The prediction of heart failure on the basis of the Prothrombin Time, Ejection Fraction, Serum Creatinine, etc.

Awnish Shankar

Abstract

Heart failure is a life-threatening cardiovascular ailment that has a major effect on global public health. Effective preventative approaches and timely therapies depend on early detection of people at risk of developing heart failure. In order to determine the likelihood of developing heart failure, this study investigates the predictive value of some of the important clinical markers like serum creatinine (SC), ejection rate (ER) , and prothrombin time (PT). Using the model prediction, we can predict whether the subject is prone to heart failure.

Background/ Literature Review

Heart failure is a clinical illness brought on by the heart's inability to pump enough blood to support venous return or meet metabolic needs. Globally, cardiovascular diseases cause myocardial infarctions and heart failure, which account for over 17 million deaths each year. In the **US**, it affects more than 5 million people. Ischemic heart disease, hypertension, diabetes, cardiomyopathies, valve disease, infections, toxins, and medications are some of the common causes. Dyspnoea (difficulty in laboured breathing), weariness (added stress on the heart), nausea, and palpitations, rales, gallop, oedema, and an enlarged liver are symptoms. Classified by NYHA(New York Heart Association) functional class I-IV based on severity of symptoms or ACC/AHA stages [1] A-D based on disease progression. Reduced cardiac output results from impaired contraction in systolic dysfunction. In systolic dysfunction, impaired contraction leads to reduced cardiac output. In diastolic dysfunction, impaired relaxation reduces ventricular filling. Compensatory mechanisms initially help maintain blood pressure but eventually worsen heart failure: Frank-Starling mechanism, neurohormonal activation (RAAS- a hormonal system that helps in regulating blood pressure, electrolyte balance and fluid volume , SNS- Sympathetic nervous system responsible for “fight and flight” mechanism, and control heart rate), ventricular remodelling. Assessments include ECG, chest x-ray, echo, labs. Treatments include lifestyle changes, medications (diuretics, ACE inhibitors, beta blockers), devices (CRT- pacemaker like devices), surgery (Coronary Artery Bypass Grafting, SurgicalVentricular Restoration, Ventricular Assist Device, transplant). Despite understanding mechanisms and treatments, heart failure has a 50% mortality chance within a span of 5-years. [2][3]

In this literature review we are trying to examine the correlation between different parameters like Prothrombin, Ejection fraction, Serum Creatine and other predictors which can help us predict heart rate failure. But in order to do that, first we need to understand how these parameters are related to heart failure. **Ejection fraction** value is the proportion of blood pumped out of the heart during a single contraction, given as a percentage with physiological values ranging between 50% and 75%. Two types of heart failure can be based on ejection fraction. One is heart failure due to reduced ejection fraction (HFrEF), previously known as heart failure due to left ventricular (LV) systolic dysfunction or systolic heart failure and characterised by an ejection fraction smaller than 40% and second one is heart failure with preserved ejection fraction (HFpEF), formerly called diastolic heart failure or heart failure with normal ejection fraction. The left ventricle contracts normally during systole, but the ventricle is stiff and fails to relax normally during diastole, thus impairing filling. On the other hand, **Serum creatinine** is a waste product generated by creatine, when a muscle breaks down. Especially, doctors focus on serum creatinine in blood to check kidney function. If a patient has high levels of serum creatinine, it may indicate renal dysfunction. Now, Sodium is a mineral that serves for the correct functioning of muscles and nerves. The serum sodium test is a routine blood exam that indicates if a patient has normal levels of sodium in the blood. An

abnormally low level of sodium in the blood might be caused by heart failure. In a study[4], 100 patients were included with congestive heart failure. Ejection fraction was measured by echocardiography. Liver function was assessed by tests for bilirubin, AST, ALT, and alkaline phosphatase. Kidney function was assessed by blood urea and creatinine levels. 55% of patients had elevated bilirubin levels. 92.5% had elevated AST and ALT. 12% had elevated alkaline phosphatase. These abnormalities were more common in patients with ejection fraction ≤40%. For kidney function, 57.5% had elevated urea and 62.5% had elevated creatinine in the ejection fraction ≤40% group. Kidney dysfunction was less common in the ejection fraction >40% group. The results show that as ejection fraction decreases in heart failure, indicating worse pump function, there are parallel increases in liver enzyme elevations and kidney dysfunction. The authors conclude that liver and kidney tests can help assess the severity and duration of heart failure. Treating the underlying heart failure can potentially reverse liver and kidney abnormalities. Regular monitoring of these organ functions is recommended in heart failure patients. In summary, this study demonstrates correlations between worsening ejection fraction, liver injury, and kidney dysfunction in heart failure patients. As heart pump function worsens, damage to other organs is more likely to occur. [4]
So, to understand the Heart failures in detail and prediction, there has been a systematic review of studies published between March 2013 and May 2018 that developed risk prediction models for heart failure outcomes in adults [3]. 40 studies were identified that reported on 58 distinct risk prediction models with sample sizes ranging from 43 to 33,349 patients. The most commonly predicted outcomes were all-cause mortality (17 models), cardiovascular mortality (9 models), heart failure hospitalisation (15 models), and composite endpoints (14 models). The 12 most commonly used predictors were **NT-proBNP** (N-terminal pro-brain natriuretic peptide.**), age, diabetes, male sex, systolic blood pressure, blood urea nitrogen, creatinine, heart rate, left ventricular ejection fraction, sodium, BMI** (Body Mass Index)**, and NYHA class** [5]. Model discrimination was generally moderate or good, with C-statistics ranging from **0.59 to 0.84** for the major (C-statistic of 0.5 indicates the model has no discrimination, while 1.0 is perfect discrimination. The models in this review generally had moderate (C-statistic 0.6-0.7) to good (C-statistic >0.7) discrimination for the major outcomes like mortality.). Calibration and validation were less frequently reported. Only 11 models (19%) were rated as low risk of bias based on the **PROBAST** (Prediction model Risk Of Bias Assessment Tool)) tool. The majority had concerns around validation, calibration, and handling of missing data.
In another research paper[7], where the objective is to develop models for predicting mortality and heart failure (HF) hospitalisation for outpatients with HF with preserved ejection fraction (HFpEF) and emphasise the importance of health status data in determining prognosis. This study doesn't cover HF with reduced ejection fraction (HFrEF) as few such cases have assessed the risks of death and hospitalisation in patients. Following 5 modelling methods were used to train models for assessing risks of mortality and hospitalisation. They are logistic regression with a forward selection of variables, logistic regression with a lasso regularisation for variable selection, random forest (RF), gradient descent boosting and support vector machine. Also, they have made use of 5-fold cross validation. Among those 5 models, the Random Forest model was the best performing model predicting mortality (Brier score: 0.17), and 0.76 (95% CI: 0.71 to 0.81) for HF hospitalisation.

Among the features present, blood urea nitrogen levels, body mass index, and Cardiomyopathy subscale scores were strongly associated with Mortality prediction. Whereas, Haemoglobin level, blood urea nitrogen, time since previous HF hospitalisation, and Cardiomyopathy subscale scores were most significant for HF Hospitalisation prediction. Subsequently[2], machine learning was used to predict survival of patients with heart failure from serum creatinine and ejection fraction alone. They have used feature ranking to identify the most relevant features of heart failure from a set of 13 feature variables. Feature ranking approaches clearly identify serum creatinine and ejection fraction as the two most relevant features and the model was built on those two factors alone. The two-feature models show that serum creatinine and ejection fraction are good enough to predict survival of heart failure patients from medical records. To prove the same, they have created a model wit all features in the entire

dataset, but the accuracy of prediction was low compared to just using two high ranked features. So from the above literature review we can see that more recent models show improvements, especially if the survival outcome is coupled with additional targets (for example, hospitalisation). Although scientists have identified a broad set of predictors and indicators, there is no shared consensus on their relative impact on survival prediction. This situation is largely due to a lack of reproducibility, which prevents drawing definitive conclusions about the importance of the detected factors. The lack of reproducibility strongly affects model performances. So there is always room to modify scope and make more accurate models.

Research Gaps:

The literature review reveals several challenges in the current understanding and predictive modeling of heart failure outcomes. Firstly, there exists a limited consensus on the relative impact of identified predictors and indicators, highlighting the need for a more comprehensive understanding of their roles in survival prediction. Reproducibility issues plague existing models, impairing their reliability and preventing definitive conclusions about the importance of detected factors. Despite recent advancements, there is a recognized scope for improvement in model performance, necessitating ongoing research to refine predictive capabilities. The study underscores inconsistencies in model calibration and validation across reviewed studies, emphasizing the importance of addressing these issues for robust and reliable predictive models. In summary, addressing these challenges is crucial to advancing the accuracy and applicability of predictive models for various heart failure cases, ultimately contributing to more effective clinical interventions and improved patient outcomes.

Feature	Explanation	Measurement	Range	Type
Anaemia	Decrease of red blood cells or haemoglobin	Boolean	0, 1	Categorical
Death event (Target)	If the patient died during the follow-up period	Boolean	0, 1	Categorical
Diabetes	If the patient has diabetes	Boolean	0, 1	Categorical
High blood pressure	If a patient has hypertension	Boolean	0, 1	Categorical
Gender	Woman or man	Binary	0, 1	Categorical
Smoking	If the patient smokes	Boolean	0, 1	Categorical
Age	Age of the patient	Years	[40...95]	Continuous
Creatinine phosphokinase (CPK)	Level of the CPK enzyme in the blood	mcg/L	[23...7861]	Continuous
Ejection fraction	Percentage of blood leaving the heart at each contraction	Percentage	[14...80]	Continuous
Platelets	Platelets in the blood	kilo platelets/mL	[25.01...850.00]	Continuous
Serum creatinine	Level of creatinine in the blood	mg/dL	[0.50...9.40]	Continuous
Serum sodium	Level of sodium in the blood	mEq/L	[114...148]	Continuous
Time	Follow-up period	Days	[4…...285]	Continuous

Exploratory Data Analysis

The dataset is taken from Government College University, Faisalabad-Pakistan, for the study of heart failure prediction, the study was approved by Institutional Review Board and ethical principles were followed to supply guidance to physicians and other participants in medical research involving human subjects (i.e., Helsinki Declaration were followed). The dataset has 299 heart patients, 194 of them were men and 105 of whom were women. Each patient belonged to NYHA classes III and IV, had left ventricular systolic failure, and was older than 40. The average follow-up period was 130 days, ranging from 4 to 285 days. The diagnosis of the illness was made using the ultrasound of the heart. Potential factors that could explain the mortality caused by congestive heart failure (CHD) were age, serum creatinine, gender, smoking, blood pressure (BP), Ejection Fraction (EF), anaemia, platelets, creatinine phosphokinase (CPK), and diabetes. While Anaemia, Diabetes, BP, Gender, and Smoking were considered categorical variables, age, CPK, EF, platelets, Serum creatinine, serum sodium, and time are continuous variables. Note, our response variable is death due to Heart failure of the patients within follow-up period.

From the **frequency distribution** of categorical variables, we got the following summary: Out of the 299 individuals in the study, 96 of them experienced heart failure before follow-period over, while the remaining 203 did not have heart failure. In terms of smoking status, 203 participants did not smoke, and 96 were smokers. Regarding diabetes, 174 individuals did not have diabetes, while 125 had diabetes. Blood pressure data showed that 194 participants did not have high blood pressure, and 105 had high blood pressure and 170 Anaemic and 129 non anaemic patients were there. We can conclude that as all categorical variables have considerable number of counterparts, so we cannot drop any categorical

Outliers' detection for the continuous variables:

Age: The boxplots shows that age is almost normally distributed and there are no outliers present in it.
Ejection Fraction: The data is slightly negatively skewed and there are two outliers, but we cannot remove this outlier, as each outlier is within the output value 0% to 100% of echocardiogram (machine used to measure ejection fraction).

Sodium: Distribution is negatively skewed as median is near to the quartile 3, there are also some outliers below minimum value.

Creatinine: The data is positively skewed, and there are lot of outliers as the normal range of Creatinine serum is typically from 0.74 to 1.35 mg/dL for adult men, while for adult women it is 0.59 to 1.04 mg/dL. Creatinine above 5.0 denotes severe kidney failure, and person may need dialysis. So, we can say that the observations visible as outliers cannot be dropped as those values within the range of serum creatinine level which can be found in human body.

Platelets: Distribution is negatively skewed, the normal range for platelets is between 1,50,000 – 4,00,000 /mL but it can go below 10,000. The condition when platelets below 1,50,000 is called Thrombocytopenia, but maximum value can go as high as 9 $*10^5$ and still does not indicate any significant health issues, thus also considered as normal. Therefore, it is okay to consider the outliers here too.
CPK: The CPK normal range for a male is between 39 – 308 U/L (units per litre), while in females the

CPK normal range is between 26 – 192 U/L. It has been found that when there is injury or stress in the brain or heart, the value of CPK reaches to higher level, therefore we cannot ignore the outliers here. Here the data is moderately positively skewed, so the feature can be tried after transforming by taking square root transformation.

Following are the plots, showing correlation between the response variables death event due to heart failure, where 'Yes' means there was death due to heart failure and 'No' means 'no-death' happened within follow-up time.

From the above plot we can see that the chances of getting death due to heart failure is more as the age increase. Also, we can see there is significant chances of heart failure when the Ejection fraction is greater than 75 units and below 30 units.

Lower value sodium shows a significant relation with the heart failure. From Creatinine plot we see that above 5.0 units of Creatinine there is a significant chance that the person may suffer from heart failure. From the below two plots we can interpret that there is no significant relation with platelets and heart failure, so we might drop this variable before training the model. It is clear from the plot of CPK that elevated levels of CPK have significant relation with heart failure. It can be seen from the follow-up time plot; death was more certain for which experts provide less follow-up time as they were higher risk patients and vice-versa for the person with higher follow-up time.

Research Methodology

In research methodology we are going through following steps such as, data processing, important feature identification and modelling in order to evaluate the likelihood of getting a heart attack.

Data Pre-processing:

The dataset is publicly available and taken from Government College University, Faisalabad-Pakistan, for the study of heart failure prediction. Initially in the dataset, we looked up the missing values at the pre-processing stage of the data because they have significant impact on the bias, variability, and stability of the model. Nevertheless, the dataset contained no missing values. Multicollinearity is another issue that affects the model performance. When two or more independent variables in a regression model have a strong linear relationship and are highly correlated, this is known as multicollinearity. This could affect the model in many ways, such as Increased Standard Errors, Unstable Coefficient, Difficulty in Identifying Significant Variables, Model Interpretability Issues.

So, to address the multicollinearity, we plotted the correlation plot and we found that all the predictors have a correlation value less than 0.25, which is a not concerning value. So, the dataset does not have a Multicollinearity issue.

Model Performance Considering Outliers:

We also checked how the model performed with and without outliers. For that we selected Logistic Regression to evaluate the performance and we found out that AUC curve with outliers was found to be 0.849 which was lower than without outliers, i.e.,0.86. Even though with clean data the AUC is better but, it results in giving lesser accuracy then original data. Also after thoroughly analysing the individual outliers, its clear that, values are important for our dataset. Hence, we will be proceeding rest of the processing with full data including outliers.

Feature selection:

Finding key features is crucial for effective data modelling. Feature selection involves evaluating variables based on their relevance to the target outcome. This process enhances model interpretability, reduces dimensionality, and mitigates overfitting, leading to more accurate and efficient predictive models in various domains, including healthcare and finance. To achieve this, we have used the Best Subset Method. The best subset method systematically evaluates all combinations of predictors to find the most predictive subset for a model. It selects the subset with the highest model performance, aiding in feature selection and optimising model complexity, particularly in situations with a moderate number of predictors and we have only 12 predictors. Based on the plotting number of variables with respect to Bayesian Information criterion, R^2 and Adjusted-R^2, the best subset for each of these plots have 4, 12, and 6 respectively, Now, in order to validate the performance of these to find the best subset. So, in conclusion we are trying to evaluate the model performance for all three subset datasets to find the optimal best subset.

On evaluating, model performance with the test dataset, we are able to understand that, the lowest BIC subset having four variables is giving the best results. In the context of a small dataset observations, the Bayesian Information Criterion (BIC) is advantageous for mitigating overfitting. BIC penalizes overly complex models by factoring in the number of parameters, making it well-suited for situations with limited data. With a smaller dataset, the risk of overfitting is heightened, as models with excessive complexity may fit noise rather than genuine patterns. BIC's preference for simpler models helps guard against overfitting, promoting the selection of more parsimonious and robust models that are likely to generalize better to new data in the constrained environment of a small dataset.

Using best subset selection method, we found that value of BIC is minimum for 4 predictor model. R^2 is showing maximum when the number of predictors is highest. Which is expected as R-square value increases with number of predictors as it will cover maximum variance, but this cannot be considered for our selection as it will overfit the model due to high number of predictors. Therefore, we use Adjusted R-square as it is an advanced version of R-square. We find that Adjusted R-square is highest for model with 6 predictors. The table at the right shows the ranking of the predictors as per the best subset we performed, here we are going to take the best 4 features based on the model performance evaluation performed above, i.e., ***Time, Ejection Fraction, Creatinine Serum*** and ***Age*** as our model Features for the prediction.

Model Selection:

We are currently targeting three models for our dataset: *Logistic regression, Decision tree and Random Forest*. The reason behind this is logistic regression has its widespread application in binary

classification, offering clear interpretability and estimating odds ratios. It oversees both continuous and categorical variables. Decision trees, with their simplicity and intuitive graphical representation, aid in understanding decision logic and key influencing factors. Random forest, chosen for its robustness, noisy data handling, addressing imbalances and providing variable importance measures for categorical variables. Their adaptability of handling diverse data types, and ease of modification contribute to our modelling strategy for predicting heart failure. The model selection done by using bootstrapping. For smaller datasets, bootstrapping is preferable over K-fold cross-validation or LOOCV as it creates surrogacy validation sets by resampling data with replacement. As a result, a more reliable assessment of the model's performance can be achieved thereby making the most of the available data. On the other hand, because of the decreased sample size in each fold, K-fold cross-validation may lead to a significant loss of information as we will be holding out one of the fold each time. The results after performing bootstrapping with different iterations got the accuracy result as follows:
For all the iterations for bootstraps, we found out that the accuracy was found to be very similar throughout the iterations. For Random Forest, we found out to have the highest accuracy of 94.1%, followed by Decision tree and then Logistic Regression.

Iterations	100 iterations	1000 iterations	5000 iterations
Logistic	0.836304	0.837827	0.8377205
Decision Tree	0.862125	0.864384	0.8646732
Random Forest	0.942857	0.9417054	0.9412884

Final Model:

As we saw that for the final model was found to be Random Forest as per the bootstrap accuracy. We again performed the hyperparameter tuning for the no. of variables selected on best subset, .i.e., time, creatinine, ejection fraction and age.

We did the hyperparameter tuning for the following variables, i.e., no. of variables, no. of trees and the no. of nodes.
We can infer form the graph, the highest accuracy was found to be with mtry 1, 100 no. of trees and optimal values of nodes to be 10.

Before Hyperparameter Tuning **After Hyperparameter Tuning**

Therefore, we are going to use a Random Forest for our dataset. We attached a file of the output of the model to the model which shows that Random Forest has an accuracy of 81.33% on test data. And we found that for the random forest Importance of variables is in order Time > Creatine > Ejection fraction > Age.

Discussion and conclusion:

The study investigates heart failure prediction using machine learning and clinical markers. Notably, serum creatinine, ejection fraction, and time emerged as crucial predictors. The Random Forest model, after hyperparameter tuning, displayed an impressive 81.33% accuracy on test data. This underscores the model's robustness and predictive power. The findings align with existing literature, emphasizing the importance of these markers in heart failure prognosis. However, ongoing research should explore additional predictors for a comprehensive understanding. Addressing outliers' impact, model comparisons, and meticulous feature selection contribute to the study's methodological rigor.

In conclusion, the study leverages statistical learning to enhance heart failure prediction. The Random Forest model, optimized for key predictors, showcases promising accuracy. These results contribute to advancing early detection and risk assessment in cardiovascular health. Future research may delve deeper into refining models, exploring novel predictors, and validating findings across diverse datasets. Overall, the study provides valuable insights into leveraging data-driven approaches for addressing the critical global health concern of heart failure.

Future research:

While this study makes significant strides in predicting heart failure using clinical markers and predictive models, there remains ample room for future research to enhance our understanding and improve predictive accuracy. One avenue for exploration involves incorporating advanced machine learning techniques, such as deep learning algorithms, to uncover complex patterns and relationships within vast datasets. Additionally, there is a need for longitudinal studies that track patients over extended periods, allowing for a more comprehensive analysis of the dynamic nature of heart failure progression. Integrating genetic and genomic data into predictive models could offer personalized risk assessments, contributing to more targeted interventions. Furthermore, collaborative efforts to establish standardized datasets and evaluation metrics across studies would facilitate reproducibility and comparability of predictive models. As emerging technologies continue to evolve, including wearable devices and real-time monitoring, future research can leverage these innovations to capture more frequent and nuanced data, thereby refining predictive models for heart failure with increased precision. Addressing these areas of inquiry holds the promise of advancing our capabilities in predicting, preventing, and managing heart failure, ultimately improving patient outcomes and global public health.

References

1. Heart Failure Classification - Stages of Heart Failure and Their Treatments. Website- (link)
2. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone (link)
3. The pathophysiology of heart failure. (link)
4. Correlation between Ejection Fraction and Hepatic and Renal functions in Heart failure patients. (link)
5. Classes and Stages of Heart Failure. Website (link)
6. Selection of potential predictors of worsening heart failure rational and design of the SELENE HF study(link)
7. Machine Learning Prediction of Mortality and Hospitalisation in Heart Failure With Preserved Ejection Fraction(link)