



Analysis of the heart failure data

12/02/2023

Presented by:

Suvam Bhadra and Mahshid Kalantari

Introduction

Cardiovascular diseases (CVDs) are the most important cause of death all over the world, with 31% of all deaths worldwide. Heart failure is a common event caused by CVDs and this dataset contains 12 features to predict mortality by heart failure. There are some patients that have high risk because they have more risk factors like diabetes, blood pressure, etc. It is crucial for the early detection of these patients, identifying the factors, and decreasing the rate of death for these patients. This study analyzed the data to find the most important factors related to death events for reducing this rate.

Methods

The data includes four tables of information related to the patients and the factors that affect heart failure. First, the tables were joined to make one uniform table for the analysis. Then, the statistical analysis was performed. The factors and outliers were checked for outliers by boxplots. The outliers were not removed from the data. In the next step, the distribution of data for each factor was analyzed using histograms. Then the frequency and the percentage of each part of the factors were plotted. To know the relationship between the death event and the other factors and the category of it, bar charts were designed. For the next step, the analysis included the relationship between age, death events, smoking category, and other factors using the violin plots. Finally, for deep analysis, the heat map was applied to find the most important features of heart failure in the data collected by plotting the correlation between each of the columns. By Plotting Kernel Density Estimation of time and age factors, the analysis was completed.

Conclusions

It can be interpreted that the minimum age is 40 and the maximum age is 95. Most of the patients are above the age of 60 which means the data patients are more toward the senior age group.

With the data at hand and the analysis performed I find that for the diagnosis of heart failure the most vulnerable can be patients belonging to the age group between 55-75 years which can be called the old adults.

After analyzing the distribution of factors, the two factors, creatinine phosphokinase, and serum creatinine showed a positively skewed distribution because of the many numbers of outliers. All the other

factors almost showed the normal distribution. The skewness in the distributions can be also seen in the statistical analysis.

Finding the relationship between death event and the other features, the age over 80 years old, anemia, high blood pressure, the low number of platelets, high-risk range of creatinine phosphokinase and serum creatinine and nonnormal range of serum sodium were the most effective factors on death events. On the other hand, smoking, diabetes, and gender did not influence death events. And for ejection fraction, the borderline and low numbers had the most effect on death events.

The mean serum creatinine level ranges between 0.74 – 1.35 mg/dL in adult men, while it should lie between 0.59 -1.04 mg/dL in healthy adult women. We have seen in the data that there is a significant count of patients above this range. This means there is a higher chance of stroke, and the serum creatinine level could be a major contributing factor in the mortality due to heart failure.

Usually, the normal range of CPK levels falls anywhere between 10 to 120 micrograms per liter (mcg/l). When the total CPK level is very high, it most often means there has been injury or stress to muscle tissue, the heart, or the brain. The pattern and timing of a rise or fall in CPK levels can be significant in making a diagnosis. This is particularly true if a heart attack is suspected. We see the median content among the patients in the data is about 250 mcg/l which is an alarming sign of a critical factor for heart failure.

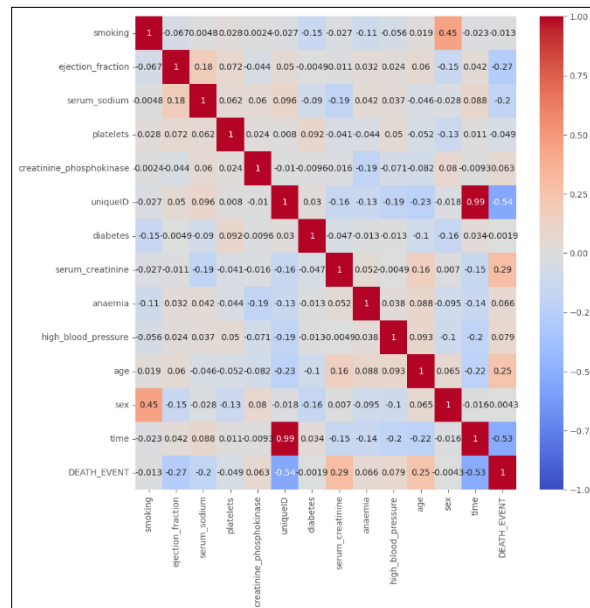
It is important to add that there are few patients who have an ejection fraction of more than 80% between the age of 45-50 years. A patient with an ejection fraction higher than 70 percent can indicate a heart condition like hypertrophic cardiomyopathy, a common cause of sudden cardiac arrest. There are patients with critically low ejection fraction (less than 20 %) which means the heart's ability to pump oxygen-rich blood out to the body has decreased well below the normal level. This can indicate previous heart damage from a heart attack or cardiomyopathy. Since it is very uncommon at the age of 40-50 years to have such low ejection fractions, these patients should be taken special care of.

In the same age group of 55-65 years, we see the percentage of anaemic patients is also higher than that of the non anaemic ones, in addition to being prone to diabetes. Anemia increases cardiac output, may lead to eccentric left ventricular hypertrophy, activation of the sympathetic nervous system, and stimulation of the renin-angiotensin-aldosterone system, and is closely associated with chronic inflammation and increased oxidative stress thus resulting in heart failure. Thus, making this age group more prone to heart failure. Thus, this age group needs significant attention for the diagnosis of heart failure.

In addition, time had a negative effect on death events and in fewer follow-up days the death event was higher than the higher numbers of follow-up days. It was the most important factor as it would have been very crucial to get diagnosed early with heart issues to get timely treatment, thus reducing the chances of any fatality.

The result showed that the range of age for male death is higher than the range of age for females however the range of age for males and females that are alive was the same. The age of dead patients

with each factor of smoking, diabetes, high blood pressure, and anemia was higher than that of alive patients.



The result from the heat map indicated that time was the most important factor in reducing the chances of any fatality and it had the inverse relationship with the death event. The second important factor was serum creatinine which could affect the death event by increasing in the blood. The third factor was ejection fraction, which significantly influenced death events. Also, serum sodium influences death events and both factors had a negative relationship with death events. The last effective factor in death events was the age which had a positive relationship with death events. Considering sex and smoking, there was a positive relationship between them, and it indicated with increasing age the effect of smoking increases.

Recommendations

As seen the outliers in most of the factors data, decided to remove the outliers to get a better analysis of the data. Especially in some columns, many outliers were found such as creatinine phosphokinase. However, removing the outliers did not change the result except for ejection fraction and platelets. Therefore, the outliers were not removed for getting a precise result and decision. In addition, removing the outliers decreases one-third of the data which causes missing a large part of the data.

For early detection, considering time is a vital factor for the patients. After that, checking the amount of serum creatinine, ejection fraction, and serum sodium is recommended which is very essential to decrease the rate of death. And the final recommendation is related to the range of age, which shows with seniors age the risk of heart failure and death events is increasing and must be considered.