



The Analysis Performance of Heart Failure Classification by Using Machine Learning Techniques

Nurul Farhana Hamzah¹, Nazri Mohd Nawi^{1*}, Abdulkareem A. Hezam¹

¹Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia (UTHM), Parit Raja, 86400 Batu Pahat, Johor, MALAYSIA

*Corresponding Author

DOI: <https://doi.org/10.30880/jscdm.2021.02.02.009>

Received 15 June 2021; Accepted 01 October 2021; Available online 15 October 2021

Abstract: Heart failure means that the heart is not pumping well as normal as it should be. A congestive heart failure is a form of heart failure that involves seeking timely medical care, although the two terms are sometimes used interchangeably. Heart failure happens when the heart muscle does not pump blood as well as it can, often referred to as congestive heart failure. Some disorders, such as heart's narrowed arteries (coronary artery disease) or high blood pressure, eventually make the heart too weak or rigid to fill and pump effectively. Early detection of heart failure by using data mining techniques has gained popularity among researchers. This research uses some classification techniques for heart failure classification from medical data. This research analyzed the performance of some classification algorithms, namely Support Vector Machine (SVM), Decision Forest (DF), and Boosted Decision Tree (BDT), to classify accurately heart failure risk data as input. The best algorithm among the three is discovered for heart failure classification at the end of this research.

Keywords: Social media, machine learning, decision forest, neural network, Support Vector Machine (SVM)

1. Introduction

In general, "Heart Failure is a clinical syndrome characterized by symptoms of breathlessness and fatigue, with signs of fluid retention and supported by objective evidence of cardiac dysfunction (systolic/ diastolic)" [1]. Heart failure, known as congestive heart failure, occurs when the heart muscle does not pump blood as well as it should. Certain conditions, such as narrowed arteries in your heart or high blood pressure, will gradually make the heart weak or stiff enough to fill and pump efficiently [2]. The end-stage of most diseases of the heart in heart failure (HF). The prevalence of HF varies from 3 to 20 per 1000 population, although it could be as high as 100 per 1000 population in people over the age of 65 years [3]. A total of 18,267 people in Malaysia died in 2018 from ischemic heart disease, averaging 50 deaths a day [4]. This comprised 12,510 men and 5,757 women.

The 2018 rate of cerebrovascular diseases, including stroke and death-causing aneurysms, increased slightly from 7.1% in 2017 to 7.8% in 2018 [5]. The most common underlying causes of heart failure in adults are coronary heart disease and hypertension [6]. All the conditions that lead to heart failure cannot be healed, but treatment can improve the signs and symptoms of heart failure and help for live longer [7]. Lifestyle changes such as exercise, salt reduction in your diet, stress management, and weight loss will also enhance the quality of life [8]. Identifying the early symptoms of heart failure is not easy since it requires an expert from the medical field to identify and suggest based on the clinical record provided [9]. Early detection with machine learning can help doctors be aware,

and at the same time, it can save time and life. Figure 1 shows the condition of a healthy and unhealthy heart for a normal person.

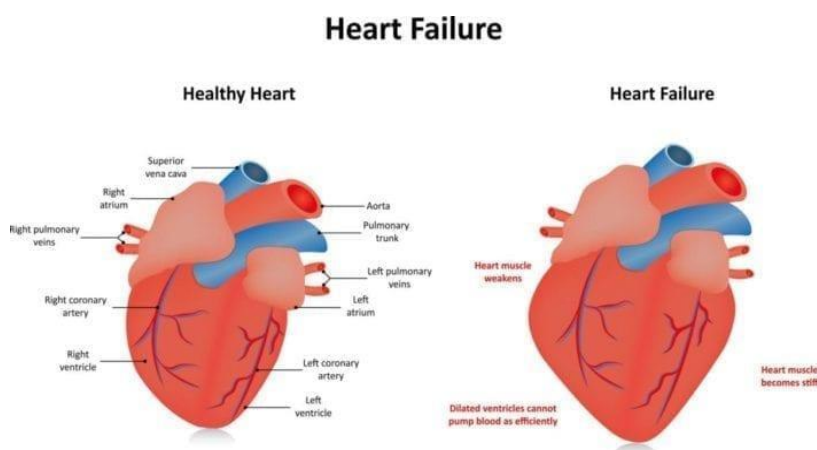


Fig. 1 - Healthy heart and heart failure [10]

2. Related Work

This study was performed to understand. Heart failure is the state in which muscle fades and enlarges in the heart wall, limiting the pumping of blood by the heart. The heart ventricles may become inflexible and do not fill in between beats properly [11]. Through the passing of time, the heart ceases to meet the proper demand for blood in the body, and, as a result, the individual begins to have trouble breathing. Coronary heart disease, diabetes, high blood pressure, and other disorders such as HIV, substance addiction or cocaine, thyroid disorder, excess vitamin E in the body, radiation or chemotherapy are the key factors behind heart failure [12]. As WHO Coronary Heart Disease (CHD) has shown, 31 percent of deaths worldwide are now the top cause [13].

Classification is a function of data mining that assigns objects to target groups or classes in a database. For each case in the data, the aim of classification is to predict the target class [14] accurately. For example, a classification model could be used to classify loan applicants as low, medium or high credit risk. Classification models were evaluated in a test data set by comparing the expected values to known goal values. Usually, the historical data for a classification project is split into two data sets, one to construct the model and the other to validate the model. It has numerous uses in industry modeling, marketing, credit analysis, and biomedical and opioid reaction modeling for patient segmentation [14].

This section will describe the various articles that have been analyzed using different classification algorithms, such as boosted decision tree, decision forest, and decision tree, and figure out which approach has the highest accuracy. The first article is about Evaluating forecasting methods by considering different accuracy measures [15]. This article is about a decision-making model that allows researchers to identify the superiority of a forecasting technique over another by considering several accuracy metrics concurrently. 10- fold cross-validation approach was used to evaluate the performance of the algorithms. The highest accuracy among these two algorithms is Random Forests (91.32%), while SVM (85.11%).

The second article is about Performance Analysis of Classification Algorithms on Parkinson's Dataset with Voice Attributes [16]. This dataset is to classify the records into two categories which are Parkinson affected or not, which are the values of the Class Label 1 or 0. The algorithm that has the highest score for accuracy is Random Forest (78.56%), Decision Tree (77.63%), AdaBoost (76.56%), and SVM (72.76%).

Deep Synergy: Predicting Anti-Cancer Drug Synergy with Deep Learning is the following article by [17]. The random forest approach has the highest accuracy (92%), although GBM has an accuracy of (87%), and SVM has an accuracy of (76%), as seen in the table above.

The fourth article [18] employs Data Mining Techniques. This study aims to find the optimal instructor model utilizing two decision tree methods, C4.5 (J48) and Random Forest (RF). The findings reveal that Model 2 of the J48 algorithm outperformed the Random model with an accuracy of (98.86%). Table 1 shows the summary of the performance of three selected data mining techniques by previous researchers.

Table 1 - Comparison of accuracy rate between various articles

References	Related	Method	Accuracy
[15]	Title: Evaluating forecasting methods by considering different accuracy measures.	Random Forest	91.32%
		SVM	85.11%
[16]	Title: Performance Analysis of Classification Algorithm on Parkinson's dataset with voice attributes.	Decision Tree	77.63%
		Random Forest	78.56%
		Ada Boost	76.56%
		Linear Support Vector Machine	72.76%
[17]	Title: Predicting anti-cancer drug synergy with deep learning.	Gradient Boosting Machine	87.00%
		Random Forests	92.00%
		Support Vector Machine	76.00%
[18]	Title: Excellence Teacher Modelling Using Data Mining	Random Forest	98.84%
		J48	98.86%
[19]	Title: Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection	SVM	75.00%
		RF	99.00%

3. Methodology/ Framework

Selecting a proper methodology is crucial in completing this project in evaluating the classification of heart failure. This research selects the Cross-Industry Standard Process for Data Mining (CRISP-DM) for the appropriate methodology. CRISP-DM is a technique that offers a structured approach to project planning for data mining. Parts of these problems were addressed in the CRISP-DM projects by defining a process model that provides a framework for implementing data mining projects independent of both the industry sector and the technology used. The CRISP-DM process model aims to make large, less expensive, more reliable, more repeatable, more manageable, and faster data mining projects. CRISP-DM refers to the process model that provides a basis for implementing data mining projects [20]. The CRISP-DM process model is designed to do large data mining projects that are more efficient, less costly, more repeatedly, quicker, and more manageable. This model was chosen because it offers an overview of a data mining project's life cycle [21]. The CRISP-DM reference model provides an overview of the life cycle of a data mining project for data mining. It contains the phases of their respective tasks and outputs in a project. A data mining project's life cycle is broken down into six phases. The phase sequence is not strict. The arrows indicate only the most significant and frequent dependencies between phases, but it depends on the results of each phase in a particular project, which phase or phase task must be performed [22].

Based on Fig. 2, the first phase is business understanding which this phase explains the objective of this project. This phase is to investigate the classification method of heart failure in the literature, classify the factor of heart failure using three algorithms, and evaluate the classification ability using accuracy, recall, and precision. The second phase is about understanding the dataset. The dataset gets from the UCI Machine Learning website. Firstly, list the obtained datasets are including clinical information and medical history. In this data set, they have 299 instances and 13 attributes. The characteristics of the attribute are an integer and real (Boolean) only. The data set has null or missing values. The next phase is data preparation. This phase reconsiders data selection criteria, decides which dataset will be using, correct, remove or ignore the noise, and decide how to deal with unique values and meaning. The next phase is modeling. In this phase, the techniques used in this project are the support vector machine, random forest, and gradient boosting machine. This phase is also to separate the heart failure dataset into two, which is into train and test sets, build the model on the train set and estimate its quality on the separate test set. Each model will do 10-fold validation to see the performance for each method used.

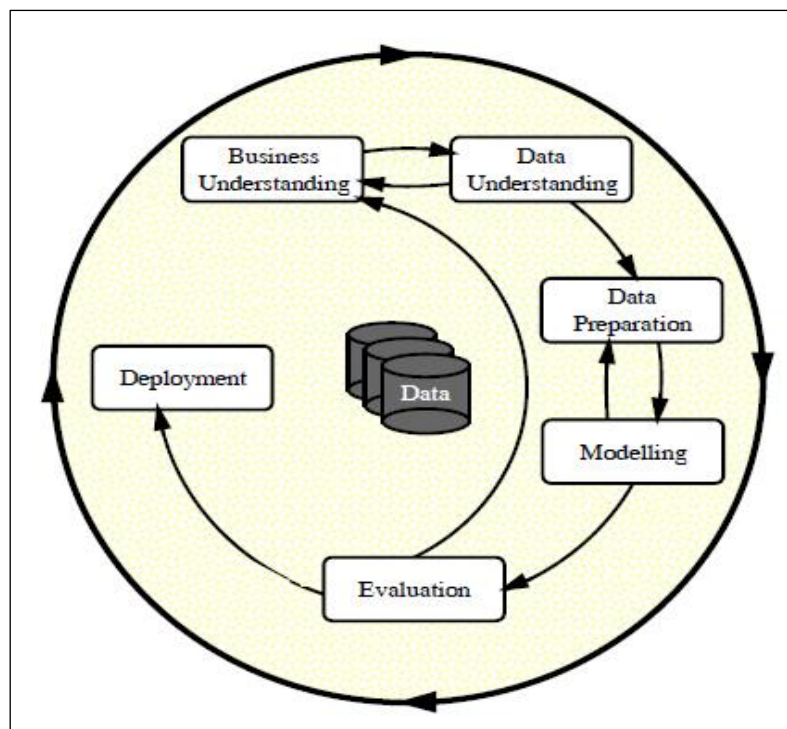


Fig. 2 - The CRISP-DM Process methodology [22]

3.1 Datasets

The dataset of heart failure was chosen from the UCI Machine Learning websites for this project. This data about the heart failure clinical record for 299 patients who had heart failure was collected during their follow-up period, where each patient profile has 13 clinical features. Table 2 shows the dataset attributes.

Table 2 - Dataset attributes

No.	Name of Features	Types of Features
1	Age	Integer
2	Anemia	Boolean
3	High Blood Pressure	Boolean
4	Creatinine Phosphokinase (CPK)	Integer
5	Diabetes	Boolean
6	Ejection Fraction	Boolean
7	Platelets	Boolean
8	Sex	Binary
9	Serum Creatinine	Integer
10	Serum Sodium	Boolean
11	Smoking	Boolean
12	Time (days)	Integer
13	Death event (target)	Boolean

3.2 Algorithms

This research selected three classification techniques for the experiments in this project which are Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosted Machine (GBM).

- SVM is a classification algorithm used in machine learning to improve prediction accuracy while minimizing data over-adjustment [23]. SVMs are divided into two categories linear SVMs and non-linear SVMs. Linear SVM is a hyperplane separation classifier. Non-linear SVM, on the other hand, occurs when SVM locates and classifies points within a certain hyperplane in the domain of

the function. It does not necessary to define the space explicitly. Instead, it can do so by defining a kernel function. Vapnik's SVM method addresses data interpretation and regression problems. SVM is one of the classification algorithms that, after training, will categories future data into several categories. Learning models are generated during the training phase by grouping original data into distinct groups based on their names.

- Random forests are a combination of tree predictors. Each tree depends on an independently sampled random vector's values and the same distribution for all forest trees [24].
- Boosting decision trees is a well-known method of classification. The predictive performance of trees is sometimes not nearly as strong on unseen data as that obtained on the training data [25]. This phenomenon is often described as overfitting, where the tree is too specialized to the training data. This will arise due to the existence of high variance in the data. The variance can be greatly reduced by inducing multiple decision trees from the same data.

3.3 Evaluation Metrics

The classification of preprocessing is carried out based on all the values of taken 13 attributes. A comparative study of classification accuracy in the support vector machine, random forest, and gradient boosting machine algorithm is carried out in this work. The evaluation metrics used in the experiments are accuracy, precision, and recall represented as a confusion matrix [23], [26].

- **Precision:** Precision is used to measure the positive patterns that are correctly predicted from the total predicted patterns in a positive class.

$$\text{Precision}, P = \frac{TP}{TP+FP} \quad (1)$$

Where TP=True Positive, FP = False Positive.

- **Recall:** Recall is used to measure the fraction of positive patterns that are correctly classified.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

Where FN = False Negative rate.

- **Accuracy:** The accuracy metric measures the ratio of correct predictions over the total number of instances evaluated.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Where TN = True Negative.

4. Experimental Results

The purpose of the experiments is to select the best classification algorithm for a heart failure diagnostic. The performance of Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting Machine (GBM) on the factor of heart failure dataset using a classification in data mining are evaluated and compared in terms of accuracy, precision, and recall. The accuracy result is shown in Table 3. While the precision results shown in Table 4 and Table 5 shows the result of the recall.

The result of accuracy for Support Vector Machine (SVM), Decision Forest (DF), and Gradient Boosted Machine (GBM) are shown in Table 3. Experiments of different data allocation of cross-validation folds were conducted using the three algorithms: SVM: SVM, DF, and GBM. In test number one, where the data were split into 90% training and 10% testing, the test result indicates that the highest accuracy was found in SVM, and the highest accuracy score was 85.5%. In test eight, while the lowest accuracy is where the data is divided into 30% training and 70% testing, and the score is 64.6%. Generally, the Support Vector Machine's classification results tend to be lower than the other methods in which the average accuracy score is 75.96%, and the result of the standard deviation for SVM is 6.683.

The test result also demonstrates the best accuracy for GBM shown in test number one, when the data was divided into 90% training and 10% testing. The highest accuracy value is 93.3% for GBM. As a result, the lowest accuracy was discovered in test number ten, where the data was separated into 10% training and 90% testing. Despite that, the lowest accuracy score is 71.4%. In general, the classification results for GBM tend to be higher than other algorithms in which the average accuracy is 82.55%. In contrast, the standard deviation results for GBM is 5.376.

Table 3 - The accuracy result of the SVM, DF, and GBM

Test	Split data	SVM	RF	GBM
1	90:10	76.7	86.7	93.3
2	80:20	85.0	73.3	83.3
3	77:33	85.5	78.3	82.6
4	70:30	78.9	78.9	78.9
5	60:40	71.7	76.7	81.7
6	50:50	79.9	79.9	82.6
7	40:60	75.4	84.4	83.8
8	30:70	64.6	82.8	84.2
9	20:80	72.8	79.5	83.7
10	10:90	69.1	72.5	71.4
Average		75.96	79.3	82.55
Standard Deviation		6.683	4.516	5.376

The test results also show that the maximum accuracy for DF is obtained in the first test when the data is split into 90% training and 10% testing. The highest accuracy rating is 86.7%. The lowest accuracy is seen in test nine when the data is split between 10% training and 90% testing. 72.5% is the lowest accuracy score. In general, the classification results of Decision Forests are slightly lower than those of Gradient Boosted Machine. The average accuracy score for Decision Forests is 79.3%. In contrast, the standard deviation for DF is 4.516.

Figure 3 shows the accuracy results for SVM, DF, and GBM that. The accuracy of DF, SVM, and GBM when split data (70:30) is the same, where the accuracy is 79.9%. However, the split data (30:70) is higher than the other two algorithms when the GBM accuracy (84.2%) is higher than the other two algorithms. While the accuracy of DF (82.8%) and SVM (64.6%). While when the data split (77:33), the accuracy of GBM (82.6%) and SVM (85.5%) increased while the accuracy of DF (78.3%) have dropped. The highest accuracy (93.3%) for GBM is when the data is divided into 90% training and 10% testing. In contrast, the lowest accuracy is when the data of SVM (64.6%) is divided into 30% training and 70% testing.

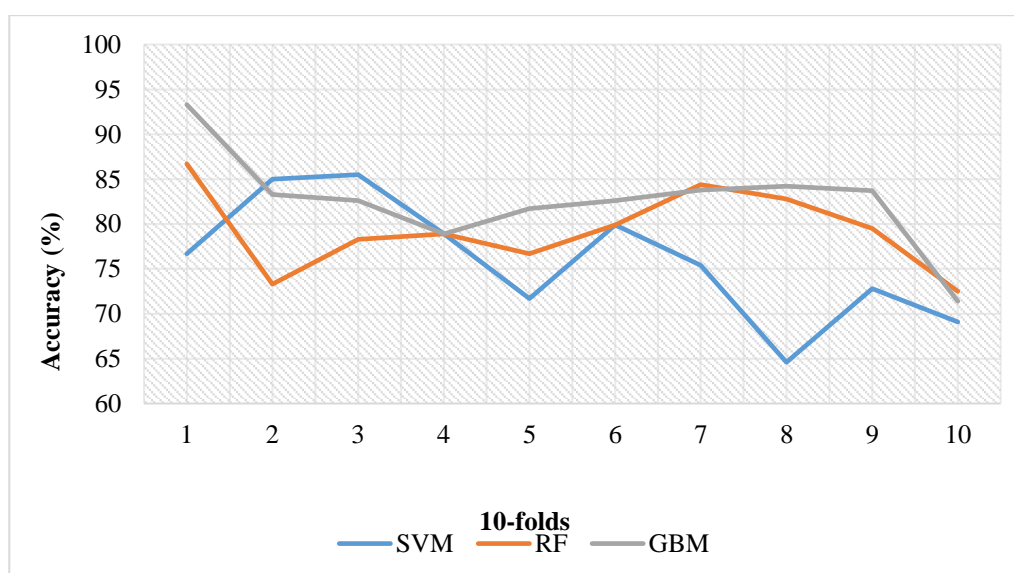
**Fig. 3 - The accuracy result of the SVM, DF, and GBM**

Table 4 shows the test result of precision for SVM, DF, and GBM. The findings were obtained by running ten tests with varying data allocation of cross-validation folds using the three algorithms, SVM, DF, and GBM. The results of the tests demonstrate that the highest precision for SVM is obtained in test two, where the data is divided into 80% training and 20% testing, respectively, with the greatest accuracy score of 0.800. The worst precision is in eight when the data is divided between 30% training and 70% testing, with 0.417. In general, the

classification results of the SVM are the lowest than those of the other approaches, with an average precision score of 0.574 and a standard deviation of 0.133 for SVM precision.

The test result also shows that the best precision for DF is found in test number one, where the data is dividing into 90% training and 10% testing. The highest precision score is 0.875. Subsequently, the worst precision is in test ten, where the data is dividing into 10% training and 90% testing, and the score is 0.557. Generally, the classification results of the DF tend to be second higher than SVM, in which the average precision score is 0.700, while the standard deviation for DF precision is 0.084. The test result shows that the best precision for GBM is found in test number one, where the data is divided into 90% training and 10% testing. The highest precision score is 1. In contrast, the worst precision is in test ten, where the data is divided into 10% training and 90% testing. The lowest score for precision is 0.530. Generally, the classification results of the GBM tend to be slightly higher than SVM and DF, in which the average precision score is 0.743, while the standard deviation for GBM precision is 0.116.

Table 4 - The precision result of the SVM, DF and GBM

Test	Split data	SVM	RF	GBM
1	90:10	0.714	0.875	1.000
2	80:20	0.800	0.647	0.789
3	77.33	0.760	0.682	0.762
4	70:30	0.704	0.720	0.677
5	60:40	0.581	0.690	0.718
6	50:50	0.653	0.660	0.702
7	40:60	0.582	0.737	0.741
8	30:70	0.417	0.767	0.769
9	20:80	0.604	0.662	0.737
10	10:90	0.506	0.557	0.530
Average		0.574	0.700	0.743
Standard Deviation		0.133	0.084	0.116

Figure 4 shows the precision results for GBM and DF that, when split data (90:10), the best precision results are GBM with the score of 1 and DF 0.875. While for the SVM, the best precision score is when the split data (80:20) is 0.800. DF and GBM have worse precision scores when the split data at (10:90), the score is 0.557 for DF and 0.530 for GBM, while SVM has the worse precision score among the others, which is 0.417 when the split data is at (30:70).

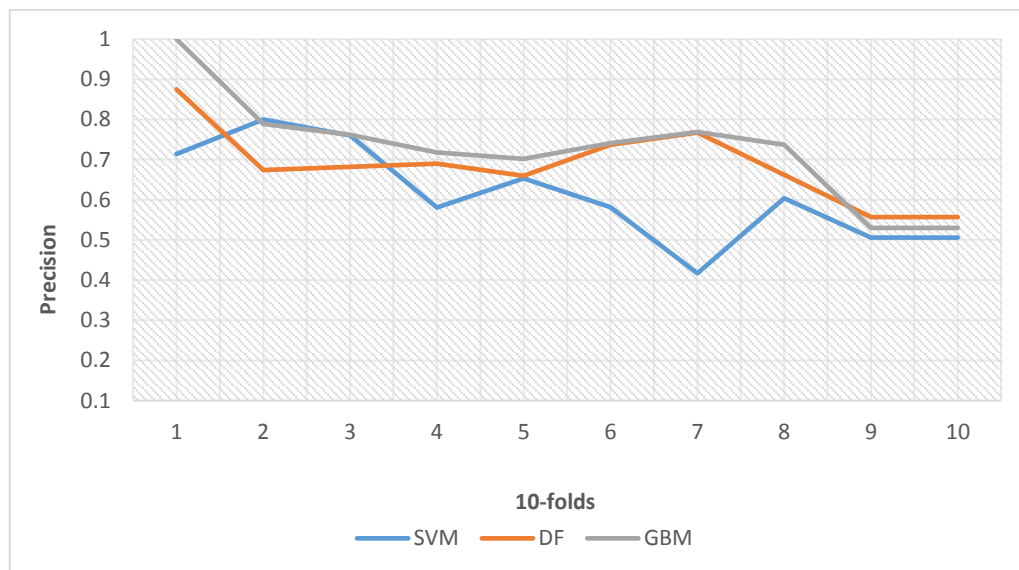


Fig. 4 - The precision result of the SVM, DF, and GBM

Table 5 shows the test result in which the best recall for SVM is found in test number seven, where the data is divided into 30% training and 70% testing with the best recall score of 0.221. At the same time, the worst recall is in test three, where the data is divided into 77% training and 33% testing. The worst score recall for SVM is 0.826. Generally, the Support Vector Machine's classification results tend to be better than the other method in which the average recall score is 0.574 while the standard deviation for SVM recall is 0.1883.

The test result also shows that the best recall for DF is found in test number two, where the data is divided into 80% training and 20% testing. The best recall score is 0.524. Subsequently, the worst recall is in test six, where the data is divided into 40% training and 60% testing, and the worst score is 0.764. Generally, the Decision Forest classification results tend to be slightly worse than Support Vector Machine, in which the average recall score is 0.454, and the standard deviation for DF recall is 0.278.

The test result shows that the best recall for GBM is found in test number three, where the data is divided into 77% training and 33% testing. The best recall score is 0.696. Subsequently, the worst recall is in test one, where the data is divided into 90% training and 10% testing, and the worst score is 0.800. Generally, the Gradient Boosted Machine classification results are the best than the other methods in which the average recall score is 0.723 while the standard deviation for GBM recall is 0.0304.

Table 5 - The recall result of the SVM, DF and GBM

Test	Split data	SVM	RF	GBM
1	90:10	0.500	0.700	0.800
2	80:20	0.762	0.524	0.714
3	77:33	0.826	0.652	0.696
4	70:30	0.633	0.600	0.700
5	60:40	0.462	0.513	0.718
6	50:50	0.711	0.689	0.733
7	40:60	0.709	0.764	0.727
8	30:70	0.221	0.676	0.735
9	20:80	0.387	0.707	0.747
10	10:90	0.524	0.583	0.726
Average		0.574	0.641	0.723
Standard Deviation		0.1883	0.0828	0.0304

Fig. 5 shows the results of recall for SVM, DF, and GBM that when split data (90:10), the recall of SVM, DF, and GBM have different values which 0.500 for SVM, 0.700 for DF, and 0.800 for GBM. But when the split data at (30:70) in test seven, the recall of SVM has the best recall results, which are 0.221, while DF 0.676 and GBM 0.735.

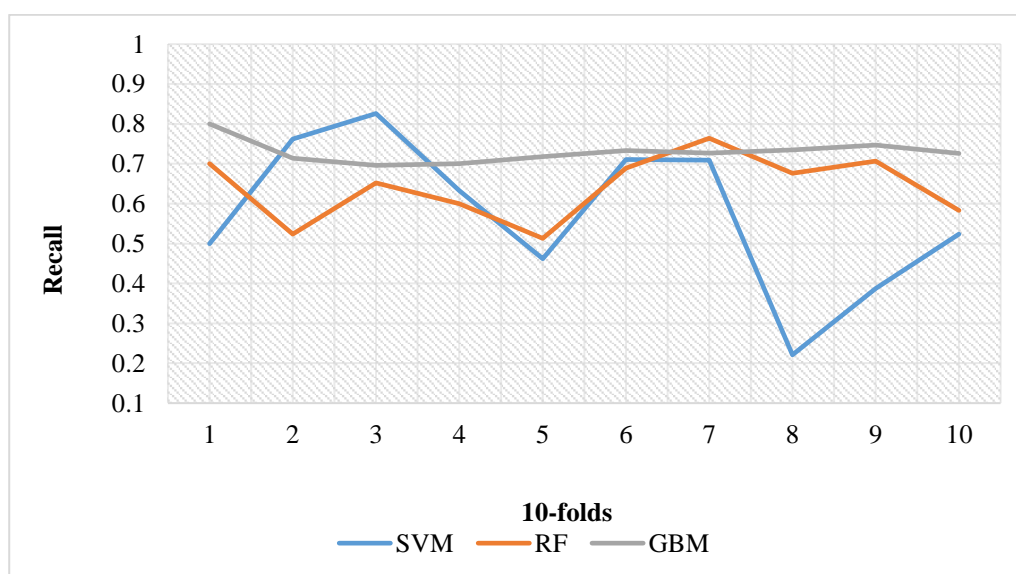


Fig. 5 - The Recall result of the SVM, DF, and GB

5. Conclusion

This research presents the analysis comparison for the performance of three selected data mining techniques for classifying heart failure data. The performance of support vector machine (SVM), random forest (RF), and gradient boosting machine (GBM) were compared based on evaluation metrics such as accuracy, precision, and recall on the given dataset to estimate the performance of each classification technique. As a result, SVM shows the highest rate of accuracy as compared to DF and GBM. The results show that GBM (82.55%) is slightly better as compared to DF (79.3%) and SVM (75.96%) in terms of accuracy. In future work, the use of some other well-known feature selection methods that can select useful features that can be used together with the classification algorithm to improve its performance in terms of its accuracy. Some other well-known classification techniques such as Naïve Bayes, Neural Network, and Decision Tree can be tested on the same dataset for comparing their performance in terms of its accuracy.

Acknowledgment

We would like to thank the Universiti Tun Hussein Onn Malaysia's Faculty of Computer Science and Information Technology for their assistance.

References

- [1] Ahmad, T., Munir, A., Bhatti, S. H., Aftab, M., & Raza, M. A. (2017). Survival analysis of heart failure patients: A case study. *PloS one*, 12(7), e0181001
- [2] Elhoseny, M., Mohammed, M. A., Mostafa, S. A., Abdulkareem, K. H., Maashi, M. S., Garcia-Zapirain, B., ... & Maashi, M. S. (2021). A new multi-agent feature wrapper machine learning approach for heart disease diagnosis. *Comput. Mater. Contin.*, 67, 51-71
- [3] Van Bakel, Adrian B., and Geoffrey Chidsey. (2002). Medical Management of Advanced Heart Failure. *Clinical Cornerstone* 4(6):42–52
- [4] Belavagi, M. C., & Muniyal, B. (2016). Performance evaluation of supervised machine learning algorithms for intrusion detection. *Procedia Computer Science*, 89, 117-123
- [5] Ciampi, Q., & Villari, B. (2007). Role of echocardiography in diagnosis and risk stratification in heart failure with left ventricular systolic dysfunction. *Cardiovascular ultrasound*, 5(1), 1-12
- [6] Cowie, M. R., Anker, S. D., Cleland, J. G., Felker, G. M., Filippatos, G., Jaarsma, T., ... & López- Sendón, J. (2014). Improving care for patients with acute heart failure: before, during and after hospitalization. *ESC heart failure*, 1(2), 110-145
- [7] familydoctor.org. 2021. "Heart Failure." Familydoctor.Org
- [8] Govil, S. R., Weidner, G., Merritt-Worden, T., & Ornish, D. (2009). Socioeconomic status and improvements in lifestyle, coronary risk factors, and quality of life: the Multisite Cardiac Lifestyle Intervention Program. *American journal of public health*, 99(7), 1263-1270
- [9] Halim, N. H. A., Hamdan, A. R., Othman, Z. A., & Jantan, H. (2017). Pemodelan guru cemerlang KPM menggunakan perlombongan data. *Journal of ICT in Education*, 4, 21-34
- [10] De Hert, M., Detraux, J., & Vancampfort, D. (2018). The intriguing relationship between coronary heart disease and mental disorders. *Dialogues in clinical neuroscience*, 20(1), 31
- [11] Jamee Shahwan, A., Abed, Y., Desormais, I., Magne, J., Preux, P. M., Aboyans, V., & Lacroix, P. (2019). Epidemiology of coronary artery disease and stroke and associated risk factors in Gaza community–Palestine. *PloS one*, 14(1), e0211131
- [12] Jin, Z., Shang, J., Zhu, Q., Ling, C., Xie, W., & Qiang, B. (2020, October). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. In *International Conference on Web Information Systems Engineering* (pp. 503-515). Springer, Cham
- [13] Saravanapriya, K., & Bagyamani, J. (2017). Performance Analysis of Classification Algorithms on Diabetes Dataset. *Int. J. Comput. Sci. Eng.*, 5(9), 15-20
- [14] Kesavaraj, G., & Sukumaran, S. (2013, July). A study on classification techniques in data mining. In *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)* (pp. 1-7). IEEE
- [15] Marjudi, S., Setik, R., Ahmad, R. M. T. R. L., Hassan, W. A. W., Harun, W., & Ismail, S. Cardiovascular Disease Risk Factors among White-Collar Workers towards Healthy Communities in Malaysia. *Studies*, 15, 17
- [16] McBrien, J. A. (2003). Assessment and diagnosis of depression in people with intellectual disability. *Journal of Intellectual Disability Research*, 47(1), 1-13
- [17] McMurray, J. J., & Stewart, S. (2002). The burden of heart failure. *European Heart Journal Supplements*, 4(suppl_D), D50-D58
- [18] Mehdiyev, N., Enke, D., Fettke, P., & Loos, P. (2016). Evaluating forecasting methods by considering different accuracy measures. *Procedia Computer Science*, 95, 264-271

- [19] Moro, S., Laureano, R., & Cortez, P. (2011). Using data mining for bank direct marketing: An application of the crisp-dm methodology
- [20] Vishwanathan, S. V. M., & Murty, M. N. (2002, May). SSVM: a simple SVM algorithm. In Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290) (Vol. 3, pp. 2393-2398). IEEE
- [21] Vorhies, W. (2016). CRISP-DM—a Standard Methodology to Ensure a Good Outcome. Data Science Central.
- [22] Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1). London, UK: Springer-Verlag
- [23] Mostafa, S. A., Mustapha, A., Mohammed, M. A., Hamed, R. I., Arunkumar, N., Abd Ghani, M. K., ... & Khaleefah, S. H. (2019). Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. *Cognitive Systems Research*, 54, 90-99
- [24] Mostafa, S. A., Mustapha, A., Mohammed, M. A., Ahmad, M. S., & Mahmoud, M. A. (2018). A fuzzy logic control in adjustable autonomy of a multi-agent system for an automated elderly movement monitoring application. *International journal of medical informatics*, 112, 173-184
- [25] Li, Q., Wen, Z., & He, B. (2020, April). Practical federated gradient boosting decision trees. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 04, pp. 4642-4649)
- [26] Dong, W., Cao, X., Wu, X., & Dong, Y. (2019). Examining pedestrian satisfaction in gated and open communities: An integration of gradient boosting decision trees and impact-asymmetry analysis. *Landscape and urban planning*, 185, 246-257