



Predicting survival of patients with heart failure from serum creatinine and ejection fraction

October 2020

[ML tool](#)

Predicting patients risk of heart failure based on easy to obtain medical data

Boris Henriksen
bh@dimaps.com

1. Introduction

1.1 Background

According to the World Health Organization (WHO) cardiovascular diseases (CVD) are the number 1 cause of death globally. An estimated 17.9 million people died from CVDs in 2016, accounting for 31% of all global deaths. [1]

Most cardiovascular diseases can be prevented by addressing risk factors. One type of heart failure is based on the ejection fraction value, which is expressed as a percentage, of how much blood is pump out with each contraction [2]. A normal heart's ejection fraction may be between 50% and 70%. A value below 40% is a heart failure. Creatinine is a waste product that forms during metabolism and is removed from the body by the kidneys. Normal levels range from 0.8 to 1.4 mg/dL. When kidney function decrease, creatinine levels increase. The level of creatinine is important in determining the risk for CVD. Detecting high levels of creatinine is important as high levels alone may cause no symptoms.[3][4]

1.2 Problem

The heart is a vital organ and medical teams may fail to see a patient's risk of heart failure.

1.3 Audience

Medical doctors could benefit highly from accurate predictions of future heart failure related events from a patient's medical record.

Machine learning applied to medical records can be an effective tool to predict the survival of patients with heart failure symptoms. An important task is to identify the most important risk factors that may lead to heart failure. To use patient data from electronic medical records could help doctors detect symptoms otherwise undetected.

2. Data

2.1 Data source

Heart failure clinical data set can be found on UCIs machine learning repository.

<https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records>

2.2 Description

The dataset contains 299 medical records of heart failure patients collected at the Faisalabad Institute of Cardiology and at the Allied Hospital in Faisalabad, Pakistan, during April-December 2015. Dataset is provided from Davide Chicco, Giuseppe Jurman.[5]

First line in dataset contains labels.

Filname: heart_failure_clinical_records_dataset.csv

Filtype: CSV

The dataset contains 13 features:

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

Feature description with label names and position in file:

age: Ranges from 40 to 95.

anaemia: Decrease of red blood cells or hemoglobin. (0=No, 1=Yes).

creatinine_phosphokinase: Levels of creatine phosphokinase enzyme in blood. High levels might indicate injury or heart failure

diabetes: Does patient have diabetes (0=No,1=Yes)

ejection_fraction: Percentage of how much blood is pumped with each contraction

high_blood_pressure: High blood pressure (0=No, 1=Yes). Actual values not in dataset.

platelets: Level of platelets, also called thrombocytes, in blood. This is important to prevent bleeding.

serum_creatinine: Level of creatinine in blood. Waste product when muscle breaks down.

serum_sodium: Level of sodium in blood. A mineral that is important for correct functioning of muscles and nerves

sex: Woman og man. (0=Woman, 1=Man)

smoking: Is patient a smoker (0=No, 1=Yes)

time: Number of days for follow-up

DEATH_EVENT: This is the result we are seeking.

2.3 Data used

The dataset contains several features that exist in a patient's medical record. Features relevant for predicting that a patient may die are serum_creatinine and ejection_fraction values. In addition it is relevant to investigate other features in the dataset have and if they correlate.

The dataset is checked for correct number of records and features. All data columns have correct data type and there are no null values.

References

1. WHO Cardiovascular diseases (CVDs). ([https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)))
2. Heart.org: Ejection Fraction Heart Failure Measurement. (<https://www.heart.org/en/health-topics/heart-failure/diagnosing-heart-failure/ejection-fraction-heart-failure-measurement#:~:text=What%20is%20%E2%80%9Cejection%20fraction%E2%80%9D%3F%20Ejection%20fraction%20%28EF%29%20is,left%20ventricle%20is%20pushed%20out%20with%20each%20heartbeat.>)
3. Why are blood creatinine levels checked?
(https://www.medicinenet.com/creatinine_blood_test/article.htm)
4. Learn why doctors test creatinine levels after heart attack. (<https://www.healthguideinfo.com/heart-attack/p65166/#heart-attack--creatinine>)
5. : Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making 20, 16 (2020)
(<https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1023-5#citeas>)
6. BMC Dataset. (<https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1023-5#Sec2>)