Internal and Emergency Medicine
https://doi.org/10.1007/s11739-022-02996-w

IM - ORIGINAL

Machine learning and LACE index for predicting 30-day readmissions after heart failure hospitalization in elderly patients

Hernan Polo Friz[1] · Valentina Esposito[2] · Giuseppe Marano[3] · Laura Primitz[1] · Alice Bovio[4] · Giovanni Delgrossi[5] · Michele Bombelli[6] · Guido Grignaffini[7] · Giovanni Monza[8] · Patrizia Boracchi[3]

Received: 4 December 2021 / Accepted: 20 April 2022

Abstract

Machine learning (ML) techniques may improve readmission prediction performance in heart failure (HF) patients. This study aimed to assess the ability of ML algorithms to predict unplanned all-cause 30-day readmissions in HF elderly patients, and to compare them with conventional LACE (Length of hospitalization, Acuity, Comorbidities, Emergency department visits) index. All patients aged ≥ 65 years discharged alive between 2010 and 2019 after a hospitalization for acute HF were included in this retrospective cohort study. We applied MICE (Multivariate Imputation via Chained Equations) method to obtain a balanced, fully valued dataset and LASSO (Least Absolute Shrinkage and Selection Operator) algorithm to get the most significant features. Training (80% of records) and test (20%) cohorts were randomly selected. Study population: 3079 patients, 394 (12.8%) presented at least one readmission within 30 days, and 2685 (87.2%) did not. In the test cohort AUCs (IC95%) of XGBoost, Ada Boost Classifier, Random forest, and Gradient Boosting, and LACE Index were: 0.803 (0.734–0.872), 0.782 (0.711–0.854), 0.776 (0.703–0.848), 0.786 (0.715–0.857), and 0.504 (0.414–0.594), respectively, for predicting readmissions. A SHAP analysis was performed to offer a breakdown of the ML variables associated with readmission. Positive and negative predicting values estimates of the different ML models and LACE index were also provided, for several values of readmission rate prevalence. Among elderly patients, the rate of all-cause unplanned 30-day readmissions after hospitalization due to an acute HF was high. ML models performed better than the conventional LACE index for predicting readmissions. ML models can be proposed as promising tools for the identification of subjects at high risk of hospitalization in this clinical setting, enabling care teams to target interventions for improving overall clinical outcomes.

Keywords Machine learning · Heart failure · Prognosis · Aged · Readmission

Introduction

Heart failure (HF) is a common, serious condition, with a prevalence of 1–2% of adult population which rises to 10% in the elderly [1–3]. Acute HF represents a leading cause of mortality and 1–2% of all hospitalizations worldwide [3–7], especially in older subjects, who frequently presents comorbidities and are often hospitalized in general medicine wards [1, 7]. Hospital care accounts for about two-thirds of HF direct cost. Beyond monetary costs, readmission may be an indicator of poor quality of care and entails a low quality of life [3–6].

Numerous studies have addressed the efficacy of different programs for reducing readmissions and death in HF patients [4, 5, 8]. Those programs may be prohibitively expensive when applied to an entire patient cohort, but become cost-effective whether selectively applied to patients at high

✉ Hernan Polo Friz
hernanemilio.polofriz@asst-brianza.it; polofriz@libero.it

1 Internal Medicine, Medical Department, Vimercate Hospital, Azienda Socio Sanitaria Territoriale (ASST) della Brianza, Via Santi Cosma e Damiano 10, 20871 Vimercate, MB, Italy

2 Almaviva Digitaltec S.R.L, Naples, Italy

3 Department of Biomedical and Clinical Sciences "L. Sacco", University of Milan, Milan, Italy

4 Almaviva S.P.a, Rome, Italy

5 Information Technology Department, ASST della Brianza, Vimercate, Italy

6 Internal Medicine, Medical Department, Desio Hospital, ASST della Brianza, Desio, Italy

7 Director for Health and Social Care, ASST della Brianza, Vimercate, Italy

8 Medical Director, ASST della Brianza, Vimercate, Italy

Published online: 04 June 2022

risk for readmission. Therefore, different predictive models have been developed to identify patients at high risk for hospital readmissions [9–13], mainly based on conventional statistical approaches and with a simple model structure. These conventional methods allow a direct interpretation of variables contribution from regression coefficients and prognostic scores can be easily obtained by model linear predictors [14]. Moreover, tutorials for correctly applying logistic regression models using standard software are available [15]. Notwithstanding that, different limitations of these conventional prognostic tools have been pointed out, such as the risk of information loss and/or biased estimates [16–19].

Machine learning (ML) techniques can account for non-linear and higher dimensional relationships between a multitude of variables [18] and a model structure that enable the analysis of non-linear patterns and complex interactions. This may improve the prediction performance in a complex clinical condition as HF. However, the prognostic performance of ML over a conventional statistical approach is controversial [19, 20]. In addition, data on the performance of ML for predicting re-hospitalization in an elderly population with HF are lacking [21], as well as direct comparisons between ML algorithms and conventional statistics-based prognostic tools.

The aim of this study was to assess the ability of ML algorithms to predict unplanned readmissions within 30 days after hospitalizations for acute decompensated HF, in elderly patients. We also compared the prognostic performance of these algorithms with a conventional predictive tool, the clinically validated LACE (Length of hospitalization, Acuity, Comorbidities, Emergency department visits) index.

Methods

Study design

This retrospective observational cohort study was conducted in the Vimercate Hospital, an Italian 500-bed general hospital that serves a population of approximately 200,000 inhabitants in Lombardy, Northern Italy. Vimercate Hospital electronically registers and tracks all clinical information through a commercial electronic health record (EHR) system, Tabula Clinica ™ [22].

Inclusion and exclusion criteria, endpoints and cases selection process.

Vimercate EHR is active since 2010. Thus, in this study, we included all patients who were discharged alive between 1st January 2010 and 31st July 2019, after an hospitalization with a discharge diagnosis of acute HF, identified using the International Classification of Diseases (ICD)-9 Codes: 428.xx or DRG (diagnosis-related group): 127.

Hospitalizations with a length of stay less than 1-day, readmission within 24 h and admissions with in-hospital death were excluded. We also excluded patients under the age of 65 and those who died within 30 days after the index event. If the patient faced multiple readmissions within the 30 days only the first readmission episode was considered, to obtain a sample statistically independent and mutually exclusive across the two classes of admission: non-readmitted and readmitted patients.

The primary endpoint was all-cause readmissions within 30 days from the index hospitalization event. The selection process flow chart is described in Fig. 1A.

Ethical approval was required to the Institutional Research Ethics Committee (Comitato Etico Brianza), and informed consent was waived given the retrospective non-interventional, observational design.

Dataset and data management

The following structured data elements were extracted from Vimercate Hospital's EHR: demographics, medical history, physical examinations, diagnoses, procedures, labs, and medications. The following variables were considered: length of stay (days), age (years), sex (0:F, 1:M), hospitalization type, p-glucose (mg/dL), urea (mg/dL), total bilirubin (mg/dL), protein electrophoresis (g/dL), albumin (%), sodium (mmol/L), p-potassium (mmol/L), chlorin (mmol/L), urate (mg/dL), creatinine (mg/dL), estimated glomerular filtration test (mL/min/1.73mq), aspartate aminotransferase (U/L), alanine aminotransferase (U/L), proBNP (pg/mL), C-reactive protein (mg/L), leukocyte (10^9/L), erythrocyte (10^{12}/L), haematocrit (%), average body volume (fl), conc.Hb body average (g/dL), platelet (10^9/L), pH, P.T. (prothrombin time, INR), APTT (ratio), APTT (sec), diastolic pressure (mmHg), systolic pressure (mmHg), cardiac frequency: beats per minute (bpm), temperature (°C),), previous hospitalizations (total number), previous emergency accesses (number in the 6 months), drugs for acid related disorders (0: absence, 1: presence), drugs for constipation (0: absence, 1: presence), antithrombotic agents (0: absence, 1: presence), blood substitutes and perfusion solutions (0: absence, 1: presence), cardiac therapy (0: absence, 1: presence), diuretics (0: absence, 1: presence), agents acting on the renin–angiotensin system (0: absence, 1: presence), corticosteroids for systemic use (0: absence, 1: presence), antibacterials for systemic use (0: absence, 1: presence), antigout preparations (0: absence, 1: presence), analgesics, psycholeptics (0: absence, 1: presence), drugs for obstructive airway diseases (0: absence, 1: presence), cough and cold preparations (0: absence, 1: presence), all other therapeutic products (0: absence, 1: presence), Barthel score (Measure of physical disability relating to activities of daily living), Charlson

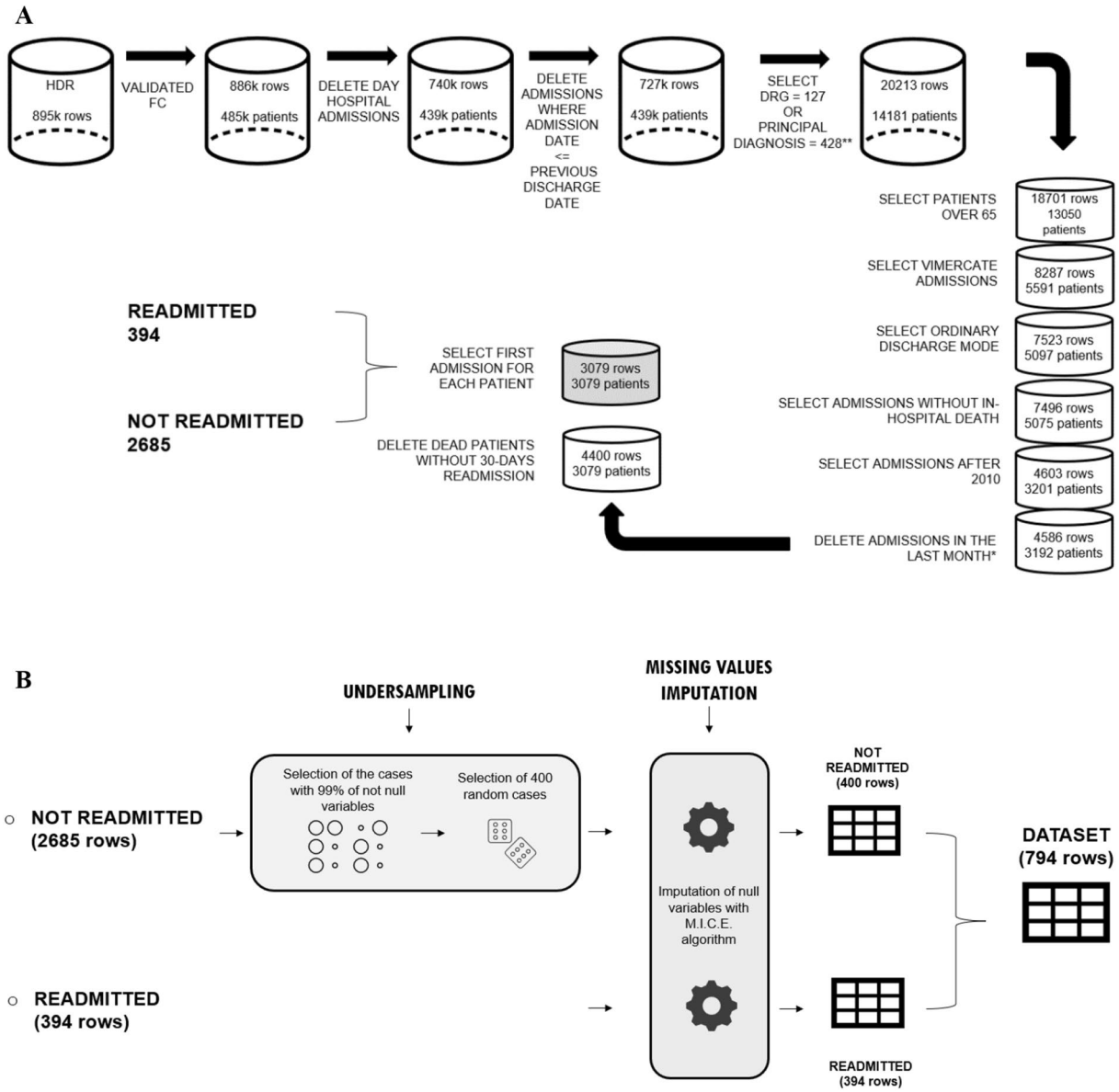

Fig. 1 Selection of study sample (**A**) and pre-processing workflow (**B**). *DRG* diagnosis-related group, *HDR* Hospital Discharge Register. *FC* social security number

Comorbidity Index: calculated using medical claims data (medical discharge reports). The values of lab results and vital parameters were collected during admission (first 48 h) and before discharge (last 48 h).

The administered drugs were grouped in therapeutic subgroups, following the Anatomic Therapeutic Chemical (ATC) second level classification.

Process of feature engineering, data pre-processing and feature selection.

As a first step, features with more than 20% of missing values were eliminated. Since the dataset was very unbalanced, it was performed an undersampling of the more populated class (not readmitted: 2685 records). The samples with at least 99% of non-missing values were selected and of these we randomly extracted 400 records. To address the residual missing values, on each class we applied the MICE (Multivariate Imputation via Chained Equations) method [23] to impute them, obtaining a balanced and fully valued dataset with 794 records. Finally, we applied the LASSO (Least Absolute Shrinkage and Selection Operator) algorithm [24] to get the most significant features. The pre-processing workflow is reported in Fig. 1B.

Development of ML prediction models

The data aforementioned were used to create train datasets for developing prediction models, using the following workflow: (a) creation of a row dataset; (b) dataset pre-processing: aimed to obtain a dataset free of missing values and redundant information; (c) feature selection (d) model creation and validation: the prepared dataset was split in two portions: training set or cohort and test set or cohort. The training set has been created through a random selection of 80% (635 records) of the rows while the test set is composed of the remaining records (159). The ML algorithms were trained on the training set, and the resulting models were tested on the test set.

The following ML algorithms were trained and tested: Ada Boost (Adaptive Boosting) Gradient Boost (Gradient Boosting) XGBoost (eXtreme Gradient Boosting) Random Forest. All the selected algorithms belong to the ensemble learning family, i.e. they use multiple models (weak learners) to get a better predictive performance than the one obtained with a single one of them. In all the four cases, the weak learners are Decision Trees.

The LACE index has been validated to identify patients at risk for readmission and/or death within 30 days after hospital discharge in both medical and surgical patients. It considers four parameters: length of hospitalization stay ("L"), acuity of the admission ("A"), i.e. Emergency Department vs. an elective admission, comorbidities of patients ("C") by incorporation the Charlson Comorbidity Index (CCI), and the number of Emergency Department visits within the last 6 months ("E"). LACE scores range from 1–19. Score: 0–4 = Low; 5–9 = Moderate; and ≥ 10 = High risk of readmission [11].

Comparison between readmitted and no readmitted patients.

Continuous variables were reported as mean and standard deviation, and median and interquartile range (IQR). The distribution of each variable was compared by the *t* test, except for the variables with a high asymmetric distribution; in these cases comparison was performed by the Mood median test. Categorical variables were reported as count and percentages. The distributions were compared using the test for proportions, or by the Fisher exact test, when there were with very few cases in one category ($n <= 5$).

Performance of the discrimination ability of the ML models and LACE

The performance of ML models and the conventional prediction LACE index to predict all-cause readmission within 30 days was compared using the area under the receiver operating characteristic curve (AUC) [20]. The cut-off we applied to identify patients at risk of re-hospitalization for LACE index was ≥ 10 [25].

Black box and SHAP analysis

Machine learning algorithms can produce accurate predictions but the processes that lead to these predictions may be not completely well described and understood, a phenomenon called "black box" [26]. Shapley additive explanations (SHAP) is a game theoretic approach to explain the output of any ML models. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions. We applied SHAP analysis to improve the global and local interpretability [27], and to achieve a better understanding of the learning process during the training phase of ML and a more complete description of the impact of variables on ML models prognostic performance.

Evaluation of predictive values

Since the models should be applied to predict the readmission of a patient within 30 days on the base of its baseline clinical characteristics, the predictive values were calculated using sensitivity and specificity of the previously reported ML models. The positive predictive value (PPV) is intended as the probability of readmission within 30 days given ML prediction was "readmission" and the negative predictive value (NPV) is intended as the probability of no readmission within 30 days given ML prediction was "non-readmission". Moreover, false reassurance value (i.e. the probability of readmission given ML prediction was "non-readmission") and false positive prediction (i.e. the probability of no readmission given ML prediction was "readmission") were also reported.

It is well known that the predictive values depend on the prevalence of the outcome in the case series in which the model classification is applied. For allowing clinicians to assess how the prevalence readmissions could affect the predictive performances of ML models and LACE index, we estimated PPV and NPV using different values of prevalence of readmission (between 5 and 50%), by the method adopted in Fagan Nomogram [28].

Results

Between 1st January 2010 and 31st July 2019, 3079 patients aged ≥ 65 years have been discharged from our hospital with a diagnosis of acute HF and were alive after 30 days from discharge, representing the study population. Among them, 394 (12.8%) presented at least one

readmission within 30 days (R), and 2685 (87.2%) did not presented readmissions within 30 days (NR). The mean age of the patients was about 81 years, without significant differences between R and no NR patients ($p = 0.6200$). The proportion of patients with CCI score ≤ 1, of ED visits = 0 were higher in the NR group with respect to R ($p = 0.0031$ and $p = 0.0011$). The proportion of patients with LACE > = 10, liver disease or any tumor were higher in the R group ($p < 0.0001$, 0.0442 and 0.0196). No significant differences were found for the other variables. Baseline characteristics of the study population, according to 30-day readmission status, is detailed in Table 1.

By applying MICE method a balanced and fully valued dataset with 794 records was obtained. No relevant differences emerged in clinical characteristic between training and test cohorts, Table 2.

When using data from the test cohort, the LACE Index scored an AUC of 0.504 (95% CI 0.414–0.594). Among ML models, the XGBoost achieved an overall AUC of 0.803 (95% CI 0.734–0.872). The AUCs (IC95%) were 0.782 (0.711–0.854), 0.776 (0.703–0.848), and 0.786 (0.715–0.857) for Ada Boost Classifier, Random forest, and Gradient Boosting, respectively. Thus, XGBoost showed an improvement of 23% in the discrimination performance against the LACE score (AUCs 0.803 and 0.504, respectively). ROC curves comparing the four ML models with the LACE score are presented in Fig. 2. Values of Sensitivity, Specificity (accuracy) for LACE index, Ada Boost Classifier, Gradient Boosting Classifier, XGBoost and Random forest, were 0.54, 0.59 (0.59); 0.78, 0.71 (0.75); 0.72, 0.69 (0.70); 0.78, 0.75 (0.77); 0.71, 0.70 (0.70), respectively.

A SHAP analysis of the XGBoost model with data from the training dataset was performed to better understand how

Table 1 Baseline characteristics of patient discharged with a diagnosis of decompensated HF (index event) according to all-cause readmissions within 30 days

	Total *N* = 3079	Readmitted *N* = 394	Not-readmitted *N* = 2685	*P*
Age (years), mean, SD Median, IQR	81.5 (7.5) 82.0 (11.0)	81.3 (7.4) 82.0 (11.0)	81.5 (7.5) 82.0 (11.0)	0.6200
Female *n* (%)	1703 (55.3%)	201 (51.0%)	1502 (55.9%)	0.0747
Length of hospitalization stay (days) mean, SD Median, IQR	9.9 (6.0) 8.0 (6.0)	10.3 (6.0) 9.0 (6.0)	9.8 (6.0) 8.0 (6.0)	0.1505
CCI score: mean, SD◆ Median, IQR	1.4 (0.8) 1.0 (1.0)	1.5 (0.9) 1.0 (1.0)	1.4 (0.8) 1.0 (1.0)	0.1048
CCI score ≥ 1	2198 (71.4%)	256 (65.0%)	1942 (72.3%)	0.0031 *
Number of ED visits: mean, SD° Median, IQR	0.3 (0.8) 0.0 (0.0)	0.5 (1.1) 0.0 (0.8)	0.3 (0.7) 0.0 (0.0)	
Number of ED visits = 0	2494 (81.0%)	295 (74.9%)	2199 (81.9%)	0.0011 *
LACE score, mean, SD Median, IQR	9.6 (1.8) 9.0 (2.0)	10.0 (1.8) 10.0 (2.0)	9.5 (1.7) 9.0 (2.0)	
LACE score ≥ 10, n (%)	1300 (42.2%)	211 (53.6%)	1089 (40.6%)	< 0.0001 *
Comorbidities				
Previous myocardial infarction	160 (5.2%)	24 (6.1%)	136 (5.1%)	0.4621
Cerebrovascular disease	38 (1.2%)	5 (1.3%)	33 (1.2%)	> 0.9999
Peripheral vascular disease	10 (0.3%)	1 (0.3%)	9 (0.3%)	> 0.9999
Diabetes without complications	263 (8.5%)	35 (8.9%)	228 (8.5%)	0.8704
Diabetes with end organ damage	4 (0.1%)	0 (0.0%)	4 (0.1%)	> 0.9999
Chronic pulmonary disease	5 (0.2%)	1 (0.3%)	4 (0.1%)	> 0.9999
Mild liver or renal disease	225 (7.3%)	39 (9.9%)	186 (6.9%)	0.0442*
Any tumor (including lymphoma or leukemia)	9 (0.3%)	4 (1.0%)	5 (0.2%)	0.0196*
Dementia	15 (0.5%)	1 (0.3%)	14 (0.5%)	0.7093
Connective tissue disease	2 (0.1%)	1 (0.3%)	1 (0.0%)	0.2396
AIDS	0 (0.0%)	0 (0.0%)	0 (0.0%)	> 0.9999
Moderate or severe liver or renal disease	0 (0.0%)	0 (0.0%)	0 (0.0%)	> 0.9999
Metastatic solid tumor	3 (0.1%)	1 (0.3%)	2 (0.1%)	0.3370

HF heart failure, *SD* standard deviation, *IQR* interquartile range, *CCI* Charlson Comorbidity Index, *ED* emergency department

*$p < 0.05$, ◆ test for comparison was the Mood median test ° due to the high frequency of subjects within one category (i.e. 0 visits) the test was performed only for the proportions of subjects with no ED visits

Table 2 Clinical variables distribution in the training and test cohorts

	Training cohort $N=635$	Test cohort $N=159$
Age (years), mean, SD	81.9, 7.3	82.7, 7.1
Median, IQR	83.0, 11.0	83.0, 10.0
Females: *n* (%)	341 (53.7%)	95 (59.7%)
Length of hospitalization stay (days) mean, SD	10.5, 6.0	10.4, 6.2
Median, IQR	9.0, 5.0	9.0, 7.0
CCI score: mean, SD	1.5 (0.8)	1.4 (0.8)
Median, IQR	1.0 (1.0)	1.0 (1.0)
CCI score ≤ 1	446 (70.2%)	111 (69.8%)
Number of ED visits: mean, SD	0.4 (0.9)	0.4 (0.8)
Median, IQR	0.0 (0.0)	0.0 (0.0)
Number of ED visits = 0	504 (79.4%)	122 (76.7%)
LACE score, mean, SD	9.9 (1.7)	9.8 (1.6)
Median, IQR	9.0 (2.0)	9.0 (2.0)
LACE score ≥ 10, *n* (%)	314 (49.4%)	77 (48.4%)
Comorbidities *n* (%)		
Previous myocardial infarction	26 (4.1%)	8 (5.0%)
Cerebrovascular disease	10 (1.6%)	0 (0.0%)
Peripheral vascular disease	2 (0.3%)	0 (0.0%)
Diabetes without complications	54 (8.5%)	11 (6.9%)
Diabetes with end organ damage	1 (0.2%)	0 (0.0%)
Chronic pulmonary disease	1 (0.2%)	0 (0.0%)
Mild liver or renal disease	51 (8.0%)	12 (7.5%)
Any tumor (including lymphoma or leukemia)	6 (0.9%)	0 (0.0%)
Dementia	3 (0.5%)	0 (0.0%)
Connective tissue disease	1 (0.2%)	0 (0.0%)
AIDS	0 (0.0%)	0 (0.0%)
Moderate or severe liver or renal disease	0 (0.0%)	0 (0.0%)
Metastatic solid tumor	1 (0.2%)	0 (0.0%)

SD standard deviation, *IQR* interquartile range, *CCI* Charlson Comorbidity Index, *ED* emergency department

individual variables in the ML models impact outcome prediction. Lower values of heart rate within the first 48 h from hospitalization ("bpm-in") were associated with a higher risk of readmission within 30 days, and a lower number of previous hospitalizations with a lower risk of readmission. The results of SHAP analysis for these and other variables are shown in Fig. 3.

To allow a complete evaluation of predictive values, Fig. 4 shows the expected values of PPV and NPV calculated for ML predictors and LACE index, for values of 30-day all-cause readmission rate prevalence between 5 and 50%. For example, since the prevalence of readmission in Vimercate Hospital is 13%, the expected PPV was about 35%; consequently, using results from XGBoost classifier provides a gain in prediction of + 22% for predicting readmissions with respect to using only the prevalence. The corresponding false positive prediction was 65%, the expected NPV was 96%, and the corresponding false reassurance was 4%, Fig. 4.

Discussion

Readmissions represent a significant clinical and economic burden in patients with HF, especially in the elderly, and the development of models to identify subjects at high risk for early hospital readmission may allow specific interventions with potential benefits for individuals, the health system and the whole community [29]. Although ML techniques have been deemed to improve the prognostic performance of traditional risk models by addressing the higher order complex interactions between risk factors, direct comparison between both approaches, especially in elderly patients, are lacking.

In the present study, we found that, among elderly patients who were alive at 30 days after an acute HF hospitalization event, the rate of all-cause unplanned readmissions within 30 days was high (13%). The ML models performed moderately well for predicting risk of readmission, with AUCs of 0.79, 0.75, 0.70, and 0.70 for XGBoost Classifier, AdaBoost Classifier, Gradient Boosting Classifier and Random Forest),

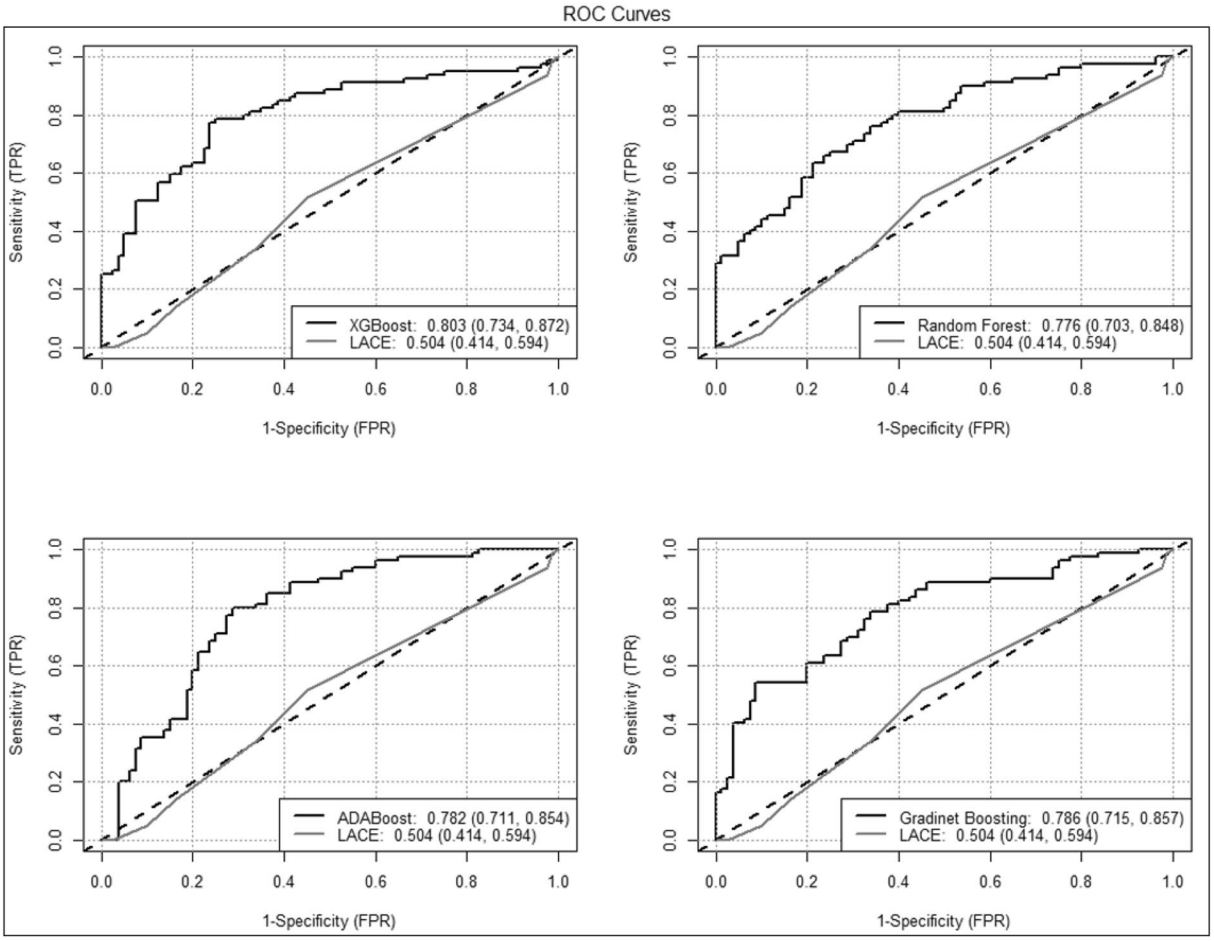

Fig. 2 Receiver operating curves for LACE index and four machine learning approaches in the prediction of 30-day all-cause readmissions after decompensated heart failure hospitalization in elderly patients

respectively. These figures represent a better performance than that achieved by the LACE index (AUC: 0.56), a validated tool developed using a conventional methodological approach.

Some points are worth to be discussed.

First, in our study ML showed a performance for predicting 30-day readmissions in older subjects in line with figures reported for younger populations. Analyzing data of HF patients admitted within a large healthcare system, Golas et al. found that deep unified networks (DUNs) model had the best performance, with an accuracy of 76.4% and an AUC of 0.705, against 0.664, 0.650 and 0.695 for logistic regression, gradient boosting, and maxout networks, respectively [30]. Mahajan et al. explored two ML methods for predicting risk of 30-day readmissions using 56 predictors from electronic health records data of 1778 unique HF patients from 31 hospitals across the United States Achieved AUCs were 0.719 and 0.621 using boosted trees and spike-and-slab regression methods, respectively [31]. In another study, the best performing ML model to predict 30-day and 180-day readmissions was Random Forests, which provided a prediction improvement of 17.8% over logistic regression [32]. Our results show a good predictive performance of ML models in an elderly population.

Second, the prognostic performance of the clinical validated and widespread used LACE index, is poorer than that achieved by ML models. Our results are consistent with those from other studies reporting a limited prediction ability of LACE index in HF patients and in other patient populations [10, 18, 25, 33]. For example, in 253 patients discharged after an acute HF exacerbation included in a retrospective study by Wang et al., with 24.50% of unplanned readmissions to hospital within 30 days after discharge, the C-statistic of logistic regression for the LACE index was 0.5610 (95% CI 0.4771–0.6447) [33]. Among adult medical patients discharged alive from 6 hospitals in Toronto, Canada, 12.6% were readmitted to hospital within 30 days, with high-risk patients (LACE ≥ 10) accounting for 34.0% of the sample but only representing 51.7% of the patients who were readmitted within 30 days. That is, the LACE index

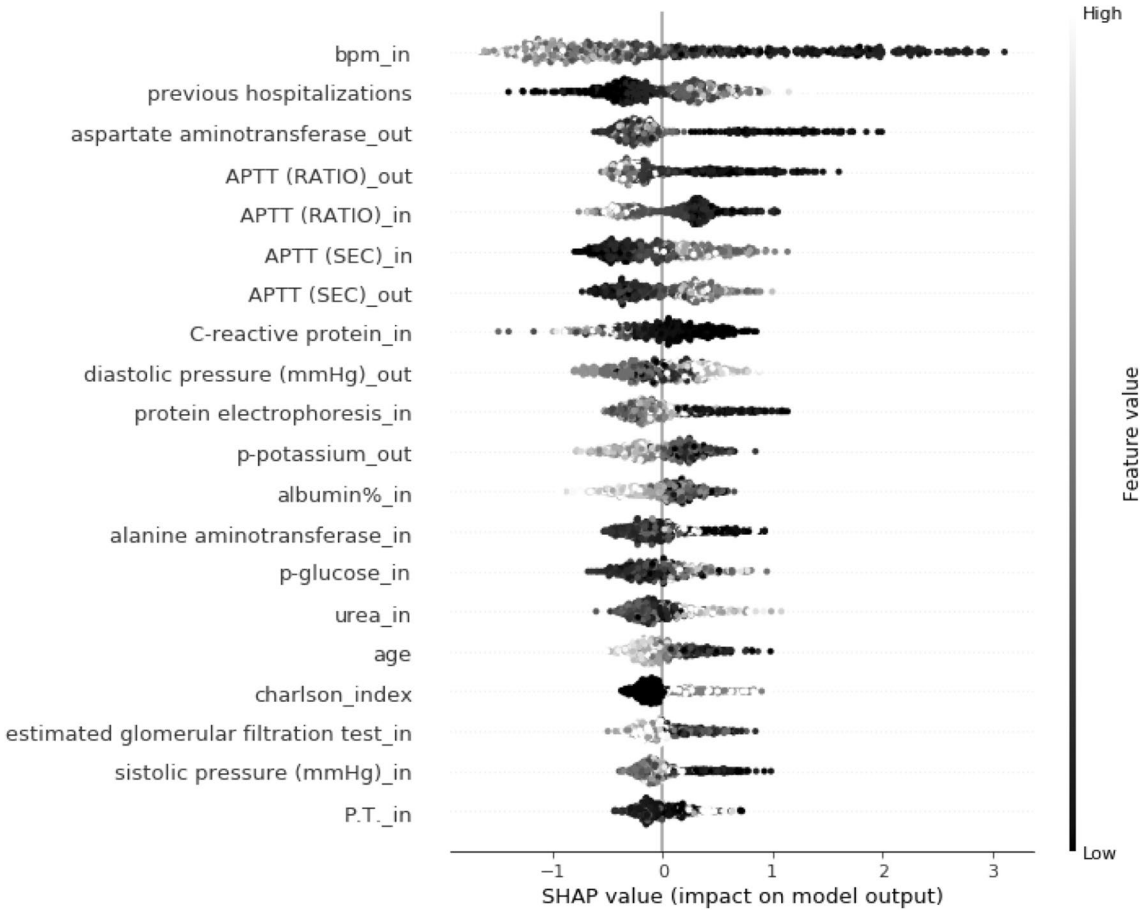

Fig. 3 SHAP value summary plot of the impact on XGBoost model output, for main variables from the training dataset. "in": variable measured within the first 48 h of hospitalization. "out": variable measured within 48 h before discharge. Feature value (vertical axis): variables are ranked in descending order according to their effect on the output (30-day readmissions). Every variable is depicted in lighter grey for higher values and darker grey for lower values. On the horizontal axis are represented the SHAP values, with positive numbers representing a positive prediction effect on the output, and negative values representing a negative effect. For example, lower values (darker grey points) of heart rate within the first 48 h from hospitalization (variable "bpm-in") were associated with positive SHAP values, representing a positive prediction effect on readmissions

was able to identify only half of all discharged patients who were readmitted within 30 days [25]. Our study extends the knowledge on the predicting performance of LACE index to elderly HF subjects, showing a poor ability to identify patients at high risk of all-cause readmissions in this population. These results highlight the better prognostic performance of ML techniques, which appears a valid alternative to some traditional tools for assessing readmission risk in the elderly.

Third, the clinical utility of ML in medicine may be limited by the reduced interpretability and black box nature of algorithms. Therefore, we applied SHAP analysis [26, 27], summarized in Fig. 3, to investigate the impact of variables from the training dataset on XGBoost model output (all-cause hospitalizations within 30 days), allowing clinicians to peek inside the black box and to get a deeper understanding of the most important features from ML models. This may help gaining trust in the predictions and confidence in applying them to clinical care, favoring the clinical translation of ML [34]. We found an increased risk of readmissions among patients with low hear rates within the first 48 h of hospitalization. Interestingly, in keeping with our results and in contrast to patients with chronic heart failure, other studies have shown that a lower hear rate at admission increases in-hospital mortality in patients with acute HF [35]. Lower levels of CRP within the first 48 h of hospitalization, a higher number of previous hospitalizations, or a higher comorbidities burden (measured by CCI) were also associated with a higher risk of 30-day readmissions, Fig. 3.

Finally, a strength of the present study is the comparison of the prognostic performance of ML models with a clinically validated widely used tool, the LACE index. Most studies investigating the ability to predict readmissions in HF patients of ML methods compared their performance with

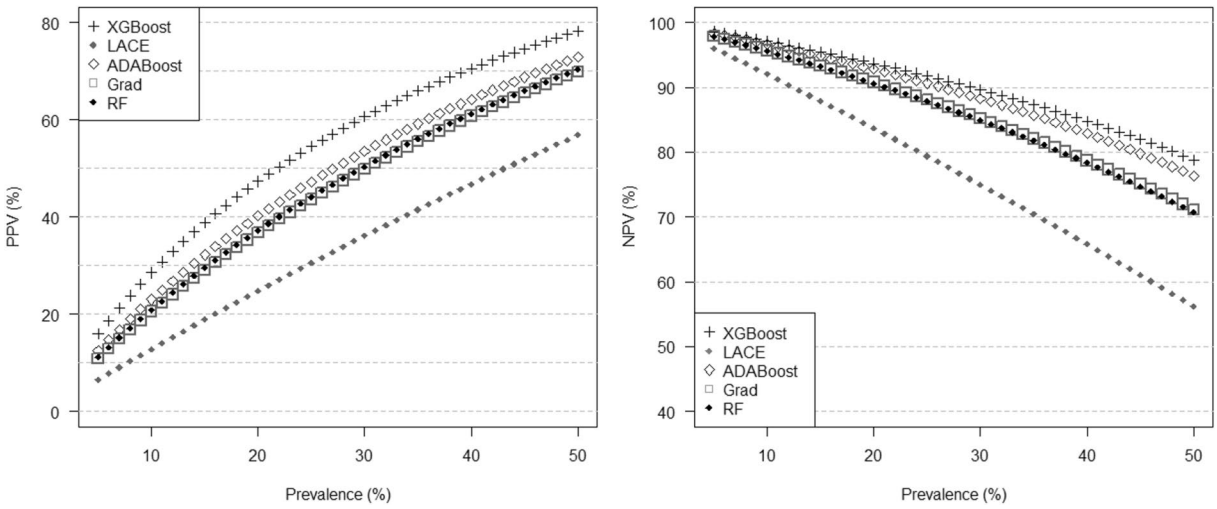

Fig. 4 Expected predictive values calculated for ML algorithms and LACE, as a function of prevalence of all-cause 30-day readmissions. Left panel: Positive predictive value (PPV); right panel: Negative predictive value (NPV)

traditional statistical approaches (e.g. logistic regression), but direct comparisons with validated prognostic tools are scarce. In addition, we developed the ML predictive models using as features common clinical and laboratory data used in the clinical practice, that is, without entering new variables or scores, favoring simplicity and availability, and avoiding time consuming data entering processes. Furthermore, in Fig. 4, estimated PPV–NPV expected values of ML models and LACE index for different values of 30-day all-cause readmission rate prevalence, were provided. This may be a contribution for tailoring the estimates to the actual readmission prevalence of a given institution, to improve the clinical implementation of ML models.

On the other hand, several limitations of the present study should be acknowledged. This is a retrospective, single hospital, internal validation study. Moreover, we included patients with HF after a hospitalization due to acute HF exacerbation. This may represent a kind of selection bias because patients with HF may have been discharged after hospitalization due to other reasons (e.g. infections, trauma, no HF cardiologic diseases). In addition, we did not document variables to describe post-discharge care, such as general practitioners visits or home care nurses activities, which may condition the readmissions rate. Finally, CCI was calculated using medical claims data which may result in under reporting chronic conditions.

Conclusions

Among elderly patients discharged alive after hospitalization due to an acute HF exacerbation event, the rate of all-cause unplanned readmissions within 30 days was high, nearly 13%. ML models performed moderately well for predicting risk of readmission, with XGBoost method showing the higher achieved values of sensitivity, specific and accuracy (0.78, 0.75 and 0.77). The traditional prognostic tool LACE index presented a lower prognostic ability (0.54, 0.59 and 0.53, respectively).

For opening the so-called black box, we performed a SHAP analysis to provide a breakdown of the main variables from the training dataset which were associated with all-cause hospitalizations within 30 days when applying XGBoost ML model. In addition, PPV e NPV values of the different ML models and LACE index for several values of readmission rate prevalence were estimated.

Machine learning models can be proposed as promising tools for the identification of elderly HF patients at high risk of hospitalization, thus enabling care teams to target interventions to improve overall clinical outcomes.

Author contributions HPF: Conceptualization, methodology, formal analysis, investigation, data curation, writing review and editing, supervision, project administration. VE: Conceptualization, ML models development, data curation, writing review and editing. GM: Methodology, formal statistical analysis, data curation, writing review and editing. LP: Investigation, data curation, interpretation of results, writing review and editing. AV: ML models development, data curation, writing review and editing. GD: Conceptualization, investigation, interpretation of results, writing review and editing. MB: Conceptualization, investigation, interpretation of results, writing review and editing. GG: Conceptualization, methodology, interpretation of results, writing review and editing. GM: Conceptualization, methodology, interpretation of results, writing review and editing. PB: Conceptualization, methodology, formal statistical analysis, data curation, writing review and editing, supervision.

Declarations

Conflict of interest The author(s) declare that they have no conflict of interest.

Human and animal rights statement and Informed consent Ethical approval was required to the Institutional Research Ethics Committee (Comitato Etico Brianza), and informed consent was waived given the retrospective non-interventional, observational design.

Writing assistance None to declare. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The work has not been published previously and it is not under consideration for publication elsewhere.

Authorship Authors declare that all have made substantial contributions to the conceptualization and design of the work, the acquisition, analysis, and the interpretation of data. All authors revised critically the draft adding important intellectual content. All authors approve the final version to be published.

References

1. Correale M, Paolillo S, Ruocco G, Palazzuoli A, Nodari S; from Italian Society of Cardiology (SIC) Working Group on Heart Failure: Comprehensive heart failure assessment: a challenge to modify the course of heart failure. Author's reply. Eur J Intern Med. 74:125-126. https://doi.org/10.1016/j.ejim.2020.01.020. Epub 2020 Feb 9. PMID: 32046942
2. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS et al (2016) ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2015:ehw128
3. Bergethon K, Ju C, De Vore A, Hardy NC, Fonarow GC, Yancy CW, Heidenreich PA, Bhatt DL, Peterson ED, Hernandez AF (2016) Trends in 30-day readmission rates for patients hospitalized with heart failure. Circulation. 9:e002594. https://doi.org/10.1161/CIRCHEARTFAILURE.115.002594 (**originally published June 14, 2016**)
4. Ponikowski P, Anker SD, AlHabib KF, Cowie MR, Force TL, Hu S et al (2014) Heart failure: preventing disease and death worldwide. ESC Hear Fail 1:4–25
5. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M et al (2014) The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. J Am Coll Cardiol Elsevier Inc 63:1123–1133. https://doi.org/10.1016/j.jacc.2013.11.053
6. Ricciardi E, La Malfa G, Guglielmi G et al (2020) Characteristics of current heart failure patients admitted to internal medicine vs. cardiology hospital units: the VASCO study. Intern Emerg Med 15:1219–1229. https://doi.org/10.1007/s11739-020-02304-4
7. Belfiore A, Palmieri VO, Di Gennaro C et al (2020) Long-term management of chronic heart failure patients in internal medicine. Intern Emerg Med 15:49–58. https://doi.org/10.1007/s11739-019-02024-4
8. Piñeiro-Fernández J, Fernández-Rial Á, Suárez-Gil R et al (2021) Evaluation of a patient-centered integrated care program for individuals with frequent hospital readmissions and multimorbidity. Intern Emerg Med. https://doi.org/10.1007/s11739-021-02876-9
9. Senni M, Parrella P, De Maria R, Cottini C, Bohm M, Ponikowski P, Filippatos G, Tribouilloy C, Di Lenarda A, Oliva F, Pulignano G, Cicoira M, Nodari S, Porcu M, Cioffi G, Gabrielli D, Parodi O, Ferrazzi P, Gavazzi A (2011) Predicting heart failure outcome from cardiac and comorbid conditions: the 3C-HF score. Int J Cardiol. https://doi.org/10.1016/j.ijcard.2011.10.071
10. Averbuch T, Lee SF, Mamas MA, Oz UE, Perez R, Connolly SJ, Ko DT, Van Spall HGC (2021) Derivation and validation of a two-variable index to predict 30-day outcomes following heart failure hospitalization. ESC Heart Fail. https://doi.org/10.1002/ehf2.13324 (**Epub ahead of print. PMID: 33932113**)
11. Van Walraven C, Dhalla IA, Bell Ch, Etchells E, Stiell IG (2010) Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ 6:51–557
12. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C et al (2011) Risk prediction models for hospital readmission: a systematic review. JAMA 306(15):1688–1698
13. Pocock SJ, Ariti CA, McMurray JJV, Maggioni A, Køber L, Squire IB et al (2013) Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. Eur Heart J 34:1404–1413. https://doi.org/10.1093/eurheartj/ehs337 (**PMID: 23095984**)
14. Van Calster B, Verbakel JY, Christodoulou E, Steyerberg EW, Collins GS (2019) Statistics versus machine learning: definitions are interesting (but understanding, methodology, and reporting are more important). J Clin Epidemiol 116:137–138. https://doi.org/10.1016/j.jclinepi.2019.08.002 (**Epub 2019 Aug 16 PMID: 31425736**)
15. Zhang Z, Zhang H, Khanal MK (2017) Development of scoring system for risk stratification in clinical medicine: a step-by-step tutorial. Ann Transl Med 5(21):436. https://doi.org/10.21037/atm.2017.08.22.PMID:29201888;PMCID:PMC5690964
16. Sun GW, Shook TL, Kay GL (1996) Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol 49(8):907–916. https://doi.org/10.1016/0895-4356(96)00025-x (**PMID: 8699212**)
17. Bagley SC, White H, Golomb BA (2001) Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. J Clin Epidemiol 54:979–985 (**PMID: 11576808**)
18. Levy WC, Anand IS (2014) Heart failure risk prediction models: what have we learned? JACC Heart Fail 2:437–439 (**[PubMed: 25194289]**)
19. Grant L, Joo P, Nemnom MJ et al (2021) Machine learning versus traditional methods for the development of risk stratification scores: a case study using original Canadian Syncope Risk Score data. Intern Emerg Med. https://doi.org/10.1007/s11739-021-02873-y
20. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B (2019) A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 110:12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004 (**Epub 2019 Feb 11 PMID: 30763612**)
21. Eapen ZJ, Liang L, Fonarow GC, Heidenreich PA, Curtis LH, Peterson ED, Hernandez AF (2013) Validated, electronic health record deployable prediction models for assessing patient risk of 30-day rehospitalization and mortality in older heart failure patients. JACC Heart Fail 1:245–251. https://doi.org/10.1016/j.jchf.2013.01.008
22. Dedalus spa. Tabula Clinica. http://www.dedalus.eu (Accessed December 2019).
23. van Buuren S, Boshuizen HC, Knook DL (1999) Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 18:681–694
24. Tibshirani R (1996) Regression Shrinkage and Selection via the lasso. J R Stat Soc Ser B (Methodol) Wiley 58 (1): 267–88 (JSTOR 2346178)

25. Gruneir A, Dhalla IA, van Walraven C, Fischer HD, Camacho X, Rochon PA, Anderson GM (2011) Unplanned readmissions after hospital discharge among patients identified as being at high risk for readmission using a validated predictive algorithm. Open Med 5(2):e104–e111 (**Epub 2011 May 31**)
26. Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160
27. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 30:4765–4774
28. Hayden SR, Brown MD (1999) Likelihood ratio: a powerful tool for incorporating the results of a diagnostic test into clinical decisionmaking. Ann Emerg Med 33(5):575–580. https://doi.org/10.1016/s0196-0644(99)70346-x (**PMID: 10216335**)
29. Andrès E, Talha S, Hajjam M, Hajjam J, Ervé S, Hajjam A (2018) Experimentation of 2.0 telemedicine in elderly patients with chronic heart failure: a study prospective in 175 patients. Eur J Intern Med. 51:e11–e12. https://doi.org/10.1016/j.ejim.2018.02.022 (**Epub 2018 Mar 7. PMID: 29525507**)
30. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, Hisamitsu T, Kojima G, Felsted J, Kakarmath S, Kvedar J, Jethwani K (2018) A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC Med Inform Decis Mak 18(1):44. https://doi.org/10.1186/s12911-018-0620-z
31. Mahajan SM, Mahajan AS, King R, Negahban S (2018) Predicting risk of 30-day readmissions using two emerging machine learning methods. Stud Health Technol Inform 250:250–255
32. Mortazavi BJ, Downing NS, Bucholz EM et al (2016) Analysis of machine learning techniques for heart failure readmissions. Circ Cardiovasc Qual Outcomes 9:629–640
33. Wang H, Robinson RD, Johnson C, Zenarosa NR, Jayswal RD, Keithley J, Delaney KA (2014) Using the LACE index to predict hospital readmissions in congestive heart failure patients. BMC Cardiovasc Disord 14:97. https://doi.org/10.1186/1471-2261-14-97
34. Wongvibulsin S, Wu KC, Zeger SL (2020) Improving clinical translation of machine learning approaches through clinician-tailored visual displays of black box algorithms: development and validation. JMIR Med Inform 8(6):e15791. https://doi.org/10.2196/15791.PMID:32515746;PMCID:PMC7312245
35. Kajimoto K, Sato N, Keida T, Sakata Y, Asai K, Mizuno M, Takano T, Investigators of the Acute Decompensated Heart Failure Syndromes (ATTEND) registry (2014) Low admission heart rate is a marker rather than a mediator of increased in-hospital mortality for patients with acute heart failure syndromes in sinus rhythm. Int J Cardiol. 171(1):98–100. https://doi.org/10.1016/j.ijcard.2013.11.087 (**Epub 2013 Dec 4. PMID: 24342405**)