Heart Failure Mortality Analysis

Dexter Schincke

Bellevue University

DSC 680

Amirfarrokh Iranitalab

9 February 2025

Heart Failure Mortality Analysis

**Topic:**

This project analyzes heart failure patient data to identify key risk factors associated with mortality. By leveraging machine learning techniques, the study aims to develop predictive models to assess patient survival based on clinical and demographic features.

**Business Problem:**

Cardiovascular diseases (CVDs) are the leading cause of death globally, accounting for approximately 31% of all deaths worldwide. Heart failure, a major consequence of CVDs, can be managed more effectively with early detection and intervention. This project aims to analyze factors contributing to heart failure mortality, helping healthcare professionals identify high-risk patients and improve treatment strategies.

* Research Questions:
  + What are the most significant predictors of heart failure mortality?
  + How do demographic factors such as age, sex, and smoking status influence survival rates?
  + What role do clinical variables (serum creatinine, ejection fraction, blood pressure) play in predicting mortality?
  + Can machine learning models accurately predict patient outcomes based on available data?
  + How can insights from this analysis improve early detection and intervention strategies?

# Dataset:

The dataset consists of patient records containing 12 clinical and demographic features relevant to heart failure. These features include biomarkers, lifestyle factors, and medical history, which can be used to develop predictive models for patient survival.

Features:

* age: Patient’s age
* anaemia: Decrease of red blood cells or hemoglobin (boolean)
* creatinine\_phosphokinase: Level of the CPK enzyme in the blood (mcg/L)
* diabetes: If the patient has diabetes (boolean)
* ejection\_fraction: Percentage of blood leaving the heart at each contraction (percentage)
* high\_blood\_pressure: If the patient has hypertension (boolean)
* platelets: Platelets in the blood (kiloplatelets/mL)
* serum\_creatinine: Level of serum creatinine in the blood (mg/dL)
* serum\_sodium: Level of serum sodium in the blood (mEq/L)
* sex: Woman or man (binary)

**Methods:**

The project will begin with data cleaning and preprocessing, including handling missing values, normalizing numerical features, and encoding categorical variables. Exploratory data analysis (EDA) will involve visualizing distributions, correlations, and patterns among different features to identify key risk factors.

For predictive modeling, classification algorithms such as Logistic Regression, Random Forest, and XGBoost will be evaluated to determine the most effective model for predicting heart failure mortality. Performance metrics such as accuracy, precision, recall, and F1-score will be used to assess model effectiveness. Feature importance analysis will also be conducted to identify the most influential predictors of patient outcomes.

### Ethical Considerations:

The study will ensure patient anonymity and compliance with ethical guidelines for handling medical data. Findings will be presented transparently, highlighting potential biases and limitations in the dataset. The project aims to provide general insights rather than individual risk assessments to prevent misinterpreting results.

### Challenges/Issues:

One challenge involves potential data imbalances, as mortality events may be less frequent in the dataset. Addressing this issue may require techniques such as oversampling, undersampling, or using weighted loss functions in machine learning models. Additionally, the interpretability of complex models must be considered, as healthcare professionals need clear insights rather than black-box predictions. Another challenge is ensuring the dataset's representativeness, as results may vary based on demographics and medical histories not covered in the dataset.

References

**Projects/Papers:**

Mustanger. (2022, December 1). heart failure: Eda and prediction. Kaggle.

<https://www.kaggle.com/code/eisgandar/heart-failure-eda-and-prediction#3%7CLOAD->DATASET

RICCIARDI, C. (2025, January 22). Heart failure prediction. Kaggle. https://www.kaggle.com/code/anacpricciardi/heart-failure-prediction

**Datasets:**

Davide Chicco, Giuseppe Jurman: 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://www.kaggle.com/code/eisgandar/heart-failure-eda-and-prediction#3%7CLOAD->DATASET