Integrating Clinical Chart and Laboratory Data for Predicting Heart Failure Recurrence

Zuzana Koscova[1], Eniko Vargova[1], Jan Pavlus[1], Radovan Smisek[1], Veronika Bulkova[2], Pavel Jurak[1], Filip Plesinger[1]

[1] Institute of Scientific Instruments of the CAS, Brno, Czechia
[2] Medical Data Transfer, Brno, Czechia

Abstract

Heart failure (HF) has the highest 30-day rehospitalization rate among all medical and surgical conditions. This study proposes a method for predicting HF recurrence using longitudinal chart data and laboratory measurements from patients in the critical care unit (CCU).

We analyzed data from 8,720 HF patients in the MIMIC-III database, extracting time series of five physiological measurements and deriving 11 statistical features from each, along with demographic and laboratory data, totaling 67 features. We employed a Random Forest (RF) classifier to predict HF recurrence over four time intervals (week, month, half a year, and a year) and conducted survival analysis using HF-recurrence probabilities from the trained model.

For the test set, we achieved a mean F1 score of 0.57±0.09, sensitivity of 0.77±0.03, and specificity of 0.77±0.04.

Our findings suggest that machine learning algorithms applied to chart data and laboratory measurements could potentionally help guide patient discharge decisions and reduce HF recurrence.

1. Introduction

HF is a chronic condition characterized by the heart's inability to pump blood efficiently, leading to inadequate blood flow to meet the body's needs. One of the critical challenges in managing patients with heart failure is predicting recurrence [1]. HF recurrence refers to the repeated episodes or exacerbations of symptoms after an initial diagnosis or after a period of stability. Early recurrence typically occurs within 30 days after discharge from a hospital for an initial heart failure event. It is often related to suboptimal management, patient non-adherence to medication, or inadequate follow-up care. Late recurrence occurs after a longer period and may be due to disease progression, new comorbidities, or insufficient long-term management strategies [2, 3]. In this study we propose a machine learning approach to predict recurrence of HF using vital-signs from the critical care unit (CCU) together with laboratory measurements and demographics to predict both early and late HF-recurrence. Accurate prediction of HF recurrence could potentionally allow healthcare providers to implement timely interventions, adjust treatment plans, and prevent hospital readmissions.

2. Dataset

For the prediction of heart failure recurrence we leveraged a MIMIC-III clinical database [4] containing deidentified health-related data associated with patients who stayed in CCU of the Beth Israel Deaconess Medical Center within years 2001-2012. We focused on patients whose primary reason for hospitalization was heart failure. Patients diagnosed with HF were retrieved from the database based on the ICD9 diagnosis codes [5]. We specifically extracted data from their first HF admission, as some patients had multiple admissions for heart failure. Together we collected medical information on 8,720 patients hospitalized with HF.

We extracted a time series of their vital sign measurements made at the bedside (approximately 1 data point per hour) together with their last laboratory test results recorded before discharge and demographics. The extracted time series made at bedside comprises of heart rate [bpm], respiratory rate [breaths/min], systolic and diastolic blood pressure [mmHg], and blood oxygen saturation [%]. These measurements are nurse-verified and documented hourly.

The extracted lab measurements included serum creatinine [mg/dl], sodium [mEq/l], blood urea nitrogen (BUN) [mg/dl], chloride [mEq/l], glucose [mg/dl], hematocrit [%], magnesium [mmol/l], phosphate [mg/dl] and potassium [mEq/l]. The average time when the laboratory measurements were taken is 10.82 ± 10.74 days from hospital admission, while the average length of hospital stay is 11.68 ± 10.08 days.

The demographic information includes sex with prevalence of 53% for males and 47% for females and age (72.93 ± 13.43). Subjects in the MIMIC-III database who reached age 89 at any point during the study were marked as 300 years old to make their age obscure. These subjects were converted to age 90 (prevalence of 10%) to work with the real-life data when building a final model for HF recurrence.

This data was complemented by information on mortality and HF recurrence. For each patient, we recorded the time to HF recurrence and all-cause death from the initial HF admission. For patients who did not experience HF recurrence or death, we noted the time of their last CCU visit.

3. Preprocessing

During the preprocessing stage, we implemented a filtering process for each of the vital signs time series features that were recorded hourly. This filtering utilized predefined thresholds based on typical physiological ranges to account for potential inaccuracies or errors from bedside monitor measurements [5]. The thresholds were deliberately set with some flexibility to minimize the exclusion of clinically relevant data. The specific lower and upper bounds used were as follows: SBP 40 to 250 mmHg, DBP 20 to 180 mmHg, HR 30 to 250 beats per minute, RR 6 to 50 breaths per minute, and SaO2 70% to 100%. Additionally, we excluded any admissions with fewer than 10 measurements for any of these bedside features, resulting in a final dataset consisting of 8,446 subjects.

From each of the five time series features, 11 statistical features were extracted, including the mean, median, max, min, std, skewness, kurtosis, max gradient, min gradient, mean gradient, and std of the gradient. This extraction yielded a total of 55 vital signs features. These features, combined with 11 laboratory measurements and 2 demographic characteristics, comprise the final dataset.

We then divided our data into training, validation, and test sets, containing 4,750, 1,584, and 2,112 subjects, respectively. During the random splitting process, we employed stratification to ensure balanced representation of outcomes (HF-recurrence/all-cause death versus HF-free) across all sets.

4. Methods

Our analysis is divided into two main parts. The first part aims to predict the recurrence of heart failure (HF) across four different time windows: one week, one month, six months, and one year. The second part involves utilizing time-to-event data to perform survival analysis using the Kaplan-Meier plot and employing the Cox proportional hazards regression model to study the effect of different variables on the outcome.

4.1. Heart failure recurrence prediction

To predict heart failure recurrence within the specified time windows, we trained four separate random forest (RF) classifiers, each corresponding to a different time frame. The dataset was labeled into two categories: HF-free and HF-recurrence. These labels were assigned based on the final outcome and the duration from the patient's first HF admission to either their last CCU visit (for patients who remained HF-recurrence-free) or the time to HF recurrence or all-cause death. The number of subjects with HF recurrence or all-cause death, as well as those who were HF-free across different time windows is captured in Table 1.

Table 1. Number of subjects for 2 categories for 4 different time-windows.

Time window	HF-free	HF-recurrence
Week	7,837	609
Month	6,997	1,449
6 months	5,916	2,530
12 months	5,420	3,026
Overall	3,527	4,919

We identified the optimal parameters for each of the four RF classifiers through a grid search with 5-fold cross-validation on the training set. Once the optimal parameters were determined, we trained the models on the full training set. The output of the RF classifiers is a probability of HF-recurrence. We established the optimal probability decision threshold by selecting the operating point of the receiver operating characteristic (ROC) curves on the validation dataset.

Finally, we constructed an ensemble model by averaging the probabilities of heart failure recurrence predicted by the four individual random forest classifiers. For each classifier, we assessed feature importance using Gini impurity and subsequently computed the average feature importance across all four models.

To verify the ability of our model to predict the HF in time we created Kaplan-Meier plot. This analysis was performed at a follow-up time equal to half the duration of the study (5.6 years). Patients who did not have an event during this time and had no further information about them in the study were censored. We split patients into the high risk group and low risk group for HF -recurrence based on the mean threshold out of 4 RF models found on the validation set (0.31).

4.2. Cox proportional regression

To assess the impact of predictor variables on survival time, we utilized the Cox proportional hazards model with the Breslow baseline estimation. Our dataset comprises 68 features, with 55 being statistical summaries derived from vital signs. Due to the high correlation among these features and the fact that Random Forest (RF) classifiers are resilient to multicollinearity while the Cox regression model is not, we needed to reduce dimensionality. Therefore, we selected a subset of the most important features based on the Gini feature importance from the RF ensemble model. For each vital sign, only the feature with the highest importance was included in the Cox regression analysis. We used a p-value threshold of 0.001 to determine the statistical significance of variables, ensuring that only those with strong evidence of association are considered significant.

5. Results

The results of HF prediction over four time intervals on the validation dataset are illustrated in Figures 1 and 2, which show the ROC and precision-recall curves (PRC), respectively. Table 2 presents the classification results of the Random Forest (RF) models across four different time-windows on the test set. Figure 3 depicts the feature importance of the RF ensemble, along with the feature importance from the four submodels, represented by colored dots. Additionally, Figure 4 displays the Kaplan-Meier plot, which categorizes patients into low-risk and high-risk groups for HF-recurrence based on the probability predicted by the RF ensemble models. Figure 5 presents the hazard ratios along with their 95% confidence intervals (CIs) from the Cox regression analysis. This analysis examines the impact of predictor variables on the time to HF recurrence. The hazard ratios and their 95% CIs are color-coded according to the p-values: variables with p-values higher or equal to 0.005 are shown in pink, while significant variables with p-values below this threshold are depicted in blue.

Table 2. Results of RF classification for 4 different time-windows.

Time window	Sensitivity	Specificity	Precision
Week	0.82	0.83	0.28
Month	0.77	0.81	0.45
6 months	0.74	0.75	0.56
12 months	0.74	0.71	0.59

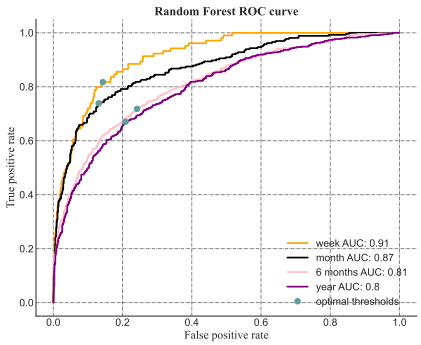

Figure 1. Receiver operating characteristics curve performed on the validation dataset for 4 different time windows together with their operating points.

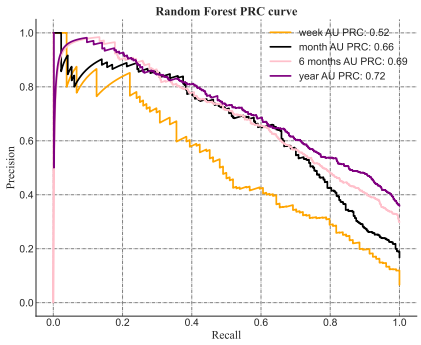

Figure 2. Precision recall curve performed on the validation dataset for 4 different time windows.

6. Discussion and conclusion

Our analysis showed that the RF classifier effectively predicts HF recurrence over various time intervals, achieving a mean F1 score of 0.57±0.09, sensitivity of 0.77±0.03, and specificity of 0.77±0.04. The model could potentionally assist in discharge decisions and optimize follow-up care. Incorporating statistical features from time series data improved performance compared to average values alone [6].

Feature importance analysis highlighted that the standard deviation of SaO2 is crucial for short-term HF recurrence prediction, while blood urea nitrogen is important for longer-term forecasts due to its link with renal impairment. Cox regression further confirmed SaO2's significance (p-value 2.61×10^{-74}, Hazard ratio 1.27, 95% CI 1.24-1.31). Other significant variables include glucose, magnesium, chloride, blood urea nitrogen, phosphate, mean RR intervals, minimal SBP, and age.

The Kaplan-Meier plot (Figure 4) demonstrates the

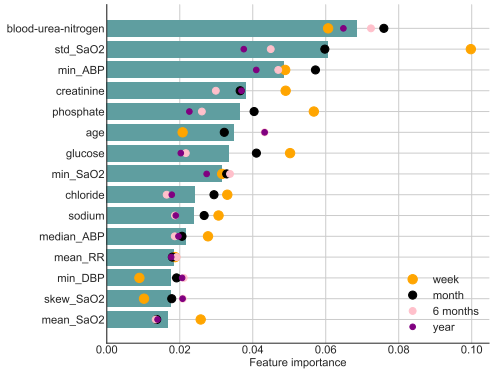

Figure 3. Random Forest feature importance for 4 different time periods represented by color dots and for overall ensemble model represented by bar plot.

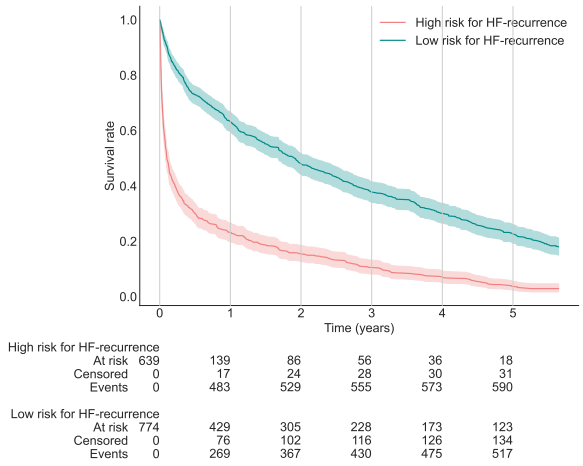

Figure 4. Kaplan-Meier plot categorizing subjects into high-risk and low-risk group for HF recurrence. The log-rank test for the comparison of the 2 categories classified by ensemble RF model shows p-value of 1.23×10^{-65} with hazard ratio of 2.83 (95% CI 1.16-2.51).

model's capability to stratify patients into high- and low-risk groups for HF recurrence based on the posterior probability generated by the ensemble RF model (p-value of 1.23×10^{-65}). The median survival time in the high-risk group for HF recurrence is 0.10 years (36.5 days), compared to 1.9 years in the low-risk group, highlighting the predictive performance of our model.

Our future goal is to transition this model from CCU settings to at-home monitoring. By investigating the effects of missing variables on HF prediction [7], we may either exclude lab measurements or use those from the patient's last CCU visit, combined with smart watch vital signs, to predict HF after discharge. Additionally, we plan to incorporate recurrent neural networks to better leverage the time-series nature of the data, rather than relying solely on

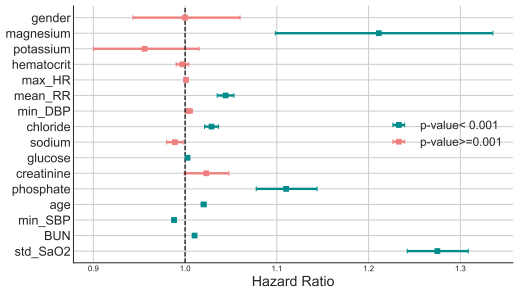

Figure 5. Hazard ratios together with their 95 % confidence intervals colored by p-value obtained from Cox proportional hazard model.

statistical feature extraction, which may result in the loss of temporal information.

Acknowledgments

The research was supported by the Czech Technological Agency grant number FW06010766 and the project RVO:68081731 by the Czech Academy of Sciences.

References

[1] Gupta A, et al. The hospital readmissions reduction program: Learning from failure of a healthcare policy. European Journal of Heart Failure August 2018;20(6):1169–1174.

[2] Heidenreich PA, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circulation Heart Failure 2013;6(3):606–619.

[3] Ponikowski P, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. European Heart Journal 2016;37(27):2129–2200.

[4] Johnson A, et al. MIMIC-III, a freely accessible critical care database. Scientific Data 2016;3(160035).

[5] Williams NJM, et al. Vital signs and reference ranges for adults: A review. Clinical Medicine Research 2019; 17(2):85–100.

[6] Koscova Z, et al. Predicting readmission of heart failure patients. In Proceedings of the 2023 Computing in Cardiology (CinC), volume 50. IEEE, October 2023; 1–4.

[7] Plesinger F, et al. The effect of missing data when predicting readmission in heart failure patients. In Proceedings of the 2023 Computing in Cardiology (CinC), volume 50. IEEE, October 2023; 1–4.

Address for correspondence:

Zuzana Koscova
Kralovopolska 147 Brno, Czech republic
koscova.zuzanka@gmail.com