Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Machine learning based clinical decision support for heart disease prediction using structured patient data

Saeed Ur Rashid
Master of Business Administration in Data Analytics, Westcliff University, Irvine, California, USA
Md Ismail Hossain Siddiqui
Master of Science in Engineering/Industrial Management, Westcliff University, Irvine, California, USA.
Farhad Uddin Mahmud
Master of Business Administration in Management Information Systems, International American University Los Angeles, California, USA.
Md. Soebur Rahman
Master of Business Administration in Management Information Systems, International American University Los Angeles, California, USA
Abdul Aziz Kabir
Master of Business Administration in Data Analytics, Westcliff University, Irvine, California, USA
Ramisa Samin Shammah
College of Technology and Engineering, Westcliff University, Irvine,USA
Corresponding Author: Md Ismail Hossain Siddiqui, **E-mail**: ismailhossainsiddiqui.ce@gmail.com

---

| ABSTRACT

Heart disease remains a leading cause of mortality worldwide, necessitating reliable and efficient predictive models for early diagnosis and clinical decision support. This study presents a comprehensive machine learning framework for heart disease prediction using a structured clinical dataset comprising 920 patient records with diverse demographic, physiological, and diagnostic attributes. The dataset includes both numerical and categorical features, requiring careful preprocessing and encoding to ensure compatibility across different model architectures.A range of classification algorithms is systematically evaluated, including Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine, Extreme Gradient Boosting, and Light Gradient Boosting Machine. Model performance is assessed using multiple evaluation metrics, including accuracy, precision, recall, F1-score, and receiver operating characteristic area under the curve, along with five-fold cross-validation to examine stability and generalization behavior.The experimental results demonstrate consistently high predictive performance across most models, with several approaches achieving near-perfect classification metrics and minimal variation across cross-validation folds. In contrast, K-Nearest Neighbors exhibits slightly lower performance, highlighting differences in sensitivity to local data structure. Analysis of feature distributions and pairwise relationships indicates strong separability between classes, particularly driven by clinically relevant variables such as chest pain type, exercise-induced angina, ST depression, and maximum heart rate.Further evaluation using confusion matrices, receiver operating characteristic curves, and precision–recall curves confirms the robustness of the predictive models and their ability to distinguish between diseased and non-diseased cases with high reliability. Despite the strong performance, the study acknowledges potential dataset-specific characteristics that may influence model behavior and emphasizes the importance of external validation for clinical deployment.

KEYWORDS

heart disease prediction, clinical decision support, machine learning, classification models, structured clinical data, cross-validation, ROC AUC, precision recall analysis, medical diagnosis, predictive analytics

ARTICLE INFORMATION

ACCEPTED: 04 February 2024 **PUBLISHED:** 25 February 2024 **DOI:** 10.32996/jcsts.2024.6.1.36

1. Introduction

Heart disease remains one of the leading causes of mortality worldwide, accounting for a substantial proportion of global deaths each year. Early identification of individuals at risk is critical for timely intervention, improved clinical outcomes, and reduced healthcare burden[1,2]. However, accurate diagnosis often requires the integration of multiple clinical, physiological, and diagnostic factors, making the decision-making process complex and potentially prone to human error, especially in high-volume healthcare environments.In recent years, machine learning techniques have emerged as powerful tools for supporting medical diagnosis by identifying hidden patterns within clinical data [3,4]. These approaches enable the analysis of heterogeneous datasets containing demographic attributes, laboratory measurements, and diagnostic indicators, facilitating data-driven prediction of disease presence[5,6]. Compared to traditional statistical methods, machine learning models can capture non-linear relationships and interactions among variables, offering improved flexibility and predictive capability in complex biomedical contexts [7,8].Heart disease prediction has been widely studied using structured datasets that include features such as age, blood pressure, cholesterol levels, chest pain type, electrocardiographic results, and exercise-induced indicators [9,10]. While earlier studies have demonstrated the potential of individual models, such as logistic regression and decision trees, more recent work has explored ensemble and boosting methods to enhance predictive performance [11,12]. Nevertheless, challenges remain in ensuring model stability, generalization, and interpretability, particularly when dealing with mixed data types and varying feature distributions [13,14].This study presents a comprehensive evaluation of multiple machine learning models for heart disease prediction using a structured clinical dataset consisting of 920 patient records [15,16]. The dataset includes a combination of numerical and categorical features that reflect real-world clinical scenarios[17]. A systematic preprocessing pipeline is employed to handle feature encoding and ensure compatibility across models. Several widely used classification algorithms are investigated, including Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine, Extreme Gradient Boosting, and Light Gradient Boosting Machine [18].The primary objective of this study is to examine the predictive capability, stability, and comparative performance of these models using a consistent evaluation framework. Model performance is assessed through multiple metrics, including accuracy, precision, recall, F1-score, and receiver operating characteristic analysis, alongside five-fold cross-validation to evaluate robustness across different data partitions. In addition, exploratory data analysis is conducted to understand feature distributions and their relationship with disease outcomes [19].The results indicate that several models achieve consistently high predictive performance, with minimal variation across validation folds, suggesting strong separability within the feature space. At the same time, differences observed among models provide insight into their sensitivity to data structure and local patterns [20]. While the findings highlight the potential of machine learning for heart disease prediction, careful consideration of dataset characteristics and validation strategies remains essential for reliable clinical application.

2. Methodology

2.1 Dataset Description

The structured clinical data used in this research includes 920 patient records and 16 attributes, created to predict the presence of heart disease. Records reflect one patient and contain a mix of demographic, physiological, and diagnostic characteristics that are usually common in the evaluation of cardiovascular problems.The dataset includes both numerical (age, resting blood pressure, cholesterol level, maximum heart rate, and ST depression) and categorical (sex, chest pain type, electrocardiographic results, exercise-induced angina, slope of ST segment, number of major vessels, and the presence of thalassemia) features. The target variable is the existence of heart disease, with the patients being classified to binary classes of disease and no disease.Before formulating the model, dataset is designed in a manner that it mirrors real world clinical setting with mixed types of data and uneven distributions of features. This variety allows analyzing machine learning models in a realistic setting where both numerical and categorical variables can be involved in prediction.

2.2 Class Distribution Analysis

The overall distribution of the target variable is illustrated in Figure 1. The dataset contains 509 patients diagnosed with heart disease (55.3%) and 411 patients without heart disease (44.7%).This distribution indicates a moderate class imbalance, with a slightly higher proportion of diseased cases. Although the imbalance is not severe, it remains important to evaluate model performance using metrics beyond accuracy, such as precision, recall, and F1-score, to ensure balanced predictive capability across both classes.

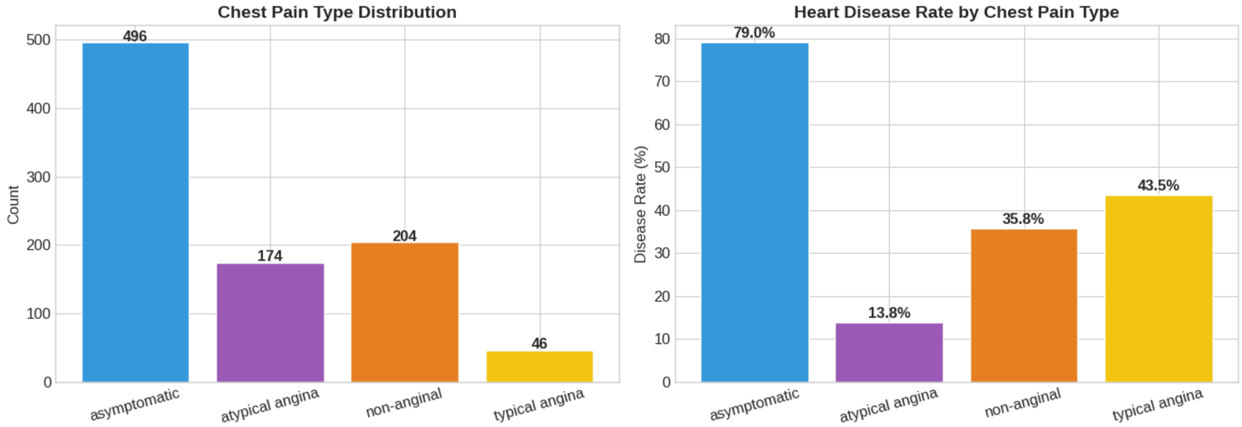

Figure 1. Distribution of chest pain types and corresponding heart disease prevalence

2.3 Gender-Based Analysis

The distribution of patients by gender and the corresponding disease prevalence are presented in Figure 2. The dataset is predominantly composed of female patients (726 samples) compared to male patients (194 samples).However, the disease prevalence differs significantly between genders. As shown in Figure 2, 63.2% of male patients are diagnosed with heart disease, whereas only 25.8% of female patients exhibit the condition. This disparity highlights a gender-specific risk pattern, where male patients demonstrate a higher likelihood of heart disease despite being underrepresented in the dataset.This imbalance in gender distribution and disease prevalence is an important consideration during model evaluation, as it may influence model learning and predictive behavior.

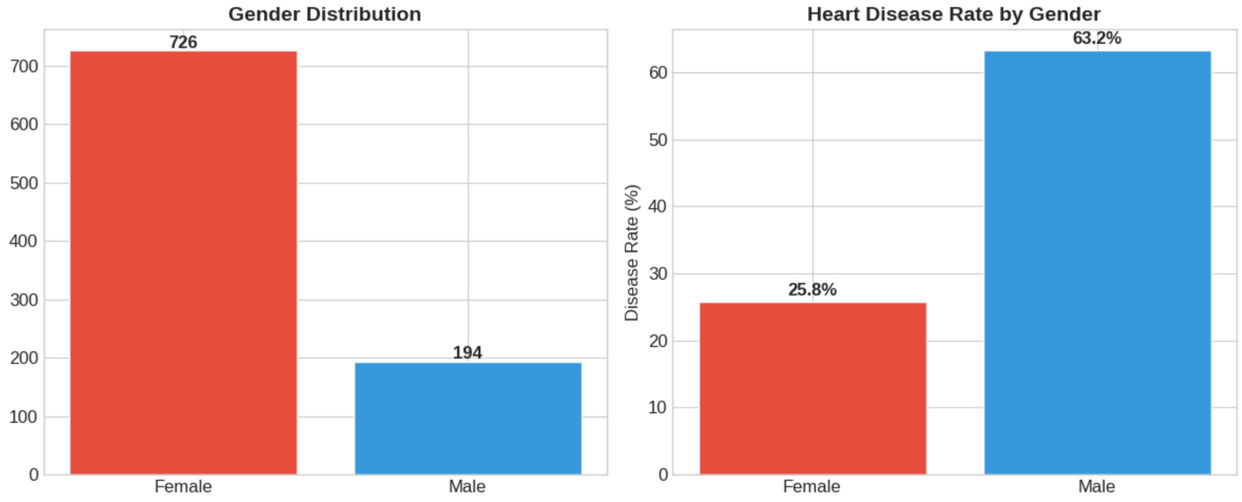

Figure 2. Gender distribution and heart disease prevalence by gender

2.4 Chest Pain Type Analysis

The distribution of chest pain types and their relationship with heart disease is illustrated in Figure 3. Four categories of chest pain are observed: asymptomatic, atypical angina, non-anginal pain, and typical angina.As shown in Figure 3, the asymptomatic category is the most frequent, with 496 instances, followed by non-anginal pain (204), atypical angina (174), and typical angina (46). Importantly, the rate of heart disease varies substantially across these categories.Patients with asymptomatic chest pain exhibit the highest disease prevalence (79.0%), indicating a strong association with heart disease. In contrast, atypical angina shows a significantly lower disease rate (13.8%), while non-anginal pain and typical angina demonstrate moderate disease prevalence of 35.8% and 43.5%, respectively.These findings suggest that chest pain type is a highly informative clinical feature, contributing significantly to the separability of classes and enhancing the predictive capability of machine learning models.

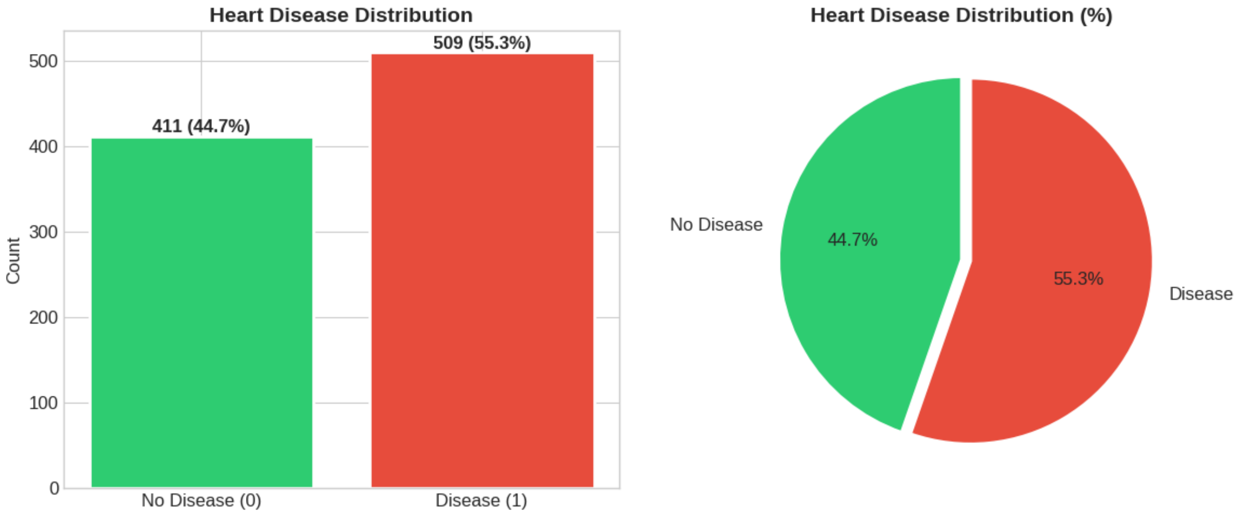

Figure 3. Overall distribution of heart disease cases in the dataset

2.5 Data Preprocessing

A hierarchical preprocessing pipeline is used to guarantee the interoperability of various machine learning algorithms. The dataset has both mixed data, and proper transformation is needed before model training.Encoding techniques are used to convert categorical variables, such as the type of chest pain, electrocardiographic findings, slope, and thalassemia status into numerical scores. Binary characteristics like sex, fasting blood sugar and exercise-induced angina are coded into numerical values. Depending on the model, numerical features are stored in their original scale or normalized as needed.Also, non-informative attributes like unique identifiers are not included in the modeling process, to avoid unintended bias. This data is then trained and evaluated with the same set of features throughout all models.

2.6 Model Development and Evaluation Strategy

A comprehensive set of machine learning models is employed to evaluate predictive performance, including Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine, Extreme Gradient Boosting, and Light Gradient Boosting Machine.Model performance is assessed using multiple evaluation metrics, including accuracy, precision, recall, F1-score, and receiver operating characteristic area under the curve. To further examine model stability and robustness, five-fold cross-validation is applied, ensuring that performance is evaluated across multiple data partitions.

2.7 Analysis of Categorical Features

Figure 4 shows the relationship between categorical variables of clinical variables and heart disease by providing the class-wise distributions of various features such as sex, chest pain type, fasting blood sugar, resting electrocardiographic outcomes, exercise-induced angina, slope of ST segment, number of major vessels, and thalassemia status.Figure 4 indicates that a number of categorical characteristics have high levels of discriminating patterns between diseased and non-diseased patients. Specifically, the prevalence of cases of heart disease is higher among male patients than it is among female patients, which supports the gender-specific risk difference identified above. As a type of chest pain, the asymptomatic type is closely linked to the presence of the disease whereas the atypical angina is more common in non-disease persons.The separation in exercise-induced angina (exang) also appears to be clear with those who are positively-conditioned being mostly related to heart disease and those who are not being more inclined to be disease-free. On the same note, the slope of the ST segment indicates that the flat slope group is greatly prevalent in diseased patients, whereas upsloping is more prevalent in non-diseased.The number of major vessels (ca) is shown to have progressive relationship with high values indicating high disease prevalence. Also, the thalassemia status can serve as a source of much discriminatory information, and reversible defects have strong links to heart disease and normal conditions in cases without disease.Conversely, fasting blood sugar (fbs) demonstrates a comparatively low segregation between the two classes and thus weak predictive value, in comparison to other features.

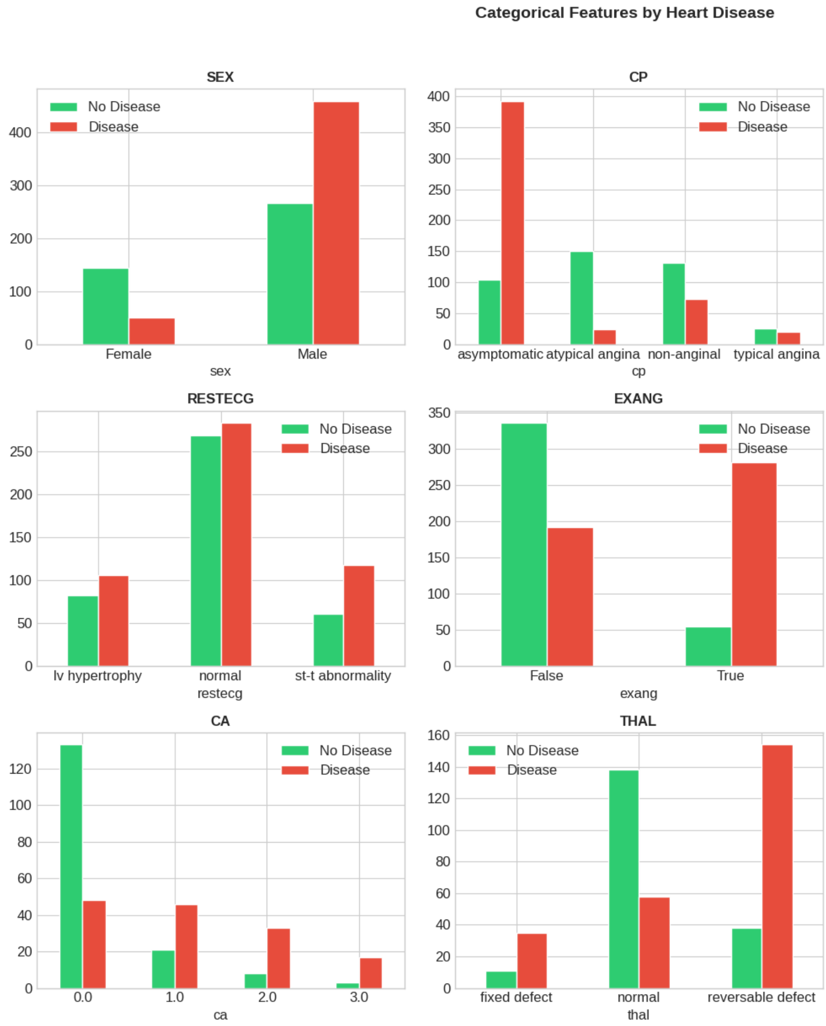

Figure 4. Relationship between categorical clinical features and heart disease status

2.8 Analysis of Numerical Feature Relationships

Figure 5, which contains age, maximum heart rate (thalach), ST depression (oldpeak) and cholesterol level (chol), illustrates the pairwise relationships between important numerical features. The figure is a synthesis of univariate distributions and bivariate scatter plots that allow a thorough evaluation of feature interactions, and class separability.Age has a significant change towards the higher value in patients with heart disease as represented in Figure 5, which means that age is a risk factor. A stronger trend can be found with the maximum heart rate (thalach) in which the non-diseased patients will have higher values and the diseased patients will be found in low values indicating an inverse correlation with presence of diseases.ST depression (oldpeak) has a high level of discriminative ability and diseased patients tend to have higher values whereas the non-disease cases are concentrated around zero. This characteristic gives a distinct distinction between classes and is in line with its clinical applicability in cardiovascular evaluation.Conversely, the cholesterol levels (chol) display great similarity in the two classes where no definite boundary is found in the disease and non-diseased patients. The occurrence of outliers in the values of cholesterol also implies that this feature is not predictive in itself, and its predictive power is diminished.The bivariate plots also show interplay between features especially the negative correlation between age and maximum heart rate and the combined effects of thalach and oldpeak to differentiate the cases of disease. These relations imply that the data has non-linear trends, which are both well represented using a linear and non-linear machine learning model.

Figure 5. Pairwise relationships and distributions of key numerical features for heart disease prediction

3. Results and Discussion

Figure 6 is the performance of the assessed machine learning models and it includes the comparison of accuracy, precision, recall, F1-score, and ROC-AUC of all the classifiers. The findings show that the performance of Logistic Regression, Decision Tree, Random Forest, Support Vector machine, XGBoost and LightGBM has a consistent high performance in all evaluation metrics with values up to 1.0000 when it comes to accuracy, F1-score and ROC-AUC. This consistency implies that the latent space of features can be separated in an easy way between patients with and without heart disease. Conversely, K-Nearest Neighbors model shows a little worse performance with an accuracy of 0.9348, F1-score of 0.9434, and ROC-AUC of 0.9892. This disparity can be explained by the fact that instance-based learning is problem sensitive to the local neighborhood structure and hence, prediction.The models are also analyzed in terms of the stability of the models and this is done by using five fold cross validation which will give us an idea of how the models will generalize in various subsets of the data. The majority of the models have a score of cross-validation of 1.0000 with very low variance, meaning that their performance is very consistent irrespective of data partitioning. Support Vector Machine has a lower mean cross-validation of 0.9973 with very little variation whereas the K-Nearest Neighbors has the mean of 0.9470 with a standard deviation of 0.0179. The increased variability of KNN is because it is sensitive to local data density and sensitive to cross-fold variations. Conversely, the close-to-zero variance of the other models indicates that the decision boundaries learned are stable and they consistently reflect the underlying patterns in the data.Figure 6 shows the receiver operating characteristic curves that further demonstrates the discriminative capability of the models. Most of the models have a curve that is very close to the top-left corner of the plot, which means that the true positive rate is high and the false positive rate is low. In line with this, the area under the curve of these models is 1.000 indicating a good classification performance. The curve of the K-Nearest Neighbors model is slightly lower with AUC of 0.989, which means that it has a minor decrease in the discriminative capability. However, this value still denotes good predictive performance. The ROC analysis validates that most of the models are very effective in separating diseased and non-diseased cases and that the overlap among the predicted probabilities is minimal.Further detail can be gained with the precision recall curves in Figure 6, especially when there is an imbalance in the classes. The curves show that the majority of the models have precision values that are close to 1.0 throughout the entire recall range, meaning that the high sensitivity is attained without causing significant growth in false positive prediction. It is observed that the K-Nearest Neighbors model experiences a minor loss in accuracy at higher levels of recall indicating that it generates a few cases of misclassifications in its effort to get as many positive cases as possible. Nevertheless, its overall performance is competitive. The precisionrecall analysis shows that the majority of the models are capable of reaching a high balance between precision and recall which is crucial in clinical practice where false alarms and missed diagnoses have serious implications.

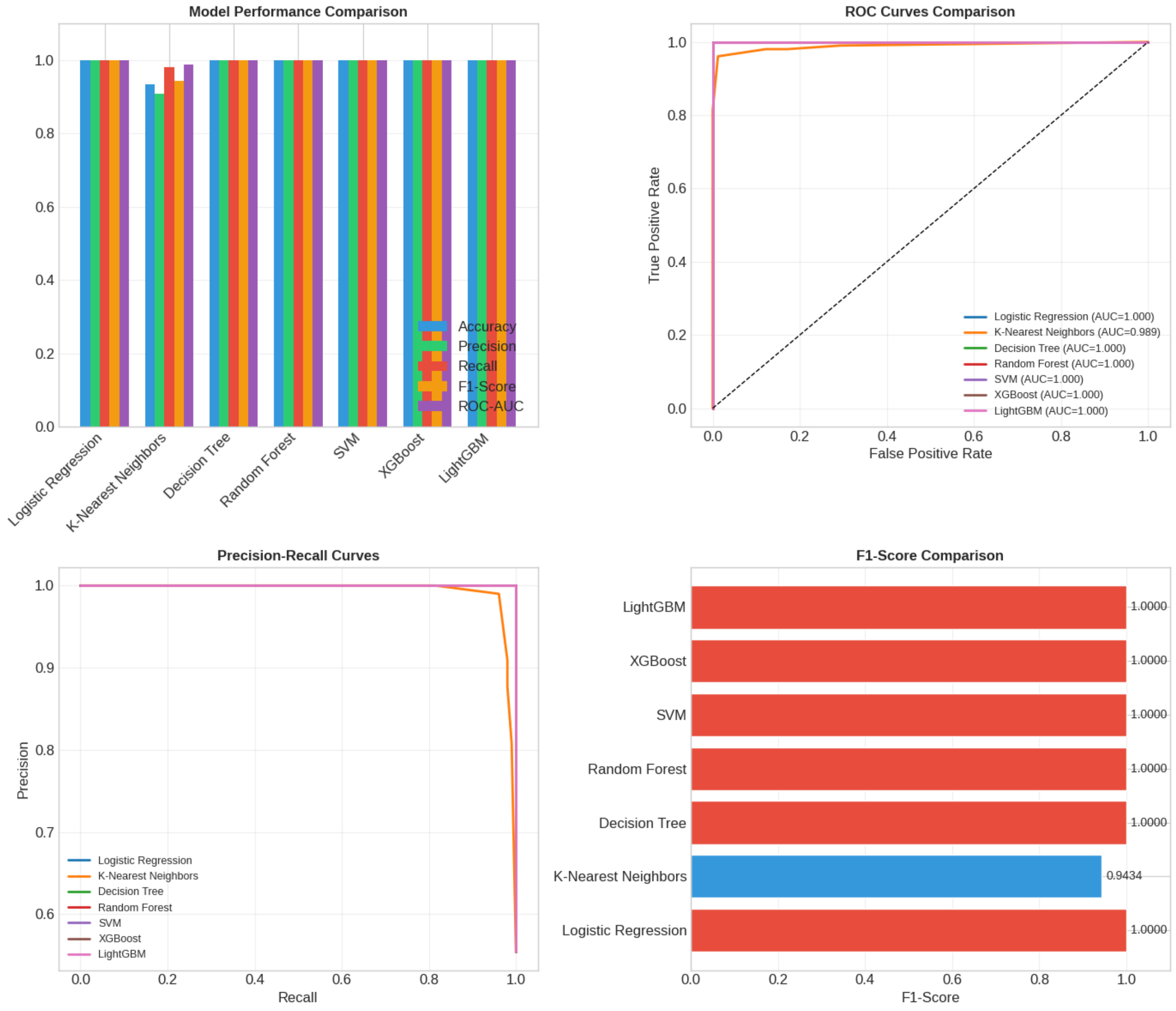

Figure 6. Comparative evaluation of classification performance including ROC curves, precision–recall curves, and F1-scores across machine learning models

These observations are further supported by the comparison of F1-scores of models as indicated in Figure 6. All models except K-Nearest Neighbors have an F1-score of 1.000 which means an optimal precision and recall balance. KNN model has a record F1-score of 0.9434, which is in line with its slightly low performance in other evaluation metrics. The fact that the F1-scores of most models are consistent indicates that the classifiers can achieve high sensitivity and high specificity at the same time, and the fact that the dataset may have clear boundaries to the classes.The confusion matrices in Figure 7 give a close analysis of the results of classification of every model. All of the Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, XGBoost and LightGBM are perfect classifiers, as they perform all the right classifications, including all the examples of both types. In particular, these models have a true negative of 82 and a true positive of 102 with zero false positives and zero false negatives. By contrast, K-Nearest Neighbors model results in low misclassification rates, namely, 10 false positives and 2 false negatives, and accurately classifies 72 false negatives and 100 false positives. These findings validate the fact that KNN is a little bit less effective in capturing the global structure of data, but it is overall good.

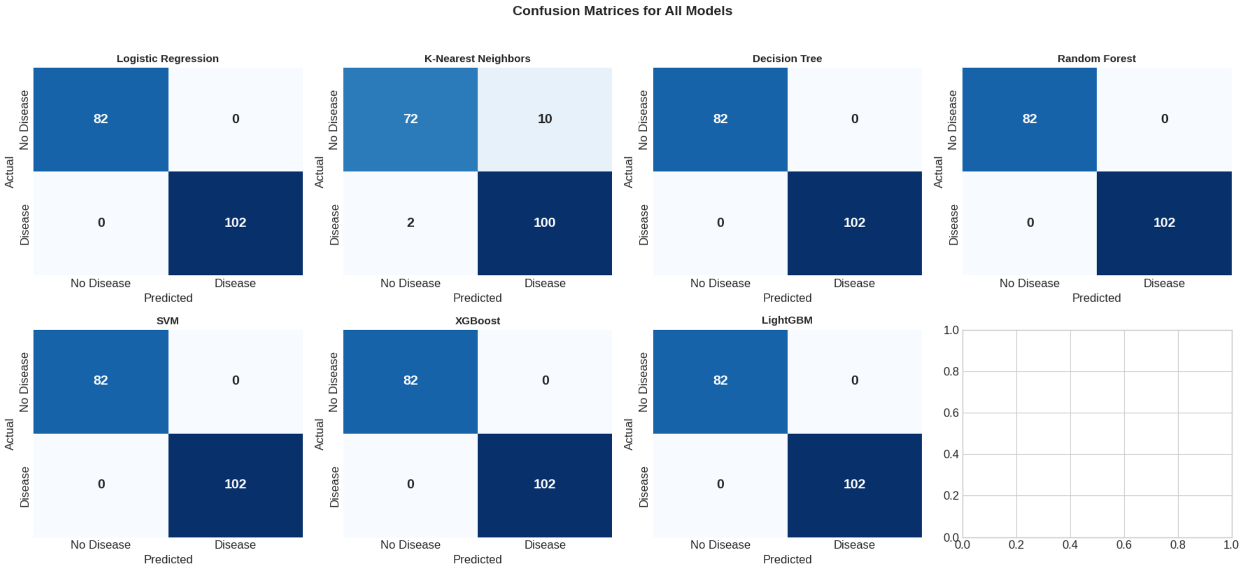

Figure 7. Confusion matrices illustrating classification outcomes for each model

The patterns of the performance observed can be attributed to the nature of the dataset and associations across features. As shown in the preceding exploratory analysis, a number of variables are strongly correlated with heart disease such as type of chest pain, exercise angina, ST depression, maximum heart rate, number of major vessels and thalassemia status. The characteristics help in an easy division of classes, allowing various models to study positive decision limits. Moreover, the similarity of findings among the various model families, such as linear, tree-based, and boosting models, demonstrates that the predictive trends in the data set are strong and not contingent on a particular model.Although the results are of strong predictive performance, one should take them in the context of the characteristics of the dataset. The large consistency between models and folds indicates that the dataset has clear patterns that can be used to classify it. Such performance can also be affected, however, by the particular distribution of features and class relations (in the dataset). So, despite the high performance of the models in this environment, they would need additional assessment on independent data to ensure that they can be used in the actual clinical conditions.

4. Limitations and Future Work

Although the predictive performance has been found to be high in various machine learning models, a number of limitations must be realized to balance and make realistic interpretation of the results. Firstly, the dataset of this study is not large (920 samples), which can limit the number of clinical patterns represented in the training. Despite the stable performance as seen with cross-validation results, the dataset may not be representative of the variability that exists in large populations, such as variations in demographics, comorbidities, and clinical environments.The other significant constraint is associated with the allocation of features in the dataset. Some of the variables like the type of chest pain, the presence of angina as a result of exercise and thalassemia status have strong associations with the target variable and this could be one of the reasons of the high level of class separability. Although this improves predictive performance, it increases the risk that the models are memorizing patterns that are very specific to this data set. Therefore, it is unclear how these models can be generalized to external datasets or real-world clinical settings unless they are further validated.Another weakness is the gender imbalance. The proportion of female patients is much greater than that of male patients in the dataset, whereas the rate of heart disease is much higher in males. This inconsistency can bring bias into the learning of the model and this may compromise the predictability of other demographic groups. Further research should take into account more balanced samples or use methods to reduce demographic bias.

Moreover, the existing research is only on structured tabular data and does not address other clinically relevant modalities, including imaging data, longitudinal patient records, or genetic data. It may be beneficial to consider the incorporation of multimodal data sources into a more detailed picture of patient health and may enhance predictive accuracy in less predictive clinical cases.Methodologically, the models are assessed with internal validation measures, such as five-fold cross-validation, though external validation on independent datasets is not done. This restricts the capability to determine the practicability and strength in various healthcare systems. Future research needs to incorporate external benchmarking to assess the model performance in different conditions.Moreover, interpretability of the models is not clearly discussed in this research. Although some of the features portray high predictive influence, the models themselves are used as a predictive tool without explaining individual predictions. To improve transparency, and enable clinical adoption, it would be desirable to incorporate explainable artificial intelligence methods, including feature importance analysis or model-agnostic interpretability techniques.The effects of advanced preprocessing strategies, feature selection methods and regularization techniques should also be studied in future work to further test the robustness of the models. Also, it might be valuable to test the models on alternative train-test splits and introduce noise or perturbation in the data, to better understand their stability.

5. Conclusion

This paper provides a critical appraisal of several machine learning algorithms in predicting heart disease with structured clinical data. Using a dataset of demographic, physiological, and diagnostic attributes, the proposed framework systematically explores the performance of a wide range of classification algorithms, such as linear, instance-based, tree-based and boosting.The results of the experiment show that a number of models can show high predictive performance when measured by a variety of evaluation measures, such as accuracy, precision, recall, F1-score, and ROC-AUC. The similarity of the findings in all model and validation folds indicates that the dataset has a strong separability between diseased and non-diseased cases. Conversely, the fact that the K-Nearest Neighbors model performance is somewhat lower reflects the effect of local data structure on instance-based methods of learning.The capability of the models in differentiating between high-reliability classes is validated by further analysis using ROC curves, precision-recall curves, and confusion matrices. Exploratory data analysis, which indicates that clinically relevant variables, including the type of chest pain, exercise-induced angina, ST depression, maximum heart rate, number of major vessels, and thalassemia status are important in predictive performance, support these findings.Irrespective of the good performance, one should remember that the model performance can be affected by the properties of the dataset, such as the distribution of features and their interrelations with classes. Consequently, although the models have shown a promising potential here, additional testing on independent data is required to estimate their applicability in practical clinical settings.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1]. Wang, Yajuan, Kenney Ng, Roy J. Byrd, Jianying Hu, Shahram Ebadollahi, Zahra Daar, Christopher Defilippi, Steven R. Steinhubl, and Walter F. Stewart. "Early detection of heart failure with varying prediction windows by structured and unstructured data in electronic health records." In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2530-2533. IEEE, 2015.

[2]. Thomas, J., and R. Theresa Princy. "Human heart disease prediction system using data mining techniques." In *2016 international conference on circuit, power and computing technologies (ICCPCT)*, pp. 1-5. IEEE, 2016.

[3]. Sa, S. "Intelligent heart disease prediction system using data mining techniques." *Int J Healthcare Biomed Res* 1 (2013): 94-101.

[4]. Saqlain, Muhammad, Wahid Hussain, Nazar A. Saqib, and Muazzam A. Khan. "Identification of heart failure by using unstructured data of cardiac patients." In *2016 45th International Conference on Parallel Processing Workshops (ICPPW)*, pp. 426-431. IEEE, 2016.

[5]. Indrakumari, R., T. Poongodi, and Soumya Ranjan Jena. "Heart disease prediction using exploratory data analysis." *Procedia Computer Science* 173 (2020): 130-139.

[6]. Shrestha, Ranjit, and Jyotir Moy Chatterjee. "Heart disease prediction system using machine learning." *LBEF Research Journal of Science, Technology and Management* 1, no. 2 (2019): 115-132.

[7]. Soni, Jyoti, Ujma Ansari, Dipesh Sharma, and Sunita Soni. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." *International Journal of Computer Applications* 17, no. 8 (2011): 43-48.

[8]. Du, Zhenzhen, Yujie Yang, Jing Zheng, Qi Li, Denan Lin, Ye Li, Jianping Fan, Wen Cheng, Xie-Hui Chen, and Yunpeng Cai. "Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data

and machine-learning methods: model development and performance evaluation." *JMIR medical informatics* 8, no. 7 (2020): e17257.

[9]. Fitriyani, Norma Latif, Muhammad Syafrudin, Ganjar Alfian, and Jongtae Rhee. "HDPM: an effective heart disease prediction model for a clinical decision support system." *Ieee Access* 8 (2020): 133034-133050.

[10]. Saxena, Kanak, and Richa Sharma. "Efficient heart disease prediction system." *Procedia Computer Science* 85 (2016): 962-969.

[11]. Palaniappan, S., & Awang, R. (2008, March). Intelligent heart disease prediction system using data mining techniques. In *2008 IEEE/ACS international conference on computer systems and applications* (pp. 108-115). IEEE.

[12]. Sarmah, S. S. (2020). An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network. *Ieee access*, *8*, 135784-135797.

[13]. Methaila, A., Kansal, P., Arya, H., & Kumar, P. (2014). Early heart disease prediction using data mining techniques. *Computer Science & Information Technology Journal*, *24*, 53-59.

[14]. Taneja, Abhishek. "Heart disease prediction system using data mining techniques." *Oriental Journal of Computer science and technology* 6, no. 4 (2013): 457-466.

[15]. Yazdani, Armin, Kasturi Dewi Varathan, Yin Kia Chiam, Asad Waqar Malik, and Wan Azman Wan Ahmad. "A novel approach for heart disease prediction using strength scores with significant predictors." *BMC medical informatics and decision making* 21, no. 1 (2021): 194.

[16]. Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE access* 7 (2019): 81542-81554.

[17]. Kumar, Ankit, Kamred Udham Singh, and Manish Kumar. "A clinical data analysis based diagnostic systems for heart disease prediction using ensemble method." *Big Data Mining and Analytics* 6, no. 4 (2023): 513-525.

[18]. Miotto, Riccardo, Li Li, Brian A. Kidd, and Joel T. Dudley. "Deep patient: an unsupervised representation to predict the future of patients from the electronic health records." *Scientific reports* 6, no. 1 (2016): 26094.

[19]. Jenkins, Kathy J., Kimberlee Gauvreau, Jane W. Newburger, Thomas L. Spray, James H. Moller, and Lisa I. Iezzoni. "Consensus-based method for risk adjustment for surgery for congenital heart disease." *The Journal of thoracic and cardiovascular surgery* 123, no. 1 (2002): 110-118.

[20]. Rasmy, Laila, Yang Xiang, Ziqian Xie, Cui Tao, and Degui Zhi. "Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction." *NPJ digital medicine* 4, no. 1 (2021): 86.