robust tool for predicting heart disease, comparable to the high accuracy of the DT model, while maintaining consistent predictive strength for both positive and negative cases.

Despite the promising outcomes of the hybrid model, potential areas for future research could include:

- Refining the Model's Parameters: The optimization of the hybrid model's parameters could further enhance its predictive performance.
- Incorporating Additional Features and Advanced Architectures: The integration of more comprehensive features and the exploration of sophisticated neural network architectures could augment the model's predictive capabilities.
- Exploring Dynamic Ensemble Techniques: The investigation of ensemble methods that dynamically combine models based on specific characteristics of instances could potentially lead to even more accurate predictions.

References

1. Gandhi Monika, Shailendra Narayan Singh (2015) Predictions in heart disease using techniques of data mining. Futuristic Trends in Computational Analysis and Knowledge Management (ABLAZE), 2015 International Conference on. IEEE.
2. Abdar Moloud, Sharareh R Niakan Kalhori, Tole Sutikno, Imam Much Ibnu Subroto, Goli Arji (2015) Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases. International Journal of Electrical and Computer Engineering (IJECE) 5: 1569-1576.
3. Purusothaman G, Krishnakumari P (2015) A survey of data mining techniques on risk prediction: Heart disease." Indian Journal of Science and Technology 8.
4. Asmi Shabana P, Justin Samuel S (2015) An analysis and accuracy prediction of heart disease with association rule and other data mining techniques. Journal of Theoretical and Applied Information Technology 79: 254.
5. Bandyopadhyay Sunayan, Julian Wolfson, David M Vock, Gabriela Vazquez-Benitez, Gediminas Adomavicius, et al. (2015) Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular

risk from electronic health record data. Data Mining and Knowledge Discovery 29: 1033-1069.
6. Andreu-Perez Javier, Carmen CY Poon, Robert D. Merrifield, Stephen TC Wong, Guang-Zhong Yang (2015) Big data for health." IEEE Journal of biomedical and health informatics 19: 1193-1208.
7. Bahrami Boshra, Mirsaeid Hosseini Shirvani (2015) Prediction and Diagnosis of Heart Disease by Data Mining Techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST) 2: 164-168.
8. Makwana Ankur, Jaymin Patel (2015) Decision Support System for Heart Disease Prediction using Data Mining Techniques." International Journal of Computer Applications 117.