Leveraging Machine Learning to Explore and Predict the Mortality Risks for Heart Failure Patients

Technical Team Challenge: Group Three

Project Team

Uche Agwuh

Ifeanyi Chukwu

Victoria David

Abigail G-Medhin

Camila Llanes-Kidder

Laban Masunga

Sean Muchenje

Martha Shaka

Evans Zoyah

Contents

1. Introduction

Cardiovascular disease (CVD) is a disorder that impacts the heart. Inclusive of this broad category are heart attacks, strokes and heart failure(Chicco & Jurman, 2020). According to the World Health Organisation (WHO), 17.9 million deaths worldwide result from CVD, accounting for 31% of all deaths (World Health Organisation, 2021). In the UK, there are 163 thousand CVD deaths annually (British Heart Foundation, 2021). Heart failure is a significant cause of illness and mortality in both men and women and affects at least 26 million globally, with increasing prevalence worldwide (A & J, 2003). There have been observed differences in response to treatment and quality care between men and women. Women tend to receive milder medication therapy or are more likely to show more adverse effects (López-Vilella et al., 2021). Individuals of the South Asian population are particularly susceptible to cardiovascular disease due to an excess risk for high blood sugar, high triglycerides, hypertension, or low HDL cholesterol (SA et al., 2016). In 2017 Ahmad *et al.* used a Cox regression model to identify significant variables and their contributing risk. Age is predicted to be the most considerable variable (P-Val = 0.00), with the hazard of death increasing 4% per year. This observation was closely followed by ejection fraction, serum creatinine and anaemia(Ahmad et al., 2017).

Ejection fraction (EF) is the percentage of blood that leaves the heart each time it contracts (squeezes). For example, an ejection fraction of 30% means that 30% of the total amount of blood in the heart's left ventricle is ejected with each heartbeat. It indicates how well the heart pumps with each heartbeat and an average ejection fraction is between 50-70%. (Chuang et al., 2013). The Cox regression produced by Ahmad *et al.* revealed that an EF $\leq$ 30 indicates a lower survival rate, with a hazard rate of 67% and 59%; thus, 30<EF<45 and EF$\leq$ 45, respectively. According to an ongoing cardiovascular cohort study known as the Framingham heart study, EF increases with age (Chuang et al., 2013). Measuring a person's EF is not always a good predictor for heart failure, as a person with a regular EF measurement can still have heart failure and be diagnosed with heart failure with preserved EF (HFpEF). However, some people with heart failure have a markedly reduced EF and are then known to have heart failure with reduced EF (HFrEF). An elevated EF is more common in older heart failure patients. The number of patients with preserved EF in heart failure increases as the general population ages in developed countries (JB & EG, 2012).

Ahmad *et al.* identified that Serum Creatinine was significant in determining CVD risk (p-value = 0.0026). Results indicated that as Serum Creatinine increases, death hazard more than doubles (Ahmad et al., 2017). Such trends have been observed in other studies. In females over the age of 40 years and males over the age of 60 years, it observed that, as we age,

serum creatinine levels steadily increase (JY et al., 2002). Widely used to assess renal function, serum creatinine is measured in mg of creatinine to an mg/dl of blood. Creatine is located within muscles, and when it is broken down, it produces the waste product creatinine. As creatinine is typically eliminated from the body as a waste product and other toxins, if found in high concentrations in the blood, it can indicate that there is kidney damage(A S Levey et al., 2003). Furthermore, elevated SC has shown correlations to an increased risk in CVD mortality. It has been demonstrated that renal dysfunction acts as an independent prognostic marker of CVD mortality, particularly for heart failure and myocardial infarction (MI)(Jose et al., 2006).

Anaemia is an illness that occurs when the body is deficient in the number or quality of red blood cells (RBC) or haemoglobin in the body. Red blood cells function to carry oxygen and haemoglobin around the body. Anaemia is assessed by measuring the percentage of haematocrit level your body is made up of RBCs. Ideal levels range between 41-50% in men and 36-48% in women(Salive et al., 1992). The Cox regression model produced by Ahmad *et al.* showed that anaemic levels could have significant (p-value = 0.0096) contributions to CVD mortality risk, with a 76% increased mortality risk(Ahmad et al., 2017).

Furthermore, age has strong correlations with anaemia, with more significant impacts on men than women after adjusting for demographic characteristics and health status. Therefore, it is said to be due to a decline in haemoglobin levels as a person ages. Causes of anaemia in the older population can be attributed to an arrangement of variables, with chronic diseases and iron deficiency being the most common. However, vitamin B12 deficiency, folate deficiency, gastrointestinal bleeding and myelodysplastic syndrome can also lead to anaemia(Smith, 2000).

Objective

Using the data from the Ahmad *et al.* study, the primary objective of this review is to reanalyse the differences in mortality between males and females with heart failure in a South Asian population and the impact of time. Therefore, using the data and results collected, a model is created to identify predictive death variables. The findings of this report could have implications for the personalisation of care for patients belonging to intersections of society.

2. Data Analysis

The dataset is composed of numerical and categorical. It has six numerical values: age, creatinine_phosphokinase, platelets, serum_creatinine, serum_sodium, and time. And anaemia, diabetes, high_blood_pressure, sex and smoking as categorical values. Figure 1 below displays the first five columns of the dataset

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
0	75.0	0	582	0	20	1	265000.00	1.9	130	1	0	4	1
1	55.0	0	7861	0	38	0	263358.03	1.1	136	1	0	6	1
2	65.0	0	146	0	20	0	162000.00	1.3	129	1	1	7	1
3	50.0	1	111	0	20	0	210000.00	1.9	137	1	0	7	1
4	65.0	1	160	1	20	0	327000.00	2.7	116	0	0	8	1

Figure 1. The first five-column of the dataset with values.

2.1 Preliminary Data Exploration:

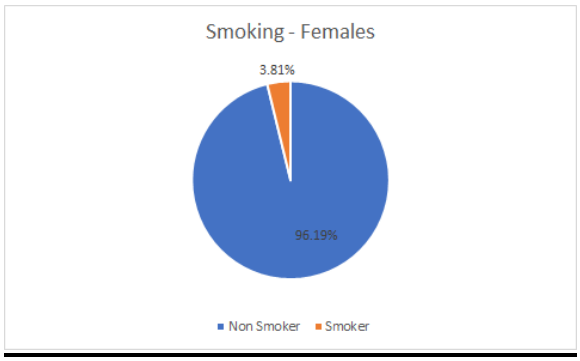
The tables below provided summary statistics for the variables assessed within the dataset, focusing on those that showed the significant impact. The results are measured for 285 days.

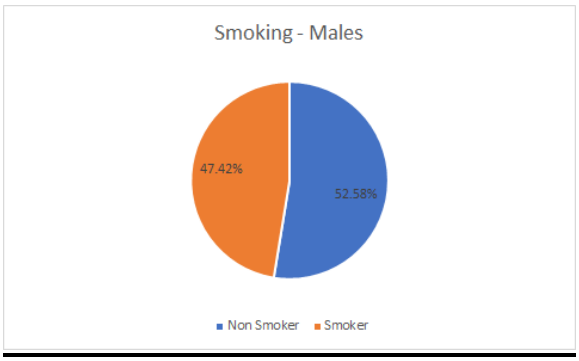
Summary statistics	Age years	Creatinine Phosphokinase units/L	Ejection Fraction %	Serum Creatinine mg/dl
Minimum	40.00	23.00	14.00	0.50
Medium	60.00	250.00	38.00	1.10
Mean	60.83	581.80	38.08	1.39
Maximum	95.00	7861.00	80.00	9.40

2.2 Data Composition about no. Male vs Female

Smoking

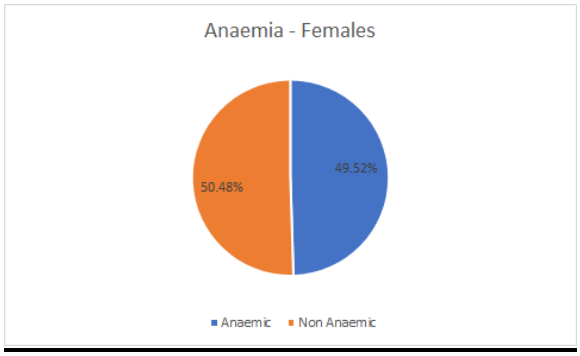
In this dataset, nearly all females did not smoke. Roughly 50/50 men smoked.

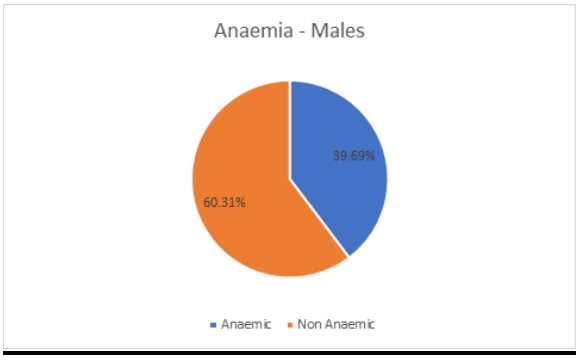




Anaemia

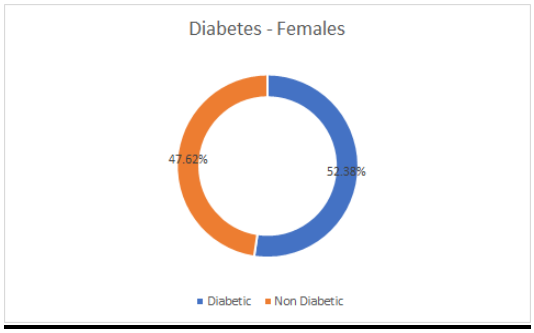
Anaemia was equally distributed amongst females. Whereas in Males, there were more who were non-anaemic.

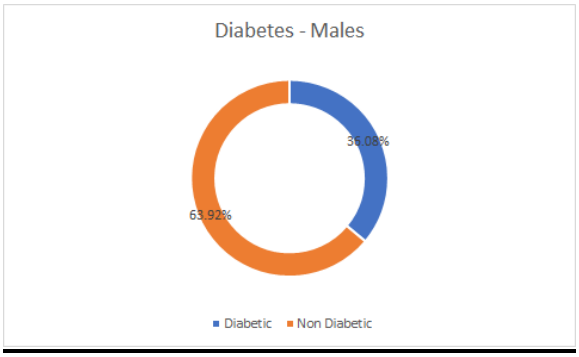




Diabetes

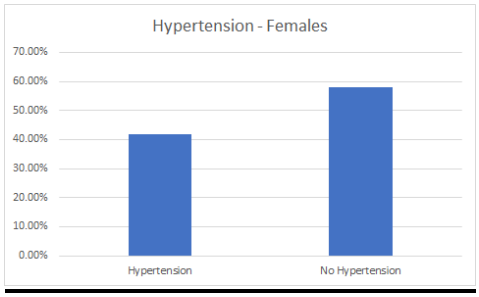
The number of females who had diabetes were equally distributed. However, more males tended not to be diabetic - Almost 2 in 3 Males.

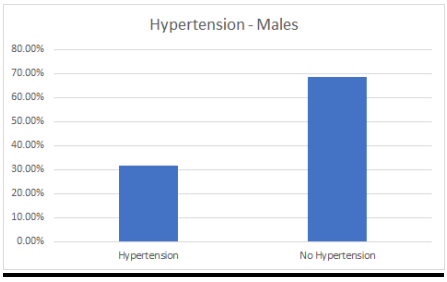




Hypertension

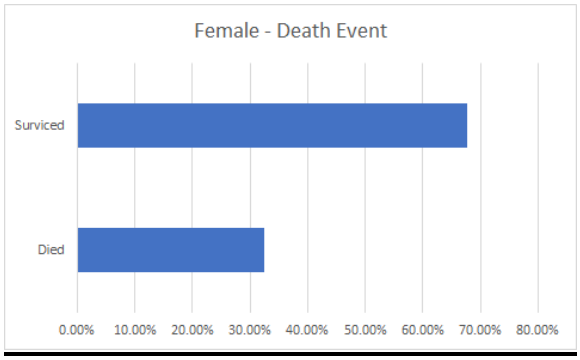
There was a slight majority of females that did not have high blood pressure. Additionally, most males did not have high blood pressure.

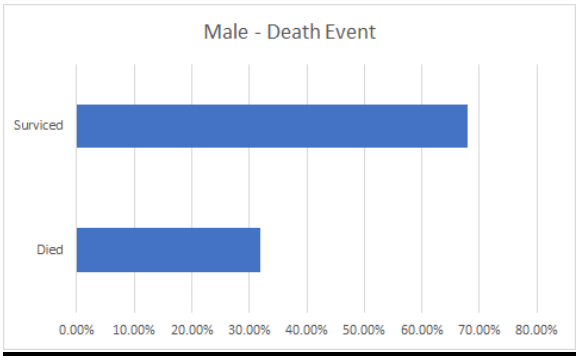




Death Event

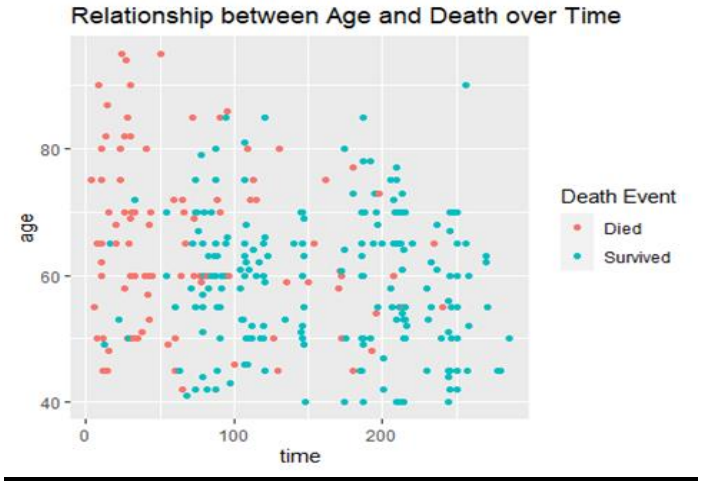
Gender did not affect the death event. Roughly 68% of both males and females survived across the dataset.





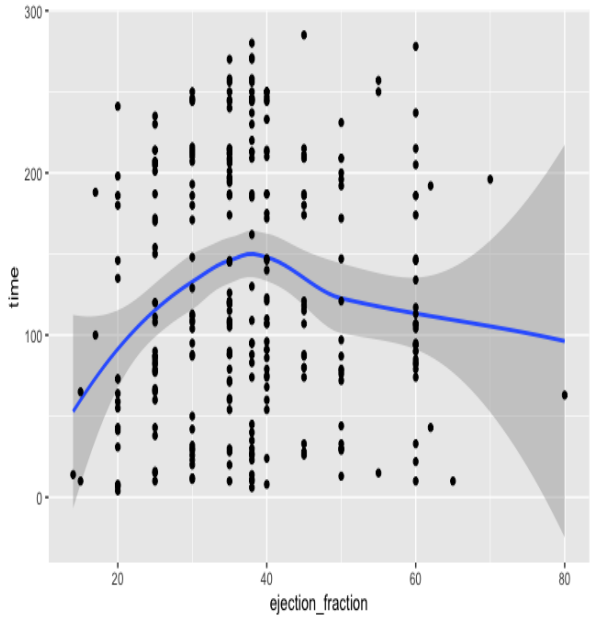
2.3 Age + Death Over Time

Overall, those that died tended to die earlier in the timeline. It could suggest that the longer time went on, the more likely you were to survive. There was no direct pattern with age in this instance. However, the plots are trending (loosely) negatively, so this could suggest the older people tended to die sooner rather than later.



2.4 Ejection fraction versus time spent in study

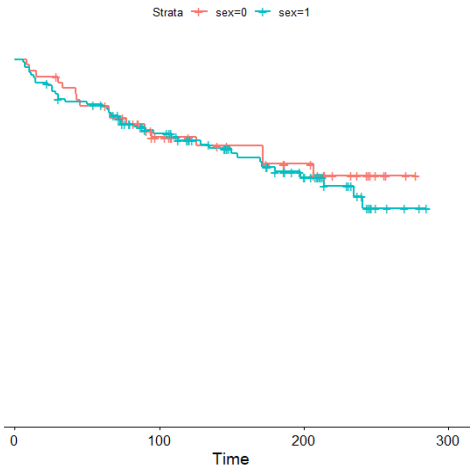
Participants with an ejection fraction of around 40 per unit appear to have the longest time in the study. Beyond 40, increasing ejection fraction appears to be associated with decreasing time spent in the study.



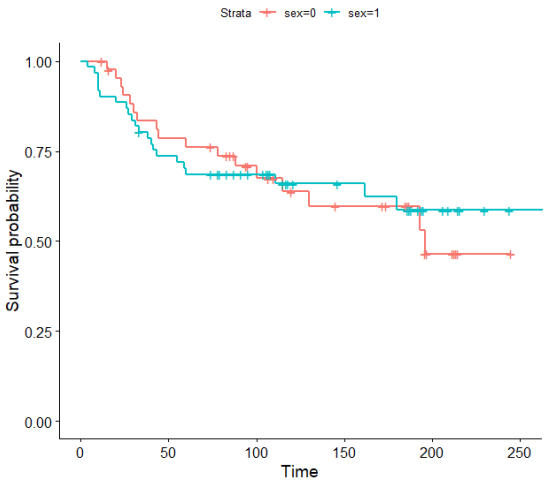
2.5 Case 1

The dataset was partitioned based on BP, which was the only variable showing the significant difference between those who experienced versus those who didn't. Then each set was assessed individually based on sex, the results didn't show any significant difference, implying that female (coded as 0) and male (coded as 1) in both cases didn't have any difference.

Without BP (p-value = 0.6)

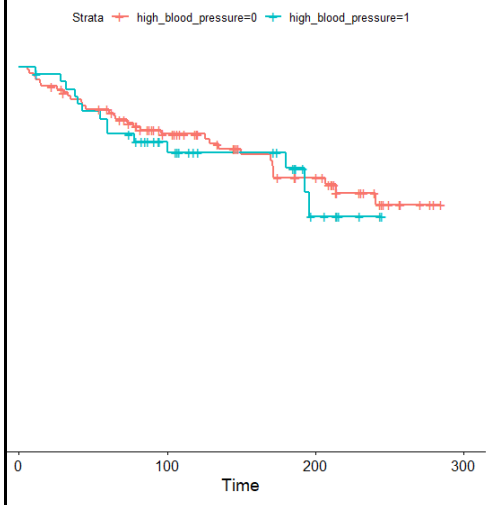
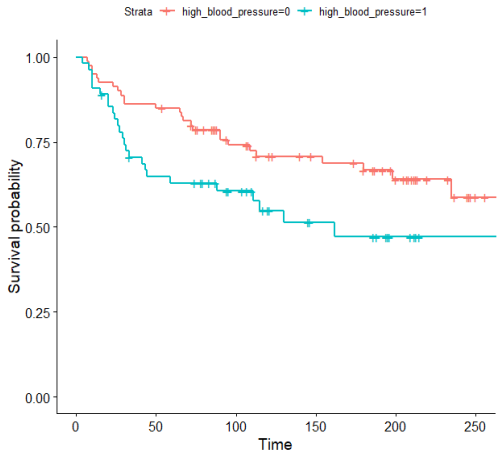


With BP (p-value =0.9)



2.6 Case 2

Reviewing the correlation between blood pressure (BP) and age, it was noted that those under the mean sample age of 60.8 years showed no significant difference (at the 5% significance level) between high and average BP in correlation to survival probability. However, those over the mean sample age of 60.8 years did indicate a significant difference (at the 5% significance level) in survival between those with high and average BP. That means the significant difference between those who experienced BP and the ones who didn't lie between the older aged population in the sample.

Below mean age (p-value = 0.7)	Equal & above mean age (p-value = 0.03)
	

2.7 Case 3

Below is the Cox model with all predictors onboard but by assessing the p-values the ones with red dots are all insignificant.

	coef	exp(coef)	se(coef)	z	p	
high_blood_pressure1	4.757e-01	1.609e+00	2.162e-01	2.201	0.0278	
age	4.641e-02	1.048e+00	9.324e-03	4.977	6.45e-07	
platelets	-4.635e-07	1.000e+00	1.126e-06	-0.412	0.6806	·
creatinine_phosphokinase	2.207e-04	1.000e+00	9.919e-05	2.225	0.0260	
serum_sodium	-4.419e-02	9.568e-01	2.327e-02	-1.899	0.0575	..
anaemia1	4.601e-01	1.584e+00	2.168e-01	2.122	0.0338	
diabetes1	1.399e-01	1.150e+00	2.231e-01	0.627	0.5307	·
ejection_fraction	-4.894e-02	9.522e-01	1.048e-02	-4.672	2.98e-06	
serum_creatinine	3.210e-01	1.379e+00	7.017e-02	4.575	4.76e-06	
sex1	-2.375e-01	7.886e-01	2.516e-01	-0.944	0.3452	·
smoking1	1.289e-01	1.138e+00	2.512e-01	0.513	0.6078	·

A backward stepwise selection was performed. The results were surprising because initially, serum_sodium was partially insignificant but after removing all other predictors which were

highly insignificant, its performance improved a bit close to the margin (at 5% significance level) which is reasonable to be carried forward in the modelling process. The final model only with variables that have an effect on survival time can be examined as summarised below;

```
                            coef  exp(coef)   se(coef)      z        p
high_blood_pressure1      4.965e-01  1.643e+00  2.137e-01  2.324   0.0201
age                       4.357e-02  1.045e+00  8.831e-03  4.934 8.05e-07
creatinine_phosphokinase  2.101e-04  1.000e+00  9.825e-05  2.138   0.0325
serum_sodium             -4.569e-02  9.553e-01  2.336e-02 -1.956   0.0505
anaemia1                  4.460e-01  1.562e+00  2.150e-01  2.074   0.0380
ejection_fraction        -4.747e-02  9.536e-01  1.027e-02 -4.621 3.82e-06
serum_creatinine          3.139e-01  1.369e+00  6.895e-02  4.552 5.31e-06

Likelihood ratio test=80.58  on 7 df, p=1.048e-14
n= 299, number of events= 96
```

3. Model Prediction and Analysis

3.1 Overview

This section will describe the preparation of the dataset and create a machine learning model that can estimate the likelihood of death due to heart failure events. This estimation is useful to hospitals in assessing the severity of patients with cardiovascular diseases. The study explored two sets of experiments. The first experiment investigated different machine learning models that best fit the selected dataset, while the second experiment investigated the model performance based on gender.

3.2 Choosing the evaluation metric

The data exploration analysis shows that the target class of predicting if someone will survive or die of heart failure has an imbalance, with 68% of the data being a negative class, and 23% is positive. Therefore, it is not feasible to use accuracy as an evaluation metric since it does not consider recall and precision. For example, when a trained model consistently predicts someone will die, it will result in an accuracy of above 68%, which is relatively high for a model that can't predict a positive class outcome. As a result, we have used f1-score as an evaluation metric that considers the recall and precision. Also, we have used the confusion matrix to evaluate the mistakes of our trained models.

3.3 Data preparation

The data processing for the numerical values required scaling so that the dataset can be represented in one scale range fit the dataset well as required by some classifiers such as SVM. Therefore, the numerical pipeline transformation involved finding the polynomial

features, standard scaling, and a simple LogisticRegression to determine the best 12 features. On the other hand, the categorical values were represented using One_Hot encoding. All the transformation is implemented using the Sklearn library. After data processing, the final number of features is 22. The final dataset was randomly shuffled and split into training and test set at a ratio of 80:20.

3.4 Model Training

In this project, we trained and evaluated the performance of six machine learning classifiers. The first step is to train the models on the training set and predict the probabilities and the target class on the test set. After predicting the possibilities, we calculated the False positive rates, True positive rates, and AUC scores while the predicted labels calculate the recall, f1 score, precision and accuracy.

4. Results and discussion

After running several experiments, the final results show that the Random Forest Classifier has the best performance with an F1 score of 0.89. In contrast, Support Vector Machine (SVC) has a worse performance score of 0. An interesting result is using accuracy as an evaluation metric, which shows that SVC with an accuracy of 70% thus outperforms the K-Nearest Neighbour (KNN) classifier with a 60% accuracy, as shown in figure 2. However, when looking at the error analysis from the prediction using the confusion matrix, SVC has more mislabel 18 patients by predicting that they will die will they survive the heart attack. This error indicates that the model cannot generalise well with the dataset and learn how to predict the positive outcome of patients compared to the KNN classifier that can at least predict three patients who will survive correctly, as figure 3. This type of misleading performance using accuracy metrics is anticipated because of the class imbalance in our dataset.

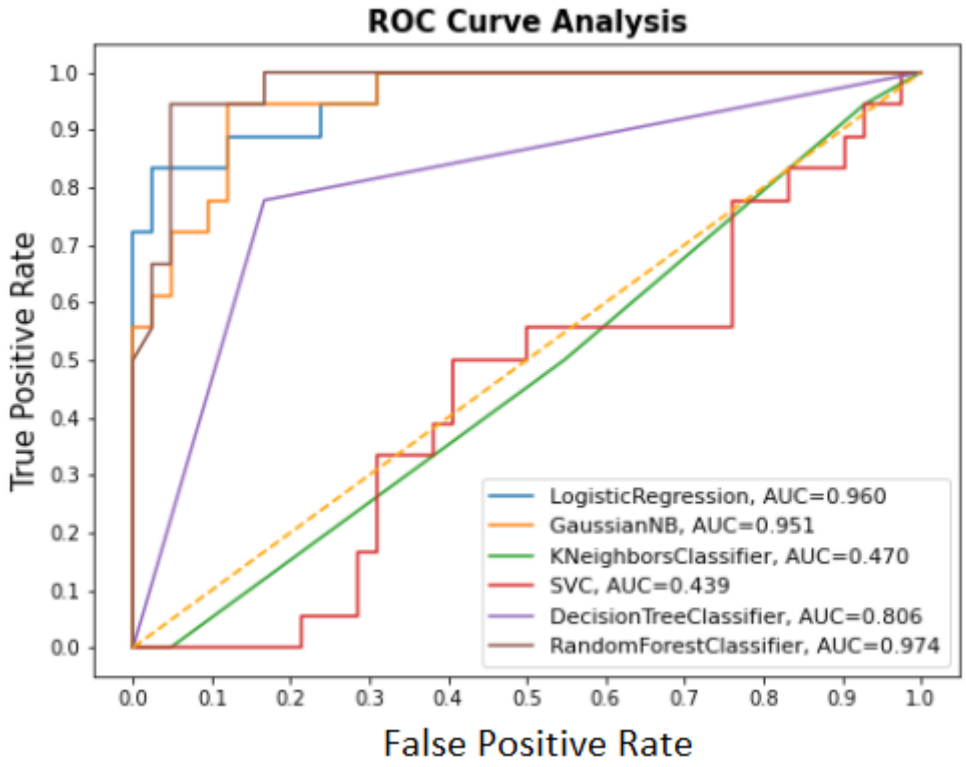
classifiers	accuracy	precision	recall	f1score
LogisticRegression	0.883333	0.761905	0.888889	0.820513
GaussianNB	0.866667	0.812500	0.722222	0.764706
KNeighborsClassifier	0.600000	0.250000	0.166667	0.200000
SVC	0.700000	0.000000	0.000000	0.000000
DecisionTreeClassifier	0.816667	0.666667	0.777778	0.717949
RandomForestClassifier	0.933333	0.850000	0.944444	0.894737

Figure 2: A tabular result of the six classifiers on the test set.

Figure 3 shows the confusion matrix for all six classifiers.

Therefore, the best evaluation matrix is f1-score that has considered both the recall and precision of the model. As a result, the Random Forest Classifier indeed performs better in the classification of the two classes and makes fewer mistakes. Also, using the ROC curve analysis stills shows that the Random Forest classifier performs the best and SVC the worse since it has the smallest Area Under the Curve (AUC).

Furthermore, patients need to be correctly classified in this kind of project because such predictions can help to save lives. Compared to the misclassification of healthy participants as patients, the worst that can happen is undergoing more tests.



To analyse the model prediction performance based on gender, we created a separate test set for males and females. The evaluation of the ROC and Confusion matrix is as shown in Figures 3 and 4, respectively. The results of the seconds' experiment show that the trained random forest model has higher performance on a test set composed of male patients compared to a test set of all female patients. Thus, this suggests the presence of biases in the dataset.

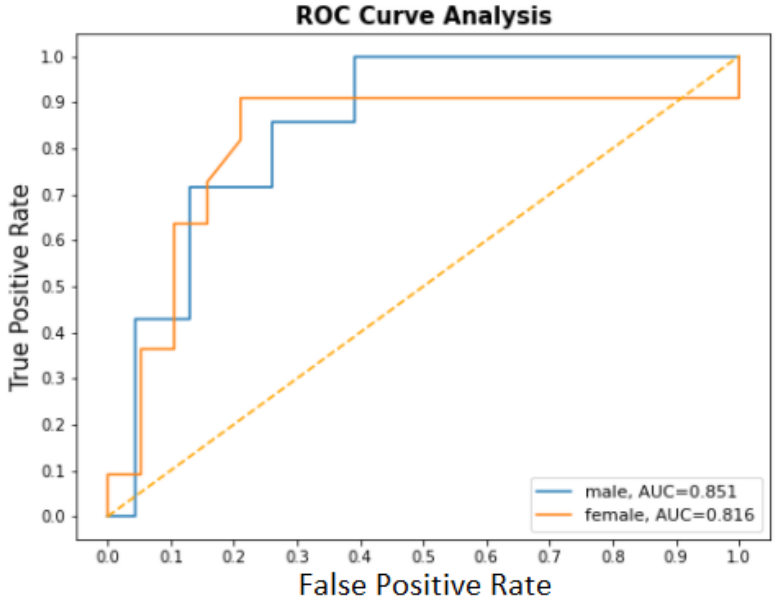


Figure 3. Shows the ROC of both female and male test set

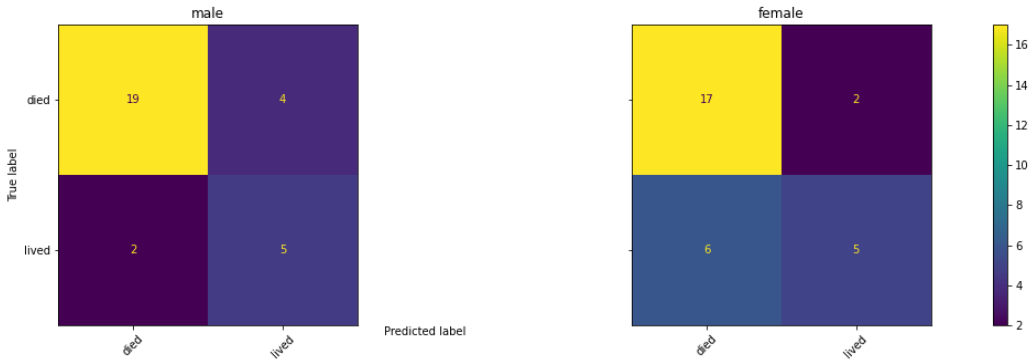


Figure 4. Shows the confusion matrix of both female and male test set

5. Discussion and Conclusion

The study focuses on the clinical factors and health outcomes of patients with heart failure and tries to describe the association between the variables: gender, age, high blood pressure, anaemia, ejection fraction and mortality. Analysis of this dataset has highlighted that despite the disparity in lifestyle factors and health parameters affecting patient outcomes in heart failure, the sex of a patient was not a determining factor for mortality in heart failure. However, patients with an ejection fraction above 40%, also known as heart failure with preserved ejection fraction (HFpEF), spent less time in the study. This could highlight that patients with HFpEF were more likely to be too ill to continue in the study and thus could identify an unmet need in the healthcare received by patients with HFpEF. However, death or exit from the study was not distinguished in measurement for length of time in the study. Older people with heart failure had a poorer prognosis as they died earlier from the time of their heart failure diagnosis, compared to younger people with heart failure. An explanation for this could be a higher incidence of frailty in this age and thus poor resilience to acute and chronic disease.

The study also shows that using the critical heart failure risk indicators; we can train a machine learning model to predict the likelihood of death due to heart failure events. First, however, it is crucial to collect a balanced dataset, and secondly, the dataset should be unbiased for the model to be used by the community.

References

A, S., & J, M. (2003). Gender differences in patients with heart failure. *European Journal of Cardiovascular Nursing : Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, *2*(1), 7–18. https://doi.org/10.1016/S1474-5151(03)00002-1

A S Levey, R D Perrone, and, & Madias, N. E. (2003). Serum Creatinine and Renal Function. *Https://Doi.Org/10.1146/Annurev.Me.39.020188.002341*, *39*, 465–490. https://doi.org/10.1146/ANNUREV.ME.39.020188.002341

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. https://doi.org/10.1371/JOURNAL.PONE.0181001

British Heart Foundation. (2021). *Facts and figures - Information for journalists | BHF*. https://www.bhf.org.uk/what-we-do/news-from-the-bhf/contact-the-press-office/facts-and-figures

Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*, *20*(1), 1–16. https://doi.org/10.1186/s12911-020-1023-5

Chuang, M. L., Gona, P., Hautvast, G., Salton, C. J., Breeuwer, M., O'Donnell, C. J., & Manning, W. J. (2013). Association of age with left ventricular volumes, ejection fraction and concentricity: the Framingham heart study. *Journal of Cardiovascular Magnetic Resonance 2013 15:1*, *15*(1), 1–2. https://doi.org/10.1186/1532-429X-15-S1-P264

JB, S., & EG, L. (2012). Aging-associated cardiovascular changes and their relationship to heart failure. *Heart Failure Clinics*, *8*(1), 143–164. https://doi.org/10.1016/J.HFC.2011.08.011

Jose, P., Skali, H., Anavekar, N., Tomson, C., Krumholz, H. M., Rouleau, J. L., Moye, L., Pfeffer, M. A., & Solomon, S. D. (2006). Increase in Creatinine and Cardiovascular Risk in Patients with Systolic Dysfunction after Myocardial Infarction. *Journal of the American Society of Nephrology*, *17*(10), 2886–2891. https://doi.org/10.1681/ASN.2006010063

JY, T., JB, S., JR, M., & MM, L.-B. (2002). The effect of age on serum creatinine levels in an aging population: relevance to vascular surgery. *Cardiovascular Surgery (London, England)*, *10*(5), 445–451. https://doi.org/10.1016/S0967-2109(02)00056-X

López-Vilella, R., Marqués-Sulé, E., Laymito Quispe, R. del P., Sánchez-Lázaro, I., Donoso

Trenado, V., Martínez Dolz, L., & Almenar Bonet, L. (2021). The Female Sex Confers Different Prognosis in Heart Failure: Same Mortality but More Readmissions. *Frontiers in Cardiovascular Medicine*, *0*, 130. https://doi.org/10.3389/FCVM.2021.618398

SA, P., R, S., MK, A., RM, A., M, D., D, K., D, K., G, R., V, M., KM, N., MM, K., Z, F., D, P., & N, T. (2016). Is the “South Asian Phenotype” Unique to South Asians?: Comparing Cardiometabolic Risk Factors in the CARRS and NHANES Studies. *Global Heart*, *11*(1), 89-96.e3. https://doi.org/10.1016/J.GHEART.2015.12.010

Salive, M. E., Cornoni-Huntley, J., Guralnik, J. M., Phillips, C. L., Wallace, R. B., Ostfeld, A. M., & Cohen, H. J. (1992). Anemia and Hemoglobin Levels in Older Persons: Relationship with Age, Gender, and Health Status. *Journal of the American Geriatrics Society*, *40*(5), 489–496. https://doi.org/10.1111/J.1532-5415.1992.TB02017.X

Smith, D. L. (2000). Anemia in the Elderly. *American Family Physician*, *62*(7), 1565–1572.

World Health Organisation. (2021). *Cardiovascular diseases*. https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1