Journal of Transactions in Systems Engineering
https://journals.tultech.eu/index.php/jtse.
ISSN: 2806-2973
Volume 4, Issue 1
DOI: https://doi.org/10.15157/JTSE.2026.4.1.522-537
Received: 18.11.2025; Revised: 19.12.2025; Accepted: 29.12.2025

Research Article

Performance Evaluation of Logistic Regression, Random Forest, and SVM Models in Heart Disease Prediction

Dhoha Raad Hussein[1], Ali Subhi Alhumaima[2], Hussein Alkattan[3,4]*, Mostafa Abotaleb[5]*

[1] College of Engineering, University of Diyala, Diyala, Iraq
[2] Electronic Computer Centre, University of Diyala, Diyala, Iraq
[3] Department of System Programming, South Ural State University, Chelyabinsk, Russia
[4] Directorate of Environment in Najaf, Ministry of Environment, Najaf, Iraq
[5] Engineering School of Digital Technologies, Yugra State University, Khanty Mansiysk, Russia
***alkattan.hussein92@gmail.com; abotalebmostafa@bk.ru**

Abstract

Early identification of high-risk patients for cardiovascular disease is critical for reducing morbidity and improving treatment outcomes. This study applies supervised machine learning techniques to predict heart disease using the publicly available Kaggle heart failure dataset, which comprises 918 observations with demographic, clinical, and laboratory attributes, including age, resting blood pressure, cholesterol level, fasting blood sugar, maximum heart rate achieved, ST depression induced by exercise (Oldpeak), and electrocardiographic and chest pain characteristics. The dataset was pre-processed using a unified pipeline that standardized numerical features and encoded categorical variables via one-hot encoding. The data were split into training and testing sets using an 80/20 stratified approach. Three classification algorithms like Logistic Regression, Random Forest, and Support Vector Machine (SVM) with a radial basis function kernel were evaluated using accuracy, precision, recall, F1-score, and ROC–AUC metrics, complemented by confusion matrices and ROC curves. All models demonstrated strong predictive performance, achieving test accuracies of approximately 0.88. The SVM model exhibited the highest discriminative capability, with a ROC–AUC of approximately 0.95, while Logistic Regression achieved the highest recall (≈ 0.93), making it particularly suitable for applications where minimizing false negatives is critical. Correlation analysis identified Oldpeak, maximum heart rate, age, and fasting blood sugar as key factors associated with heart disease. These findings suggest that relatively simple machine learning models, when combined with appropriate preprocessing, can serve as effective decision-support tools for heart disease risk stratification in clinical settings.

Keywords: Heart Disease Prediction; Machine Learning; Support Vector Machine; Logistic Regression; Random Forest; Medical Decision Support.

INTRODUCTION

Cardiovascular diseases (CVDs) stay one of the top causes of morbidity and mortality globally, being also a significant clinical and economic burden to healthcare systems. Overview of publications on epidemiology from cardiovascular affairs bodies show that

heart disease and stroke are leading causes of many premature deaths and disability-adjusted life years, which calls for early risk stratification and preventive interventions [1]. In addition to classical risk factors, for example hypertension, dyslipidaemia and diabetes, however, recent studies affirm the interaction of lifestyle behaviours, diet and global health, whereby modifiable behaviour can significantly influence the natural history of chronic neurological and cardiometabolic disease [2]. Simultaneously, infectious and latent viral disorders — e.g., cytomegalovirus have been implicated in long-term effects on population health, complicating risk-based management in heterogeneous patient groups [3].

Data-driven approaches have emerged as a key solution to convert heterogeneous clinical information from patient states into predictive actionable knowledge within this context. Data-driven machine learning (ML) methods have become commonly used in healthcare for outcome prediction tasks such as prognosis, diagnosis and treatment response modeling. Deep neural architectures comprising both convolutional and recurrent building blocks, paired with self-attention [4], have demonstrated success in capturing complex temporal and multivariate signal structures from electronic health records predictive of disease. Traditional supervised learning techniques have also been successfully utilized, e.g., for breast cancer survivability studies comparing multiple algorithms across structured clinical datasets [5] or the diabetes prediction framework which uses both enhanced decision-tree–based classifiers as well as deep learning approaches [6, 7]. These examples demonstrate how even small and somewhat sophisticated medical datasets, given ample engineering and careful model selection and validation procedures, can be improved significantly.

ML in particular is spreading quickly among multiple domains, not only in health and biomedicine. ML has played a pivotal role in the physical sciences, specifically in the area of materials discovery where efforts utilize simulations on the large scale and assemble experimental databases to speed up the design of solid-state materials with targeted properties [8]. One domain in which semi-supervised classifiers incorporating spatial and spectral constraints have been used to improve land-cover mapping accuracy in situations where labelled data are scarce, is in remote sensing [9]. The majority of context-aware pipelines in agricultural engineering integration of textural feature extraction and soft-computing methods have they could automatically grade agricultural products including mango fruits [10]. In energy engineering, on the other hand, gradient boosting machines have also been applied to commercial building energy use prediction, consistently showing superior performance on complex nonlinear load shapes [11]. This broad utility is also seen in the economic and financial applications: for instance, genetic algorithms can be applied to optimize hybrid regression models for the inflation-rate prediction task, and for macroeconomic scenarios, such models are can outperform traditional models regarding predictive accuracy [12].

Education and learning analytics is another lively area where ML and data mining approaches are used to analyse and predict student performance. Much research has

suggested models to predict students at risk of failure early in the semester from demographic, behavioural and performance data [13–19]. Within this literature, there has also been a shift towards examining how learners engage with educational materials, including the use of learner-generated questions as a resource in teaching and learning, in science education [20-26]. These examples from disparate domains should illustrate the generality of the supervised learning paradigm and the expectation of predictive models to be accurate, robust, interpretable and well-calibrated to the future production seen in real-world settings.

In clinical neuroscience, ML is used to combine imaging and clinical measurements to predict treatment outcomes. Multi-modal models have been created--for example, to predict the therapeutic benefit of pharmacological drugs on response inhibition in Parkinson's disease utilizing neuroimaging and clinical features [17]. This type of work highlights the ability of ML to aid in personalized clinical decision making. Yet the generalisability of clinical prediction models beyond the data on which they have been developed is crucial to their clinical applicability. Recent methodological overviews about ML-based clinical prediction modelling highlight that overfitting, distribution shift and model variance must be managed appropriately with validation strategies, regularization and control of model complexity [18].

Cross-validation became one of the most important techniques for prediction error estimation and guiding model selection as it was introduced in foundational statistical research [19]. Later work has developed the extension of these ideas into more difficult cases, in which out-of-time cross-validation strategies are created to deal with the effect of dataset shift when temporal or longitudinal classification tasks have to be done [3, 20]. Stratified K-fold cross-validation and class-balancing strategies have been shown to produce more reliable and fair performance estimates on classification tasks of class-imbalance, especially in high-capacity ensemble models [21]. When using cross-validation aggregation to combine autoregressive neural network forecasts, more stable and accurate forecasts can be obtained when applying it for time-series forecasting [22]. In addition to cross-validation, resampling and bootstrap methods offer universal techniques to evaluate variable relevance and uncertainty about empirical models [23], and recent work has even probed the statistical behaviour of validation performance itself, treating it as a random variable and analysing its extremes [24].

Not coincidentally, these advances are also connected to contemporary views on the bias–variance trade-off in machine learning. While classical statistical intuition would argue that better models always overfit at some point, modern analyses of high-dimensional regimes have shown more subtle behaviours — such as the double-descent phenomenon that provide resonance between practical success in deep-learning and theoretical understanding [25]. These insights are all the more relevant in clinical prediction where the sacrifice between model complexity and interpretability and generalizability is keenly felt.

In this methodological context, predicting heart-disease risk using structured clinical data presents a unique and appealing case study. CVDs represent the cardinal cause of morbidity and mortality maintainable by epidemiological evidence worldwide [2], and this, in turn, sparks the need for computational tools that assist clinicians at the front-line in identifying high-risk patients based on routinely collected variables [1]. Concurrently, an increasing amount of ML work has emerged targeting disease prediction including for cancers, metabolic disorders, and other chronic diseases [4–7,18]. Based on these developments, the current work explores supervised learning models to predict heart-disease using a publicly available heart-failure dataset consisting of demographic, physiological and laboratory features. We also apply a unified preprocessing pipeline that standardizes numerical features and encodes categorical variables for common algorithms logistics regression, random forest and support vector machine.

The aim is twofold. We assess these models by comparing them on a stratified training–testing split and diverse performance metrics (accuracy, precision, recall, F1-score and area under receiver-operating characteristic curve with confusion matrices and ROC curves giving more insight into the specific types of errors). The second, building upon the literature on cross-validation and generalization [19–23], provides discussion of how methodological choices such as stratification, resampling and class-balance impact performance estimates in medical classification. We hope to contribute a clear and reproducible baseline analysis to allow easy extension of more sophisticated architectures and larger multi-centre datasets in subsequent work by placing our heart-disease prediction analysis in the context of machine learning applications and validation theory [8–12,18–25].

The main objective of this study is to evaluate and compare of three traditional Machine Learning models Logistic Regression, Random Forest, and Support Vector Machine (SVM) with RBF kernel for heart disease prediction tasks on the Kaggle Heart Failure Prediction data. Specifically, the study aims to:

- Build a consistent and transparent preprocessing pipeline that scales numerical features and one-hot encodes categorical ones for all models.
- Evaluate each model under stratified validation (i.e., in order to maintain the true and false positive rate, train our models on one subset of data and validate on another) with a rich set of performance metrics (accuracy, precision-recall, ROC–AUC, ROC curves and confusion matrix) that capture clinically relevant trade-offs between false positives and negatives.
- Determine the most informative clinical factors associated with heart disease and discuss model performance using known risk factors for cardiac diseases.
- Develop a ready-to-use benchmark for heart-diseases prediction & decision-support systems, such as the one being used in clinical and real-world scenarios.

DATA AND METHODOLOGY

Dataset

The data used in this study is a dataset named Heart Failure Prediction24 uploaded on the Kaggle platform by Soriano [27]. The dataset is composed of concatenation of data collected as part of clinical routine, from patients referred for the suspicion of cardiovascular disease (CVD) and it was created for supervised learning on heart-disease prediction. It has 918 observations on each row representing the data of an individual patient and columns indicate whether this patient has a "positive" or a "negative" heart disease. It is commonly used as a benchmark for classification models in medical decision support because of its clean and moderate size, and the relevance to clinical practice [27-31].

The predictor variables are a mix of continuous and categorical variables indicating demographic, physical measurements and results of diagnostic tests. The quantitative features are age (age, years), resting blood pressure (RestingBP, mmHg), serum cholesterol (Cholesterol, mg/dL), fasting blood sugar (FastingBS), basal exercise heart rate (YzFreshr), maximum heart rate achieved during peak exercise (MaxHR,bpm) and ST segment depression induced by exercise relative to rest (Oldpeak). These factors reflect important components of cardiovascular health and functional capacity that are established correlates for risk of both coronary heart disease and failure [27]. Furthermore, the dataset consists of some categorical variables such as: sex (Sex), chest pain type (ChestPainType), resting electrical cardiogram (RestingECG), exercise induced angina (ExerciseAngina) and slope generated by peak exercise ST segment (ST_Slope). All these variables explain symptom patterns and non-invasive diagnostic findings that are usually accessible in clinical practice of cardiology [27].

A preliminary overview of the Kaggle dataset reveals that while the target variable remains fairly imbalanced with more patient labelled as having heart disease compared to those who do not. Explanatory analysis of the descriptive statistics and displayed distribution reveals distinct differences between both outcome groups, especially for MaxHR, Oldpeak and age which make them promising candidate for classification. Correlation analysis again suggests that Oldpeak and MaxHR are highly correlated with the target label, while age and fasting blood sugar are moderately separated, which is in line with clinical knowledge for patients at higher cardiovascular risk [27].

For model building, stratified sampling on the target variable was used to split the 918 original samples into a training set (80% of samples) and a test set (20%). The training set were employed for fitting the machine learning models and internal validation, the held-out test sets were only used to appraise final performance. We did not perform more preprocessing in the test. Numerical features were scaled such that they had zero mean and unit variance based on training data and categorical ones were transformed by one-hot encoding with a reference category left out (the most frequent value) to avoid redundancy. This process generates a fully numeric feature matrix suitable for the selected

algorithms and ensures that the analysis is reproducible, consistent with best practice for use of Kaggle heart-failure dataset in supervised learning experiments [27].

The distributions of six major clinical variables: Age, RestingBP, Cholesterol, FastingBS, MaxHR and Oldpeak stratified by HeartDisease status (HeartDisease = 0 vs. HeartDisease = 1) are shown in Figure 1. Each subplot contains histograms and smoothed density curves of those without heart disease (yellow) and with heart disease (blue). The Age panel illustrates that patients diagnosed for heart-disease are older, with their density skewed toward higher ages. RestingBP and Cholesterol are two features of partially overlapping distributions with slightly higher values in the case group, implying a modest risk association. The distribution of fastingBS is strongly skewed; most patients have small values but a visibly higher ith fraction of those with high fasting blood sugar among the heart diseased. This is as well seen in the MaxHR subplot, where a greater number of heart-disease patients fall towards the area of low maximum heartrate (maximum exercise capacity). On the other, the Oldpeak (ST-segment depression) is typically significantly greater in heart disease patients, reflecting more severe ischemic response during exercise testing. All the panels combined, are used to show that distinctions in distribution between the groups two are highest for Age, MaxHR and Oldpeak, thereby establishing their importance as discriminating predictors.

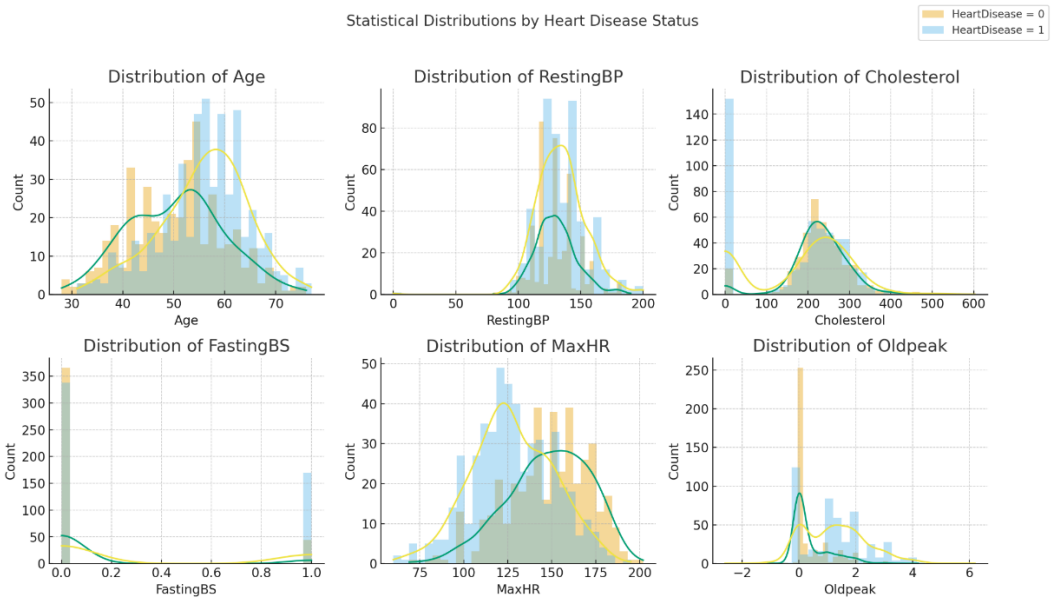

Figure 1. Distribution of Demographics and Clinical Characteristics by HD status.

Figure 2 shows a heatmap of the correlation between Age, RestingBP, Cholesterol, FastingBS, MaxHR and Oldpeak and the binary response variable HeartDisease. The Pearson correlation coefficient is displayed in each cell, and color of the cells indicates both strength and direction of association. Down the final row and column, HeartDisease is positively correlated with Age ($r \approx 0.28$), FastingBS ($r \approx 0.27$) and Oldpeak ($r \approx 0.40$), but negatively correlated with MaxHR ($r \approx -0.40$) b-indicating that older age, higher fasting blood sugar and greater ST-segment depression are associated with increased risk,

whereas higher exercise capacity is protective/ preventive. Cholesterol has only a mild negative association with HeartDisease in this data, and RestingBP is hardly related positively. The correlations among the predictors are mostly moderate, hinting at non-significant multicollinearity and also that the variables carry partially different information to the classifier.

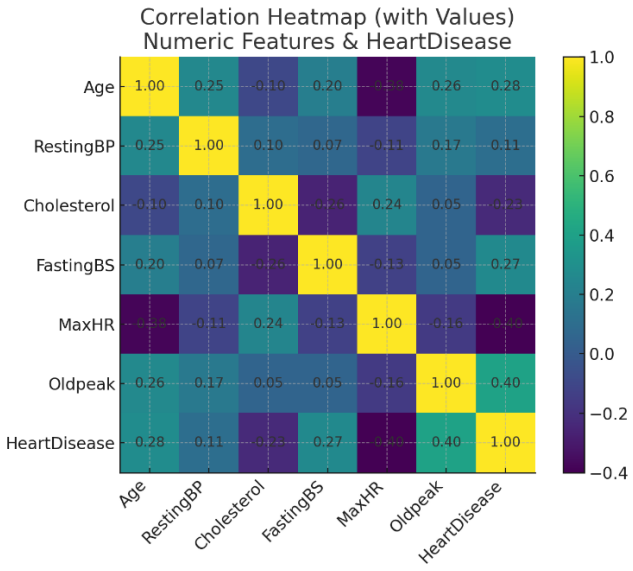

Figure 2. Correlation Heatmap of Numerical Variables with Heart Disease Result.

Logistic Regression

Logistic regression is a GLM that models the probability of being in the heart-disease class as logistic (sigmoid) transformation of a linear combination of predictors. In the current study, all z-score transformed continuous variables (Age, RestingBP, Cholesterol, FastingBS, MaxHR Oldpeak) along with one-hot encoded categorical variables (Sex,ChestPainType, RestingECG ExerciseAngina ST_Slope) are entered linearly into our model. The estimated coefficients provide the logarithm of the odds in favor of each predictive feature; this can be used to gauge how strongly the variable is associated with the outcome. A positive coefficient raises the log-odds of heart disease, a negative lowers it.

Logistic regression models the probability of heart disease directly via equations (1) and (2):

$$P(y = 1 \mid \mathbf{x}) = \frac{1}{1 + \exp\left(-z(\mathbf{x})\right)} \tag{1}$$

Were

$$z(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \tag{2}$$

- β_0 is the intercept.
- β; are the regression coefficients learned from the data.

The predicted class is:

$$\hat{y} = \begin{cases} 1, & \text{if } P(y = 1 \mid \mathbf{x}) \geq \tau \\ 0, & \text{if } P(y = 1 \mid \mathbf{x}) < \tau \end{cases}$$

with decision threshold $\tau = 0.5$ in this work.

Random Forest

Random forest classifier is an ensemble learning approach that combines the predictions from a set of decision trees, each trained with bootstrap samples based on the training data. Each tree performs a splitting of the feature space into decision regions by selecting single variables, in a way that reduces most impurities (e.g., Gini index), recursively. Randomness in the generation of a forest: During training, we use two sources of randomness to create an RF: random sampling with replacement (i.e., bootstrapping or bagging) to generate different training subsets for each tree; and random selection of a subset of features at each split. This juxtaposition yields de-correlation among the trees and hence improves ensemble stability and generalization.

A random forest is an ensemble of T decision trees. Each tree $h_t(\mathbf{x})$ outputs a class label (0 or 1).

The final prediction is obtained by majority voting, see equation (3):

$$\hat{y} = \arg \max_{c \in [(0,1)} \sum_{t=1}^{T} \mathbf{1}(h_t(\mathbf{x}) = c) \tag{3}$$

Were:

- T is the number of trees,
- $\mathbf{1}(\cdot)$ is the indicator function that equals 1 if the condition is true and 0 otherwise.

Probabilities can be estimated as equation (4):

$$\hat{P}(y = c \mid \mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}(h_t(\mathbf{x}) = c), c \in \{0,1\}. \tag{4}$$

Support Vector Machine RBF Kernel

The support vector machine (SVM) is a margin-based classifier in the transformed feature space which estimates decision boundary that maximizes the gap between classes. At simplest form, SVM will define a `hyperplane' with the biggest possible separation between positives and negatives using only some training points (the support vectors) to calculate it. To capture the non-linear relationship between predictors and outcome, the model incorporates a kernel function that non-linearly maps input features to high-dimensional space. In this study, Radial Basis Function (RBF) kernel is adopted to quantify the similarity between examples as a Gaussian function of the squared Euclidean distance of two examples.

For the binary SVM classifier, the decision function can be written as equation (5):

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{5}$$

Were

- $(\mathrm{x_i}, \mathrm{y_i})$ are training samples with labels $\mathrm{y_i} \in \{-1, +1\}$
- (in practice we map heart disease $0 \rightarrow -1$ and $1 \rightarrow +1$).

- $\alpha_i \geq 0$ are learned Lagrange multipliers,
- b is the bias term,
- $K(\because)$ is the kernel function.

For the RBF kernel used in this work:

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2\right) \quad (6)$$

- with $\gamma > 0$ controlling the kernel width.

The predicted class is then:

$$\hat{y} = \begin{cases} 1, & \text{if } f(\mathbf{x}) \geq 0 \\ 0, & \text{if } f(\mathbf{x}) < 0 \end{cases}$$

after mapping back from labels $\{-1, +1\}$ to $\{0,1\}$.

RESULTS

Table 1 depicts a slightly imbalanced distribution of patients, with 55.3% labeled as the disease and 44.7% not having the disease. This mild class imbalance is another good reason to use stratified splitting and careful handling of evaluation criteria (precision, recall, F1) rather than accuracy.

Table 1. Distribution of Heart Disease Outcome in the Kaggle Dataset

Heart Disease	Count	Percentage (%)
0 (No disease)	410	44.7
1 (Disease)	508	55.3

The general distribution of the numerical clinical variables is presented in Table 2. Patients have a mean age of 53.5 years with an average resting blood pressure of approximately 132 mmHg and cholesterol levels around 199 mg/dL. MaxHR ranges from 60 to 202 bpm, indicative of a wide range of exercise capacity, and Oldpeak (ST _ depression) from −2.6 to 6.2, demonstrating substantial diversity in the ischemic response to stress.

Table 2. Descriptive Statistics of Numerical Predictors.

Variable	Mean	Std. Dev.	Min	Max
Age	53.51	9.43	28.0	77.0
RestingBP	132.40	18.51	0.0	200.0
Cholesterol	198.80	109.38	0.0	603.0
FastingBS	0.23	0.42	0.0	1.0
MaxHR	136.81	25.46	60.0	202.0
Oldpeak	0.89	1.07	−2.6	6.2

Table 3 presents pair wise Pearson correlation coefficients for the numerical predictors with the binary outcome. Oldpeak (r = 0.404) and MaxHR (r = −0.400) were the strongest associated variables, validating that larger ST-segment depression and lower maximum heart rate are associated positively with higher risk of having a heart disease. Age and FastingBS have weak or moderate positive associations, whereas RestingBP has only a slight association, and Cholesterol seems to have slightly negative.

Table 3. Correlation between Numerical Features and Heart Disease.

Variable	Correlation with Heart Disease
Age	0.282
RestingBP	0.108
Cholesterol	−0.233
FastingBS	0.267
MaxHR	−0.400
Oldpeak	0.404

Best model test-set performance Comparison of the test-set performances between Logistic Regression, Random Forest and SVM with RBF kernel are presented in Table 4. Models perform similarly with high accuracy (~0.88). Logistic Regression shows the highest recall (0.931), which is desirable in cases where reducing missed heart-disease patients is more important. At the sample level, the SVM model provides the highest ROC–AUC (0.946), achieving thereby best overall discriminative capabilities, whereas Random Forest results in very balanced precision and recall (both 0.892).

Table 4. Performance Metrics of the Three Machine Learning Models (Test Set).

Model	Accuracy	Precision	Recall	F1-score	ROC–AUC
Logistic Regression	0.886	0.872	0.931	0.900	0.931
Random Forest	0.880	0.892	0.892	0.892	0.930
SVM (RBF)	0.886	0.886	0.912	0.899	0.946

The receiver operating characteristic (ROC) curves of the three classifiers—i.e., Logistic Regression, Random Forest and SVM with RBF kernel—are plotted in Figure 3 for comparison on test set. All models are substantially above the diagonal reference line (baseline), as their discriminative power is high. SVM curve is on a straight line closest to the top left which has AUC value in the order of >0.946, next comes LR (AUC ≈ 0.931) and RF (AUC ≈ 0.930) verifying that SVM made best overall segregation between these two groups of patients those with and without heart disease.

The confusion matrix of the Logistic Regression model is represented in Figure 4. The classifier correctly identifies 68 non–heart-disease cases (true negative) and 95 heart-disease cases (true positive), but also makes 14 false positives and only 7 false negatives. This is characteristic of a high recall for detection of heart disease, which is desired in screening tasks were failing to detect positive cases would be especially costly.

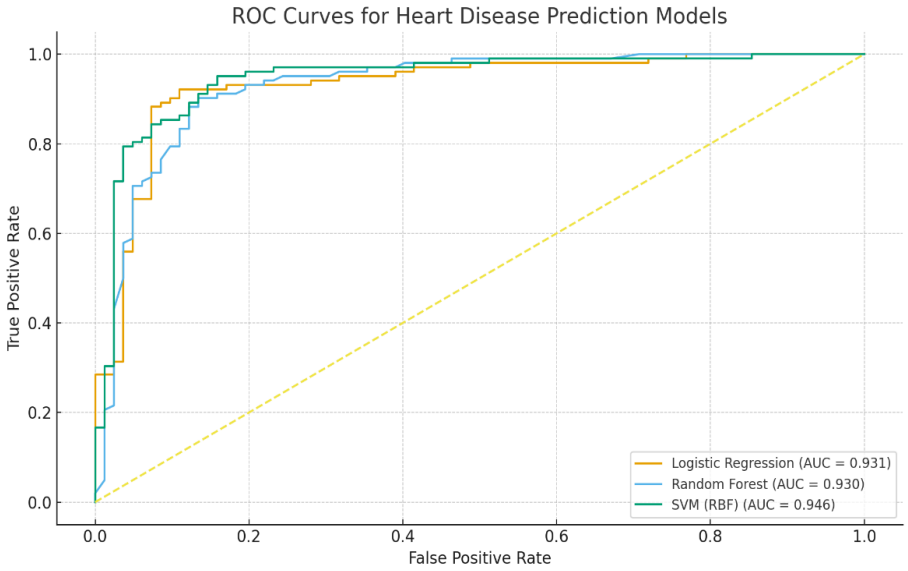

Figure 3. ROC Curves for predicting heart disease Model.

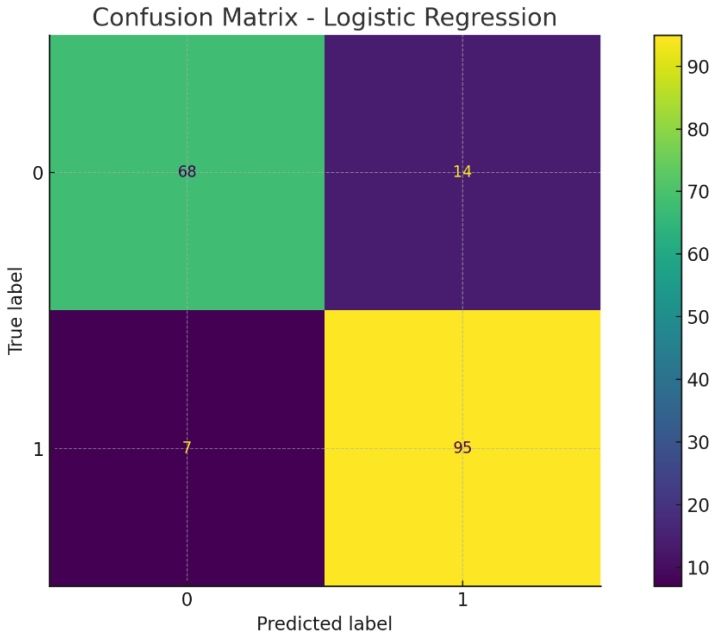

Figure 4. Confusion Matrix for Logistic Regression.

Confusion matrix of the Random Forest classifier is shown in Figure 5. The model accurately predicts 71 non-disease and 91 disease cases, with 11 false positives, and 11 false negative classification. Logistic Regression a slight better balance between false positives and negatives is achieved by Random Forest with similar precision and recall values and robust, symmetric performance in both classes.

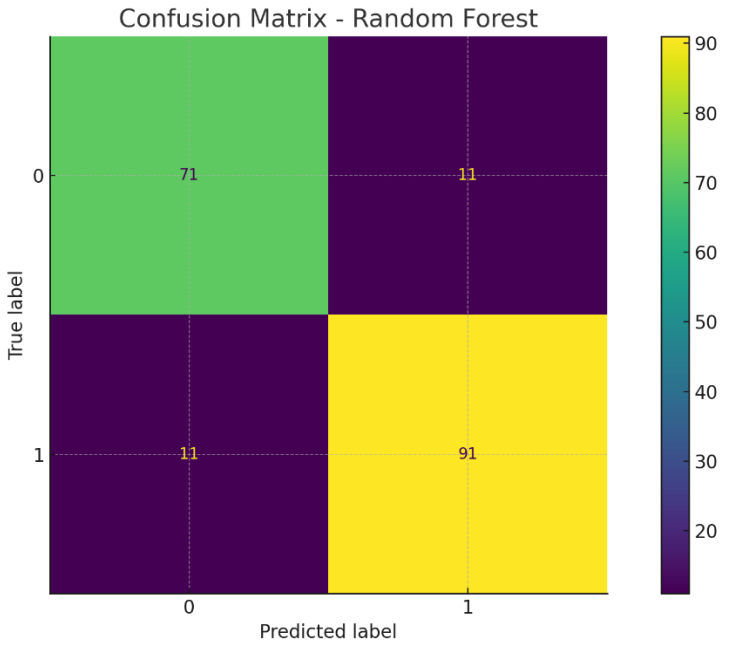

Figure 5. Confusion Matrix for Random Forest.

Confusion matrix of SVM RBF kernel is shown in Figure 6. SVM has correctly predicted 70 non–heart-disease cases and 93 heart-disease cases, having misidentified just 12 non-disease case as diseased and 9 disease cases as none-disease. These results indicate a trade-off between the behaviors of Logistic Regression and Random Forest with sufficiently low error rates in both class and consistent with x axis in Figure 3.

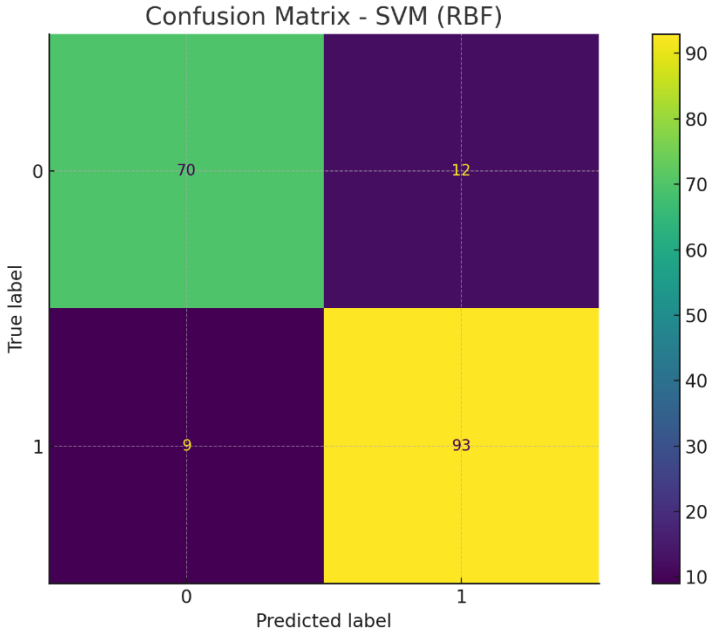

Figure 6. Confusion Matrix for SVM (RBF).

The overall performance for all the three models is summarized with a bar graph under Figure 7. All classifiers perform very close to each other, with accuracies all around 0.88 on test set, which suggests that the Logistic Regression, Random Forest and SVM are able to give good consistent predictions on this dataset. Despite their small difference's bars, an accurate model is not enough to separate accurately all the models and other criteria are needed.

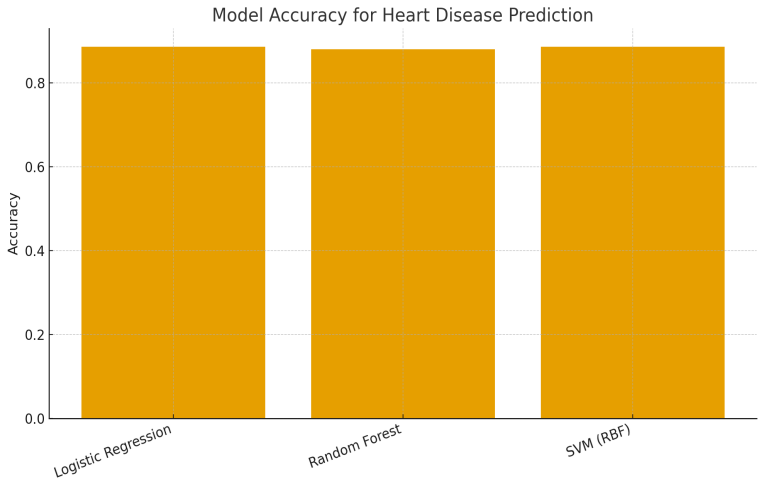

Figure 7. Heart Disease Prediction ML Model Accuracy.

Precision, recall and F1-score for each model is depicted in Fig. 8. Logistic Regression has the highest recall, indicating that it is more effective at identifying heart-disease patients correctly than having false negatives. Random Forest has almost equal precision, recall, and F1 to present balanced performance. The SVM model exhibits both precision and F1 score almost comparable to Logistic Regression, however it has slightly lower recall at the cost of better ranking (measured by ROC–AUC). Taken together, these three bars demonstrate the trade-offs among the three classifiers, and validate that using several measures will be necessary when choosing a model for clinical decision support.

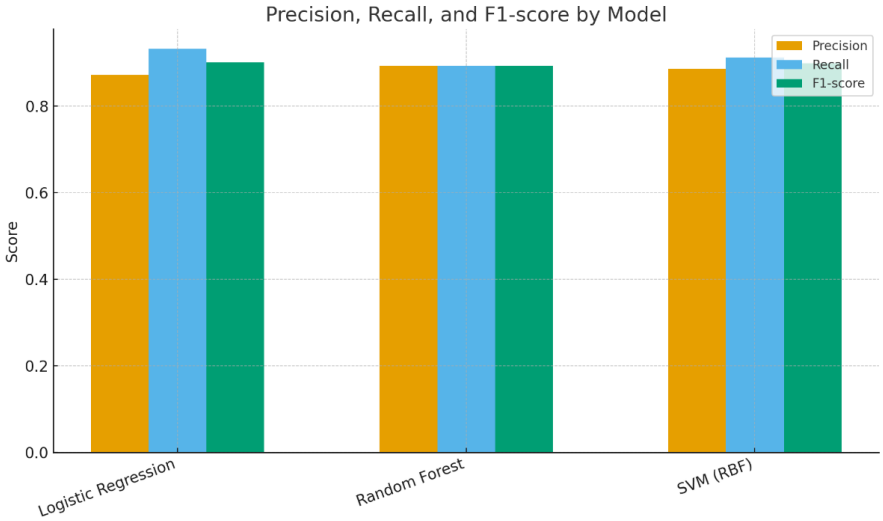

Figure 8. Precision, Recall and F1 by model.

CONCLUSION

This paper is a systematic study considering Logistic Regression, Random Forest, SVM with RBF kernel-the most popular machine learning models in practice today for predicting heart-disease using Kaggle Heart Failure Prediction dataset. In a combined pipeline where both standardized numerical features, one-hot encoded categorical variables, and stratified data splitting are used, they perform comparably well in predicting in the high 88% range of accuracy with strong precision, recall, F1-Score as well as ROC-AUC. Logistic Regression performed best in terms of recall, Random Forest was optimal for balanced precision and recall, while SVM provided the highest overall discriminative power according to ROC–AUC.

These findings suggest that simple, classical algorithms can be used to deliver robust and interpretable decision support when trained and validated properly. The analysis substantiates the clinical plausibility of the learned patterns: Age, MaxHR, Oldpeak and FastingBS were identified important as risk indicators which is in accordance with established medical knowledge about cardiovascular diseases. From a methodological point of view, the study highlighted several good practices for preprocessing and validation, as well as multi-metric evaluation, that can be integrated within a common framework to ensure reproducibility in medical prediction tasks.

Although the experiments were performed on only one public dataset, because there is no strict constraint about the data used with this framework, it can be easily transferred to other heart-disease datasets or even enlarged to multi-centre clinical registries. Further work will validate the models on expanded and diverse populations and evaluate other computational methods and ensemble approaches, as well as make use of explainable AI to produce case-level explanations conducive to clinical process.

CONFLICT OF INTERESTS

The authors confirm that there is no conflict of interest associated with this publication.

REFERENCES

1. Benjamin, E.J., et al. heart disease and stroke statistics—2019 update: A report from the American Heart Association. *Circulation* **2019**, *139*, e56–e528.
2. Arora, S., Santiago, J.A., Bernstein, M., Potashkin, J.A. Diet and lifestyle impact the development and progression of Alzheimer's dementia. *Front. Nutr.* **2023**, *10*, 1213223.
3. Zuhair, M., et al. Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. *Rev. Med. Virol.* **2019**, *29*, e2034.
4. Usama, M., Ahmad, B., Xiao, W., Hossain, M.S., Muhammad, G. Self-attention based recurrent convolutional neural network for disease prediction using healthcare data. Comput. *Methods Programs Biomed*. **2020**, *190*, 105191.
5. Shukla, N., Hagenbuchner, M.; Win, K.T.; Yang, J. Breast cancer data analysis for survivability studies and prediction. Comput. *Methods Programs Biomed.* **2018**, *155*, 199–208.

6. Naz, H., et al. Deep learning approach for diabetes prediction using PIMA Indian dataset. *J. Diabetes Metab. Disord.* **2020**, *19*, 391–403.
7. Kaur, G., Chhabra, A. Improved J48 classification algorithm for the prediction of diabetes. *Int. J. Comput. Appl.* **2014**, *98*, 13–17.
8. Schmidt, J., Marques, M.R., Botti, S., Marques, M.A.L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater. 2019*, *5*, 83.
9. Xiong, B., Jiang, W., Zhang, F. Semi-supervised classification considering space and spectrum constraint for remote sensing imagery. *In Proceedings of the 2010 18th International Conference on Geoinformatics*; IEEE: Beijing, China, **2010**, pp. 1–6.
10. Olaniyi, E.O., Oyedotun, O.K., Ogunlade, C.A., Khashman, A. In-line grading system for mango fruits using GLCM feature extraction and soft-computing techniques. *Int. J. Appl. Pattern Recognit.* **2019**, *6*, 58–75.
11. Touzani, S., Granderson, J., Fernandes, S. Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy Build.* **2018**, *158*, 1533–1543.
12. Dharma, F., et al. Prediction of Indonesian inflation rate using regression model based on genetic algorithms. *J. Online Inform.* **2020**, *5*, 45–52.
13. Khan, A., Ghosh, S.K. Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Educ. Inf. Technol.* **2021**, *26*, 205–240.
14. Nadar, N., Kamatchi, R. A novel student risk identification model using machine learning approach. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 305–309.
15. Yousafzai, B.K., Hayat, M., Afzal, S. Application of machine learning and data mining in predicting the performance of intermediate and secondary education level student. *Educ. Inf. Technol.* **2020**, *25*, 4677–4697.
16. Smirani, L.K., Yamani, H.A., Menzli, L.J., Boulahia, J.A. Using ensemble learning algorithms to predict student failure and enabling customized educational paths. *Sci. Program.* **2022**, *2022*, 3805235.
17. Ye, Z., et al. Predicting beneficial effects of Atomoxetine and Citalopram on response inhibition in Parkinson's disease with clinical and neuroimaging measures. *Hum. Brain Mapp.* **2016**, *37*, 1026–1037.
18. Kernbach, J.M., Staartjes, V.E. Foundations of machine learning-based clinical prediction modeling: Part II—Generalization and overfitting. *In Machine Learning in Clinical Neuroscience; Staartjes, V.E., et al., Eds.; Acta Neurochirurgica Supplement 134*; Springer: Cham, Switzerland, **2022**, pp. 15–21.
19. Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 111–133.
20. Maldonado, S., López, J., Iturriaga, A. Out-of-time cross-validation strategies for classification in the presence of dataset shift. *Appl. Intell.* **2022**, *52*, 5770–5783.
21. Mahesh, T.R., Geman, O., Margala, M., Guduri, M. The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification. *Healthc. Anal.* **2023**, *4*, 100247.
22. Barrow, D.K., Crone, S.F. Cross-validation aggregation for combining autoregressive neural network forecasts. *Int. J. Forecast.* **2016**, *32*, 1120–1137.

23. Gimenez-Nadal, J.I., Lafuente, M., Molina, J.A., Velilla, J. Resampling and bootstrap algorithms to assess the relevance of variables: Applications to cross-section entrepreneurship data. *Empir. Econ.* **2019**, *56*, 233–267.
24. Dodge, J., Gururangan, S., Card, D., Schwartz, R., Smith, N.A. Expected validation performance and estimation of a random variable's maximum. arXiv **2021**, arXiv:2110.00613.
25. Belkin, M., Hsu, D., Ma, S., Mandal, S. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 15849–15854.
26. Chin, C., Osborne, J. Students' questions: A potential resource for teaching and learning science. *Stud. Sci. Educ.* **2008**, *44*, 1–39.
27. Soriano, F. Heart Failure Prediction. Kaggle. Available online: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction (accessed on 15 April 2025).
28. Shehu, B., Shehu, K., Kraja, F., & Kraja, B. Predictors of Chemotherapy-Induced Cardiotoxicity in Breast Cancer: A Real-World Cohort Study. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2025**, *8*(4), 1113–1136.
29. Chakraborty S., Mukherjee T. A Hybrid Early Warning System Based on IoT and Machine Learning for Heart Disease Prediction, *Journal of Integrated Engineering and Applied Sciences*. **2025**, *3*(2), 257-269.
30. Alhasani, A.T. Unsupervised Clustering of Multivariate Sports Activity Data Using K-Means: A Study on the Sport Data Multivariate Time Series Dataset. *Journal of Transactions in Systems Engineering*, **2025**, *3*(2), 367–381.
31. Balaji, T., Krishna, G. V., Kumar, P. R., Sameera, M. S. D., Avani, V. S., & Sowjanya, G. N. Multimodal Deep Learning for Disease Diagnosis and Risk Stratification: Integrating Genomic, Clinical, and Imaging Data. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2025**, *8*(4), 937–969.