Heliyon 9 (2023) e20942

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Prediction of 30-day unplanned hospital readmission through survival analysis

Pedro Pons-Suñer [a,*], Laura Arnal [a], François Signol [a], M. Jose Caballero Mateos [b], Bernardo Valdivieso Martínez [b], Juan-Carlos Perez-Cortes [a]

[a] ITI, Universitat Politècnica de València, Camino de Vera, s/n, 46022 València, Spain
[b] Health Research Institute of La Fe University Hospital, Fernando Abril Martorell, Torre A, s/n, 46026 València, Spain

ARTICLE INFO

Keywords:
30-day hospital readmission
Survival analysis
Discharge decision-making
Machine learning
Right censoring

ABSTRACT

Background and Objective: Unplanned hospital readmissions are a severe and recurrent problem that affects all health systems. Estimating the risk of being readmitted the following days after discharge is difficult since many heterogeneous factors can influence this. The extensive work concerning this problem proposes solutions mostly based on classification machine-learning models. Survival analysis methods could make a better match with the assessment of readmission risk and are yet to become well-established in this field.

Methods: We compare different statistical and machine learning survival analysis models trained with right-censored all-cause hospital admission data with covariates available at the moment of discharge. The main focus is on tree-ensemble regression methods based on the assumption of proportional hazards. These models are more thoroughly evaluated at a 30-day time period after discharge, although the actual prediction could be set to any time up to 90 days.

Results: The mean performance obtained by each of the proposed survival models ranges from 0.707 to 0.716 C-Index and 0.709 to 0.72 ROC-AUC at a 30-day time period after discharge. The model with the lower performance on both metrics was Cox Proportional Hazards, while the model marking the upper end on both ranges is an XGBoost Regression model with a Cox objective function.

Conclusions: Our findings indicate that survival models perform well addressing the hospital readmission problem, machine-learning models getting the edge over statistical methods. There seems to be an improvement over classification models when attempting to predict at a 30-day period since discharge, perhaps due to a better handling of cases nearing the 30-day boundary. Some preprocessing steps, such as limiting the observation period to 90 days after discharge, are also highlighted since they resulted in a performance boost.

* Corresponding author.
E-mail addresses: pedropons@iti.es (P. Pons-Suñer), larnal@iti.es (L. Arnal), fsignol@iti.es (F. Signol), maria_jose_caballero@iislafe.es (M.J. Caballero Mateos), valdivieso_ber@gva.es (B. Valdivieso Martínez), jcperez@iti.es (J.-C. Perez-Cortes).

https://doi.org/10.1016/j.heliyon.2023.e20942
Received 21 June 2023; Received in revised form 6 October 2023; Accepted 11 October 2023

Available online 16 October 2023

Table 1
Summary of related work.

Method	Performance	Comments	References
Statistical (LR)	0.58-0.75 C-statistic	Ease of use, very few features.	[4–6]
Classification (ML)	0.5-0.9 ROC-AUC, median 0.68	Rigid separation between 29 and 30-day readmission.	[7–9]
Survival analysis (ML)	0.7 C-index, 1.74 RMSE	Not as widely used in conjunction with ML.	[10,11]

1. Introduction

Hospital readmission can be defined as unplanned admission to the hospital before a particular period (typically 30 days) after the previous discharge. 30-day readmissions are often used as an indicator of healthcare quality, as they impact both patients' health and hospital economic resources.

The percentage of potentially avoidable readmissions ranges widely, with an estimated median of 27% [1]. Higher rates are reported in patients with complex diseases, such as 39% in patients with cancer [2]. Medication, diagnosis and management problems are among the main causes of these unplanned readmissions [3].

Correctly predicting readmissions could aid in improving the quality of care and optimise hospitalisation resource usage. In this regard, a decision support system could assist physicians during their decision-making process, helping them stratify patients according to their readmission risk. This tool would help experts adjust patients' treatment or post-discharge follow-up in a personalised fashion.

In this context, survival analysis could be helpful since it takes advantage of censored events. However, to the best of our knowledge, it has not been as widely used as classification or regression machine learning (ML) models. In this use case, "survival" means "time without readmission" and "censored event" means "patients who have not yet been readmitted by the end of the study".

The main contributions of this paper are twofold. On the one hand, a comparative study of survival analysis models applied to all-cause hospital readmission prediction. On the other hand, some methodological steps are highlighted since they have significantly improved the performance of the models. These steps could also be used in similar tasks, perhaps benefiting from this performance boost.

The remainder of this paper is structured as follows: the Related work subsection describes briefly other approaches to this task. Section Material and methods describes the dataset, data preparation pipeline and models used. The results section shows the performance achieved for every model, which is later discussed in the Discussion section. The final section outlines the conclusions and future work of this study.

1.1. Related work

This task has been tackled widely in the literature using both traditional statistics and machine learning models. Logistic regression is one of the most used models for this task, being LACE [4] and HOSPITAL [5,6] two of the most widely used. LACE scores readmission or death risk by measuring Length of stay, Acuity of the admission, Charlson comorbidity index score, and Emergency department use in the six months before admission. On the other hand, HOSPITAL index uses seven features to measure this risk: Hemoglobin at discharge, discharge from an Oncology service, Sodium level at discharge, Procedure during the index admission, Index Type of admission, number of Admissions during the last 12 months, and Length of stay.

Some studies try to predict hospital readmission through a classification approach, using models such as support vector machine (SVM), random forest (RF), or gradient boosting techniques (GB) [7,8]. Recent surveys [9] discovered that performance ranges widely from 0.5 to 0.9 Area Under the Receiver Operating Characteristic Curve (ROC-AUC), with a median of 0.68. However, this approach places a rigid separation: 29-day and 30-day readmission have probably similar risks but are from opposite classes in this approach.

Another approach consists of using regression analysis tools. Besides avoiding a strict separation between classes, regression methods provide more flexibility. For instance, patients could be stratified into different groups depending on the predicted days until readmission and take the corresponding mitigation measures for each case. In readmission, the preferred approach is using survival analysis, since these techniques can handle censored events.

Previous works using survival models have used random survival forests and neural networks achieving 0.7 of concordance index (C-index) [10]. In [11], a survival Bayesian additive regression kernel model is proposed for modelling 30-day hospital readmission data, reporting 1.74 root-mean-square error (RMSE) between the observed and posterior survival predicted outcome.

Table 1 summarises the information in this section. It is worth noting that comparing studies is difficult due to the differences in terms of datasets (population, features included, inclusion criteria, etc.).

2. Material and methods

2.1. Dataset

In this paper, the dataset is based on the one presented in [12], but enriched with new variables that are detailed below. This dataset consists of approximately 33000 episodes of 21800 patients. This pseudo-anonymised and non-public dataset is protected by the European GDPR and Spanish LOPDGDD laws.

The features included in our dataset can be classified into five groups:

- **Consumptions**: aggregated consumption of services and tests, such as visits to the hospital or urgency departments previous to the index episode.
- **Laboratory**: numerical results from urine and blood laboratory tests. Each episode is associated with the result of the last available test, with a maximum period of 90 days.
- **Treatments**: active principles of drugs administered up to 90 days before the hospitalisation episode and those prescribed at the moment of discharge. Treatments were grouped by Anatomical Therapeutic Chemical (ATC) Classification System codes. Only outpatient treatments are included in this study.
- **Hospitalisation**: features regarding hospitalisation episodes and their respective context, such as length of stay and the main cause of admission. This also contains patient data, such as gender or age at admission.
- **Comorbidities**: patient comorbidities diagnosed before the evaluated hospital episode. These are coded in 18 different categories according to Charlson Comorbidity Index, such as kidney disease or diabetes with chronic complications.

Compared to [12], this study includes new information:

- **Route of administration** of each prescripted treatment. This includes 27 categories such as oral, subcutaneous or nasal. Polypharmacy and medication regimen complexity have already been described in other studies ([13], [14]) as a potential cause of readmission or, at least, deserving great attention in future work.
- **Barthel index** [15]. This scale measures a patient's performance in daily activities, a higher score meaning a higher patient's independence. All 10 variables needed to compute the index and the final score are included in the models.
- **Diagnosis and procedures codes**. The dataset was extended to include not only the main diagnosis or procedure but also other secondary ones that could occur during the episode.

2.2. Data preprocessing and feature extraction

2.2.1. Discard episodes with exitus

As the date of patient exitus, for those who apply, is available to us, we can give special consideration to these cases. When considering these patients, the question of which label to assign to them arises, because these are not strictly speaking hospital readmissions but nevertheless correspond to some kind of complication. As this work focuses on readmission risk, it has been decided to remove exitus entries according to the following rules:

- Patients who have the same exitus date and discharge date. As the patient's decease is probably the cause of discharge, we consider these cases not suitable for the proposed analysis.
- Episodes whose time difference between discharge and death is less than our observed time horizon. These could be considered censored at the date of exitus, as we do not know if those patients could have been readmitted had they not died before. However, since we cannot ascertain the exitus itself being that different from the actual event of readmission, we decided to exclude these cases.
- There is an exception to the previous rule: episodes with a difference between discharge and exitus dates inferior to the observation period, but having intermediate registered episodes, are valid since we know the patient had been readmitted before their death.

Fig. 1 illustrates three hypothetical cases for three different patients with observed exitus dates where each of the rules above applies.

2.2.2. Missing data imputation and feature quality filters

Missing data are present in Barthel index and laboratory results. The optimal imputation method for each feature is automatically chosen from a list of univariate (mean, mode and zero imputing) and multivariate imputers (Iterative Random Forest and Bayesian Ridge Regression), according to the best imputation performance on cross-validation (CV) of the training set.

2.2.3. Feature extraction

After performing the previous procedures to ensure data quality, some feature extraction steps are carried out. Unique ATC codes and administration routes from each patient's treatment information are aggregated in two count variables, accounting for polipharmacy and complexity of treatment respectively. Charlson comorbidities of each patient diagnosed as of the moment of

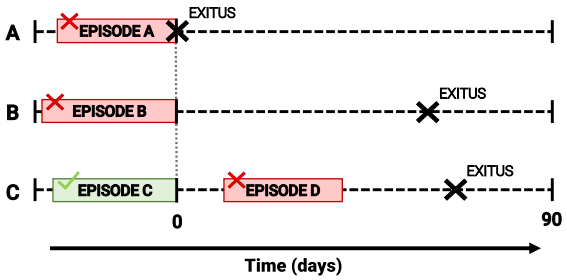


Fig. 1. Rules for removing episodes based on exitus date. The boxes represent hospitalization periods (A) exitus date and date of discharge are the same, (B) exitus happens in less than 90 days after discharge, and (C), same as B, but there is a following episode, making Episode C suitable for the study.

discharge are also summed up to create a total comorbidities count variable. Lastly, the difference in time between the evaluated hospital episode and the previous one is calculated in order to determine if the current episode is, in itself, an early readmission.

Furthermore, One-Hot Encoding is applied to the month of admission, cause and department of admission codes, and code of the assigned nursing unit.

2.2.4. Criterions for quality-based feature filtering

One feature of each pair with a Pearson Correlation greater than 0.8 is recursively removed. Besides, those with more than 99% of values equal to the mode value, deemed as quasi-constant, are filtered out after the missing data imputation.

2.2.5. Use of post-discharge information

Post-discharge information, such as discharge prescriptions, which are important to assess medication reconcilliation, can pose an availability problem when predicting at the moment of discharge. Systems that update some databases at each day's end could not grant some of this information at a given point in time, which would lead to these features being missing values. Recent studies have shown particular attention to discharge medication as a potential risk factor when attempting to predict unplanned readmissions ([16], [17]). For instance, in [7], a considerable performance drop was detected while predicting only with pre-discharge variables (0.69 ROC-AUC) versus an all-variables prediction (0.81 ROC-AUC). As the technical issues which may cause a data system to be unable to retrieve post-discharge information do not apply to every one of them, we have decided to keep these features in our dataset but evaluate the performance degradation knowing that their possible absence could lead to less accurate results.

2.3. Model design and evaluation

2.3.1. Survival models

As aforementioned in Section 1, the focus of this work is placed on a survival analysis approach at modelling time-to-readmission. In that regard, we employ both statistical and machine-learning-based models to investigate the association between predictor variables and "survival" time, that is, the time until the readmission event is observed, if any.

The Cox Proportional Hazards (CPH) model [18] is a regression model that simultaneously evaluates the effect of several covariates on the expected survival time. The CPH model aims to adjust a hazard function, denoted by $h(t)$, which can be interpreted as the probability of reaching the event (in this case, readmission) at time t. Being $x_1, x_2, \cdots, x_n$ a set of n covariates, and $\beta_1, \beta_2, \cdots, \beta_n$ their associated risk coefficients, which the model needs to estimate in order to predict the parametric risk score (ρ), the hazard (h) function for a patient at any time t is calculated as follows:

$$h(t) = h_0(t) \cdot \exp(\rho) \tag{1}$$

$$\text{In CPH: } \rho = \sum_{i=1}^{n} x_i \beta_i \tag{2}$$

In Equation (1), t is the survival time and $h_0(t)$ is the baseline hazard, common to all patients, that changes over time and functions as an intercept for each value of t. In CPH, the parametric risk score ρ is calculated as a linear combination of covariates and risk coefficients (Equation (2)). The coefficients $\exp(\beta_i)$ are the hazard ratios: if β_i is greater than one, the value of its associated i^{th} covariate is negatively correlated with survival time and positively with event hazard. This model, while simple and well-established, has its limitations. As the baseline hazard is the only time-dependent component, the hazard contribution of each covariate is assumed to be constant over time and proportional between subgroups of individuals. In other words, a single covariate's coefficient can only be positively or negatively correlated with hazard, regardless of time, and cannot interact with other covariates. The final prediction is the product of the baseline function and the exponential function of a linear combination of the risk coefficients and their corresponding covariates.

Besides this traditional statistical technique, we use other well-established machine learning models. Specifically, we use three models built upon the assumption of proportional hazards (PH) but based on tree ensembles, which could benefit from their non-linearity property and handle more complex relationships between covariates of heterogeneous sources and types. Additionally, a model based on Accelerated Failure Time (AFT) has been tested, adding another perspective different from Cox-based models. AFT

Table 2
Hyperparameter search spaces for each model.

Model	Parameter	Search space
CPH	alpha	0.001 — 1
	n_iter	10 — 100
	ties	[breslow,efron]
RSF	n_estimators	10 — 100
	min_samples_split	2 — 10
	min_samples_leaf	1 — 10
	max_depth	5 — 15
	max_features	[sqrt,log2]
XGBcox	n_estimators	10 — 200
	learning_rate	0.01 — 0.3
	max_depth	5 — 10
	grow_policy	[depthwise, lossguide]
	booster	[gbtree,dart]
	subsample	0.8 — 1
	colsample_bytree	0.8 — 1
	colsample_bylevel	0.8 — 1
	colsample_bynode	0.8 — 1
	reg_alpha	0.001 — 0.1
	reg_lambda	0.001 — 0.1
	min_child_weight	1 — 10
GBS	n_estimators	10 — 100
	learning_rate	0.01 — 0.3
	max_depth	5 — 10
	min_samples_leaf	1 — 10
	min_samples_split	2 — 10
	subsample	0.8 — 1
	max_features	[sqrt,log2]
	dropout_rate	0 — 0.2
XGBaft	num_boost_round	10 — 200
	subsample	0.8 — 1
	colsample_bytree	0.8 — 1
	colsample_bylevel	0.8 — 1
	colsample_bynode	0.8 — 1
	reg_alpha	0.001 — 0.1
	reg_lambda	0.001 — 1
	booster	[gbtree,dart]
	learning_rate	0.01 — 0.3
	max_depth	5 — 10
	grow_policy	[depthwise, lossguide]
	aft_loss_distribution	[normal,logistic, extreme]
	aft_loss_distribution_scale	0.5 — 2

models allow the risk contribution of covariates to vary through time, adding another layer of complexity to the hazard function calculation [19].

In summary, we trained and evaluated five models: Cox Proportional Hazards (CPH), Random Survival Forest (RSF) [20], Gradient Boosting Survival Analysis (GBS) [20], and Extreme Gradient Boosting Regression using the Survival Cox objective function (XGBcox) and the Survival Accelerated Failure Time objective function (XGBaft) [21].

Random Survival Forests are based on the original Random Forest models [22]. Here, the concept is adapted to survival analysis by taking into account censored data in the splitting rules, branching the data into nodes with similar survival outcomes. Gradient Boosting [24], unlike Random Forest, works by cascading many weaker decision tree estimators, where each one depends on the performance of the previous tree and tries to improve it where it failed. The list goes on until a maximum number of trees is reached or until no improvement is registered. The final prediction is a weighted collective decision made by all the weaker predictors, where the best ones contribute the greatest. Extreme Gradient Boosting [25] is an optimised implementation of Gradient Boosting, and uses different regularization techniques that usually lead to better performance with good computational efficiency.

All the models used in this study are hyper-tuned with Random Search. The hyperparameter search spaces explored by each type of model are provided in Table 2. After determining the best configuration for each model, they were trained and evaluated using 10-fold cross-validation.

2.3.2. Evaluation metric

Regarding survival-analysis models evaluation, the Harrel's Concordance-Index [26], analogous to ROC-AUC in classification tasks, measures the rank correlation between the model-predicted risks and the true observed times until event. This score also exploits censored data, using all comparable pairs in its calculation, that is, pairs of examples in which we know the event for one of them was observed before the other. However, the most popular Harrel's C-Index has been observed to skew results when the amount of censored examples is high [27]. An alternative method, Uno's C-Index, was proposed to solve this issue [27], taking both censored distributions in train and test sets into account. Since our censored data fraction is high (above 82% censored cases when the time is limited to 90 days), all C-Index results shown in Section 3 are calculated with Uno's formula.

The proportional hazards assumptions are tested in order to avoid possible biases that could harm cox-based models performance. A statistical test based on scaled Schoenfeld residuals, followed by a visual inspection of the survival curves, validates the CPH assumptions.

2.3.3. Restricting right-censored events

In survival analysis, one must define the initiating and terminating events. In our case, the initiating event is the patient's discharge from the hospital, marking the end of the episode of reference. The possible terminating event is the hospital readmission, marking the beginning of the next episode. We could then measure the probability of survival at any time t after discharge, that is, the probability of not being readmitted in the time up to t.

Due to data in this study ranging from 2015 to 2018, we can find both early and later readmissions (several months or even years after). From a practical point of view, the prediction of early readmissions is more interesting insofar as it is less probable that a late readmission could have been conditioned by a previous hospitalisation. Since the typical period to consider unexpected

Table 3
Statistical and machine learning models and the used implementations.

Algorithm	Implementation
CPH	*sksurv.linear_model.CoxPHSurvivalAnalysis*
GBS	*sksurv.ensemble.GradientBoostingSurvivalAnalysis*
RSF	*sksurv.ensemble.RandomSurvivalForest*
XGBcox, XGBaft	*xgboost.XGBRegressor*

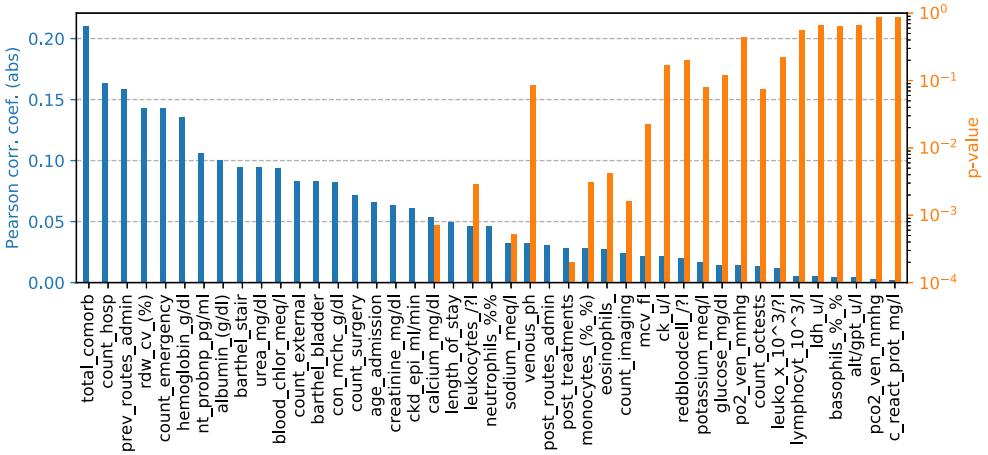


Fig. 2. Absolute Pearson Correlation Coefficient (blue) and matching p-value (orange) between the target feature (days from discharge to readmission) and each continuous predictive feature. Although lower p-values exist, the scale is limited to 10^{-4}. Only non-censored data was used in this test.

readmissions is 30 days, we conducted and compared two experiments. The first does not make any difference between early and late readmissions. The second considers an event as censored if the subsequent hospitalisation happens after 90 days, considering that later readmissions are probably totally unrelated to the previous hospitalisation. In this second setup, we expect the model to focus and make better predictions on short-term readmissions at the cost of lowering attention in long-term predictions.

2.4. Software

All preprocessing, modelling and evaluation methods proposed in this paper have been performed using tools available for Python 3.6 onwards. Notable packages used for creating the missing-data imputers and the survival-analysis models used in this work, as well as measuring their performance in various aspects, include scikit-learn [28], lifelines [29], scikit-survival [20] and xgboost [30]. Table 3 details the source of implementation of each estimator appearing in this paper.

3. Results

3.1. Dataset description

After performing the data quality assessment and feature extraction steps described in Section 2.2, the resulting dataset contains 184 features. Among these, we find:

- 106 continuous variables, 63 of which are pharmacy treatment and route of administration count variables. The remaining 43 are original and derived variables sourcing from all groups of variables save diagnosis and procedure codes. Statistical descriptors of these last features are displayed in Table 4. Additionally, they are compared against the dependent variable (days from discharge to readmission) using Pearson's Correlation Coefficient. The results of this test are presented in Fig. 2.
- The 63 pharmacy variables mentioned in the point before are treated as categorical, binary features. Although these are count variables, ranging from zero to any positive integer number, the most common values are zero and one, more being the exception. After binarising, a code receiving the unit value means that the patient was prescribed one or more treatments with the ATC or route of administration identified by said code. This set of features is described in Table 5. ANOVA tests were used to study the differences in their effects on the dependent variable, the results of which are displayed in Fig. 3.
- 78 binary features, most of which are the result of one-hot encoding original categorical features. These are described in Table 6, and the results of the ANOVA tests studying their effect on the target feature are presented in Fig. 4.

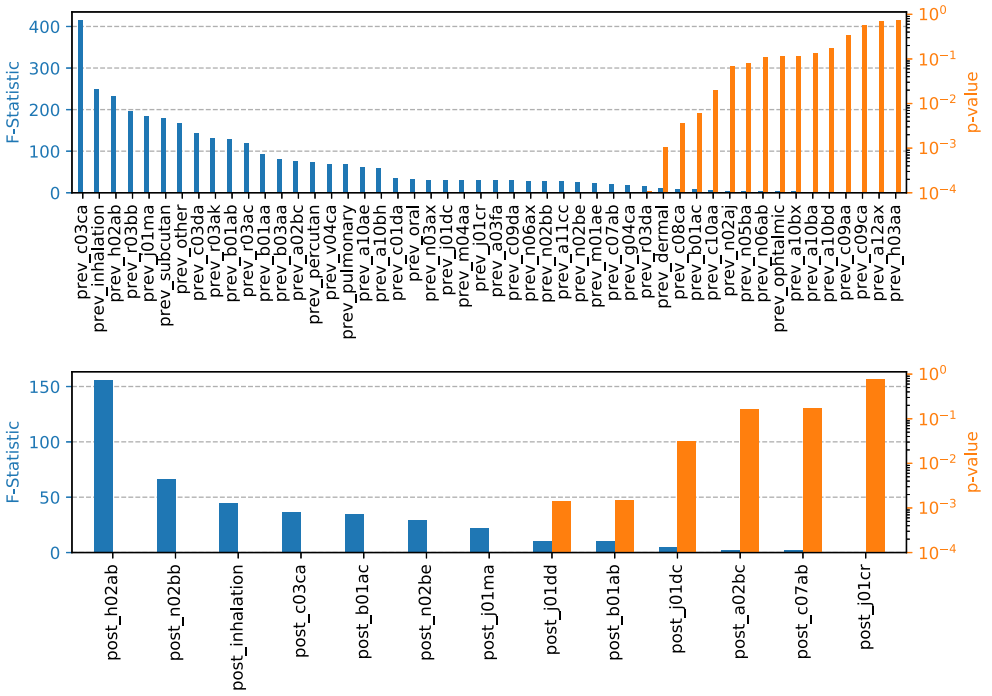


Fig. 3. F-Statistic from a One-Way ANOVA test (blue) and matching p-value (orange) between the target feature (days from discharge to readmission) and each pharmacy feature (categorised as 0 or ≥ 1 treatments). Although lower p-values exist, the scale is limited to 10^{-4}. Only non-censored data was used in this test.

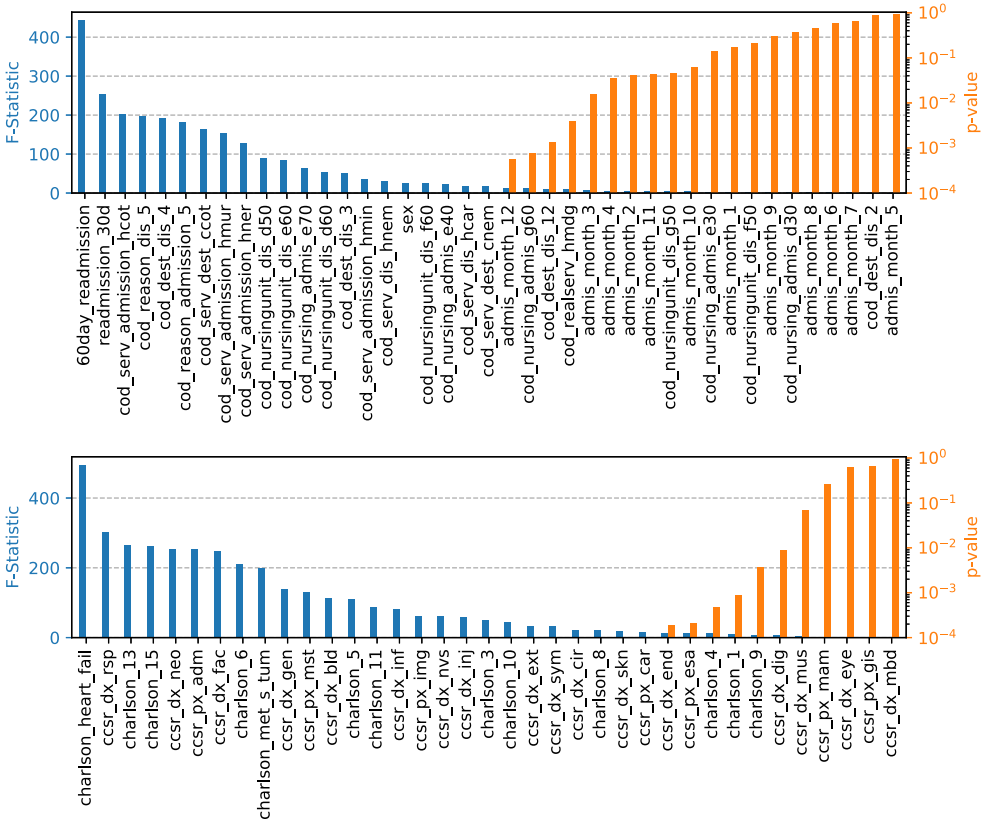


Fig. 4. F-Statistic from a One-Way ANOVA test (blue) and matching p-value (orange) between the target feature (days from discharge to readmission) and each binary predictive feature. Although lower p-values exist, the scale is limited to 10^{-4}. Only non-censored data was used in this test.

Table 4
Descriptive statistics for all continuous variables used in the final models. Descriptors for all episodes and those that end up in 30-day readmissions are presented side by side, as well as their observed 5th to 95th percentile ranges.

Feature	mean (std),All	mean (std),R30d	5%-95%
age_admission	67.6 (18.9)	72.0 (16.3)	29.0-90.8
albumin_(g/dl)	3.6 (0.5)	3.4 (0.5)	2.6-4.4
alt/gpt_u/l	32.5 (80.6)	31.7 (52.8)	7.0-92.6
barthel_bladder	7.4 (4.0)	7.1 (4.1)	0.0-10.0
barthel_stair	5.9 (4.2)	5.5 (4.2)	0.0-10.0
basophils_%_%	0.4 (0.3)	0.4 (0.3)	0.1-1.0
blood_chlor_meq/l	102.2 (4.4)	101.3 (5.0)	95.0-109.0
c_react_prot_mg/l	47.7 (64.8)	52.9 (67.5)	0.8-184.8
calcium_mg/dl	8.9 (0.7)	8.8 (0.8)	7.7-10.0
ck_u/l	165.8 (626.7)	143.5 (465.4)	19.0-470.0
ckd_epi_ml/min	78.0 (29.2)	71.2 (31.8)	25.0-121.0
con_mchc_g/dl	32.6 (1.3)	32.3 (1.3)	30.4-34.7
count_emergency	7.2 (8.5)	9.8 (10.9)	1.0-22.0
count_external	25.9 (45.2)	34.6 (62.5)	0.0-88.0
count_hosp	3.0 (4.1)	4.5 (5.1)	0.0-11.0
count_imaging	0.7 (1.8)	0.6 (1.7)	0.0-3.0
count_octests	0.3 (1.0)	0.4 (1.4)	0.0-2.0
count_surgery	0.2 (0.6)	0.1 (0.4)	0.0-1.0
creatinine_mg/dl	1.0 (1.0)	1.2 (1.3)	0.5-2.0
eosinophils_	2.3 (2.4)	2.2 (2.4)	0.0-6.5
glucose_mg/dl	117.0 (49.3)	121.5 (62.0)	75.0-208.4
hemoglobin_g/dl	12.2 (2.1)	11.5 (2.0)	8.9-15.6
ldh_u/l	296.2 (237.0)	324.9 (340.9)	145.0-578.0
length_of_stay	6.8 (7.0)	8.0 (8.1)	1.0-18.0
leuko_x_10^3/?l	8.9 (6.5)	9.2 (9.0)	4.1-15.7
leukocytes_/?l	103.7 (180.7)	124.7 (198.4)	0.0-500.0
lymphocyt_10^3/l	1.8 (3.6)	1.5 (1.1)	0.6-3.2
mcv_fl	90.2 (6.4)	90.3 (6.9)	80.1-100.0
monocytes_(%_%)	9.1 (3.7)	9.3 (4.9)	4.5-14.7
neutrophils_%%	66.7 (13.0)	68.9 (14.0)	45.0-86.6
nt_probnp_pg/ml	3269.7 (5834.5)	4468.2 (7218.9)	59.1-14050.6
pco2_ven_mmhg	42.9 (7.9)	42.8 (8.0)	31.3-56.1
po2_ven_mmhg	51.8 (31.0)	52.7 (30.6)	20.7-116.0
post_routes_admin	1.1 (0.8)	1.2 (0.8)	0.0-2.0
post_treatments	2.2 (1.9)	2.4 (2.0)	0.0-6.0
potassium_meq/l	4.2 (0.5)	4.2 (0.6)	3.3-5.0
prev_routes_admin	1.9 (1.4)	2.4 (1.4)	0.0-4.0
rdw_cv_(%)	14.4 (2.1)	15.2 (2.5)	12.1-18.5
redbloodcell_/?l	64.6 (92.0)	79.7 (100.7)	0.0-250.0
sodium_meq/l	140.1 (3.6)	139.5 (4.1)	134.0-145.0
total_comorb	2.5 (2.0)	3.4 (2.1)	0.0-6.0
urea_mg/dl	43.4 (28.8)	51.1 (35.6)	15.5-98.4
venous_ph	7.4 (0.1)	7.4 (0.1)	7.3-7.5

3.2. Survival analysis

In this section, we evaluate the performance of the statistical and machine-learning models mentioned in Section 2.3. Fig. 5a shows the Uno's C-Index score obtained with these models while Fig. 5b shows the time-dependent ROC-AUC, evaluating predictions from 10 to 90 days with 5 days increments. We focus on the 30-day prediction, as it is the usual period used in classification tasks. Fig. 6 shows greater detail of the ROC-AUC at a 30-day cutoff, also showing the score distribution across evaluated folds. On average, the best performance is achieved with XGBcox. In the following sections, unless stated, comparisons are done with XGBcox model.

Regarding running time, Table 7 shows the mean training times, as well as single-prediction times, based on the measurements for each iteration in a 10-Fold experiment.

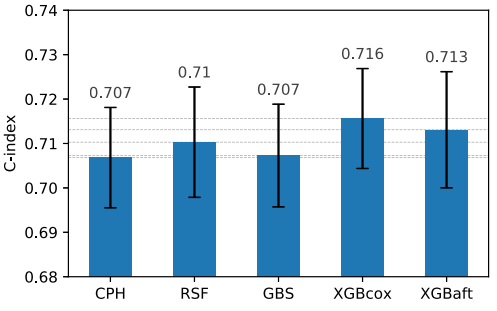
3.3. Right-censoring to 90 days

Fig. 7 shows the effect of right-censoring the observed times until the event to a maximum of 90 days on the performance score distribution at 30 days since discharge. Not only the time value is adjusted, but the "event observed" mark is also set to "censored" as explained in 2.3.3, as we deem readmissions over 90 days after discharge as probably unrelated to the last hospitalisation episode.

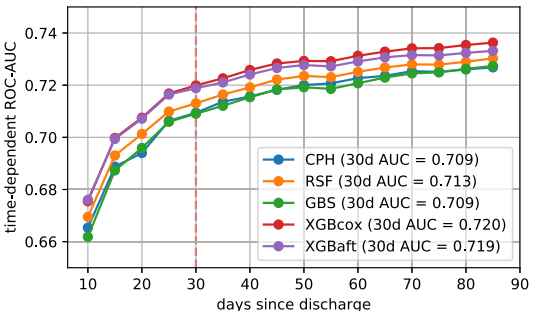
Table 5
Distributions of all pharmacy variables used in the final models. These are discrete in nature, counting the number of drugs prescribed with an active principle or route of administration. Percentage of patients with one or more drugs prescribed sharing the same code, for all episodes and those that end up in 30-day readmissions, are presented side by side.

Feature	≥ 1(%),All	≥ 1(%),R30d
post_a02bc	9.5	9.8
post_b01ab	20.0	18.7
post_b01ac	7.0	5.7
post_c03ca	7.7	11.6
post_c07ab	5.2	5.9
post_h02ab	13.7	19.1
post_j01cr	6.8	5.9
post_j01dc	5.4	5.1
post_j01dd	5.6	7.1
post_j01ma	14.3	15.9
post_n02bb	10.2	7.5
post_n02be	16.6	13.6
post_inhalation	5.8	8.1
prev_a02bc	47.6	60.6
prev_a03fa	6.1	8.5
prev_a10ae	8.5	12.4
prev_a10ba	8.1	8.8
prev_a10bd	7.4	7.8
prev_a10bh	6.5	9.9
prev_a10bx	5.5	7.2
prev_a11cc	9.0	11.4
prev_a12ax	6.9	8.3
prev_b01aa	12.8	17.6
prev_b01ab	8.0	14.0
prev_b01ac	23.8	29.1
prev_b03aa	7.1	11.1
prev_c01da	6.0	9.0
prev_c03ca	23.2	36.8
prev_c03da	6.3	10.5
prev_c07ab	17.1	20.9
prev_c08ca	14.5	19.4
prev_c09aa	13.2	14.5

Feature	≥ 1(%),All	≥ 1(%),R30d
prev_c09ca	13.6	15.8
prev_c09da	8.9	9.3
prev_c10aa	33.8	37.5
prev_g04ca	8.7	10.9
prev_h02ab	12.1	20.3
prev_h03aa	6.3	6.4
prev_j01cr	8.2	10.3
prev_j01dc	7.2	10.8
prev_j01ma	14.5	22.8
prev_m01ae	11.1	10.4
prev_m04aa	7.6	11.8
prev_n02aj	10.4	10.5
prev_n02bb	17.8	22.3
prev_n02be	31.4	38.2
prev_n03ax	9.6	13.0
prev_n05ba	29.2	34.5
prev_n06ab	10.8	12.3
prev_n06ax	10.6	13.2
prev_r03ac	9.5	14.7
prev_r03ak	11.1	15.9
prev_r03bb	9.9	16.6
prev_r03da	5.9	7.8
prev_v04ca	10.7	15.2
prev_dermal	11.8	13.6
prev_inhalation	19.3	29.4
prev_ophtalmic	10.4	12.3
prev_oral	87.2	94.2
prev_other	14.7	22.2
prev_percutan	6.9	11.5
prev_pulmonary	5.1	7.0
prev_subcutan	18.2	27.6



(a) Uno's C-Index: mean scores and standard deviation obtained with 10-Fold CV.



(b) Time-dependent ROC-AUC (average of a 10-Fold CV at each set time).

Fig. 5. Performance evaluation and comparison of the trained survival models by 10-fold cross validation.

3.4. Patient exitus

Fig. 8 compares the results obtained when keeping all cases that end with exitus in the dataset, but censoring their time-to-event to the day of decease, versus filtering exitus as explained in 2.2.1. This results in censoring or removing 7.24% of episodes in our dataset.

Table 6
Distributions of all binary variables used in the final models. Percentage of unit value for all episodes and those that end up in 30-day readmissions are presented side by side.

Feature	= 1(%),All	= 1(%),R30d	Feature	= 1(%),All	= 1(%),R30d
60d_readmission	12.6	26.1	charlson_11	11.8	16.8
admis_month_1	9.8	10.3	charlson_13	24.6	36.9
admis_month_10	8.6	7.5	charlson_15	23.9	35.9
admis_month_11	8.3	8.6	charlson_3	14.3	19.8
admis_month_12	8.6	9.3	charlson_4	21.4	25.0
admis_month_2	8.7	8.3	charlson_5	10.2	13.7
admis_month_3	9.0	8.8	charlson_6	36.6	46.3
admis_month_4	8.0	7.7	charlson_8	8.3	10.9
admis_month_5	8.3	8.5	charlson_9	12.8	15.0
admis_month_6	7.6	7.8	charlson_heart_fail	25.6	40.1
admis_month_7	7.6	7.8	charlson_met_s_tum	4.9	10.6
admis_month_8	7.5	8.1	cod_dest_disch_12	20.0	18.3
admis_month_9	7.8	7.3	cod_dest_disch_2	15.7	15.7
ccsr_dx_bld	13.1	19.9	cod_dest_disch_3	51.7	47.8
ccsr_dx_cir	67.6	75.7	cod_dest_disch_4	7.8	12.4
ccsr_dx_dig	29.6	31.9	cod_nursing_admis_d30	5.1	5.5
ccsr_dx_end	57.6	65.3	cod_nursing_admis_e30	5.2	4.6
ccsr_dx_ext	13.0	9.1	cod_nursing_admis_e40	8.4	10.4
ccsr_dx_eye	5.2	6.0	cod_nursing_admis_e70	10.0	13.1
ccsr_dx_fac	69.7	81.8	cod_nursing_admis_g60	8.1	6.6
ccsr_dx_gen	30.4	40.9	cod_nursing_disch_d50	5.5	2.9
ccsr_dx_inf	19.2	25.1	cod_nursing_disch_d60	5.9	3.3
ccsr_dx_inj	20.9	16.5	cod_nursing_disch_e60	6.5	3.6
ccsr_dx_mbd	24.0	24.3	cod_nursing_disch_f50	5.3	6.3
ccsr_dx_mus	18.2	19.0	cod_nursing_disch_f60	6.9	8.3
ccsr_dx_neo	13.8	23.5	cod_nursing_disch_g50	6.7	5.8
ccsr_dx_nvs	25.1	29.5	cod_realserv_hmdg	5.8	5.3
ccsr_dx_rsp	33.0	44.7	cod_reason_admission_5	10.1	5.1
ccsr_dx_skn	5.3	6.2	cod_reason_disch_5	9.8	15.1
ccsr_dx_sym	23.7	29.2	cod_serv_admis_hcot	10.7	5.0
ccsr_px_adm	45.6	55.3	cod_serv_admis_hmin	7.3	9.5
ccsr_px_car	6.9	6.6	cod_serv_admis_hmur	11.6	16.2
ccsr_px_esa	5.3	7.6	cod_serv_admis_hner	5.6	2.6
ccsr_px_gis	10.7	10.1	cod_serv_dest_ccot	8.0	3.9
ccsr_px_img	58.6	56.0	cod_serv_dest_cnem	7.6	8.3
ccsr_px_mam	14.7	14.8	cod_serv_dis_hcar	12.6	12.2
ccsr_px_mst	10.4	5.8	cod_serv_dis_hnem	12.4	14.1
charlson_1	11.0	12.4	readmission_30d	8.4	17.9
charlson_10	35.1	43.0	sex	46.6	42.4

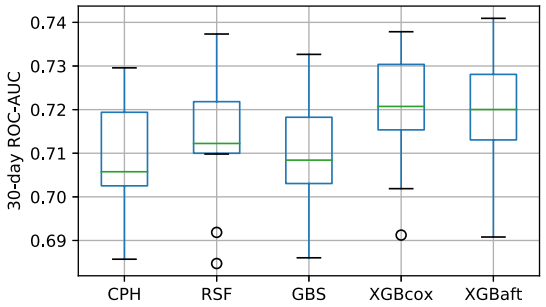


Fig. 6. Comparison of models: ROC-AUC when predicting readmission risks at 30 days since discharge. Results in boxplots are computed with a 10-fold cross-validation.

3.5. Post-discharge information

In Section 2.2.5 we discuss the decision of keeping post-discharge information (such as treatments at the moment of discharge, or health service in charge of the follow-up), evaluating the impact its absence would have. Fig. 9 shows the effect of excluding post-discharge features on our model's performance.

Table 7
Running times for each model, measured in seconds. Mean and standard deviation for both total training and mean single prediction times are calculated using the registered running time for each iteration in a 10-Fold experiment.

Model	Training (s): mean (std)	Prediction (s): mean (std)
CPH	38.56 (7.54)	$5.66 \cdot 10^{-5}$ ($1.04 \cdot 10^{-5}$)
RSF	430.33 (9.74)	$1.08 \cdot 10^{-2}$ ($1.12 \cdot 10^{-4}$)
GBS	553.0 (3.12)	$2.10 \cdot 10^{-3}$ ($1.06 \cdot 10^{-4}$)
XGBcox	6.31 (0.12)	$2.64 \cdot 10^{-3}$ ($2.78 \cdot 10^{-4}$)
XGBaft	3.47 (0.04)	$3.38 \cdot 10^{-3}$ ($3.46 \cdot 10^{-4}$)

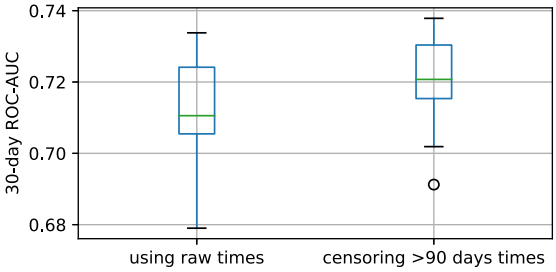


Fig. 7. 30-day ROC-AUC comparison depending on the censoring criteria. Left uses raw observed time periods, while right censors 90-day or later readmissions. Results obtained with XGBcox, 10-fold CV.

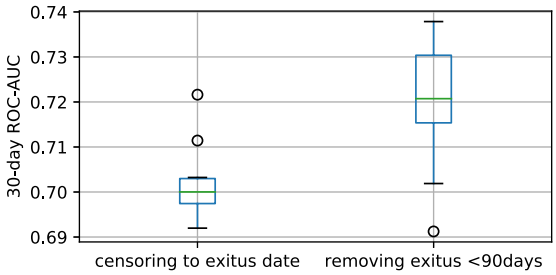


Fig. 8. Effect of filtering episodes followed by exitus. Comparison of ROC-AUC at 30 days when censoring time-to-event to exitus date versus filtering out exitus happening in less than 90 days since last discharge, 10-fold CV with XGBcox.

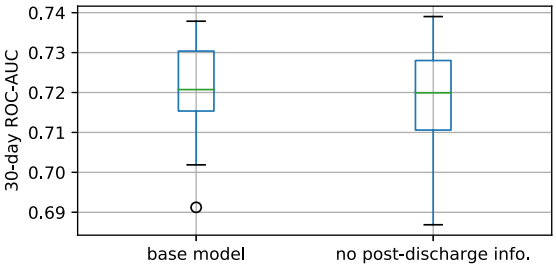


Fig. 9. Performance at 30 days since discharge when including (left) or excluding (right) post-discharge information. XGBcox model, 10-fold CV.

3.6. Risk factors

Another interesting result provided by the construction of a predictive model is the importance that the variables have on the predictions it makes. We focus here on the most efficient model, which is XGBcox.

The top influential features are chosen based on the permutation-based importance (first introduced in [22]), computed using the permutation importance method provided by scikit-learn. In short, this method randomly shuffles each feature, evaluates the model with it and compares this new performance with the one without shuffling. Fig. 10 shows the ranked list of 20 features that cause the largest drop in ROC-AUC at 30 days post-discharge when their values are shuffled. In the same decreasing order, the top nine features are:

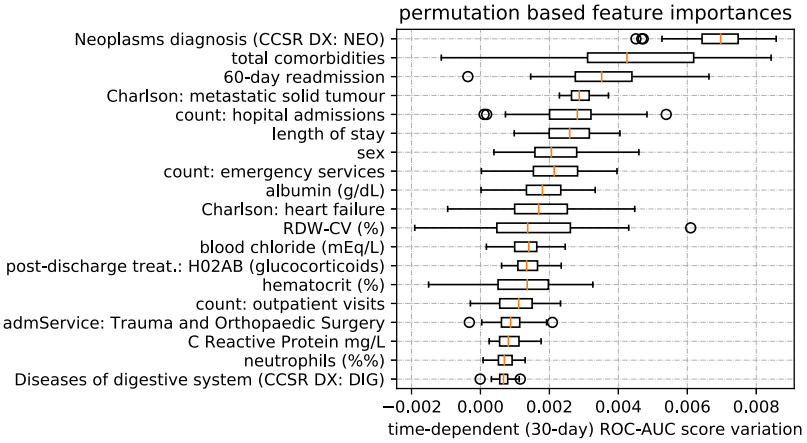


Fig. 10. Top 20 most important features obtained with permutation importance of XGBcox. For each feature, a boxplot shows de ROC-AUC score variation when evaluating the model with shuffled values of said feature across many iterations.

- *neoplasms diagnosis (CCSR DX: NEO)*: Clinical Classifications Software Refined (CCSR) diagnosis code [23] for neoplasms.
- *total comorbidities*: the total number of diagnosed Charlson comorbidities.
- *60-day readmission*: the current episode is a readmission up to 60 days after the previous discharge.
- *Charlson: metastatic solid tumour*: comorbidity caused by a metastatic solid tumour (Charlson code 17).
- *count: hospital admissions*: number of previous hospitalisation events.
- *length of stay*: days from admission to discharge.
- *sex*: sex of the patient.
- *albumin g/dL*: albumin levels in blood.
- *count: emergency services*: number of visits to emergency services.

Knowing the most influential features according to Fig. 10, we can study the mean effect of these covariates on the patient survival time (readmission-free time). Fig. 11 shows the partial dependence (PD) and the individual conditional expectation (ICE) plots for each of the top nine features. ICE shows, for each observation, the variation of model's output when changing the value of the variable at study, without changing the values of the other variables. For clarity, in Fig. 11 only 95% confidence interval of ICE is shown. The average of all ICE plots results in a PD plot.

Meanwhile, in Fig. 12, the average survival functions for the same most influential features are displayed. In each panel, a test set of data is split by a cutoff value of the corresponding risk factor, creating two groups: one with an expected low risk and the other with a higher risk of readmission. Categorical features are split using their category values directly, while continuous features are split by the median value. After splitting, survival functions are obtained for every patient in each split and averaged for each day. Since the XGBoost model cannot directly predict survival functions, we use a combinational approach using both XGBoost and CPH. The CPH model fits the baseline survival function using Breslow's method, while XGBoost predict function takes the covariates for each patient and returns a parametric risk score. This combination can be better understood by visiting Equation (1), where in this case, CPH predicts the $h_0(t)$ component, common to all patients, and XGBoost contributes with the parametric risk score (ρ) for each and every patient.

4. Discussion

The XGBoost Cox Regression model obtained both the highest C-Index and ROC-AUC at 30 days post-discharge, albeit being only slightly better than its AFT counterpart (Fig. 5). Although XGBcox and XGBaft performed similar, XGBcox is slightly more stable according to Fig. 6. CPH and GBS performed almost identically and obtained the lower performance, although with a difference of only 0.01 c-index and 30 days ROC-AUC. This suggests that, with this dataset, Cox-based models are complex enough to make a good fit between our covariates and survival time, and using more sophisticated models is not guaranteed to yield better results.

Regarding running times, which are displayed in Table 7, both XGBoost models needed much lower training times than the others, with the simpler CPH model achieving a relatively close third position. Nonetheless, prediction times could be a more critical aspect in the usual use case scenario. Here, CPH was the fastest performing model by a big margin. However, all five model are capable of making predictions fast enough for almost any scenario, even online, RSF taking the longest with approximately one centisecond of mean single prediction time.

The models subjected to test in this article are based on the assumption of proportional hazards, which needs to be checked to avoid biases and loss of predictive power [31]. The assumption of the hazards being proportional implies that the effect of each risk factor over a baseline hazard function is constant and non-evolving over time. A statistical test based on scaled Schoenfeld residuals [32], performed through all the features in our dataset, flagged some of these features as possibly non-conforming with the PH assumption. The models were therefore trained without these variables, or by discretizing them into a reduced number of categories,

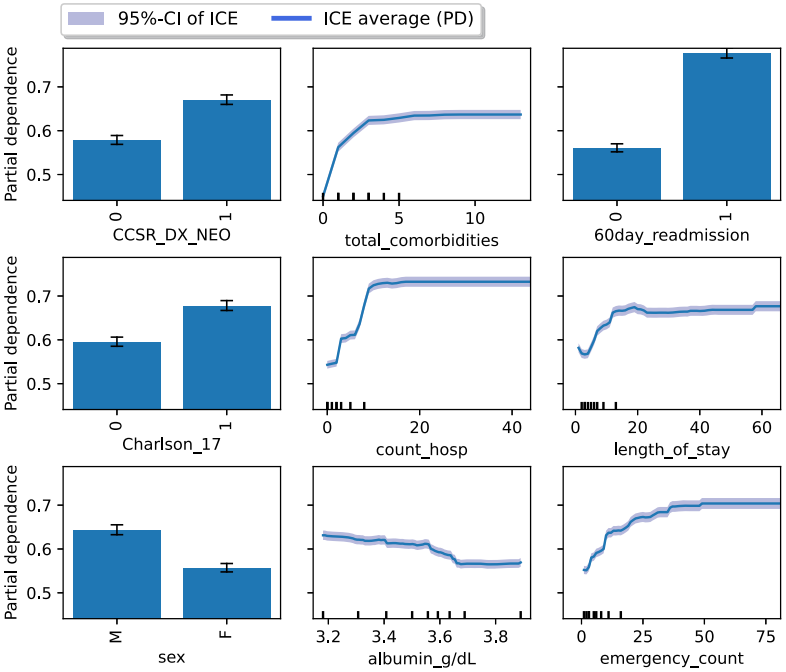


Fig. 11. Partial dependence (PD) plot of the most important features according to permutation importance. For each partial dependence line (the average of the individual conditional expectation or ICE), the 95% confidence interval of the ICE is displayed.

and in none of these experiments did the model improve. Furthermore, a visual Kaplan-Meier analysis of the least PH-compliant risk factors showed almost perfect proportionality between subgroups of patients without lines ever crossing. All these results indicate that the proportionality hypothesis is reasonably respected.

Observing Fig. 7, we can conclude that forcing the right censoring to 90 days after discharge positively impacts the survival model prediction performance. Not only did the mean ROC-AUC slightly increase, but the scores also appear to be more consistent. This may suggest that our assumption that predictions would improve by focusing the training times window around the date of interest was correct. The performance boost achieved with this simple procedure makes it one of the highlights of this study.

As our binary problem has only two possible outcomes (the patient is either readmitted or not), the inclusion of patient exitus in this definition is not straightforward. After observing Fig. 8, we assume the best temporary decision is to set exitus cases aside, pending a new whole design that can properly incorporate them. This point is further elaborated in Section 6.

It is worth noting that some features could not be fully available during the model's real use in a hospital. For instance, when deciding on patient discharge, prescripted drugs at discharge may not be available. In this paper, it is assumed that these features are available at the time of prediction. Nevertheless, contrary to the results reported in some studies mentioned earlier, if we observe the median ROC-AUC score in Fig. 9, there is not a significant drop in performance when excluding these features. It should be noted, however, that the variability of the ROC-AUC increases, suggesting that the unavailability of the post-discharge variables adds uncertainty to the prediction.

Regarding the inclusion of the new Barthel and route of administration features, no significant improvements in performance were noted. This could indicate that there is redundant information in our data, or that a performance ceiling has been reached, at least with our currently available data. The existence of an unavoidable level of uncertainty is developed in Section 6.

The list of main risk factors, sorted by descending order of permutation importance, is shown in Fig. 10. While the importance is not distributed in a balanced manner, with a subset of features accumulating the most score variation, none of the features in the list reaches 0.01 ROC-AUC variation. This could indicate that this relatively simple, univariate method for assessing feature importance is not capable of reporting hidden, more complex shared effects between variables that could be participating within the model.

Finally, the main risk factors of the XGBcox model mentioned before have been described. Fig. 11 shows that harzard risk increases with all continuous variables except for albumin. Length of stay shows a parabola-like behaviour in lower values. The presence of neoplasm, oncology comorbidities and previous readmissions also increase readmission risk. It is worth noting that PD plots show the univariate response for each feature, and a multivariate approach could highlight different behaviours. The corresponding survival functions observed in Fig. 12 suggest that the predictions made by a combination of XGBoost and the CPH model follow very similar curves to those obtained with a non-parametric Kaplan-Meier estimator, which is a strong indication that it has been possible to model the risk of readmission using the available variables. It can also be observed that the 95% confidence interval is wider in boolean covariates than in continuous values, which may be explained by the splitting method. While continuous features have been split by their median value, dividing the test set into two sets almost identical in size, boolean and categorical features can only be divided by their different categories. As many of our boolean features are very sparse, this results in the positive category being less

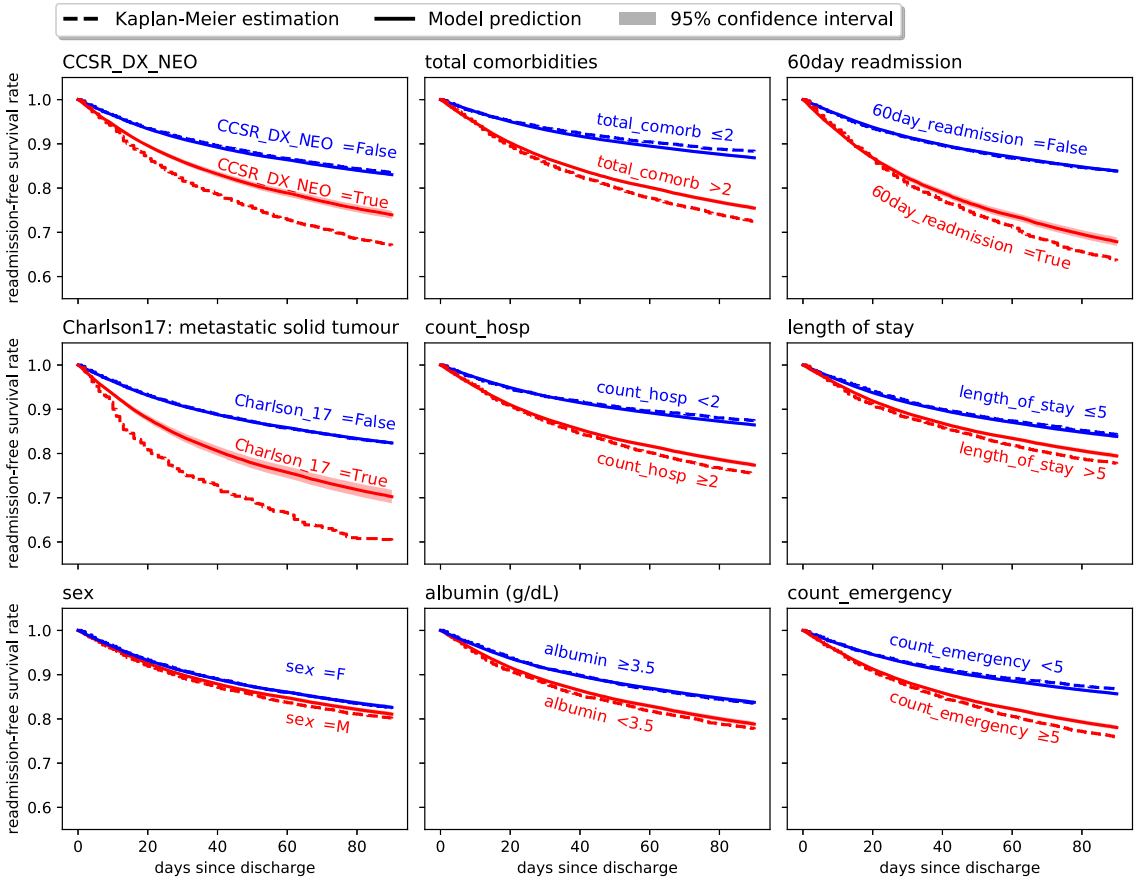


Fig. 12. Survival function when splitting test data by median value of main risk factors into low and high risk groups. Each panel compares the mean survival function obtained by a non-parametric Kaplan-Meier estimator with our model's predicted survival function, for all individuals in test set. Our model prediction combines the baseline survival function of a Cox Proportional Hazards model obtained with Breslow's method and the risks predicted by an XGBoost regression model. (dashed line) Kaplan-Meier estimation, (solid line) own model prediction, (red) population with a higher expected risk of readmission or lower readmission-free time, (blue) low expected readmission risk population.

populated than the complementary null-category set, thus explaining the wider confidence interval and the more erratic matching Kaplan-Meier curve.

4.1. Limitations

The main limitation of this study is the lack of an external validation cohort. It is also worth noting that the applicability of these models in other datasets is restricted by the availability of the features used in this study.

Since this is a retrospective study ending in 2018, some differences could arise with the current data distribution. These differences in distribution are known as *data drift* and should be assessed when any of these models are used with new data. Most of the models trained in this study are based on CPH, so they rely on the proportional hazards assumption. Although this assumption has been checked in the dataset, new data could not satisfy this assumption.

Finally, some features could not be available at the moment of discharge. In this study, we assumed that all variables will be available. Models will have to be adapted accordingly otherwise.

5. Conclusions

In the task of modelling the readmission risk at 90 days using heterogeneous variables (comorbidities, laboratory, etc.), five survival analysis models were compared. Models based on Cox assumptions have shown equal, if not better, performance than more complex survival models. Machine learning models, especially XGBcox, have shown moderately better performance than statistical models, both in terms of C-Index (0.716) and 30-days ROC-AUC (0.72).

Even when using regression models, the performance at 30 days since discharge cutoff is slightly above what had been obtained using classifier models for the same time period [12], 0.69 ROC-AUC. The risk of readmission at 29 days is probably not that

different from that at 31 days. Unlike with a classifier, a regression model does not draw a strict boundary between the patients who are readmitted before or after 30 days. This could be one of the main causes for improvement.

The main improvements in performance were caused by limiting the observed time horizon to 90 days, thus focusing on short-termed readmissions, as well as filtering out exitus cases, which cannot be directly assigned into either an event observation or censoring.

6. Future work

Following physicians' recommendations, all patients with exitus date before 90 days after the last hospital admission are removed from this study. This results in a 7.24% portion of the dataset being removed. However, an alternate approach could benefit from exitus information and incorporate it into a joint prediction system. Patient mortality could be equated with readmissions to build a single complication outcome model. An alternative approach would consist of designing a multiclass model that could either predict the patient readmission or death, if any, or even two separate, complementary models.

This study lays the groundwork for another model that could help make the discharge decision throughout a patient's stay. As we have seen throughout this study, survival models allow the evaluation of risks at different periods since discharge. Putting aside post-discharge information, a model like the ones described in this work could be used any time during the hospital stay. Comparing the risk yielded by the model on consecutive days, specialists may be able to evaluate a patient's progression. If the readmission risk decreases to a minimum, a patient could be nearing a suitable date for discharge. Vice versa, a patient with an increasing risk could be prompted to lengthen their stay at the hospital, waiting for a more suitable scenario to come. For such an evolution model to be useful in practice, the results must be calibrated so that each decision can be grounded in a meaningful and reliable probability of early, unplanned readmission.

All machine learning models used in this work, although slightly better than the traditional CPH, obtained comparable final results, as if a plateau had already been reached. Certainly, it might be worth continuing to look for other variables related to readmission that might add predictive power to the models. However, it could be advisable to study the theoretical maximum performance or, analogously, the minimum error associated with the task, also known as the Bayes error. This unavoidable error is due to several factors that introduce uncertainty, such as early readmissions caused by unexpected accidents totally unrelated to the ailments that prompted the previous episode. Knowing where these uncertainties lay could point us toward new, alternative directions in our search for making better predictions of avoidable hospital readmissions.

Ethics

This research was carried out with accordance with University and Polytechnic La Fe Hospital, which approved the study on 26th November 2019 under the name "Desarrollo de un modelo predictivo de ingresos y reingresos no programados a 30 días en el Departamento de Salud-Valencia La Fe" (Registration number: 2019-309-1). Patient privacy was maintained by using data previously pseudo anonymised with non-traceable codes and only authorised people obtained data from electronic health records. Comité de Ética de la Investigación con medicamentos (CEIm) waived the requirement for written consent since data was pseudo anonymised and the study complies with national and European legal requirements regarding data protection.

CRediT authorship contribution statement

Pedro Pons-Suñer: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Laura Arnal:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **François Signol:** Conceptualization, Methodology, Software, Validation, Writing – review & editing, Formal analysis. **M. Jose Caballero Mateos:** Data curation, Resources, Validation, Writing – review & editing. **Bernardo Valdivieso Martínez:** Conceptualization, Data curation, Resources, Supervision, Validation, Writing – review & editing. **Juan-Carlos Perez-Cortes:** Conceptualization, Formal analysis, Project administration, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The Valencia Health Care Department has imposed restrictions on the data used in this study due to its potential for use in identification of participants. The authors have been granted data access after signing specific agreements guaranteeing and demonstrating compliance with national and European legal requirements regarding data protection and after the approval of an Ethics Committee, and so are not publicly available. Pseudo-anonymised data are however available upon reasonable request by contacting the Medical research Institute of Hospital La Fe (info@iislafe.es) to any researcher wishing to use them for non-commercial purposes and who could guarantee and demonstrate compliance with national and European legal requirements regarding data protection.

Acknowledgements

This work was funded by Generalitat Valenciana - IVACE and the European Union through FEDER under grants IMDEEA-2020-55 and IMDEEA-2021-100.

This work was also funded by Generalitat Valenciana - IVACE under grant IMAMCA/2023/11.

We also acknowledge researchers at La Fe Health Department and Health Research Institute of La Fe University Hospital (València, Spain) for their valuable contribution to this work.

References

[1] C. van Walraven, C. Bennett, A. Jennings, P.C. Austin, A.J. Forster, Proportion of hospital readmissions deemed avoidable: a systematic review, Can. Med. Assoc. J. 183 (7) (2011) E391–E402, https://doi.org/10.1503/cmaj.101860, https://www.cmaj.ca/content/183/7/E391.full.pdf, https://www.cmaj.ca/content/183/7/E391.

[2] P.C. Johnson, Y. Xiao, R.L. Wong, S. D'Arpino, S.M. Moran, D.E. Lage, B. Temel, M. Ruddy, L.N. Traeger, J.A. Greer, E.P. Hochberg, J.S. Temel, A. El-Jawahri, R.D. Nipp, Potentially avoidable hospital readmissions in patients with advanced cancer, J. Oncol. Pract. 15 (5) (2019) e420–e427, https://doi.org/10.1200/JOP.18.00595, pMID: 30946642.

[3] A.M.B. van der Does, E.L. Kneepkens, E.B. Uitvlugt, S.L. Jansen, L. Schilder, G. Tokmaji, S.C. Wijers, M. Radersma, J.N.M. Heijnen, P.F.A. Teunissen, P.B.J.E. Hulshof, G.M. Overvliet, C.E.H. Siegert, F. Karapinar-Çarkit, Preventability of unplanned readmissions within 30 days of discharge. a cross-sectional, single-center study, PLoS ONE 15 (4) (2020) 1–12, https://doi.org/10.1371/journal.pone.0229940.

[4] C. Van Walraven, I.A. Dhalla, C. Bell, E. Etchells, I.G. Stiell, K. Zarnke, P.C. Austin, A.J. Forster, Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community, Can. Med. Assoc. J. 182 (6) (2010) 551–557.

[5] J. Donzé, D. Aujesky, D. Williams, J.L. Schnipper, Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model, JAMA Intern. Med. 173 (8) (2013) 632–638.

[6] R. Robinson, T. Hudali, The hospital score and lace index as predictors of 30 day readmission in a retrospective study at a university-affiliated community hospital, PeerJ 5 (2017) e3137.

[7] S. Kalagara, A.E.M. Eltorai, W.M. Durand, J.M. DePasse, A.H. Daniels, Machine learning modeling for predicting hospital readmission following lumbar laminectomy, J. Neurosurg. Spine SPI 30 (3) (2019) 344–352, https://doi.org/10.3171/2018.8.SPINE1869, https://thejns.org/spine/view/journals/j-neurosurg-spine/30/3/article-p344.xml.

[8] M. Cox, J.C. Panagides, A. Tabari, S. Kalva, J. Kalpathy-Cramer, D. Daye, Risk stratification with explainable machine learning for 30-day procedure-related mortality and 30-day unplanned readmission in patients with peripheral arterial disease, PLoS ONE 17 (11) (2022) 1–18, https://doi.org/10.1371/journal.pone.0277507.

[9] Y. Huang, A. Talwar, S. Chatterjee, R.R. Aparasu, Application of machine learning in predicting hospital readmissions: a scoping review of the literature, BMC Med. Res. Methodol. 21 (1) (2021) 1–14.

[10] J. Todd, A. Gepp, S. Stern, B.J. Vanstone, Improving decision making in the management of hospital readmissions using modern survival analysis techniques, Decis. Support Syst. 156 (2022) 113747, https://doi.org/10.1016/j.dss.2022.113747, https://www.sciencedirect.com/science/article/pii/S0167923622000185.

[11] S. Chakraborty, P. Zhao, Y. Huang, T. Dey, Semiparametric survival analysis of 30-day hospital readmissions with bayesian additive regression kernel model, Stats 5 (3) (2022) 617–630, https://doi.org/10.3390/stats5030038, https://www.mdpi.com/2571-905X/5/3/38.

[12] L. Arnal, P. Pons-Suñer, J.R. Navarro-Cerdán, P. Ruiz-Valls, M.J. Caballero Mateos, B. Valdivieso Martínez, J.-C. Perez-Cortes, Decision support through risk cost estimation in 30-day hospital unplanned readmission, PLoS ONE 17 (7) (2022) e0271331.

[13] A.C. Colavecchia, D.R. Putney, M.L. Johnson, R.R. Aparasu, Discharge medication complexity and 30-day heart failure readmissions, Res. Social Adm. Pharm. 13 (4) (2017) 857–863.

[14] M.N. Willson, C.L. Greer, D.L. Weeks, Medication regimen complexity and hospital readmission for an adverse drug event, Ann. Pharmacother. 48 (1) (2014) 26–32.

[15] F.I. Mahoney, et al., Functional evaluation: the barthel index, Maryland State Med. J. 14 (2) (1965) 61–65.

[16] R. Robinson, M. Bhattarai, T. Hudali, C. Vogler, Predictors of 30-day hospital readmission: the direct comparison of number of discharge medications to the hospital score and lace index, Fut. Healthcare J. 6 (3) (2019) 209.

[17] M. Glans, A. Kragh Ekstam, U. Jakobsson, Å. Bondesson, P. Midlöv, Medication-related hospital readmissions within 30 days of discharge—a retrospective study of risk factors in older adults, PLoS ONE 16 (6) (2021) e0253024.

[18] D.R. Cox, Regression models and life-tables, J. R. Stat. Soc., Ser. B, Methodol. 34 (2) (1972) 187–202.

[19] L.-J. Wei, The accelerated failure time model: a useful alternative to the cox regression model in survival analysis, Stat. Med. 11 (14–15) (1992) 1871–1879.

[20] S. Pölsterl, scikit-survival: a library for time-to-event analysis built on top of scikit-learn, J. Mach. Learn. Res. 21 (1) (2020) 8747–8752.

[21] A. Barnwal, H. Cho, T. Hocking, Survival regression with accelerated failure time model in xgboost, J. Comput. Graph. Stat. 31 (4) (2022) 1292–1302.

[22] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[23] Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality, Rockville, MD, Clinical Classifications Software Refined (CCSR), http://www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp. (Accessed 29 August 2023) (Last modified on 2022-09-12).

[24] J.H. Friedman, Stochastic gradient boosting, Comput. Stat. Data Anal. 38 (4) (2002) 367–378.

[25] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, et al., Xgboost: extreme gradient boosting, R package version 0.4-2 1 (4) (2015) 1–4.

[26] F.E. Harrell Jr, K.L. Lee, D.B. Mark, Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, Stat. Med. 15 (4) (1996) 361–387.

[27] H. Uno, T. Cai, M.J. Pencina, R.B. D'Agostino, L.-J. Wei, On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data, Stat. Med. 30 (10) (2011) 1105–1117.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[29] C. Davidson-Pilon, Lifelines: survival analysis in python, J. Open Sour. Softw. 4 (40) (2019) 1317.

[30] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[31] I. Kuitunen, V.T. Ponkilainen, M.M. Uimonen, A. Eskelinen, A. Reito, Testing the proportional hazards assumption in cox regression and dealing with possible non-proportionality in total joint arthroplasty research: methodological perspectives and review, BMC Musculoskelet. Disord. 22 (1) (2021) 489.

[32] P.M. Grambsch, T.M. Therneau, Proportional hazards tests and diagnostics based on weighted residuals, Biometrika 81 (3) (1994) 515–526.