Malaysian Journal of Computing, 6 (2): 872 - 884, 2021

eISSN: 2600-8238 online

MACHINE LEARNING TECHNIQUES FOR EARLY HEART FAILURE PREDICTION

Nur Shahellin Mansur Huang[1], Zaidah Ibrahim[2] and Norizan Mat Diah[3*]

[1,2,3]Faculty of Computer and Mathematical Sciences,*
Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia
[1]shahellinnur@yahoo.com, [2]zaidah@fskm.uitm.edu.my, [3*]norizan@fskm.uitm.edu.my

ABSTRACT

This paper discusses the performance of four popular machine learning techniques for predicting heart failure using a publicly available dataset from kaggle.com, which are Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR). They were selected due to their good performance in medical-related applications. Heart failure is a common public health problem, and there is a need to improve the management of heart failure cases to increase the survival rate. The vast amount of medical data related to heart failure and the availability of powerful computing devices allow researchers to conduct more experiments. The performance of the machine learning techniques was measured by accuracy, precision, recall, f1-score, sensitivity, and specificity in predicting heart failure with 13 symptoms or features. Experimental analysis showed that RF produces the highest performance score, which is 0.88 compared to SVM, NB, and LR. Further experiments with RF were also conducted to determine the important features in predicting heart failure, and the results indicated that all 13 symptoms or features are important.

Keywords*: Heart Failure Prediction, Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine.*

Received for review: 04-06-2021; Accepted: 01-07-2021; Published: 10-08-2021

1. Introduction

Electronic health records (EHRs) are common clinical data sources for medical prediction problems (Goldstein *et al*., 2017). It enables individualised prognostic evaluations for each patient and guidance for appropriate therapy. Experimental analysis between conventional statistical and machine learning models for prognosis indicates that machine learning models outperform conventional statistical models (Steele *et al.*, 2018). Conventional statistical models are usually based on testing hypotheses requiring significant human intervention to select the prognosis and variables. On the other hand, machine learning models focus on the predictive performance and generalisation of models with repetitive processes to improve the algorithm (Handelman *et al.*, 2018).

One of the critical medical predictions is heart failure (Tripoliti *et al.*, 2017). Heart failure is not a disease but a complex clinical syndrome avoiding the heart from performing the circulatory demands of the body (Tripoliti *et al.*, 2017). According to Khan *et al.* (2017), many factors lead to heart failures, such as diabetes, high cholesterol, overweight, smoking, high blood pressure, thalassemia, poor diet, and age. The World Health Organization (WHO) announced that almost 31% of global deaths are caused by heart failure (Shrivastava *et al.*, 2015). There are various types of heart failure: coronary artery disease (CAD), congenital heart defects, arrhythmia, cardiomyopathy, atherosclerosis, and heart infections

(Themistocleous *et al.*, 2017). They can be characterised by symptoms, such as irregular heartbeat, swollen legs and feet, aching in the chest and shoulder, sore gums and jaw, shortness of breath, sleeping problems, and fatigue. This paper discusses the early prediction of the unhealthy heart or potential heart failure by cholesterol measurement, blood pressure, blood sugar, electrocardiographic measurement, heart rate, and chest pain experienced or angina symptoms (Themistocleous *et al.*, 2017) using machine learning methods. Early prediction of heart failure enables the healthcare management system to build an effective disease management strategy that may inhibit the progression of the disease. Furthermore, it may improve the quality of life of the patients. This study also investigates the importance of the 13 variables or symptoms of heart failure in the experiments.

2. Related Work

Machine learning is a discipline focusing on constructing computer systems that can automatically improve based on experience (Jordan & Mitchell, 2015; Shahrel, *et al*., 2021). It has gained popularity in the medical area due to its ability to deal with vast, complex, and unequal data, and one of them is prediction (Stephan *et al.*, 2017). Various machine learning methods have been applied in predicting heart failure problems. A comparative study among six machine learning techniques, Artificial Neural Network (ANN), Support Vector Machine (SVM), Linear Regression (LR), K-Nearest Neighbour (KNN), Decision Tree (DT), and Naive Bayes (NB), has been conducted to predict heart disease; LR outperforms the other five machine learning techniques (Dwivedi, 2018). However, a comparative analysis for heart disease prediction has been conducted among DT, NB, ANN, KNN, and SVM; SVM predicts better than the other machine learning methods with 84.15% accuracy (Pouriyeh *et al*., 2017). SVM also performed well in predicting cardiovascular heart disease compared to other classification techniques, which are DT and ANN (Mohan, 2013).

Zheng *et al*. (2015) compared SVM, ANN, and Hidden Markov Model (HMM) in diagnosing congestive heart failure, and the results indicated that SVM is better than ANN and HMM. A Hybrid Random Forest (RF) has produced 88.7% accuracy in predicting heart failure (Senthilkumar *et al.*, 2019). Thota *et al*. (2018) achieved 93.0% accuracy using RF to predict whether a patient suffers from heart failure. Thus, from the previous experimental analysis, this study investigates the performance of NB, LR, RF, and SVM to predict heart failure using the publicly available dataset from kaggle.com.

2.1 Dataset and Features

The dataset is obtained from the Kaggle heart disease dataset consisting of 303 patients (https://www.kaggle.com/ronitf/heart-disease-uci). There are 14 variables in this dataset: age, sex, cp (chest pain), trestbps (resting blood pressure), chol (cholesterol), fbs (fasting blood sugar), restecg (resting electrocardiographic), thalach (maximum heart rate), exang (exercise-induced angina), oldpeak, slope, ca (number of major vessels), and thal (thalassemia). They serve as input variables or features, and the output variable indicating whether the patient with the specified symptoms has heart failure. In this experiment, an output variable, target, with the values of 0 and 1, indicates the absence and presence of heart failure, respectively. Table 1 shows the description for each variable, and Figure 1 shows some examples of the heart disease prediction data used in this study.

Table 1. Description of Features

age	Age of the person in years
sex	Sex of the person (0: female | 1: male)
cp	Chest pain experience (0: typical angina | 1: atypical angina | 2: non-anginal pain | 3: symptomatic)
trestbps	Resting blood pressure of the person (mm Hg on admission to the hospital)
chol	Cholesterol measurement of the person (in mg/dl)

fbs	Fasting blood sugar of the person (1: true if more than 120mg/dl | 0: false if less than 120mg/dl)
restecg	Resting electrocardiographic measurement (0: normal | 1: having ST-T wave abnormality | 2: showing probable or definite left ventricular hypertrophy)
thalach	Maximum heart rate achieved by the person
exang	Exercise-induced angina (1: yes | 0: no)
oldpeak	ST (positions on the ECG plot) depression induced by exercise relative to rest
slope	The slope of the peak exercise ST segment (1: upsloping | 2: flat | 3: downsloping)
ca	Number of major vessels (0 to 4)
thal	Thalassemia (3: normal | 6: fixed defect | 7: reversible defect)
target	Heart failure problem (0: no | 1: yes)

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1

Figure 1. Example of Dataset for Heart Disease Prediction

2.2 Machine Learning Techniques

a) RandomForest

Random Forest (RF) is a learning technique or classification algorithm applied for prediction, regression, classification, and behaviour analysis by constructing Decision Trees (DT). This technique works well on large datasets (Shaikhina *et al.*, 2019). In predicting heart disease, the RF technique will select random samples of data from a dataset and construct a DT for each sample. There will be many DTs in one single RF model. Then, this technique will get a prediction from each DT. A vote will be computed for each predicted result, and the most votes will be selected as the final prediction result, as shown in Figure 2.

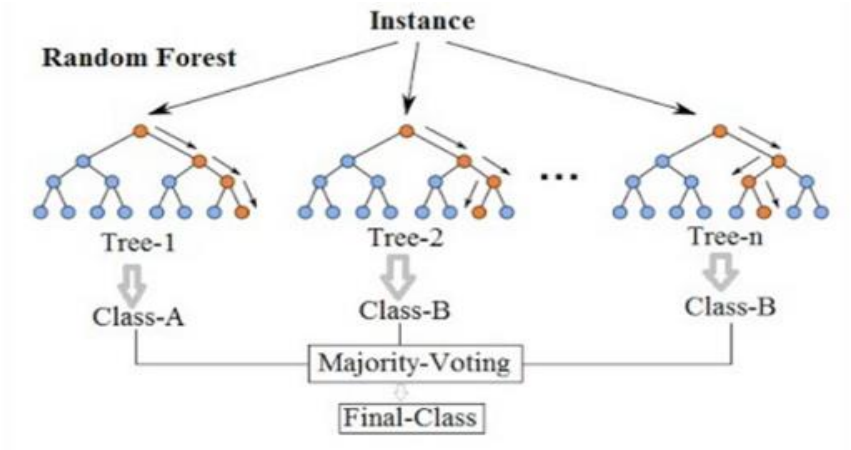


Figure 2. Random Forest Technique (Koehrsen, 2017)

For classification problems, the performance metrics used to calculate and evaluate an algorithm are accuracy, confusion matrix, precision-recall, and F1-score with the help of confusion matrix.

i. Accuracy:
Accuracy is the ratio between the number of correct predictions and the total number of predictions.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

Other metrics, such as precision, recall, and F1 score, will be examined to get more insight into the RF performance.

ii. Precision:
Precision is the ratio between the number of correct positives and the number of true positives plus the number of falsepositives.

$$Precision = TP/(TP + FP) \quad (2)$$

iii. Recall:
Recall is the ratio between the number of correct positives and the number of true positives plus the number of falsenegatives.

$$Recall = TP/(TP + FN) \quad (3)$$

iv. *F1* score:
F1 score is the harmonic mean of precision and recall. *F1* score is performed by taking the weighted average of precision and recall [2][3].

$$F1\ score = 2 * (Recall * Precision)/(Recall + Precision) \quad (4)$$

v. Sensitivity
Sensitivity evaluates the proportion of patients who have heart failure and are correctlypredicted to have heart failure.

$$Sensitivity = TP/(TP + FN) \tag{5}$$

vi. Specificity
Specificitymeasures the proportion of patients who do not have heart failure and are correctly predicted as not having heart failure.

$$Specificity = TN/(TN + FP) \tag{6}$$

According to the mathematical expression [1][2][3][5][6], *TP* is represented as True Positive, *FN* as False Negative, *FP* as False Positive, and *TN* as True Negative, respectively.

b) Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning technique used for prediction, regression, and classification. It works by creating a hyper-plane or a line separating the data into separate classes. The line will act as a decision boundary. An example is shown in Figure 3, in which each class is identified by red = 1 and green = 0.

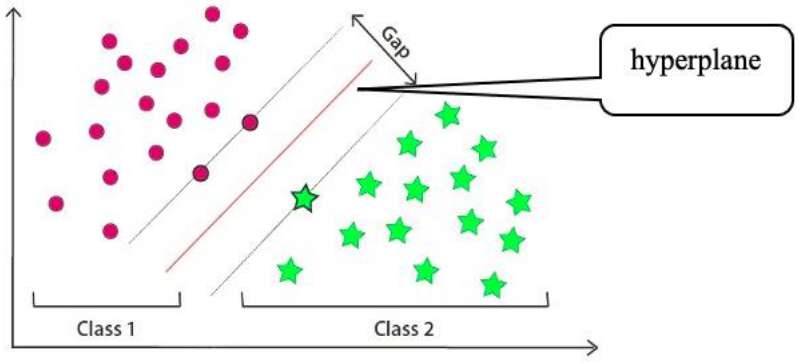


Figure 3. Support Vector Machine (SVM) (Data Flair, n.d.)

Categorising datasets into classes can be done by separating them using the linear function (hyper-plane) defined by equation [7] below:

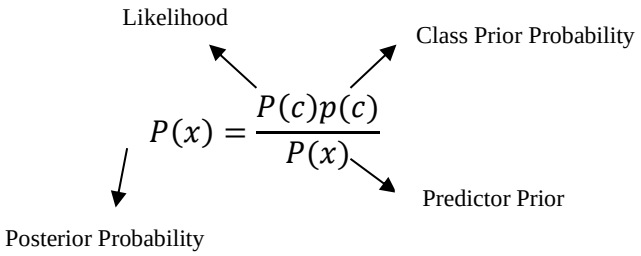
$$\hat{y}_n = \mathbf{w}^\intercal \mathbf{x}_n + b, \qquad \hat{y}_n \in \{1, -1\}$$
$$\text{where} \quad n \in \{1, ..., N\} \tag{7}$$

where Y_n as a prediction of the model, *W* as a vector of weights, X_n donates a data point, **b** as bias, and **N** as datasets. In predicting heart failure, SVM predicts either the patient has heart failure (1) or not (0). The SVM performance is influenced by its kernel function and three of the popular kernel functions are Linear*(K(x,xi) = sum(x*xi))*, Radial Basis Function *(K(x,xi) = exp(-gamma*sum((x – xi^2)))* where gamma is a parameter, ranging from 0 to 1, and Polynomial*(K(x,xi) = 1 + sum(x*xi)^d)* where d is the degree of the polynomial.

c) Naive Bayes

According to Tsangaratos & Ilia (2016), Naive Bayes (NB) is an effective algorithm in the classification technique using Bayes Theorem. This algorithm can predict the probability of different classes based on various attributes. Classification problems with multiple classes can use NB to solve the problems. In this experiment, NB can be served in the diagnosis of heart disease patients. This algorithm can detect the presence of heart disease based on patients'

previous records. Equation [8] shows Bayes Theorem used in NB.



$$P(X) = P(c) \times P((c) \times \dots \dots \times P((c) \times P(c) \quad (8)$$

d) Logistic Regression

Logistic Regression (LR) measures the relationship between the dependent variables (patient will have heart failure or not) and the independent variables (the thirteen input variables) by estimating probabilities using its underlying logistic function. Then, these probabilities are transformed into binary values for the predicted output using the sigmoid function (Hosmer & Lemeshow, 2004). The equation [9] for LR is shown below:

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots$$
$$h(\theta) = g(z)$$
$$g(z) = \frac{1}{1+e^{-z}} \quad (9)$$

Where the predicted output *z* is a linear function of features and Θ coefficients. Each Θ is randomly initialised at the beginning of the training process. However, during the training process, the Θ corresponding to each feature or variable is updated until the loss function value is minimised. The value of $h(\Theta)$ corresponds to $P(y=1|x)$, the probability of output being binary 1, given input x. $P(y=0|x)$ is equal to $1-h(\Theta)$. When the value of *z* is 0, $g(z)$ is 0.5. Whenever *z* is positive, $h(\Theta)$ is greater than 0.5, and the output is 1. The same goes when *z* is negative; the value of y is 0.

3. Implementation

The experiment was implemented in Python with the scikit-learn library (https://scikit-learn.org/stable/user_guide.html). Firstly, all essential libraries and the dataset of heart failure from Kaggle were imported, as shown below:

```
dataset = pd.read_csv("heart.csv")
```

Next is the Exploratory Data Analysis (EDA) process, in which the purpose is to analysethe variable or feature with the target variable. The percentage of patients with and without heart problems wasanalysedwhere '0' indicates patients without heart problems and '1' indicates patients with heart problems. The 'target' variable was used in this process, as shown below:

```
y = dataset["target"] sns.countplot(y)
target_temp = dataset.target.value_counts()
```

Figure 4 shows the percentage of patients without heart problems (45.54%) and patients

with heart problems (54.46%). These percentages indicate that the data is a bit unbalanced, but this is common for medical data. The variable for 'sex' with '0' indicates female and '1' indicates male patients.

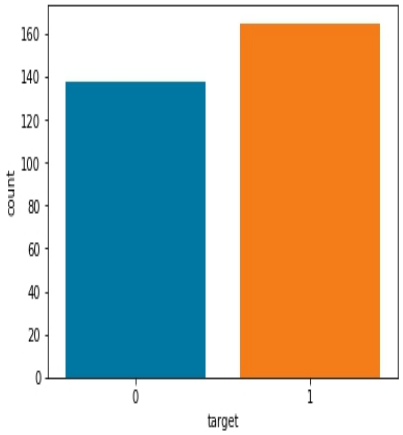


Figure 4. Patient with and without Heart Problems

```
dataset["sex"].unique()
sns.barplot(data["sex"],data["target"])
```

Figure 5 shows the percentage of female patients (31.68%) and male patients (68.32%). It shows that males have a higher possibility of having heart problems than females.

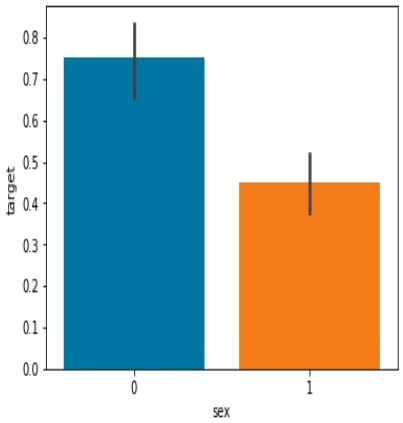


Figure 5. Female Patients and Male Patients

The 'cp' variable,the chest pain type feature,has values from 0 to 3. In 'cp', '0' is typical angina, '1' is atypical angina, '2'is non-anginal pain, and '3' is symptomatic, as shown in Figure 6. It shows that the patients with typical anginal have a low possibility to have heart failures.

```
dataset["cp"].unique()
sns.barplot(dataset["cp"],y)
```

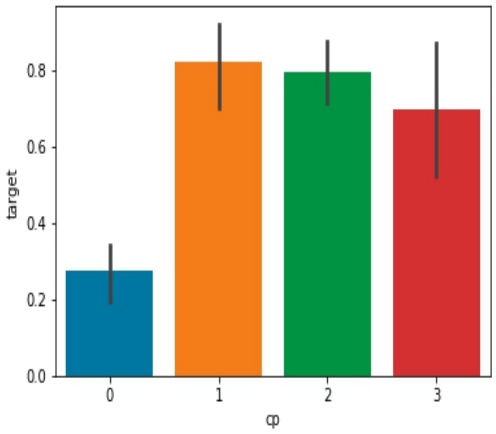


Figure 6. Patient with Typical Anginal, Atypical Angina, Non-Anginal Pain, and Symptomatic

Then, the 'fbs' variable indicates fasting blood sugar of the patients where '1' indicates true, if more than 120mg/dl, and '0' indicates false, if less than 120mg/dl, as shown in Figure 7.

```
dataset["fbs"].unique()
sns.barplot(dataset["fbs"],y)
```

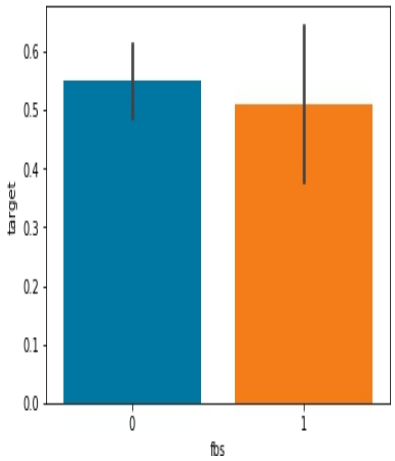


Figure 7. Fasting Blood Sugar of the Patients

Next, the 'restecg' variable represents resting electrocardiographic measurement where '0' indicates normal, '1' indicates having ST-T wave abnormality, and '2' indicates probable or definite left ventricular hypertrophy, as shown in Figure 8. It shows that patients with restecg '0' and '1' have a higher possibility to have heart disease compared to patients with restecg'2'.

```
dataset["restecg"].unique()
sns.barplot(dataset["restecg"],y)
```

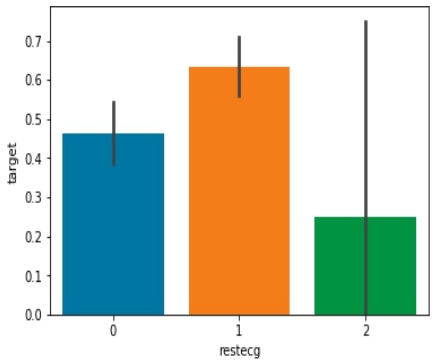


Figure 8. Resting Electrocardiographic Measurement

Then, the 'slope' of the peak exercise ST segment was analysed where '1' indicates upsloping, '2' indicates flat, and '3' indicates downsloping, as shown in Figure 9. It shows that Slope '2' causes heart pain more than Slope '0' and '1'.

```
dataset["slope"].unique()
sns.barplot(dataset["slope"],y)
dataset["ca"].unique() sns.barplot(dataset["ca"],y)
```

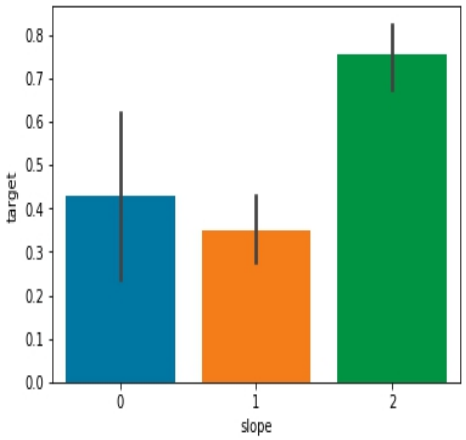


Figure 9. The Peak Exercise ST Segment

This study also analysed the 'ca' feature indicating the number of major vessels (0 to 4), in which ca '4' has a large number of heart patients, as shown in Figure 10.

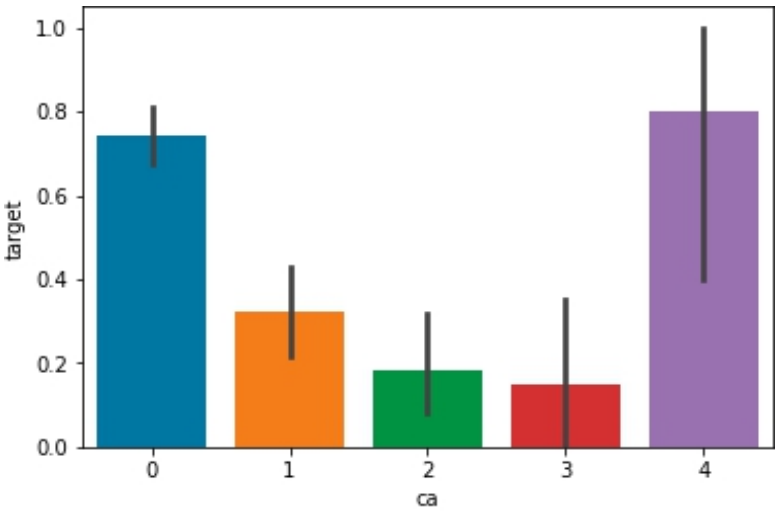


Figure 10. The Number of Major Vessels

Next is analysing the 'thal' feature for thalassemia patients, as shown in Figure 11.

```
dataset["thal"].unique()
sns.barplot(dataset["thal"],y
```

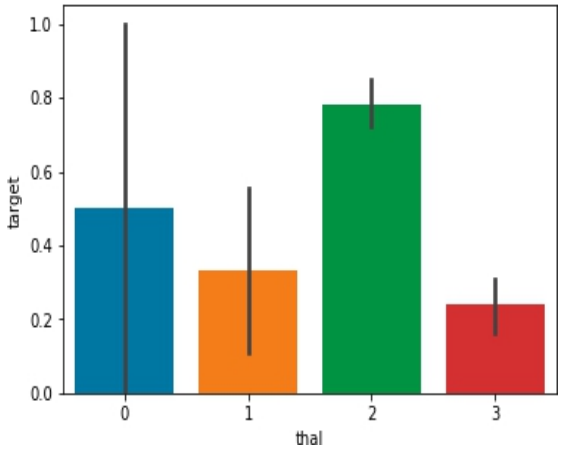


Figure 11. Thalassemia Patients

After analysing all features in the dataset, the researchers split the dataset to train and test in which training features have 242 records, 80% of the data, and testing features have 61 records, 20% of the data. The experiment started with the normalisation of the dataset to avoid over-fitting.

4. Results and Discussion

Preliminary experiments were conducted on SVM by applying the three kernel functions mentioned in the previous section: linear, RBF, and polynomial. The linear kernel function produced the best results. Thus, this result was compared with the results of the other techniques. Table 2 shows the performance of the four machine learning techniques that have been investigated in this study. Referring to Table 2, RF shows the best results for recall and sensitivity, which is 0.97, respectively. RF and LR achieve the same values of accuracy and

f1-score, which are 0.88 and 0.9, respectively. On the other hand, LR produces the highest precision and specificity, which are 0.87 and 0.81, respectively. SVM achieves the highest validation accuracy, which is 0.9. NB seems not to outperform any of the other techniques in any of the evaluation criteria. Table 3 lists the performance score of all the four techniques by computing the average for all the evaluation criteria, and it shows that RF achieves the highest average score, followed by LR, NB, and SVM.

Table 2. Results of RF, SVM, NB, and LR for heart failure prediction.

ML model/result	**Random Forest**	**Support Vector Machine**	**Naïve Bayes**	**Logistic Regression**
Accuracy	0.88	0.85	0.86	0.88
Validation accuracy	0.83	0.90	0.79	0.72
Precision	0.84	0.81	0.85	0.87
Recall	0.97	0.86	0.90	0.94
F1-Score	0.90	0.82	0.88	0.90
Sensitivity	0.97	0.85	0.90	0.94
Specificity	0.76	0.75	0.80	0.81

Table 3**.** Performance score of RF, SVM, NB, and LR.

ML model/results	**Average Performance Score**
Random Forest	0.88
Support Vector Machine	0.83
Naive Bayes	0.85
Logistic Regression	0.87

Another experiment was conducted to determine the features that are significant to predicting heart failure using RF. The result is illustrated in Figure 12, in which the least significant feature is fbs, which is fasting blood sugar.

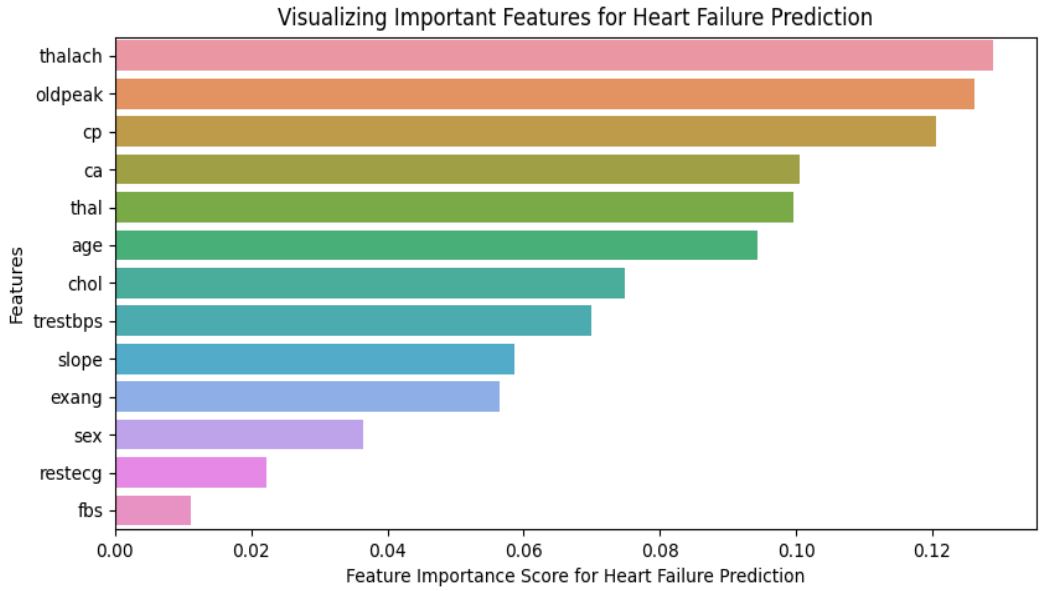


Figure 12. Visualising important features for heart failure prediction produced by Random Forest.

The following experiment was to eliminate one feature at a time, based on the least significant feature, as listed in Figure 12, and apply RF for the prediction. The five least important

features have been eliminated from the training data. The accuracy is shown in Table 4, showing that the accuracy is decreasing as the researchers eliminate the features. The results indicate that all 13 variables or features are important in making a good prediction for heart failure.

Table 4: Accuracy results produced by RF based on the different number of features.

Features	Accuracy (%)
All 13 features	88.52
12 features without fbs	85.24
11 features without fbs and restecg	83.60
10 features without fbs, restecg, and sex	83.32
9 features without fbs, restecg, sex, and exang	81.96
8 features without fbs, restecg, sex, exang, and slope	80.32

5. Conclusion

Machine learning techniques help to reduce the effort and time for medical officers to conduct early predictions for healthcare management purposes. As the number of deaths increases due to heart failures, a machine learning technique system can help predict heart failure accurately and effectively. This study shows that applying machine learning techniques in making early predictions of heart failure may have the potential to improve the healthcare management system. In this experiment, RF seems to achieve the best performance score compared to other techniques. It can lead to a promising disease management strategy that may reduce the progression of the disease. For future work, a hybrid of machine learning techniques with optimisation algorithms with more data will be examined, increasing the accuracy.

REFERENCES

Dwivedi, A.K. (2018). Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Computing & Applications*, 29(10), 685–693.

Data Flair. (n.d.). *Support Vector Machines Tutorial -Learn to implement SVM in Python.* Retrieved from https://data-flair.training/blogs/svm-support-vector-machine-tutorial/

Goldstein, B.A, Navar A.M, Pencina, M.J & Ioannidis, J.P. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc*, *24*(1), 198-208.

Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Lee, M. J. & Asadi, H. (2018). eDoctor: machine learning and the future of medicine. *J Intern Med*, *284*(6), 603-619.

Hosmer, D.W Jr & Lemeshow, S. (2004). *Applied Logistic Regression,* Wiley, New York.

Jordan, M. I. & Mitchell, T. M. (2015). Machine Learning: Trends, perspectives and prospects, *Science*, *349*(6245), 255-260.

Khan, S., Mohd, N. N., Shahzad, A., Ulah, A., Mushtaq, M., Mir, J. & Aamir, M. (2017). Comparative Analysis for Heart Disease Prediction. *JOIV: International Journal on Informatics Visualization,*1(4-2), 227-231.

Koehrsen, W. (2017, December 28). *Random Forest Simple Explanation.* Retrieved from https://www.spotx.tv/resources/blog/developer-blog/exploring-random-forest-internal-knowledge-and-converting-the-model-to-table-form/

Mohan, V. (2013). Comparative Analysis of Classification Function Techniques for Heart Disease Prediction. *International Journal of Innovative Research in Computer and Communication Engineering*, 1(3),735-741.

Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H. & Gutierrez, J. (2017). A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease, *2017. IEEE Symposium on Computers and Communications (ISCC*), pp. 204-207

Shahrel, M. Z., Mutalib, S., & Abdul-Rahman, S. (2021). PriceCop-Price Monitor and Prediction Using Linear Regression and LSVM-ABC Methods for E-commerce Platform. *International Journal of Information Engineering & Electronic Business*, *13*(1). 1-14.

Shaikhina, T, Lowe, D., Daga, S., Briggs, D., Higgins, R. & Khovanova, N. (2019). Decision Tree and Random Forest Models for Outcome Prediction in Antibody Incompatible Kidney Transplantation. *Biomedical Signal Processing and Contro, 52*, 456-462.

Shrivastava, A. K., Singh, H. V., Raizada, A. & Singh, S. K. (2015). C-reactive Protein. Inflammation and Coronary Heart Disease. *The Egyptian Heart Journal, 67*(2*).* 89-97.

Steele, A. J., Denaxas, S. C., Shah, A. D., Hemingway, H. & Luscombe, N. M. (2018). Machine Learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease, *PLoS One*, 13(8), e0202344.

Themistocleous, I., Stefanakis, M., & Douda, H. T. (2017). Coronary heart disease part I: pathophysiology and risk factors. *Journal of Physical Activity, Nutrition and Rehabilitation*, 167–175.

Thota, L., Nimmala, S., &Manasa, K. (2018). Heart Disease Prediction Using Random Forest Algorithm. *Global Journal of Engineering Science and Research*, *5*(8), 248-252.

Tsangaratos, P., & Ilia, I. (2016). Comparison of a Logistic Regression and Naive Bayes Classifier in Landslide Susceptibility Assessments: The Influence of Models Complexity and Training Dataset Size. *CATENA,* 145.164-179.

Tripoliti, E. E., Papadopoulos, T. G., Karanasiou, G. S., Naka, K. K. & Fotiadis, D. I., (2017). Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse Events Through Machine Learning Techniques. *Journal of Computational and Structural Biotechnology*, 15, 26-47.

Zheng, Y., Guo, X., Qin, J. & Xiao, S. (2015). Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics. *Computer Methods Programs Biomed*, I22, 372–83.