National Heart Centre Singapore
SingHealth

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

ANALYTIC MODEL FOR EARLY DETECTION FOR HEART FAILURE

PREDICTIVE HEALTH GROUP (PHG)

AN6003 Analytics Strategy
AY22 Group Project – Team 4
Kao, Han-Ying; Liu, Yi-Chun (Victoria); Shi, Zhuoya; Wang, Zhiyi; Zhang, Yachen (Monica); Zhang, Yibao

National Heart Centre Singapore
SingHealth

Executive Summary

Heart failure and other cardiovascular diseases is one of the most common reasons for hospital admissions and the second largest cause of death in Singapore, accounting around 32% of all death. To mediate this large endemic in Singapore, it is critical for the country to get ahead of the problem with predictive healthcare analytics and flag individuals at risk for heart failure before it happens to prevent future deterioration.

Predictive Health Group (PHG) is the solution to the problem. We are a group of doctors and scientists who believe that the future of healthcare lies is analytics. PHG has successfully created models that help predict survival rate of breast cancer patients as well as individuals at risk of liver cancer. The latest model that we have created predicts an individuals' potential of heart failure based on four main factors: age, creatinine level, sodium level, and ejection fraction.

Our vision is for our predictive model to be implemented into future health screening. As creatinine level and sodium level is examined through blood test and ejection fraction is examined through electrocardiogram, which are both already included in a typical health screening, we do not foresee any difficulty including our predictive model into future health screening.

National Heart Centre Singapore
SingHealth

Table of Contents

1. Background

1.1 Brief overview of heart failure

Heart failure is a chronic, progressive condition in which the heart muscle is unable to pump enough blood to meet the body's needs for blood and oxygen. Simply put, the heart cannot keep up with its workload (Heart.org).

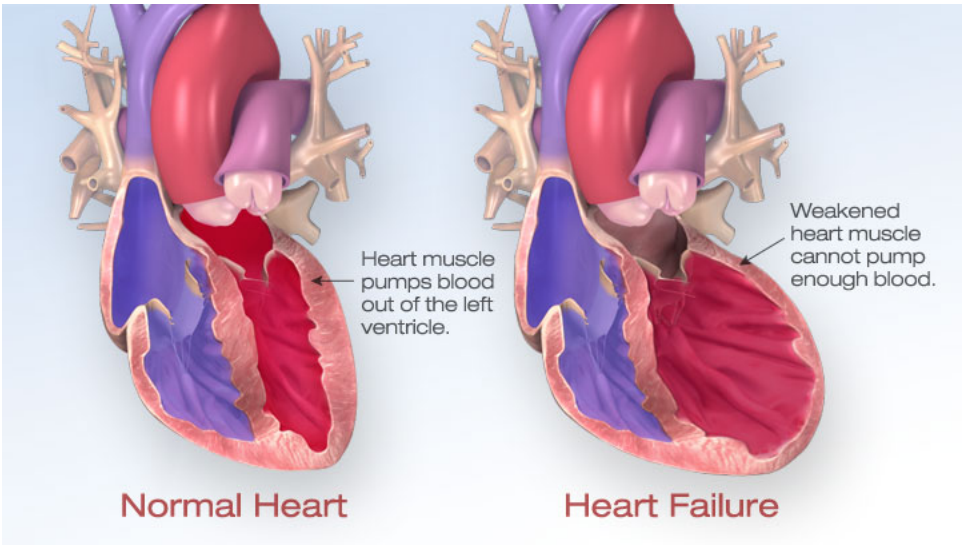


Figure-1: Heart failure

At first the heart tries to make up for this by enlarging, developing more muscle mass and pumping faster. The body also tries to compensate in other ways like narrowing blood vessels to keep blood pressure up and diverting blood away from less important tissues and organs. These temporary measures mask the problem of heart failure, but they don't solve it. Heart failure continues and worsens until these compensating processes no longer work. The body's compensation mechanisms help explain why some people may not become aware of their condition until years after their heart begins its decline.

Heart failure is extremely prevalent in Singapore. About 20% of deaths in Singapore in 2021 were due to ischaemic heart disease (moh) which is one of the most frequent causes of heart failure (Severino et al., 2020). Up to 4.5% of Singaporeans live with heart failure as compared to 1-2% in the US and Europe (Samuel et al., 2022). Though 4.5% may seem like a small percentage, it is one of the most common causes of hospital admissions in Singapore (NUHC).

1.2 Current solution: current heart failure tests

Currently, the diagnosis of heart failure is based on symptoms, physical examination and tests. Tests include but are not limited to blood tests, Chest radiograph (X-ray), electrocardiogram (ECG), echocardiography, cardiac magnetic resonance imaging (CMR), cardiac catheterisation and exercise stress test. However, a typical health screening package in Singapore only covers elementary tests related to heart diseases such as blood tests, X-ray

and ECG (CHAN 2021). For further examination, it's necessary to see a cardiologist. The following prices are ballpark ranges in private cardiology out-patient clinics in Singapore dollars (SGD) :

- First consult – $180-$300 – Price will depend on the length of your consult.
- ECG – An ECG is a simple test to look at the electrical signal of the heart. It is easy and gives the cardiologist a first glance at the health of the heart. Usually it will cost in the region of $60-$80.
- Echocardiography – An echocardiography is a heart ultrasound test. It forms the backbone of heart assessment. The exam takes about 30 minutes, and there is no preparation needed on your part. You lie on a couch and a skilled cardiac technologist will take images of your heart to look at heart and valve structure and function. It will then be reported by your cardiologist. It will cost in the region of $450-$600.
- Treadmill test – In a stress ECG treadmill test, you have ECG leads fixed to your chest with stickers, you then walk on a treadmill. Every few minutes it gets faster and the slope increases to put increasing levels of stress on your body and heart. The price ranges from $350-$500.

1.3 Problem statement

Singapore is seeing an ever-increasing number of patients living with heart failure (nuhcs) and the traditional full diagnosis of heart failure is expensive and time-consuming, preventing people from checking with the doctor. While early detection of heart failure and immediate intervention is more important than treating the disease. Once heart failure develops, there is no cure; only continuous treatments to alleviate symptoms and prevent further worsening, which creates a constant fiscal and emotional drain on society.

1.4 Proposed Solutions

The goal of our team is to create a model that can predict individuals at risk for heart failure, based on specific variables, to be used by healthcare providers during their patients' annual check-up and guarantee the accuracy of our model to acceptable levels. Once a patient is flagged for being at risk, early intervention can be provided.

2. Heart Failure Detection Model

2.1 Introductions to our solution

Based on the above analytics, we can see that heart diseases, especially heart failure is the most common cause of death. Because the disease is more treatable or remittable in its early stages, we believe it makes sense to develop a system that inputs several characteristics of the user to determine the likelihood of heart failure.

We collected some data related to heart failure from Kaggle, and each of the data contains some features. By applying some machine learning models to the above analysis, we were able to obtain some features of the user to determine the likelihood that the patient would develop heart failure. If a patient is judged by the system to be at high risk of heart failure, the physician can recommend further diagnosis to reduce the likelihood of heart failure.

For the practicality of using the model, we eventually want the input features to be obtained from the regular physical examination, such as blood sampling, ECG, or some of the user's own habits and physiological attributes, such as age, etc. We hope that this method will effectively predict the likelihood of heart failure and give early warning to those at high risk, thus maximizing the number of lives saved.

2.2 Data Exploration

One of the most frequently asked questions is that what are the main factors that contribute to heart failure. We assessed the effect of different factors on heart failure by means of data visualization. The images we plotted allow us to observe the important effects of some factors on heart failure. After the analysis of data visualization, we also ranked the importance of all the inputs through a random forest model.

2.2. Age

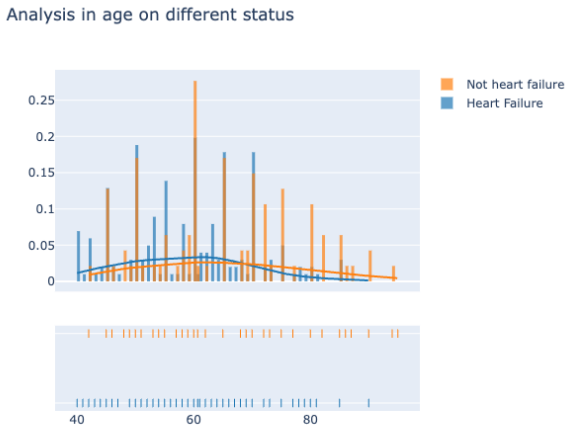


Figure-2: Analysis in age on different status

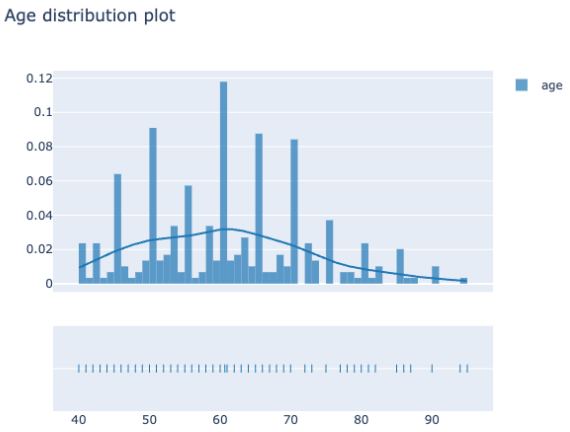


Figure-3: Age distribution

From the figure of age distribution, we can see the age is distributed from 40 to above 90. The spread is high for data ranging between 40 to 80, that of people at age below 40 and greater than 80 is very low. This implies those age above 40 is at high risk. Especially, among the people around age 60, the risk of having heart failure is at the highest level of 45%, which can be seen from the figure of analysis in age on different status.

2.2.2 Gender

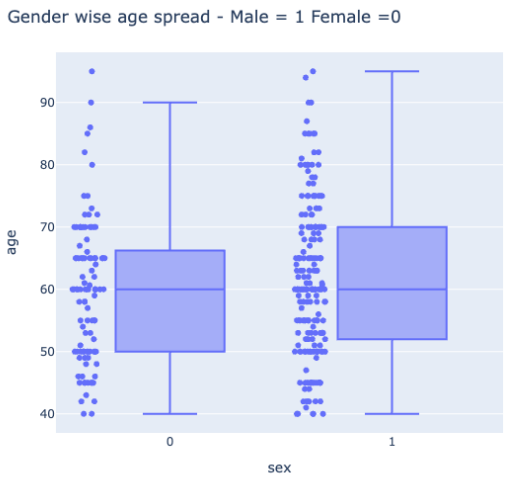


Figure-4: Gender wise age spread

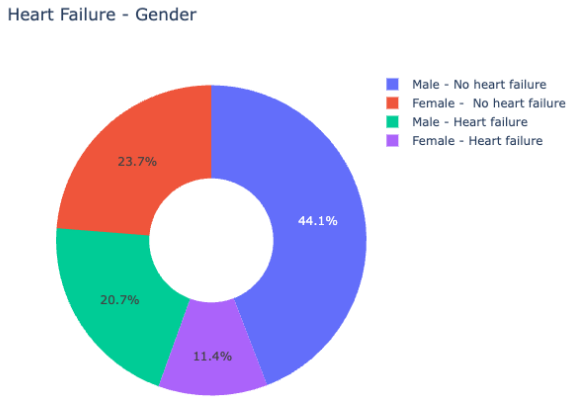


Figure-5: Heart failure – gender

From the above figures, we can see sex is not a dominating factor that causes heart failure. To be more specific, from the pie chart above, among the groups of female and male, the ratios of having heart failure to not having heart failure are both close to 1:2. Therefore, it can be concluded that sex does not make a significant impact among the risk of heart failure.

2.2.3 Smoking

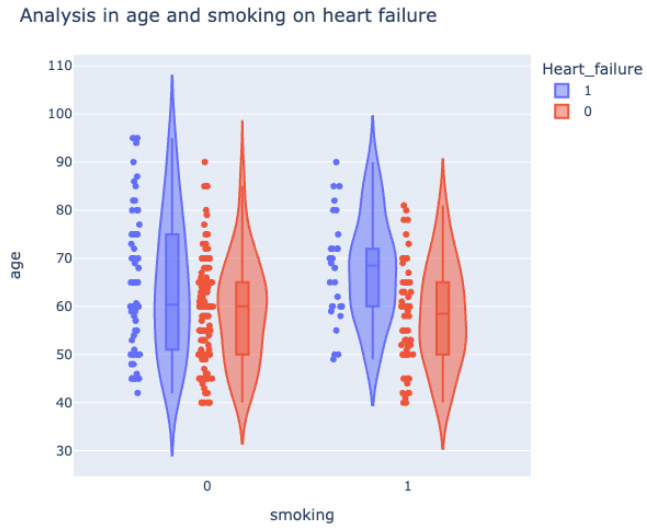


Figure-6: analysis in age and smoking on heart failure

We can observe that smoking habits make an impact on the risk of heart failure. Among two groups of people who have smoking habits and not, smoking habits lead to a shorter life expectancy and the risk of having heart failure is relatively higher among people who do not

smoke. People without smoking habits also tend to be healthier and less risky of having heart failure.

2.2.4 Diabetes

Lastly, in terms of diabetes factor, it shows that having diabetes reduces the age at risk of heart failure. That means having diabetes results in a younger age of having heart failure. People with diabetes have more risks in having heart failure and potentially leads to a lower life quality.

Figure-7: analysis in age and diabetes on heart failure

2.3 Methodology

2.3.1 Overview

The objective is to apply supervised learning algorithm to automatedly and quickly predict whether a given patient have the heart failure, given a list of cardiovascular risk factors. The response variable is "Heart_failure", and there are 11 predictors will be used to train the model after data preparation. "Heart_failure" = 0 is the baseline reference level as we are more concerned about the people who suffer from heart failure ("Heart_failure" = 1). Three models are explored using R:

- Logistic Regression
- CART (Classification And Regression Tree)
- Random Forest

Considering the relatively small dataset, potential overfitting and model computational complexity, we perform feature selection and generate two subset selections of six and four predictors respectively (Table-4). Each model is refitted using the same subset selection and the results are summarized in Table-5. **Recall (1-False Negative Rate)** is important in medical cases, as we are more interested in identifying heart failure patient and we don't want accidentally to discharge/ miss out a patient who have heart failure (false negative). Based on

overall accuracy score and recall, we decided to use **random forest with the most import four features** to identify the segments of the population who suffer from heart failure.

2.3.2 Data preparation

Several preparation processes are completed before analytical modelling to meet the objective:

- Quality check: there are no missing values in the entire dataset
- Data cleaning: drop the variable "times" given unclear definition
- Train-test-split: models are built on train set (80%) and their performance is evaluated on test set
- Oversampling: as aforementioned, the dataset is imbalanced, and we balance the data on the trainset via oversampling the group having heart failure. Model performance test is done on original unbalanced data to reflect the future data distribution, which is more similar to the original data instead of balanced data.

2.3.3 Full logistic regression

Logistic function is suitable for binary target variable, which can serve as probability function of P(Y=1) i.e., P(Heart_failure) = 1 (Have). Based on the dataset after completing the data preparation, execute full logistic regression on all 11 features. We apply the standard threshold 0.5 and predict the heart failure by comparing P(Y=1) against 0.5. The result summary is shown in Appendix-1, there are four predictors that are statistically significant as least at 5% significance level, namely age, ejection_fraction, serum_creatinine and serum_sodium.

- There is a significant positive association between age and risk of heart failure, align with the previous insights (Figure [ ]).
- There is significant negative association between ejection_fraction and heat failure.
- Some statistically insignificant predictors such as diabetes (its sign is negative) and high blood pressure don't provide much information that we could use to estimate heart failure. However, those coefficients should be interpreted with caution given the correlation between the set of predictors.

Comparing the prediction and actual value, we get the Confusion Matrix (Table-1). Given that a patient is actually having heart failure, the logistic model correctly precited heart failure 68.4% of the cases.

Table-1: confusion matrix of full logistic model

	Predict		Rates	
Actual	No Heart Failure	Heart Failure		
No Heart Failure	27	14	0.6585 (TNR)	0.3415(FPR)
Heart Failure	6	13	0.3158(FNR)	0.6842(TPR)

(Note: TNR: true negative rate, FNR: false negative rate, FPR: false positive rate, TPR: true positive rate)

2.3.4 Full CART

CART model scores on transparency and expandability, which is popular among medical scientists because it mimics the way that a doctor think to some extent (Hastie, 2009). The training process:

- Grow the maximal tree, which stratifies the patient into strata of having and not having heart failure, as shown in Appendix-2.
- Using cost-complexity pruning, search for the optimal cp (complexity parameter) region (Figure-8) and select the optimal tree via 10-fold cross validation with 1 SE rule.
- Use the optimal geometric mean cp (0.0611, see Appendix-5) to prune the large tree to the optimal size (3 splits and 4 nodes). The final tree is the simplest tree whose cross-validation error is still within the cross-validation error cap 0.5226 (see Appendix-5).
- The optimal tree achieves the better accuracy score and decreases the false negative rate by around 10% (Table-2). It's consistent with the empirical evidence that the large tree overfit the data.

Table-2: Results of decision tree

	Accuracy	False Negative Rate
Maximal Tree	0.6833	0.2632
Optimal Tree	0.7667	0.1579

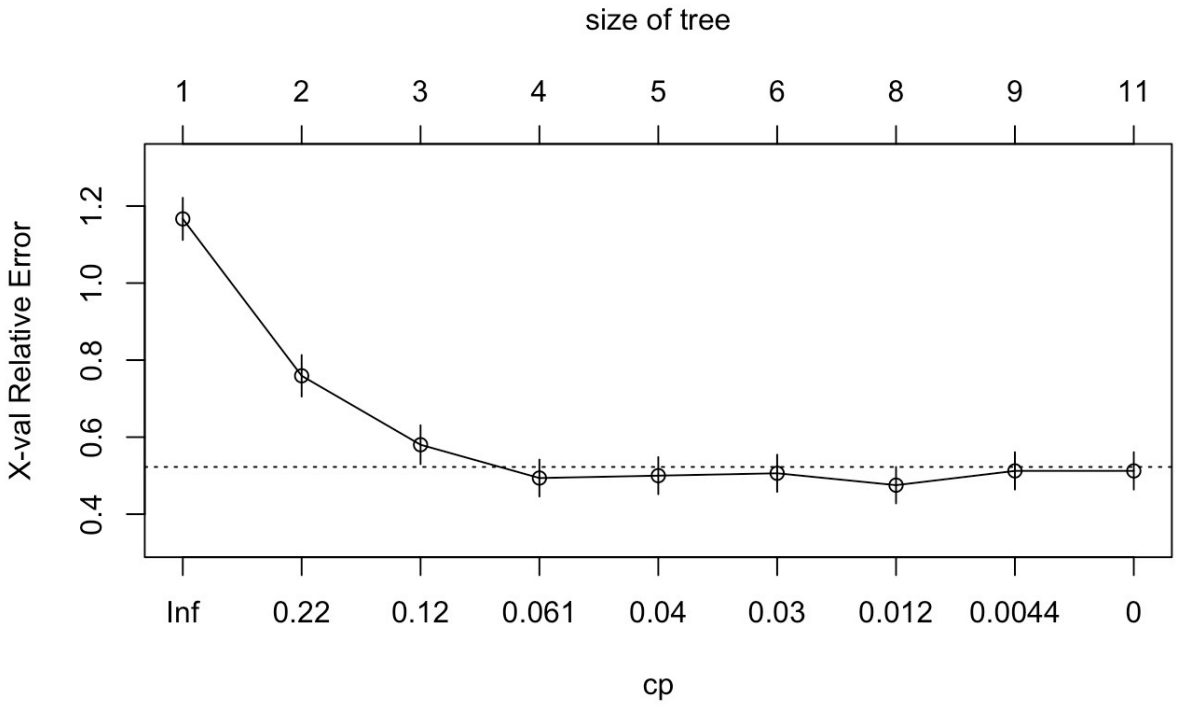


Figure-8: Complexity parameter plot

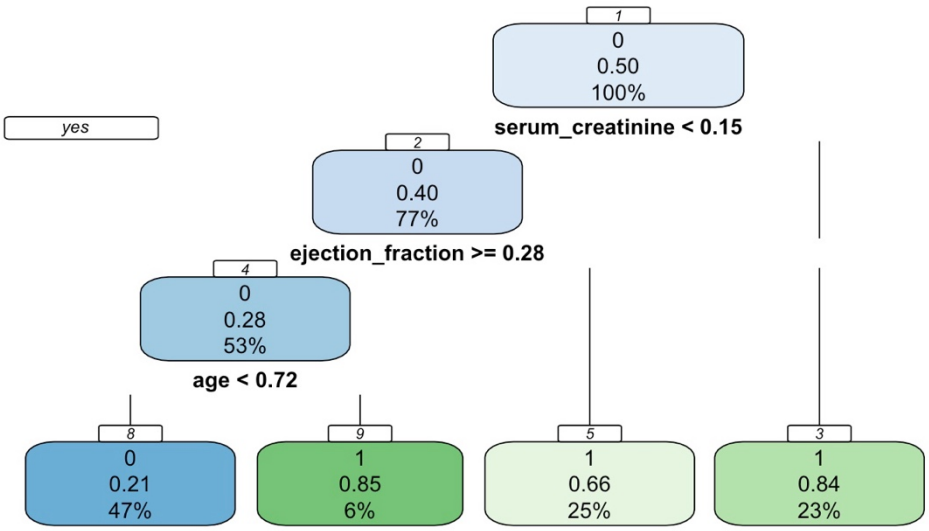


Figure-9: Optimal tree plot

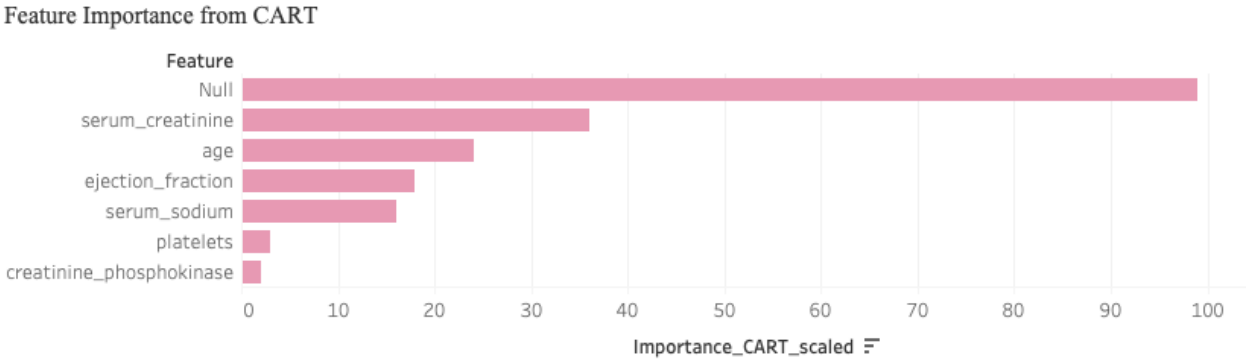


Figure-10: Feature importance from CART

We look at the most important predictors indicated by CART feature importance (Figure-10). We could see that aligning with the insights from logistic regression, serum_creatinine, age, ejection_fraction are the key risk factors of heart failure, which is further backed by the optimal tree that selects those three variables as the best split (Figure-9).

2.3.5 Full random forest

Random forest is an advanced model that applies bagging on many decision trees. The rationale of applying random forest to the dataset is that it typically helpful when predictors are correlated (James, 2017), such as ejection_fraction and serum_sodium for our case (Appendix-3). 500 trees are trained, and the error rate settles down after around 100 trees (Appendix-4). Based on Table-3, random forest decreases false positive rate substantially but doesn't improve the false negative rate. Same as logistic model, given that a patient is actually having heart failure, the logistic model correctly precited heart failure 68.4% of the

cases. Align with logistic regression, age, ejection_fraction, serum_creatinine and serum_sodium, platelets and creatinine_phosphokinase are among the top 6 important predictors indicated by feature importance of random forest (Figure-11).

Table-3: confusion matrix of full random forest

	Predict		Rates	
Actual	No Heart Failure	Heart Failure		
No Heart Failure	37	4	0.9024 (TNR)	0.0976(FPR)
Heart Failure	6	13	0.3158(FNR)	0.6842(TPR)

(Note: TNR: true negative rate, FNR: false negative rate, FPR: false positive rate, TPR: true positive rate)

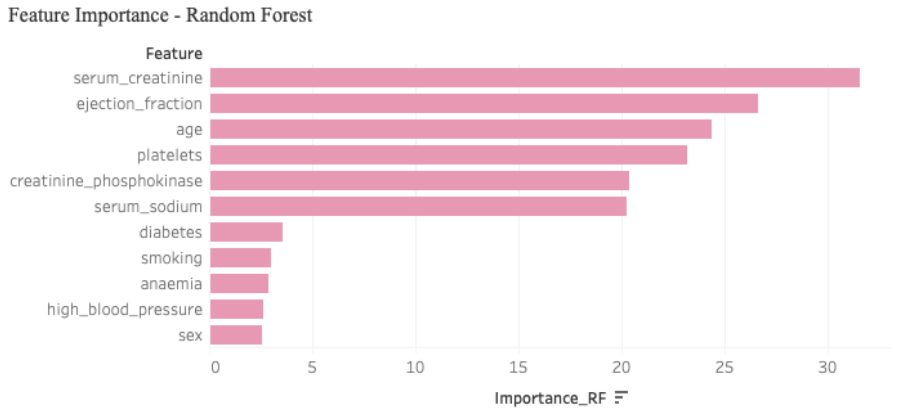


Figure-11: Feature importance from CART

2.3.6 Model selection

The motivation to perform model selection is to find s subset of the variables that are sufficient for explaining their joint effect on the heart failure (Hastie, 2009); decrease model complexity and prevent overfitting. All three models discussed are refitted with the most significant six and four features (Figure-11) and the results are summarised in Table-5. As the number of predictors is reduced form the full set, the change in the overall accuracy isn't remarkable, yet the false negative rate starts to increase (Table-5).

Table-4: Feature selection

Six Feature	age, ejection_fraction, serum_creatinine and serum_sodium, platelets and creatinine_phosphokinase
Four Feature	age, ejection_fraction, serum_creatinine and serum_sodium

Figure-12: Feature importance and selection

2.3.7 Model evaluation

From real-world implementation perspective to detect heart failure, false negative rate (1-recall) is more important. Additionally, the classes are highly imbalance in our situation, and the specificity is not very informative for these problems as it will tend to be high regardless of the quality of the classifier (most people don't have heart failure and classified as such). Hence, we determine the final model based on false negative rate.

Each model uses same number of data given no missing value. Random forest generally delivers better results than logistic regression and CART on the entire dataset or with feature selection. The simplest random forest model, in comparison, has the best generalisation performance on both category of heart failure and the highest overall accuracy score. It further supports the insights that those four factors are among the main determining factors for heart failure, and NHCS is recommended to attach greater importance to them form this.

In conclusion, advise NHCS to use **random forest with four features (age, ejection_fraction, serum_creatinine and serum_sodium)** to identify the heart failure patients.

Table-5: Model result summary

Features	Model	Accuracy	error.rate.false.positive.	error.rate.false.negative.
all	logistic	0.6667	0.3415	0.3158
all	CART	0.7667	0.2683	0.1579
all	random forest	0.8333	0.0976	0.3158
6	logistic	0.6667	0.3659	0.2632
6	CART	0.7667	0.2683	0.1589
6	random forest	0.8333	0.1463	0.2105
4	logistic	0.6667	0.3659	0.2632
4	CART	0.7667	0.2683	0.1579
4	random forest	0.8333	0.1707	0.1579

3. Practical Implementation (deployment)

3.1 Our model in the real world

We propose that NCHS and other SingHealth hospitals include our heart failure prediction model in the health screening analysis and report provided to the public. This will allow individuals at risk for heart failure to be identified early so that preventative measures can be taken to avoid further deterioration in cardiovascular health. Given the considerable cost cardiovascular disease and heart failure has on both individuals and society, it is critical to lower heart failure's burden on society.

From our model in the previous section, we have concluded that some of the key indicators of heart failure include age, creatinine levels, sodium levels, and ejection fraction. Platelets count, creatinine levels and sodium levels can all be determined through conducting a blood test; whereas ejection fraction is detected through conducting electrocardiogram (ECG).

The practical implementation method of our model would be to ensure that when blood tests and ECG is conducted during general health screening, the measurement for platelets count, creatine level, sodium levels and ejection fraction is also identified. Once identified, the physician would input the numbers into our model to predict whether the individual is at-risk of heart failure or not. If the individual is not at-risk of heart failure, then their final health screening report would indicate that they are not at risk. However, if the individual is flagged as at-risk of heart failure, then their report would indicate that they are at risk and follow-up consultation with the physician would automatically be booked. During the consultation the physician and at-risk individual can discuss and come up with a plan to prevent further deterioration of cardiovascular health.

3.2 Current Singapore health screening

Currently in Singapore, a typical health screening package includes the following tests (CHAN 2021):

- Medical health assessment by qualified physician
- Physical examination that takes height, weight, BMI and vision testing
- Blood glucose test
- Blood cholesterol test
- Blood pressure test
- ECG
- Full blood count
- Urine analysis
- Chest X-ray

As blood test and ECG is already conducted during a current typical health screening, the strength of our practical implementation method would be that barely any additional resources will be required from either individuals or the health screening provider, in our case SingHealth. SingHealth just has to ensure that the key indicators for health failure are identified along with the typical tests conducted.

3.3 Implementation constraint

The main constraint with our practical implementation method is that it relies on the Singaporean population conducting annual health checks. While Singaporean currently see a trend of increasing demand for health screening, with a 30 percentage increase in 2021, the older or those with lower socioeconomic status tend to not do health screenings as often.

4. Real World Impact (metrics)

We understand that NHCS is dedicated to providing optimal care and outcomes through continuous advancement in patient care, education and training, and research(NHCS 2022). The model can achieve NHCS's mission in all aspects. Not only can the model detect potential risks and prevent unnecessary costs but also identify key factors for further research and medical achievements. With the model, NHCS can make profound impacts.

4.1 Patient Care

NHCS's implementation of model brings great value to individuals, therefore increasing patient care. To identify risks for heart failure, people only need to do EKG and blood test which are cheap and regular health checks. After classifying the risks, people are given customized instructions for their health conditions.

4.1.1 Low risk in heart failure

People are provided with reliable data-driven evidence to decide whether or not they spend extra money and time for Cardiac Screening Packages. Please refer to 1.2 Current solution for prices for current heart failure tests. The MRI estimated price ranges from S$1,100 to S$2,800 for public hospitals and S$4,500 to S$5,000 for private hospital.

4.1.2 High risk in heart failure

People who are under high risk should do further health checks to monitor their health condition and consult with professionals to prevent heart failure. The model presents good prediction result with 83% accuracy. According to findings from the model, people are convinced to improve their lifestyles and daily habits. For example, they should improve eating habits because serum creatinine presents great importance on the model.

With the improvement in their lifestyles, people are able to prevent heart failure and save the expensive medication expenses which usually cost S$1,100 for public and S$7,700 for private clinic (EVLANOVA 2022). Furthermore, it prevents potential patients and their family from the physical as well as mental suffering that comes with painful treatments.

4.2 Education & Training

Provide solid education and training

When teaching, professors are provided with data-based evidence to support their teaching. The model serves as a preliminary understanding to educate medical students, which forms a better learning process. Furthermore, professors can put more emphasis on discussing treatments for certain groups of patients.

Cultivate future talents

We provide opportunities for medical students interested in data science to intern or shadow with senior team members. With medical students' domain knowledge, we can work together and inspire ideas for more data science possibilities in healthcare applications. NHCS can then apply the results to achieve missions.

Improve clinical training

Physicians are informed of the risks of patients. With the time saved from low-risk patients, they can pay extra time and develop knowledge on the treatment for high-risk patients, increasing their experiences in treating heart failure patients.

4.3 Research

Improve research process

Understanding key factors leading to heart failure, medical researchers are provided with tools to target patients for research, design experiments, and develop treatments for specific groups. They can effectively utilize resources and increase the chances of making achievements.

Improve national health and labor population

Applying the model improves research process and increases research achievements. With successful research, we can significantly improve national health and Singapore labor force. As mentioned, up to 4.5% of Singaporeans live with heart failure as compared to 1-2% in the US and Europe (Samuel et al., 2022). Though 4.5% may seem like a small percentage, it is one of the most common causes of hospital admissions in Singapore (NUHC). The estimated population with heart failure is 250,000 with the estimated medication expenses would be S$1,100 billion (EVLANOVA 2022). The model can prevent the human resources spent on heart failure related medical treatment and protect the labor force population.

National Heart Centre Singapore
SingHealth

Reference List:

CHAN, A. (2021). "How Much Do Health Screenings In Singapore Cost?". from https://www.singsaver.com.sg/blog/cost-of-health-screening-in-singapore.

EVLANOVA, A. (2022). "Average Cost of Cardiovascular Disease Treatment in Singapore." from https://www.valuechampion.sg/average-cost-cardiovascular-disease-treatment-singapore.

Hastie, Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning [electronic resource] : Data Mining, Inference, and Prediction, Second Edition (2nd ed. 2009.). Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Heart.org. "What is Heart Failure?". from https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure.

James, Gareth (Gareth Michael) et al. An Introduction to Statistical Learning: with Applications in R . New York, NY: Springer New York, 2013. Print.

moh. "PRINCIPAL CAUSES OF DEATH." from https://www.moh.gov.sg/resources-statistics/singapore-health-facts/principal-causes-of-death.

NHCS (2022). "Overview – National Heart Centre Singapore." from https://www.nhcs.com.sg/about-us/.

NUHC. from https://m.facebook.com/NUHCS/photos/a.288988538214009/1206592329786954/.

nuhcs. "Heart Failure Programme." from https://www.nuhcs.com.sg/Our-Services/OurCoreClinicalProgrammes/Heart-Failure-Programme/Pages/default.aspx.

Severino, D'Amato, A., Pucci, M., Infusino, F., Birtolo, L. I., Mariani, M. V., Lavalle, C., Maestrini, V., Mancone, M., & Fedele, F. (2020). Ischemic Heart Disease and Heart Failure: Role of Coronary Ion Channels. International Journal of Molecular Sciences, 21(9), 3167–. https://doi.org/10.3390/ijms21093167

National Heart Centre Singapore
SingHealth

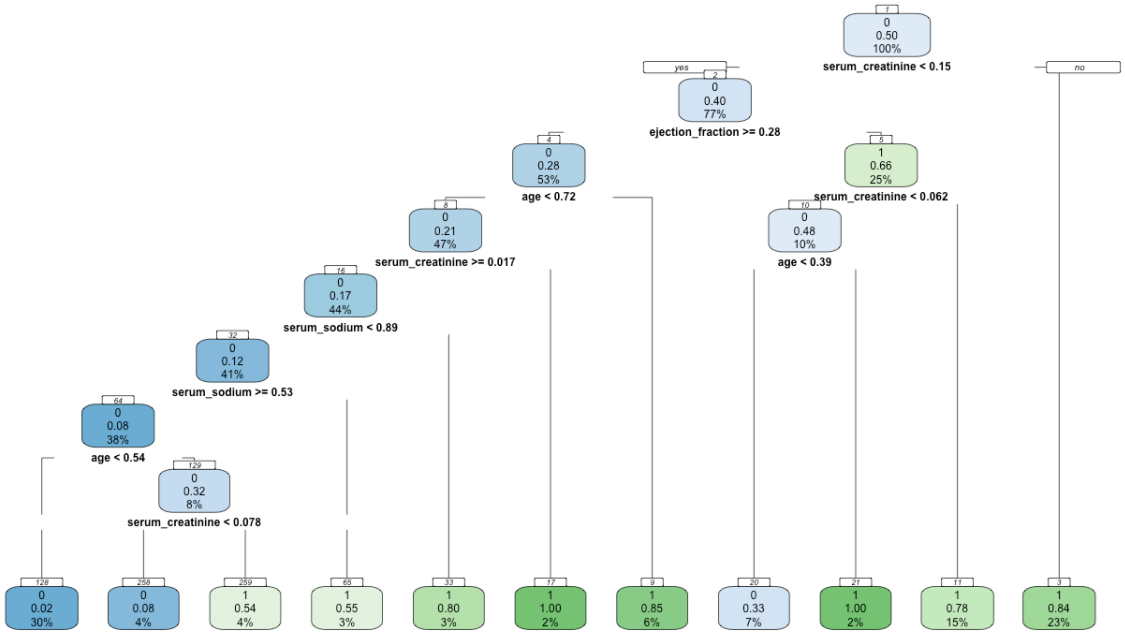
Appendix:

Appendix-1: Result of full Logistic Regression

```
Call:
glm(formula = Heart_failure ~ ., family = binomial, data = trainset)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.45201  -0.93384   0.00861   0.92593   2.08301

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)               0.85824    0.94083   0.912  0.36165
age                       2.97025    0.64249   4.623 3.78e-06 ***
anaemia1                  0.42026    0.27126   1.549  0.12130
creatinine_phosphokinase  1.31292    1.36674   0.961  0.33674
diabetes1                -0.26143    0.27544  -0.949  0.34255
ejection_fraction        -4.51207    0.86029  -5.245 1.56e-07 ***
high_blood_pressure1      0.10657    0.28471   0.374  0.70819
platelets                 1.77242    1.12873   1.570  0.11635
serum_creatinine          4.29825    1.65745   2.593  0.00951 **
serum_sodium             -2.33678    1.08523  -2.153  0.03130 *
sex1                     -0.14586    0.32464  -0.449  0.65321
smoking1                  0.03679    0.30491   0.121  0.90395
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 449.16  on 323  degrees of freedom
Residual deviance: 363.79  on 312  degrees of freedom
AIC: 387.79
```

National Heart Centre Singapore
SingHealth

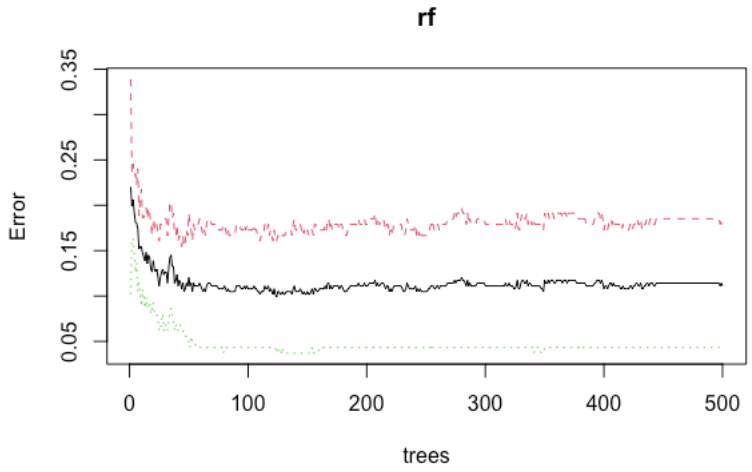
Appendix-2: Plot of full CART



Appendix-3: Predictor correlation (part)

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure
age	1.000000	0.088006	-0.081584	-0.101012	0.060098	0.093289
anaemia	0.088006	1.000000	-0.190741	-0.012729	0.031557	0.038182
creatinine_phosphokinase	-0.081584	-0.190741	1.000000	-0.009639	-0.044080	-0.070590
diabetes	-0.101012	-0.012729	-0.009639	1.000000	-0.004850	-0.012732
ejection_fraction	0.060098	0.031557	-0.044080	-0.004850	1.000000	0.024445
high_blood_pressure	0.093289	0.038182	-0.070590	-0.012732	0.024445	1.000000
platelets	-0.052354	-0.043786	0.024463	0.092193	0.072177	0.049963
serum_creatinine	0.159187	0.052174	-0.016408	-0.046975	-0.011302	-0.004935
serum_sodium	-0.045966	0.041882	0.059550	-0.089551	0.175902	0.037109
sex	0.065430	-0.094769	0.079791	-0.157730	-0.148386	-0.104615
smoking	0.018668	-0.107290	0.002421	-0.147173	-0.067315	-0.055711
time	-0.224068	-0.141414	-0.009346	0.033726	0.041729	-0.196439
Heart_failure	0.253729	0.066270	0.062728	-0.001943	-0.268603	0.079351

Appendix-4: Result of full random forest



Appendix -5: Optimal complexity parameter search

```
> ## automate search for optimal tree
> CVerror.cap = cart_max$cptable[which.min(cart_max$cptable[,"xerror"]),"xerror"]+
+   cart_max$cptable[which.min(cart_max$cptable[,"xerror"]),"xstd"]
> CVerror.cap
[1] 0.5226027
> i=1;j=4
> while (cart_max$cptable[i,j] > CVerror.cap) {
+   print(i)
+   (i=i+1)}
[1] 1
[1] 2
[1] 3
> cp.opt = ifelse(i>1, sqrt(cart_max$cptable[i,1]*cart_max$cptable[i-1,1]),1)
> cp.opt
[1] 0.06110799
```