Title: Leveraging Machine Learning Models for Predicting 1-Year Mortality in Heart Failure Patients.

Abstract

Background:
Heart failure (HF) is a leading cause of death globally, with high rates of rehospitalization and mortality within the first year of diagnosis. Accurate prediction of 1-year mortality in HF patients is essential for identifying high-risk individuals who may benefit from early interventions and personalised treatment plans. Traditional risk assessment models often lack sufficient accuracy and scope. Machine learning (ML) techniques, with their capacity to analyse complex datasets and identify patterns in patient data, offer significant potential for improving the prediction of patient outcomes. This study evaluates the performance of five different machine learning models in predicting 1-year mortality in HF patients based on a variety of clinical and demographic factors.

Methodology:
The dataset used for model development included patient demographic data and clinical features such as age, length of stay, serum potassium levels, and comorbidities. Missing data were addressed by removing variables with more than 50% missing values. The remaining missing data were imputed using the K-Nearest Neighbors (KNN) method, considering the five nearest neighbours. The imputed dataset was further processed by retaining one variable from each pair of variables with a correlation of at least 0.80. The final dataset was normalised using the Min-Max method for numerical variables and One-Hot encoding for categorical variables. After preprocessing and feature engineering, five machine learning algorithms were trained and validated: Logistic Regression (LR), Random Forest (RF), Gradient Boosting, Support Vector Machine (SVM), and Neural Networks. Model training and evaluation were performed using 10-fold cross-validation to ensure robust performance to optimise each model. The models were evaluated based on the area under the receiver operating characteristic curve (AUC-ROC).

Results:
The Gradient Boosting model demonstrated the best overall performance, achieving the highest AUC-ROC score of 0.7099, making it the most effective model for predicting 1-year mortality in HF patients. Logistic Regression followed closely with an AUC-ROC of 0.7096, and Random Forest with an AUC-ROC of 0.6934. Among the key features selected by the Gradient Boosting model were the number of outpatient visits in the past year, the number of Emergency Department visits in recent months, age at admission, and initial weight measurements, among others.

Conclusion:
This study highlights the potential of machine learning models in accurately predicting 1-year mortality in HF patients. These models can be valuable tools in clinical practice to identify high-risk patients, enabling healthcare providers to tailor interventions and improve patient outcomes. Future research should focus on refining the calibration of these models and integrating them into clinical workflows for real-time decision support. The findings demonstrate that when properly applied, machine learning has the ability to enhance

predictive accuracy in heart failure management, paving the way for more personalised and data-driven healthcare.