Machine Learning for Enhanced Heart Disease Prediction: A Comprehensive Classifier Evaluation

Tonghui Wu[a]
Information and Computer Science, Xi'an Jiaotong-Liverpool University, Jiangsu, China

Keywords: Machine Learning, Prediction Model, Heart Disease, Classifier Evaluation.

Abstract: In recent decades, the growing recognition of the importance of preventing heart disease and identifying potential issues early has become paramount. Advances in machine learning (ML) technologies, fueled by the wealth of medical data, have emerged as essential tools in accurately forecasting cardiovascular diseases. This study aims to address the challenge of predicting heart disease with greater accuracy, an endeavor critical to the field of healthcare due to heart disease being a leading cause of mortality globally. Utilizing a comprehensive dataset sourced from a reputable cardiology database enriched with features reflecting mental health states such as degrees of depression, the study diverges from traditional models by incorporating these psychosocial factors. Extensive evaluation of twelve different ML classifiers, including Logistic Regression, Decision Trees, and Neural Networks, among others, was conducted to assess their performance in accurately predicting heart disease. The evaluation metric of choice was the F1 score, selected for its balance between precision and recall, particularly pertinent in medical diagnostics. Findings reveal that Logistic Regression outperformed other classifiers regarding accuracy, precision, recall, and F1 score. This supports the hypothesis that incorporating mental health indicators can enhance predictive models for heart disease. The study underscores the importance of considering both physiological and psychological factors in heart disease prediction and highlights the efficacy of ML techniques in navigating the complexities of healthcare diagnostics.

1 INTRODUCTION

In the last few decades, there has been a growing understanding of the significance of preventing heart disease and detecting potential problems at an early stage. Cardiovascular disease is a significant global cause of death, as stated by the World Health Organisation. Annually, over 17.9 million individuals succumb to cardiovascular disease, which represents approximately 31% of all global deaths (Novidianto et al., 2021). The copious amount of extensive medical data and technological developments, particularly in ML, have evaluated the probability of cardiovascular disease, which is a pivotal area of study in the medical domain (Cao et al., 2022). Given the substantial volume of medical data inside the healthcare industry, ML approaches have become crucial for producing accurate predictions regarding cardiovascular disorders. This is mostly because of the developments in ML techniques (Slart et al., 2021).

In a study conducted by Yazdani et al. (2021), seven classification approaches, including k-Nearest Neighbours (k-NN), Naive Bayes, and Support Vector Machine (SVM), were employed to develop predictive models for cardiovascular disease (Yazdani et al., 2021). In addition, Chicco et al. (2020) emphasize the benefits of using the Matthews Correlation Coefficient (MCC) and accuracy as the standard of evaluating the classification (Chicco et al., 2020). Furthermore, Latha et al. (2019) enhanced the precision of forecasting the likelihood of cardiovascular illness by employing integrated classification algorithms (Latha et al., 2019). Prior research has produced advancements in the realm of heart disease prognosis. However, some models frequently have restrictions regarding data set collection, analysis and classifier evaluation,

[a] https://orcid.org/0009-0000-5541-6140

Wu, T.
Machine Learning for Enhanced Heart Disease Prediction: A Comprehensive Classifier Evaluation.
DOI: 10.5220/0012961000004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 633-638
ISBN: 978-989-758-713-9

underscoring the persisting hurdles in heart disease prediction (Klyachkin et al., 2014) (Suresh et al., 2022) (Assegie et al., 2022). These gaps lead to limitations in individualized heart disease risk assessment and highlight the urgency of exploring and expanding heart disease datasets to develop new models with more accurate classifiers. Furthermore, the traditional data sets used to train the model primarily focused on physiological and clinical metrics. However, such data sets rarely take into account the impact of people's mental state on the development of heart disease. Multiple research have demonstrated that psychological elements, including affective emotions and psychological well-being, might influence cardiovascular health (CVH) (Castillo-Mayén et al, 2021). Although depression, anxiety, and stress have a negative effect on patients suffering from chronic heart failure (CHF), these conditions are not being recognized and treated well in this vulnerable group (Tsabedze et al., 2021). This burgeoning field of inquiry seeks to understand how the fluctuations in one's mental well-being, often a byproduct of the fast-paced modern life, may predispose individuals to or exacerbate existing cardiovascular issues (Omasu et al., 2022).

Despite a plethora of classifiers available in machine learning libraries, there is no consensus on an optimal approach for heart disease prediction. As indicated in this research, in this study, the F1 score is selected as the prior evaluation metric to ensure a balance between precision and recall, crucial for medical case classification. Through more in-depth exploration of the data set, this study also evaluated the accuracy of 12 classifiers for this non-traditional data set. Rather than taking a single classifier for testing, it is significantly different from existing methods, which have rarely been fully explored in previous studies. Diverging from the datasets traditionally utilized in related research, this study enriches the datasets by incorporating features indicative of depressive states. The inclusion of mental health parameters, such as the degree of depression, signifies a novel approach to constructing predictive models. This expansion acknowledges the complex interplay between the mind and the body and is indicative of a holistic perspective that has been underrepresented in previous models. By doing so, a more comprehensive understanding of heart disease risks, one that encapsulates both the physical and psychological dimensions of health can be achieved.

2 METHOD

2.1 Data Preparation

In examining the provided visual data, it is imperative to consider two factors. Firstly, the distribution across different categories within the dataset is of note. Secondly, the influence of each categorical distribution on accurately predicting the existence or nonexistence of heart disease is of considerable interest.

The dataset used in this study was sourced from a reputable database in the cardiology field (Kaggle, 2024), encompassing an array of patient health indicators. It comprises approximately 304 samples (actual number to be inserted), with each sample characterized by 14 features, including age, gender, type of chest discomfort, etc. These features are categorized into several categories. Each category is associated with corresponding label names, such as " existence of heart disease" (1) and "non- existence of heart disease" (0).

The dataset underwent thorough cleaning and validation before pre-processing to ensure data quality and analytical accuracy. Initially, any incomplete or missing data entries (as shown in Figure 1) were removed. Subsequently, numerical features were normalized to eliminate discrepancies between different scales.

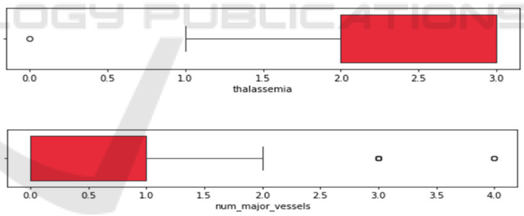

Figure 1: The box plot of dataset before data pre-processing (Photo/Picture credit: Original).

Moreover, given the potential link between heart disease occurrence and mental health, this study incorporated features reflecting psychological states, such as the degree of depression, a dimension often absents in traditional heart disease datasets. Including these mental health parameters expanded the feature set of the dataset, offering a more comprehensive perspective for examining heart disease risks that include physiological and psychosocial factors.

From Figure 2, it can be observed that the dataset is reasonably balanced. The near-equal proportion of cases ensures that the predictive modelling is less likely to be biased toward one class.

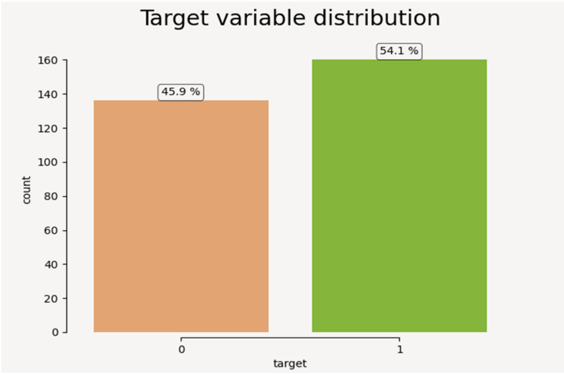

Figure 2: The target variable distribution of dataset (Photo/Picture credit: Original).

Figure 3 provides a series of density plots for cholesterol levels, maximum heart rate achieved, ST depression, etc., to depict the distribution of numerical features within the dataset, segmented by the target variable. A count plot for the number of major vessels (num_major_vessels) is also shown. These plots serve a dual purpose: they offer insight into the distributions of individual numerical features and reveal the differences between those with and without heart disease.

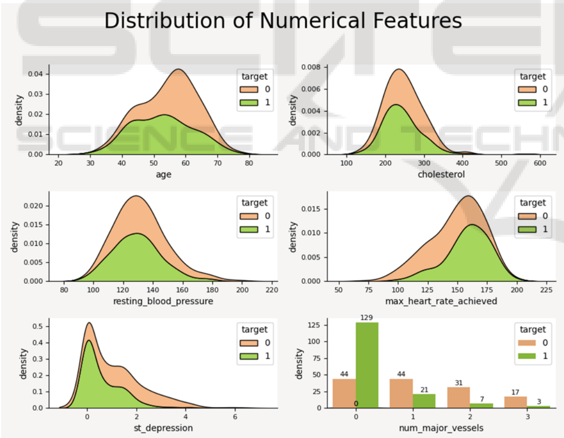

Figure 3: The distribution of numerical features (Photo/Picture credit: Original).

Through this extensive data preparation process, the study ensured that the subsequent modelling and analysis phases could accurately capture the multidimensional characteristics of heart disease risk, laying a solid foundation for in-depth exploration of heart disease prediction.

Feature selection involved identifying the most relevant features to the outcome variable, which in this case was the presence of heart disease. The point-biserial provided insight into the interactions between several attributes and the dependent variable, which can facilitate the investigation of the correlation between pertinent characteristics and the goal variable. Features with high correlation to the target variable were prioritized as they will likely have more predictive power. This heatmap (as shown in Figure 4) highlighted the interdependencies between variables and aided in detecting multicollinearity, where two or more independent variables are highly correlated.

2.2 Machine Learning Models

The quest for the most efficacious classifier within the realm of ML for heart disease prediction is pivotal and complex due to the diverse array of algorithms available. Therefore, a comprehensive evaluation of 12 distinct classifiers from the Scikit-learn library was conducted, broadly categorized into six classes.

2.2.1 Linear Models

Linear models are a class of algorithms that make the assumption that there is a linear relationship between the input and output variables. They are typically easy to implement and interpret and have fast computation times. Among the classifiers evaluated in this study, it includes:

Logistic Regression: A regression model with categorical response variable, commonly used for binary classification tasks.

Linear Discriminant Analysis (LDA): LDA discovers the linear feature combination that most effectively divides a class into two or more.

Quadratic Discriminant Analysis (QDA): QDA enables quadratic decision-making surfaces and accommodates a broader range of data structures.

2.2.2 Tree-Based Models

Tree-based models utilize decision trees as their fundamental building block and are well-suited for capturing complex relationships in data. Among the classifiers evaluated in this study, it includes:

Decision Tree: A non-parametric supervised learning technique for regression and classification applications.

Random Forest: An ensemble of decision trees, typically trained with the "bagging" method to improve the predictive accuracy and control overfitting.

AdaBoost: Short for Adaptive Boosting, it merges several weak classifiers iteratively into a single robust classifier.

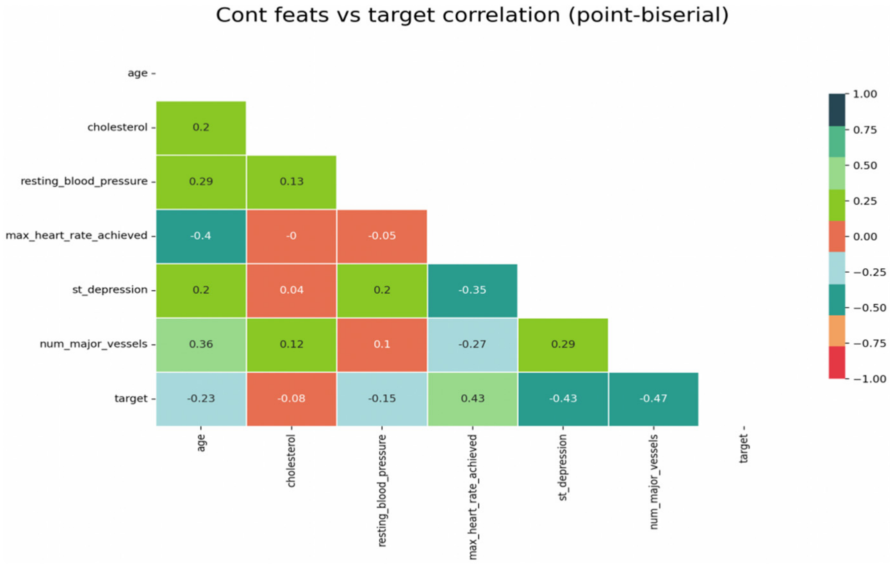

Figure 4: The point-biserial correlation between different features (Photo/Picture credit : Original).

Gradient Boosting: AdaBoost is a sequential ensemble technique that constructs decision trees incrementally, with each new tree aimed at rectifying mistakes caused by earlier trained trees.

2.2.3 Neural Networks

Neural networks were also part of the classifier arsenal explored in this study, and they have the profound capacity to model intricate relationships within large datasets demonstrated in many studies (Li, 2024; Liu, 2023; Qiu, 2022). Despite their opacity, neural networks are robust for capturing the intricate and often non-linear interplay of biomedical signals pertinent to heart disease prognosis.

2.2.4 Bayesian Methods

Bayesian models apply Bayes' theorem for classification, providing a probabilistic approach that can be highly effective. Naive Bayes used in this exploration is a simple yet surprisingly powerful algorithm that assumes independence between the predictors and is particularly useful for large datasets.

2.2.5 SVM

SVMs are types of supervised learning algorithms used for both classification and regression tasks by finding the hyperplane that best divides a dataset into classes. The Nu Support Vector Classifier (Nu SVC) is a variant of SVM that allows for increased versatility in the selection of punishments and the number of support vectors. The Support Vector Classifier (Support Vectors) is another term for the SVM algorithm when it is used explicitly for classification.

2.2.6 KNN

Instance-based learning refers to a type of learning algorithm that involves comparing new problem instances with previously encountered ones from training. These instances are kept in memory for this comparison. The k-NN algorithm retains all existing cases and categories of new cases by using a measure of similarity.

2.3 Implementation Details

2.3.1 Classifier Evaluation

After introducing the various ML classifiers for cardiovascular disease prognosis, classifiers were evaluated based on a suite of performance metrics, such as accuracy, receiver operating characteristic area under the curve (ROC_AUC), recall, precision, and the F1 score. Without utilizing accuracy as the primary evaluation standard, the standard needs to weigh the importance of false positives and negatives equally. Because this research is dealing with medical cases, it needs to consider the fact that when people consult with a hospital and do heart disease testing, the hospital needs to choose a model that reduces false positives but does not miss too much to protect both reputation and the health of their clinic. Hence, the evaluation anchored on the F1 score, serving as the key statistic, offers a valuable viewpoint on the performance of the model by calculating the harmonic mean of precision and recall.

$$\text{F1 score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (1)$$

3 RESULTS AND DISCUSSION

This study has comprehensively evaluated 12 different classifiers to predict heart disease. From Figure 5, it can be observed that Logistic Regression emerged as the top-performing classifier with the F1 score. Linear Discriminant Analysis and Quadratic Discriminant Analysis closely followed the performance. The classifiers with the most minor performance were Support Vectors and Nearest Neighbors, demonstrating considerably lower metrics in all categories.

	Classifier	Accuracy	ROC_AUC	Recall	Precision	F1
0	Logistic Regression	86.490000	0.920000	0.910000	0.820000	0.860000
9	Linear DA	85.140000	0.920000	0.890000	0.820000	0.850000
10	Quadratic DA	85.140000	0.900000	0.830000	0.850000	0.840000
5	Random Forest	83.780000	0.920000	0.830000	0.830000	0.830000
4	Decision Tree	82.430000	0.820000	0.830000	0.810000	0.820000
6	AdaBoost	82.430000	0.860000	0.910000	0.760000	0.830000
7	Gradient Boosting	82.430000	0.900000	0.890000	0.780000	0.830000
8	Naive Bayes	82.430000	0.920000	0.860000	0.790000	0.820000
3	Nu SVC	81.080000	0.910000	0.910000	0.740000	0.820000
11	Neural Net	78.380000	0.880000	0.940000	0.700000	0.800000
2	Support Vectors	64.860000	0.800000	0.890000	0.580000	0.700000
1	Nearest Neighbors	55.410000	0.600000	0.310000	0.550000	0.400000

Figure 5: The result of evaluation of 12 different classifiers (Photo/Picture credit: Original).

The Logistic Regression classifier demonstrates robust performance across various metrics, boasting an accuracy of 86.49%, a ROC_AUC of 0.92 indicating excellent discriminative ability, alongside a high recall of 0.91 and precision of 0.82, The highest F1 score signifies an equilibrated model in terms of both sensitivity and positive predictive value. On the other hand, Support Vectors and Nearest Neighbors lagged significantly behind. The ROC_AUC scores followed a similar pattern, with the top classifiers hovering around the 0.90 mark, signifying excellent discriminative ability between the positive and negative classes.

Logistic Regression performed best regarding the F1 score, which suggests it effectively manages the trade-off between Precision and Recall. Logistic Regression works well with linearly separable data, and the perched ROC_AUC value signifies the model's proficient ability to properly differentiate among the classes. Its success here suggests that the relationship between the variables and the outcome might be approximately linear or that the most significant features for the prediction are linear.

LDA and QDA performed close behind, which is common in cases where class distributions are assumed to be Gaussian. The difference between LDA and QDA performance might be due to LDA assuming the same covariance matrix for each class. QDA does not, which allows for capturing more complex relationships.

Ensemble Methods (such as Random Forest, AdaBoost, and Gradient Boosting): These methods build upon decision trees and aggregate their results to improve performance. Random Forest, for example, alleviates overfitting by computing the average of numerous deep trees that are trained on distinct segments of the dataset. AdaBoost focuses on instances that are harder to classify, and Gradient Boosting builds trees sequentially to correct previous errors. Their strong ROC_AUC scores indicate good class separation. However, slightly lower F1 scores suggest a more complex relationship in the data that these models might be overfitting or not capturing entirely.

Naive Bayes: Given its strong ROC_AUC score, it performs well at differentiating the classes. However, Naive Bayes assumes feature independence, and if this assumption does not hold (which is common in real-world data), it can affect the Precision and Recall.

Neural Networks: While Neural Networks have the potential to model complex relationships, they might require larger datasets to generalize well and might also overfit if the network is too complex or not regularized properly. The high Recall but lower Precision and F1 Score might indicate that while the network is good at identifying positive cases, it could be more effective at precision, potentially predicting too many false positives.

SVC and KNN: The poor performance of these models could be due to several reasons. SVMs might suffer if the correct kernel is not chosen, or the features are not scaled correctly. KNN might not perform well if the data has many features (high dimensionality), which can lead to the 'curse of dimensionality' or if the features have different scales. The low Recall for KNN suggests it struggles to identify actual positive cases, which could be due to an inappropriate choice of 'k', the impact of the curse of dimensionality.

4 CONCLUSIONS

This study aimed at strengthening heart disease prediction by applying ML. It concluded with an extensive assessment of twelve classifiers. Building upon the earlier identification of the intricate relationship between psychological states and cardiovascular health, this study expanded traditional datasets to include mental health indicators, creating a more holistic model of the complex interplay between mind and body. The Logistic Regression classifier yielded the most promising results, which achieved high accuracy and skillfully balanced precision and recall, vital for clinical applicability. The insights gained reinforce the necessity for classifiers that can navigate the delicate intricacies of medical diagnostics. In the future, this study intends to refine the selection of features further, delve deeper into the models' interpretability, and broaden the scope to encapsulate an enormous array of health predictors, continuing the commitment to advance predictive analytics in public health.

REFERENCES

Assegie, T. A., Kumar, R. P., Kumar, N. K., & Vigneswari, D. 2022. An empirical study on machine learning algorithms for heart disease prediction. IAES International Journal of Artificial Intelligence (IJ-AI), 11(3), 1066.

Cao, R., Rahmani, A. M., & Lindsay, K. L. 2022. Prenatal stress assessment using heart rate variability and salivary cortisol: a machine learning-based approach. Plos One, 17(9), e0274298.

Castillo-Mayén, R., Luque, B., García, S. R., Cuadrado, E., Gutiérrez-Domingo, T., Arenas, A., … & Tabernero, C. 2021. Positive psychological profiles based on perceived health clustering in patients with cardiovascular disease: a longitudinal study. BMJ Open, 11(5), e050818.

Chicco, D., & Jurman, G. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21(1).

Kaggle. Heart Disease Prediction Ensemble. https://www.kaggle.com/code/imtkaggleteam/heart-disease-prediction-ensemble/input. 2024

Klyachkin, Y. M., Karapetyan, A., Ratajczak, M. Z., & Abdel-Latif, A. 2014. The role of bioactive lipids in stem cell mobilization and homing: novel therapeutics for myocardial ischemia. BioMed Research International, 2014, 1-12.

Latha, C. B. C., & Jeeva, S. C. 2019. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, 100203.

Liu, Y. and Bao, Y., 2023. Intelligent monitoring of spatially-distributed cracks using distributed fiber optic sensors assisted by deep learning. Measurement, 220, p.113418.

Li, S., Kou, P., Ma, M., Yang, H., Huang, S., & Yang, Z. 2024. Application of Semi-supervised Learning in Image Classification: Research on Fusion of Labeled and Unlabeled Data. IEEE Access.

Novidianto, R., Wibowo, H., & Chandranegara, D. R. 2021. Clustermix k-prototypes algorithm to capture variable characteristics of patient mortality with heart failure. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control.

Omasu, F., Kawano, A., Nagayasu, M., & Nishi, A. 2022. Research on lifestyle habits caused by stress. Open Journal of Preventive Medicine, 12(09), 190-198.

Qiu, Y., Wang, J., Jin, Z., Chen, H., Zhang, M., & Guo, L. 2022. Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomedical Signal Processing and Control, 72, 103323.

Slart, R. H. J. A., Williams, M., Juárez-Orozco, L. E., Rischpler, C., Dweck, M. R., Glaudemans, A. W. J. M., … & Saraste, A. 2021. Position paper of the EACVI and EANM on artificial intelligence applications in multimodality cardiovascular imaging using SPECT/CT, PET/CT, and cardiac CT. European Journal of Nuclear Medicine and Molecular Imaging, 48(5), 1399-1413.

Suresh, T., Assegie, T. A., Rajkumar, S., & Kumar, N. K. 2022. A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model. International Journal of Electrical and Computer Engineering (IJECE), 12(2), 1831.

Tsabedze, N., Kinsey, J., Mpanya, D., Mogashoa, V., Klug, E., & Manga, P. 2021. The prevalence of depression, stress and anxiety symptoms in patients with chronic heart failure.

Yazdani, A., Varathan, K. D., Chiam, Y. K., Malik, A. W., & Ahmad, W. A. W. 2021. A novel approach for heart disease prediction using strength scores with significant predictors. BMC Medical Inform

Zhang, L., et al. 2022. Advances in machine learning techniques for pneumonia detection and classification. Journal of Medical Imaging and Health Informatics, 10(5), 1025-1032.