Computer Methods and Programs in Biomedicine 164 (2018) 49–64

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Predictive models for hospital readmission risk: A systematic review of methods

Arkaitz Artetxe[a,*], Andoni Beristain[a], Manuel Graña[b]

[a] Vicomtech-IK4 Research Centre, Mikeletegi Pasealekua 57, 20009 San Sebastian, Spain
[b] Computation Intelligence Group, Basque University (UPV/EHU) P. Manuel Lardizabal 1, 20018 San Sebastian, Spain

ARTICLE INFO

Article history:
Received 16 November 2017
Revised 3 May 2018
Accepted 5 June 2018

ABSTRACT

Objectives: Hospital readmission risk prediction facilitates the identification of patients potentially at high risk so that resources can be used more efficiently in terms of cost-benefit. In this context, several models for readmission risk prediction have been proposed in recent years. The goal of this review is to give an overview of prediction models for hospital readmission, describe the data analysis methods and algorithms used for building the models, and synthesize their results.
Methods: Studies that reported the predictive performance of a model for hospital readmission risk were included. We defined the scope of the review and accordingly built a search query to select the candidate papers. This query string was used as input for the chosen search engines, namely PubMed and Google Scholar. For each study, we recorded the population, feature selection method, classification algorithm, sample size, readmission threshold, readmission rate and predictive performance of the model.
Results: We identified 77 studies that met the inclusion criteria, out of 265 citations. In 68% of the studies ($n = 52$) logistic regression or other regression techniques were utilized as the main method. Ten (13%) studies used survival analysis for model construction, while 14 (18%) used machine learning techniques for classification, of which decision tree-based methods and SVM were the most utilized algorithms. Among these, only four studies reported the use of any class imbalance addressing technique, of which resampling is the most frequent (75%). The performance of the models varied significantly among studies, with Area Under the ROC Curve (AUC) values in the ranges between 0.54 and 0.92.
Conclusion: Logistic regression and survival analysis have been traditionally the most widely used techniques for model building. Nevertheless, machine learning techniques are becoming increasingly popular in recent years. Recent comparative studies suggest that machine learning techniques can improve prediction ability over traditional statistical approaches. Regardless, the lack of an appropriate benchmark dataset of hospital readmissions makes a comparison of models' performance across different studies difficult.

1. Introduction

Hospital readmissions are defined as admissions to a hospital within a -usually short- time span after discharge from hospital. Readmissions are frequent and costly events that impose tremendous burden on patients and healthcare systems [1,2].

Hospital readmissions are becoming a strong concern of hospitals and policy makers as a measure of the quality of care given and have been adopted by many organizations as a quality indicator [3]. Centers for Medicare and Medicaid Services (CMS) in the USA [4] and policy makers in the UK [5] have introduced financial penalties to hospitals with high readmission rates by reducing the payment for patients readmitted within 30-day of discharge. Therefore, there is a growing interest within the research community to address this problem from a data analysis perspective.

Some authors have performed bibliographic review studies with the aim of synthesizing the literature on prediction models to estimate the readmission risk. In 2011 Kansagara et al. [6] presented what is probably the most referenced systematic review paper in the field. This thorough review was mainly focused on model performance description and comparison to assess the suitability of the models for clinical or administrative use. The authors of the paper conclude that most readmission risk prediction models perform poorly and efforts to improve their performance are needed. The study also concludes that readmission risk prediction is a complex problem by nature, with many inherent limitations.

* Corresponding author.
E-mail address: aartetxe@vicomtech.org (A. Artetxe).

https://doi.org/10.1016/j.cmpb.2018.06.006
0169-2607/

Table 1
Research questions.

#	Research Question	Rationale
Q1	Which machine learning techniques were used in readmission risk prediction?	To identify which machine learning techniques are used for readmission risk prediction model construction.
Q2	What is the overall performance of models in readmission risk prediction?	To assess the discrimination ability of the models in readmission risk prediction.

In 2015 Swain and Kharrazi [7] conducted a semi-systematic review of readmission predictive factors published prior to March 2013. This review was, to some degree, based on the work by Kansagara et al. and can be considered an extension of it. This work was focused on identifying the most significant predictive factors from previous readmission prediction models.

Other studies place the focus on a certain subpopulation rather than covering all the published risk prediction models. Ross et al. [8] conducted a review of statistical models for the readmission of heart failure (HF) patients. This work included the identification of analytic models, in addition to identifying patient characteristics associated with readmission. A more recent study by Leppin et al. [9] reviewed randomized trials that assessed the effect of interventions intended to prevent 30-day hospital readmissions.

Most of the previous review studies have focused on measuring the discrimination ability of the models and identifying predictive characteristics associated with readmission. In different but related fields, review studies targeting the analysis of data analysis approaches can be found. For instance, a systematic literature review on data mining techniques applied in cardiology was recently presented in [10].

Nevertheless, to our knowledge no study regarding machine learning techniques, including feature selection and class imbalance has been presented in the field of readmission prediction. Since those are areas which may provide improvements over previous methods, we have already carried out some experiments that support our initial thoughts [11,12].

The objective of this study is to systematically review the prediction models for hospital readmission by describing the data analysis methods and algorithms used for building the models.

This paper is organized as follows: Section 2 summarizes the research methodology of our study. In section 3 we present the review results. In Section 4 we discuss the results and findings of this study. Lastly, Section 5 presents the conclusions and future work.

2. Materials and methods

In this work, we conduct a systematic review following the three stages proposed by Tranfield et al. [13], namely planning, conducting and reporting. According to this methodology, we first define the research questions. Then we define the search strategy by identifying the source databases and the inclusion and exclusion criteria. Next, we present the data extraction procedure and then present the results.

2.1. Research questions

The overall objective of our systematic review is to identify and analyse the most relevant research studies on readmission risk prediction. More precisely, this review aims to analyse the data analysis methods utilized in the topic, paying special attention to machine learning techniques. Table 1 shows the research questions that guided this review.

Given that the first research question (Q1) is broad, it was divided into three sub-questions:

- Q1.1 Classification algorithms,
- Q1.2 Feature selection techniques, and
- Q1.3 Class imbalance addressing techniques.

2.2. Search strategy

2.2.1. Search engines

PubMed and Google Scholar were the chosen search engines for retrieving the primary citations. Google Scholar was selected because of its broad coverage of general scientific publications while PubMed provided access to the more specialized MEDLINE (Medical Literature Analysis and Retrieval System Online) database.

Table 2 shows the search strings that we created to be used with the search engines. The search strings were selected with the aim of finding an appropriate trade-off between coverage and manageable size.

Additional citations retrieved from reference lists of main review articles were also included in the list of citations for further analysis.

2.2.2. Search limits

The following search limits were applied:

- Peer-reviewed journal articles in English

We limited the search to indexed journal articles written in the English language. Peer-reviewed journal articles are considered to provide a good view of accepted and validated methodologies and knowledge.

- Search scope

We performed the search using all fields available, so as not to restrict the search to only the title and abstract or to a particular subject area. Our main goal was not to disregard high impact papers due to restrictive search conditions. We excluded studies targeting readmissions after surgery as these were too specific to each surgical procedure.

- Publication date range

We did not restrict our search to a specific time frame. Citations were collected in September 2017 so only a part of the studies published in 2017 were included.

2.2.3. Data extraction

For each study included in the review, we have extracted and synthesized data associated with the research questions defined. The algorithms used for building the model were extracted in relation to Q1. Discrimination ability (Q2) of the models was synthesized using the AUC metric (Area Under the ROC Curve) if reported.

Table 2
Search strings.

Database	Search term
PubMed	((readmission*) OR (rehospitalization*)) AND ((“prediction model”) OR (“predictive model”) OR (“risk model”))
Google Scholar	((readmission) OR (rehospitalization)) AND ((“prediction model”) OR (“predictive model”) OR (“risk model”))

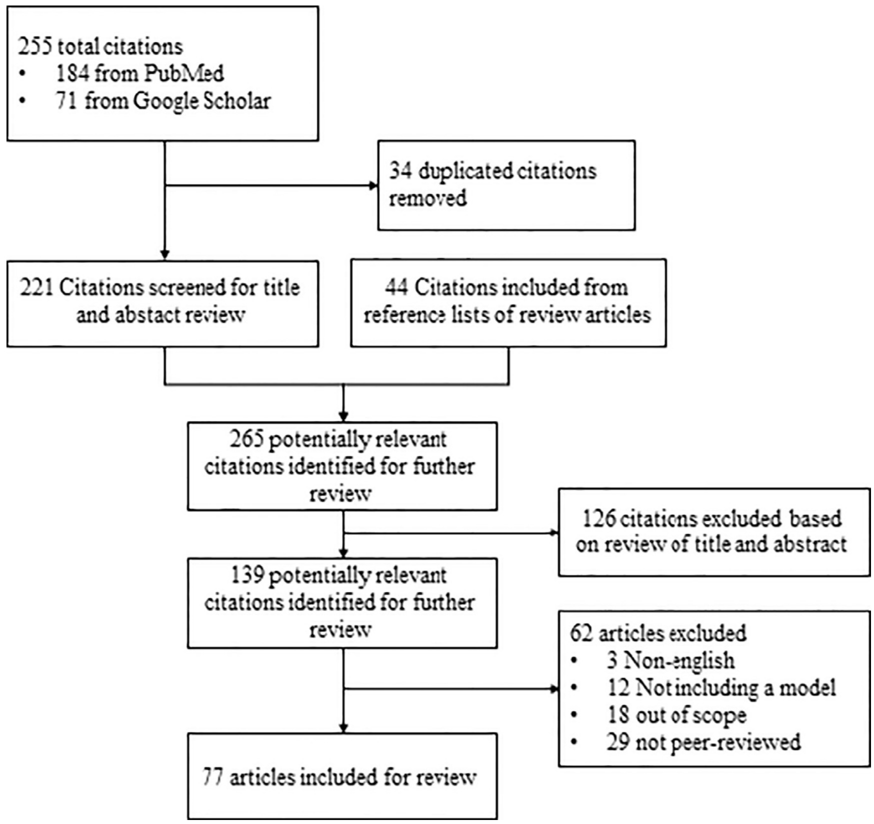

Fig. 1. Flow diagram of the selection process.

Other performance evaluation metrics such as accuracy, sensitivity, specificity or positive predictive value (PPV) were extracted if AUC was not reported. AUC was the preferred metric of choice due to being used by the majority of the selected studies (75%) in this literature review. It is considered to be the standard "single number" method to assess the overall accuracy of predictive models [91] since it "provides a richer measure of classification performance than scalar measures such as accuracy or error rate" [92].

The different sub-questions of Q1 lead us to the extraction of the following information: Feature selection technique, class-imbalance addressing procedures and readmission rate, which is directly related to the imbalance-ratio. Additionally, target population, readmission threshold (in hours, days or months) and dataset size (number of instances of the dataset) were also collected.

This work was prepared using the PRISMA (preferred reporting items for systematic reviews and meta-analyses) checklist [14].

3. Results

Table 3 describes in detail the 76 studies included for review, after the selection process. In this section, we present the results of our systematic review study. First, we present an overview of the results and then we discuss the results regarding specific research questions.

3.1. Overview of selected studies

As shown in Fig. 1, we gathered 255 eligible references from the search engines. Duplicated references (34) in the merged list retrieved from both databases were excluded. To this list, we added references extracted from the reference lists of main review articles in the literature (44 additional references). At this step, we had a dataset consisting of 265 potentially relevant references for analysis and review. In the following step, 126 references were excluded based on the review of the title and abstract. Further reviews excluded 62 articles that did not fulfil the predefined inclusion criteria. Thirty-two of them were excluded due to the language and peer-review criteria. Twelve citations were excluded for not including a readmission prediction model and 18 were discarded for being outside the review scope.

Fig. 2 shows the number of papers (only covers studies included in the review) per year. As we can see, the number of papers has increased in recent years, reaching a peak in 2015. The number of studies corresponding to 2017 are those found at the date the search was carried out, hence the apparent decrease.

3.2. Data analysis methods

Readmission risk prediction has been addressed from different perspectives. Traditionally, studies have been conducted following statistical multivariable modelling, which has been widely used in medical research. Fundamentally two different but related procedures can be distinguished among statistical modelling approaches: regression analysis and survival analysis.

Regression analysis is a process that estimates the probability of the target variable given some linear combination of the predictors. Binary logistic regression (LR) is a regression model where the target variable is binary, that is, it can take only two values, 0 or 1. It is the most utilized regression model in readmission prediction, given that the output is modelled as readmitted (1) or not readmitted (0).

Survival models, on the other hand, relate the features to the time that passes before the event (i.e. readmission) occurs.

In recent years, machine learning and data mining have emerged as approaches that can potentially improve the prediction ability of readmission risk prediction models. Those techniques

Table 3
Description of the included studies.

Identifier	Population	Feature Selection Method	Algorithm	Readmission rate (%)	No. of instances	Readmission	Discrimination (AUC)
Corrigan 1992 [74]	adults		Cox regression	30,14	4219	1-year	NR
Marcantonio 1999 [47]	65<	bivariate -> backward elimination LR	LR	50	308	30-day	NR
Coleman 2004 [30]	65<	backward elimination LR	LR	NR	1401	30-day	Administrative data: 0,77, administrative and self-reported: 0,83
Halfon 2006 [38]	all	univariate -> backward elimination Poisson R.	Poisson regression	5,1	131,809	30-day	0,67–0,72
Keenan 2008 [43]	HF	stepwise selection LR	LR	23,6	567,447	30-day	0,61
Silverstein 2008 [58]	65<	forward addition & backward elimination LR	LR	11,72	29,292	30-day	0,65
novotny 2008 [83]	all	univariate LR	LR	14	1077	2-day	sensitivity: 90,1–18.5, specificity: 16.2–87.6, PPV: 14.91–19.6
Jencks 2009 [41]	65 < or disabled[1]		Cox regression	19,6	11,855,702	30-day, 180-day	NR
amarasingham 2010 [22]	HF	univariate LR and multivariate LR	LR	24,7 (3,1)	1372	30-day	0.72 (95% CI, 0.70–0.75)
Hasan 2010 [40]	all	univariate LR	LR	17,5	10,946	30-day	0,61
Walraven 2010 [63]	all	backward stepping LR	LR	8	4812	30-day	0,68 (95% CI, 0,65–0,71)
Allaudeen 2011 [18]	all	univariate GEE	GEE	17	10,359	30-day	NR
Berman 2011 [26]	advanced liver disease	univariate -> forward stepwise LR	LR	20	554	30-day	NR
Epstein 2011 [35]	HF, Pneumonia (65 <)	univariate -> sequential removal LR	HGLM	11–32 (HF)	234,477	30, 60, 90-day	NR
lopez 2011 [45]	64<	forward stepwise LR	LR	1,3	28,430	180-day	0,76
Watson 2011 [66]	HF	univariate & multivariate LR	LR	12,75	729	30-day	0,67
Allen 2012 [19]	Systolic HF	Stepwise LR	LR	13,3	4584	30-day	all-cause: 0.64, HF: 0.59
amalakuhan 2012 [21]	COPD	RF	RF	47	106	1-year	0,72
Au 2012 [23]	HF	RF	RF	18,77	59,652	30-day	0,6 (95% CI, 0.59–0.60)
Billings 2012 [28]	all		LR	12,2	576,868	30-day	0,7 (95% CI, 0,69–0,7)
Nijhawan 2012 [52]	HIV	univariate LR	LR	25	2476	30-day	0,72 (95% CI, 0.70–0.75)
ouanes 2012 [53]	ICU	univariate	LR	3	3462	7-day	0,74 (95% CI, 0.68–0.79)
Wang 2012 [65]	HF	backward selection -> forward selection Cox	Cox regression	4,2	198,640	30-day, 1-year	0.821(95% CI, 0.814–0.827), 0.815 (95% CI, 0.812–0.818)
Zapatero 2012 [68]	all	univariate LR	LR	12,4	999,089	30-day	Accuracy: 87.6%
baillie 2013 [3]	all		NR	14,4	120,396	30-day	0,61
Dharmarajan 2013 [2]	HF, AMI, Pneumonia	Univariate Cox regression	LR	24,8 (HF)	1,330,157 (HF)	30-day	NR
Donze 2013 [33]	all	univariable LR -> backward elimination LR	LR	22,3	10,731	30-day	0,67
Garrison 2013 [37]	all	bivariate wilcoxon rank sum, ficher, chi2	LR	30,4	276	30-day	NR
Shulan 2013 [57]	Veterans	multivariate LR (stepwise)	LR	16,15	8718	30-day	0,8 (95% CI, 0.79–0.81)

(continued on next page)

Table 3 (*continued*)

Identifier	Population	Feature Selection Method	Algorithm	Readmission rate (%)	No. of instances	Readmission	Discrimination (AUC)
Singal 2013 [59]	Cirrhosis	univariate LR -> Multivariate LR	LR	27	836	30-day	0,66 (95% CI, 0.59–0.73)
Wallmann 2013 [1]	Cardiac-related disease[2]	backward elimination LR	LR	4,5	35,531	30-day	0,75
abdelrahman 2014 [16]	HF	Wrapper (final), information gain, gain ratio, symmetrical uncertainty	LR	19	2787	30-day	0,86
allison 2014 [20]	OPAT	backward selection LR	LR	26	782	30-day	0,61
Morris 2014 [50]	60 < (ED)	stepwise LR	LR	NR	585,888	90-day	NR
Nguyen 2014 [51]	COPD	univariate	GEE	18	4596	30-day	NR
pugh 2014 [56]	65<	univariate	GLM	22,7	105,450	30-day	0,65
vanDiepen 2014 [62]	cardiovascular ICU	univariate -> stepwise LR	LR	4,4	10,799	any	0,799
walsh 2014 [64]	all	LASSO	LASSO, SVM	7,16	92,530	30-day	all-cause: 0,68(95% CI: 0,66–0,7), HF: 0,92(95% CI, 0,87–0,97)
Spiva 2014 [78]	all	LR	LR	27,1	598	30-day	0,77
sfoungaristos 2014 [84]	renal colic	univariate LR	LR	18,1	452	6-week	NR
alassaad 2015 [17]	80<	PCA - colinear variables removed -> backward elimination	Cox regression	68	368	1-year	0,71
betihavas 2015 [27]	HF	backward elimination Cox	Cox regression	13	280	28-day	0,8
cui 2015 [31]	all	bivariate	LR	33,7	61,926	1-year	0,7 (95% CI, 0.696–0.705)
Deschodt 2015 [32]	75<	univariate -> backward LR	LR	18,5–29,1	442	1-month, 3-month	NR
Hao 2015 [39]	all	variance minimization criterion	Survival RF	NR	211,232	30-day	0,72
padhukasahasram 2015 [54]	HF		Cox regression, survival forest	54,3	789	any	0,69
Pereira 2015 [55]	75 < (ED)	univariate LR -> forward selection LR (& kaplan-meier, cox, gehan or wilcoxon)	LR, Cox regression	1,8/6,1/10	11,521	72-hour, 30-day, 90-day	30-day: 0,77 (95% CI, 0,71–0,83)
tsui 2015 [60]	65 < (ED)	multivariate LR	LR	7,8	1,167,521	28-day	0,82
Yu 2015 [67]	AMI, HF, Pneumonia, all		SVM, Cox regression	18,87	74,746	30-day	0.66, 0.65, 0.63, 0.74
Zheng 2015 [69]	HF		SVM, RF, NN	21,63	1641	30-day	accuracy: 78.4, sensitivity: 97.3, specificity: 8.6
Vigod 2015 [75]	acute psychiatric unit	stepwise LR	LR	9,2	65,499	30-day	0,63
Tulloch 2015 [76]	psychiatric	stepwise removal LR	Cox regression & LR	14,6	7891	90-day	0,65
shadmi 2015 [85]	all	CART, C5.0 & neural network	LR	16,8	17,334	30-day	0,7
futoma 2015 [88]	all	univariate -> multivariate	LR, RF, SVM, deep NN	19	3,295,775	30-day	0,68
baltodano 2016 [24]	ventral hernia repair	univariate LR	LR	4,7	17,789	30-day	0,71
cai 2016 [29]	all	CBFS with best-first search	Bayesian network	NR	32,634	7-day	0,82
fisher 2016 [36]	In rehabilitation & high risk	univariate	Classification Tree, HGLM	25,3	25,908	30-day	0,58–0,69
kaur 2016 [42]	Pediatric ICU	univariate LR -> forward & backward LR	LR	33	256	48-hour	0.79

(*continued on next page*)

Table 3 (*continued*)

Identifier	Population	Feature Selection Method	Algorithm	Readmission rate (%)	No. of instances	Readmission	Discrimination (AUC)
low 2016 [46]	Asian adults	univariate LR -> multivariate LR	LR	15,5	74,102	30-day	0,78 (95% CI, 0.77–0.79)
mclaren 2016 [48]	HF	univariate	LR	18	1999	30-day	0,63
mcmanus 2016 [49]	AMI (65 <)	PCA-based feature reduction	LR	13,18	804	30-day	0.65
turgeman 2016 [61]	HF	Pearson correlation	ensemble (Boosted C5.0 & SVM)	28	4840	30-day	accuracy: 0.81–0.85
Mortazavi 2016 [70]	HF		RF, SVM, Boosting, LR	14,8	1004	30-day, 180-day	30day-allcause: 0,628 (95% CI, 0.624–0.633), 180day-allcause: 0.654 (95% CI, 0.650–0.657)
Krumholz 2016 [71]	HF	RF	Cox regression	17,1	1004	30-day	0,65
Bradford 2016 [72]	HF	univariate	LR	13,3	2420	15-day/30-day/45-day	0,70/0,68/0,69
Lin 2016 [73]	65<	univariate	LR	14,6–19,1	39,156, 178,286	30-day, 1-year	inpatient: 0,655(95% CI, 0.646–0.664), outpatient: 0.642 (95% CI, 0.639–0.646)
agrawal 2016 [87]	all	backward elimination	LR, RF, GBM & ensemble	4,3/20,8	64,252	72-hour, 30-day	0,66 / 0,711
bergese 2017 [25]	Pediatric ED		Classification Tree, NN	2,2	28,341	120-hour	Accuracy: 81, sensitivity: 79.8, specificity: 97
dorajoo 2017 [34]	All	backward elimination LR	LR	45	1291	15-day	0,65
leong 2017 [44]	HF	bivariate	LR	9,8	1475	30-day	0,76
Tabak 2017 [77]	all	univariate LR	LR	11,9	1195640	30-day	0,72
Casalini 2017 [79]	all	univariate	LR	11,6	5388	30-day	PPV: 78.3, specificity: 99.1, sensitivity: 24.8
jamei 2017 [80]	all	correlation-based	NN	9,7	323,813	30-day	0,78
collins 2017 [81]	type-2 diabetes	Gini split worth	LR	17,1	63,237	30-day	0,82
fernandez 2017 [82]	HF	chi-square	LR	20	27,581	30-day	NR
greenwald 2017 [86]	all	univariate LR -> backward elimination	LR	11,7	30,000	30-day	0,74(95% CI, 0.73–0.75)
mesgarpour 2017 [98]	all	RF, SVM -> forward stepwise BPM	BPM	NR	3,893,508	1-year	0,759–0,771

Abbreviations: AUC, Area Under the (ROC) Curve; NR, Not reported; PPV, Predictive Positive Value; CI, Confidence Interval; HF, Heart failure; ICU, Intensive Care Unit; AMI, Acute Myocardial Infarction; ED, Emergency Department; COPD, Chronic Obstructive Pulmonary Disease; HIV, Human Immunodeficiency Virus; OPAT, Outpatient Parenteral Antimicrobial Therapy; RF, Random Forest; SVM, Support Vector Machine; NN, Neural Network; LR: Logistic Regression; GLM, Generalized Linear Model; GEE, Generalized Estimating Equation; HGLM, Hierarchical Generalized Linear Model; CPHM, Cox Proportional Hazards Model; LASSO, Least Absolute Shrinkage and Selection Operator; PCA, Principal Component analysis; CBFS, Correlation-Based Feature Selection; GBM, Gradient Boosting Method; BPM, Bayes Point Machine.

[1] Medicare beneficiaries

[2] Major diagnostic category 5 (MDC-5)

introduce classification algorithms widely used in multiple predictive modelling fields. These are not limited to the classifier itself, as they also encompass a wider set of techniques such as feature selection, variable discretization or missing value imputation among others.

Fig. 3 shows a simplified taxonomy of procedures carried out in the studies included in this review. The most significant methods are categorized within the chart according to the modelling approach they propose.

Fig. 4 shows the evolution of the proportion of modelling techniques according to the type of approach. A trend can be observed where machine learning techniques emerging in recent years are gaining relevance over the statistical modelling techniques.

3.3. Feature selection techniques

Feature selection (also referred to as feature subset selection or FSS) is a common practice in many data analytic fields that allows the identification of the most significant variables of a dataset. In medicine, it is of special importance since it allows the identification of the key factors associated to a disease, or a specific risk condition. Feature selection techniques are of special interest due to their ability to reduce both overfitting and the complexity of the models, as they can help with model interpretability. This is particularly important in the clinical environment, where data acquisition is often related to costly procedures and there is a need for clinicians to understand and believe the model.

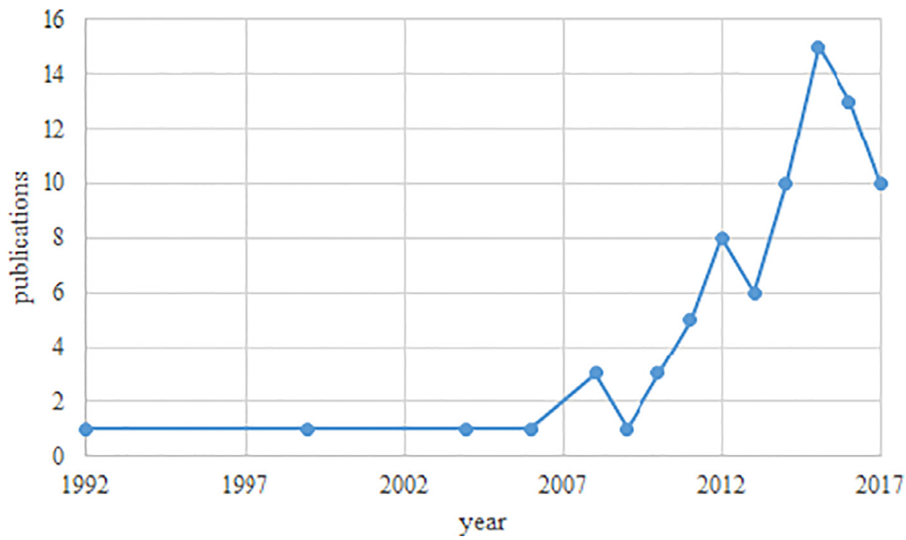

Fig. 2. Number of publications per year.

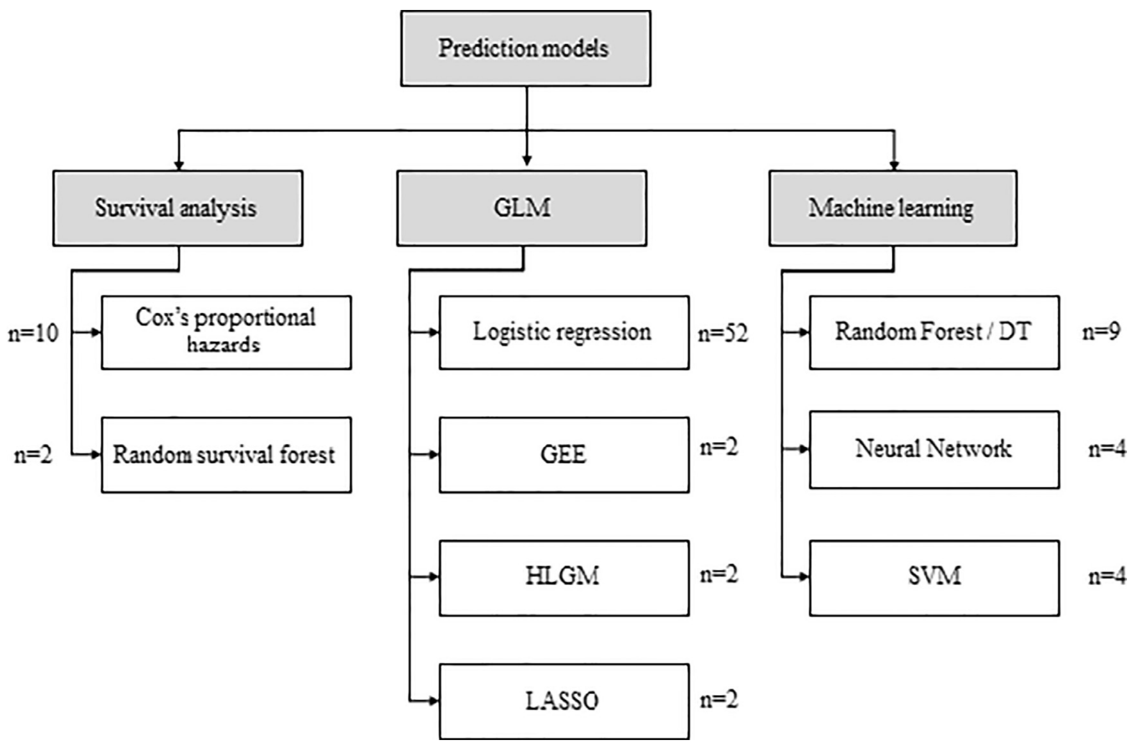

Fig. 3. Taxonomy of main data analysis methods. GLM: Generalized Linear Model, GEE: Generalized Estimated Equation, Hierarchical Generalized Linear Model, LASSO: Least Absolute Shrinkage and Selection Operator, DT, Decision Tree, SVM: Support Vector Machine.

In the context of readmission risk prediction, feature selection is closely related to the classification model used. Here we can clearly distinguish the classical approach, consisting of a regression analysis procedure preceded by a univariate parametric or model-free method that selects the most significant variables to be included in the model.

The most extended feature selection procedure for those models following the classical regression-based approach is to perform a univariate analysis by means of statistical tests such as Student's *t*-test, chi2, Wilcoxon or ANalysis Of VAriance (ANOVA). Significant predictors from the univariate analyses are then included in the final model. Variables with p-values lower than a pre-established threshold (typically 0.001 [16] although it does vary among studies) are considered significant features.

A more refined hybrid approach that includes a stepwise [15] approach is widely utilized among regression-based models. Below we describe the (logistic) regression with a multi-step heuristic approach, which consists of the following steps:

1. Univariate variable selection (optional): For every feature, a univariate logistic regression model is built. Only features with a p-value from a given statistical test below a specified threshold are retained.
2. A multivariable logistic regression is built on a stepwise fashion. There are two basic approaches:
   a. Forward selection, which initializes the model without features and iteratively adds features, retaining only those

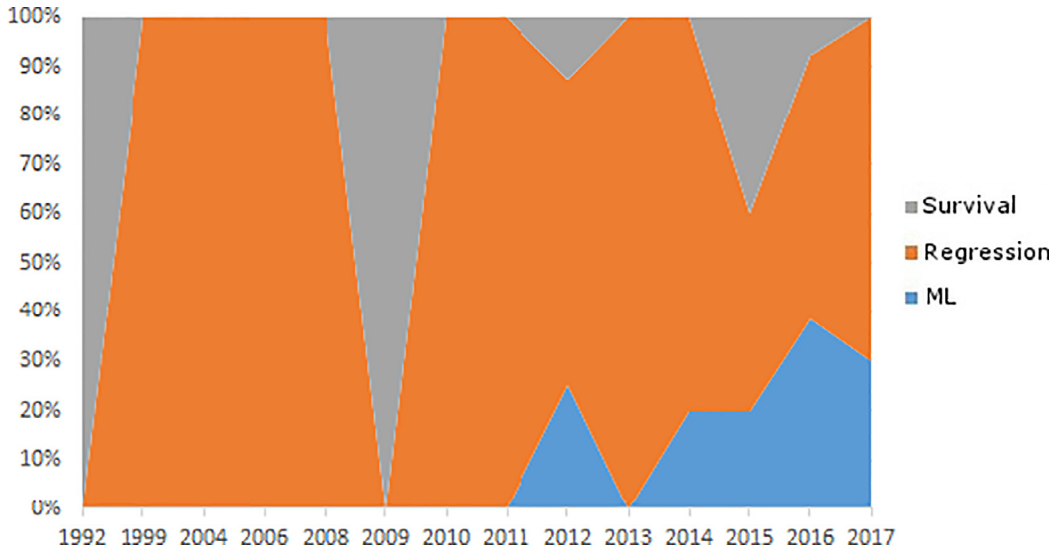

Fig. 4. Distribution of methods per type and year. ML: Machine Learning. Note that years without any publication included in the study are not present.

whose addition shows a statistically significant improvement of the fit.

b. Backward elimination, which initializes the model with the whole set of features to iteratively remove the features that do not improve (or worsen) the model fit.

3. The final logistic regression model is built using the features selected in previous steps.

Other types of regression analysis such as LASSO are also found in the literature in order to build parsimonious models [64,54,88]. Out of the studies following a machine learning approach, different feature selection techniques are utilized. Abdelrahman et al. [16] follow an approach consisting of systematically evaluating different feature selection and ranking methods such as wrapper subset selection, information gain, gain ratio and symmetrical uncertainty. Cai et al. [29] use a correlation-based feature selection (CBFS) method for selecting the most significant features. Some other authors follow an embedded feature selection approach, which consists in conducting the feature search within the classifier itself, as part of the learning process. In [85] Shadmi et al. apply decision trees (CART and C5.0) as well as neural networks to find the most significant feature set that will feed the LR model. The authors state that feature selection using data mining was especially useful, “as with large datasets statistical significance is easily reached in univariate analysis”. Bootstrapping (sampling with replacement) is a technique used for finding a good feature subset without trying every subset (exhaustive approach). This technique along with random feature selection is used in RF, which is used as an embedded feature selection method in [21,23,71,98]. This method estimates the importance of each variable based on the difference of the out-of-bag error after permuting the values of each feature [97]. Nevertheless, many of the papers that follow machine learning approaches do not report the use of any specific feature selection approach. This is probably related to the ability of the classification models to handle modest-to-high dimensional datasets (e.g. RF, SVM). Also, features included in the models may be previously selected based on clinical importance using either published models and/or domain expert criteria [64].

3.4. Class imbalance

In supervised classification, data imbalance occurs when the a priori probabilities of the classes are significantly different, i.e. there exists a minority (positive) class having relatively less number of instances than the majority (negative) class. In readmission prediction, as well as in many other fields (e.g. fraud detection or fault diagnosis), instances of the minority class are outnumbered by the negative instances. Most of the classification algorithms assume relatively balanced a priori probabilities for both classes [96], so when the training dataset is imbalanced, the resulting model is biased towards the majority class.

Readmission prediction is an intrinsically imbalanced problem. The all-cause 30-day readmission rate is estimated at 20% [41], although it varies greatly depending on multiple factors (e.g. readmission threshold, subpopulation characteristics, etc.). The level of class imbalance of a dataset is given by the imbalance ratio (IR), so that an IR of 1:10 expresses that for each sample of the positive class there are 10 samples of the negative class.

Unlike most standard classification algorithms used in machine learning (e.g. decision trees or linear discriminant analysis), linear regression is not affected by class imbalance (at least for modestly imbalanced data) [89]. According to Oomen et al. [95], when using Maximum-likelihood LR, the separability of the data is not affected by class-imbalance. The authors conclude that for LR models it is important to use a sample that has the same class distribution as the original population rather than introducing sampling bias (e.g. via resampling). Most of the current standard classification algorithms favour the larger class when the class sizes differ considerably, achieving a high accuracy in the majority class and low accuracy for the minority class [94].

On account of this issue, several correction strategies at both data and algorithm level have been developed in an attempt to address this problem. Fig. 5 depicts a simplified taxonomy of the different existing approaches for addressing class-imbalance.

As shown in Table 4, most studies using machine learning algorithms do not report the use of any procedure regarding class imbalance. Among those who do, resampling is the most utilized strategy to overcome class imbalance, either undersampling or oversampling.

In [69] the authors make use of random oversampling and conclude that the “proposed methods tend to have better generalization results when trained with the data after over-sampling”. Although Zheng et al. [69] point out the possibility of using more sophisticated resampling methods such as SMOTE [90], the studies included in this review that do resampling employ basic random over or sub sampling techniques.

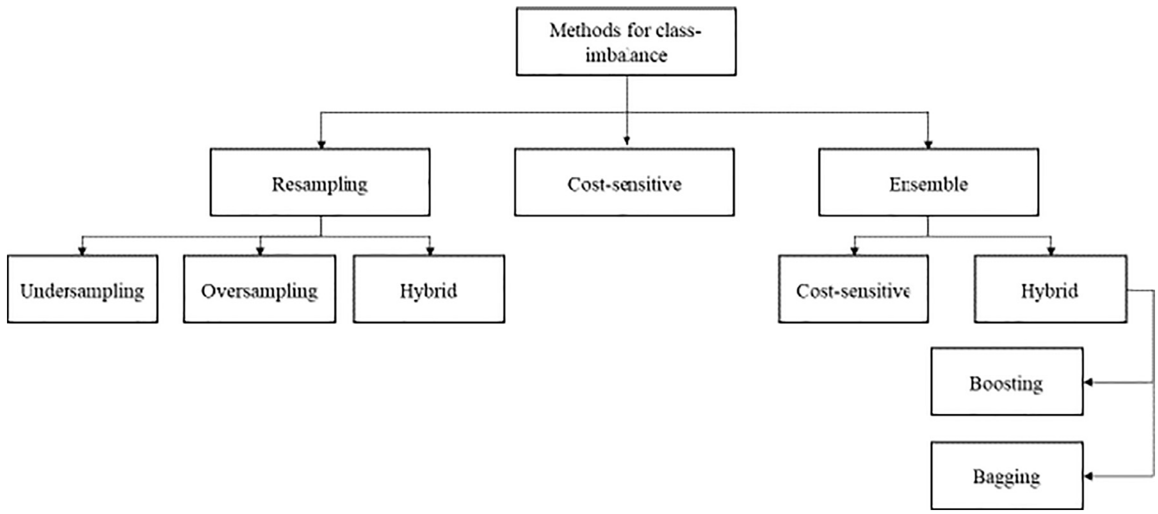

Fig. 5. Simplified taxonomy of class-imbalance addressing techniques.

Table 4
Class imbalance addressing methods in readmission risk prediction.

Source	Class imbalance correction	IR
Amalakuhan et al. [21], 2012	–	1:2.1
Au et al. [23], 2012	–	1:5.3
Walsh and Hripcsak [64], 2014	undersampling	1:14
Yu et al. [67], 2015	–	1:5.3
Zheng et al. [69], 2015	oversampling	1:4.6
Futoma et al. [88], 2015	–	1:5.2
Cai et al. [29], 2016	–	–
Fisher et al. [36], 2016	–	1:4
Turgeman and May [61], 2016	Ensemble (boosting)	1:3.6
Mortazavi et al. [70], 2016	undersampling / oversampling / cost-sensitive*	1:6.8
Agrawal et al. [87], 2016	–	1:4.8
Bergese et al. [25], 2017	–	1:45.5
Jamei et al. [80], 2017	–	1:10.3
Mesgarpour et al. [98], 2017	–	–

*cost-sensitive (weighting) chosen for final model.

In [70] different methods are compared, including resampling (oversampling and subsampling) and cost-sensitive learning (weighting). According to the authors, weighting achieved the best results and this method was included in the final model. By contrast, Turgeman et al. [61] use a mixed-ensemble model which controls the classification error of the minority class, based on a boosting meta-algorithm.

3.5. Performance of the models

Sixty-four of the included studies reported overall model performance metrics. According to that stated in Section 2.2.3, we compared the discrimination ability of the models using the AUC, when available. In respect of the discrimination ability of the models, most of them report modest AUC scores, mostly below 0.75, which matches the results presented in [6] whereas almost 19% of the models reported AUC scores above 0.75. However, the discrimination of the models is not comparable since they are greatly influenced by the population object of the study as well as factors such as readmission length threshold. Fig. 6 shows the values of the readmission risk prediction models reporting AUC.

The main problem with evaluating the performance of readmission risk prediction models is the existence of multiple factors that affect it. The choice of the readmission threshold is one important aspect that is related to the model's performance. Table 5 presents the mean, minimum, maximum and standard deviation of the reported AUC values according to the readmission thresholds (in days) used.

Table 5
Performance of models reporting AUC by readmission threshold.

Threshold	# of studies	Mean	Min	Max	Std. dev.
<3	2	0.725	0.66	0.79	0.092
7	2	0.78	0.74	0.82	0.056
15	2	0.675	0.65	0.7	0.035
30*	45	0.71	0.60	0.92	0.076
45	1	0.69	0.69	0.69	–
90	1	0.65	0.65	0.65	–
180	2	0.707	0.654	0.76	0.075
365	4	0.725	0.7	0.77	0.031

*Including 28-day.

4. Discussion

4.1. Prediction performance of the models

The Area Under Receiver Operating Characteristic (ROC) Curve or c-statistic is the standard de facto metric for measuring the discrimination ability of readmission risk prediction models. Given that the main goal of some studies is to identify predictors associated to readmission, these kind of studies often do not provide the c-statistic as the overall performance metric. In addition, a

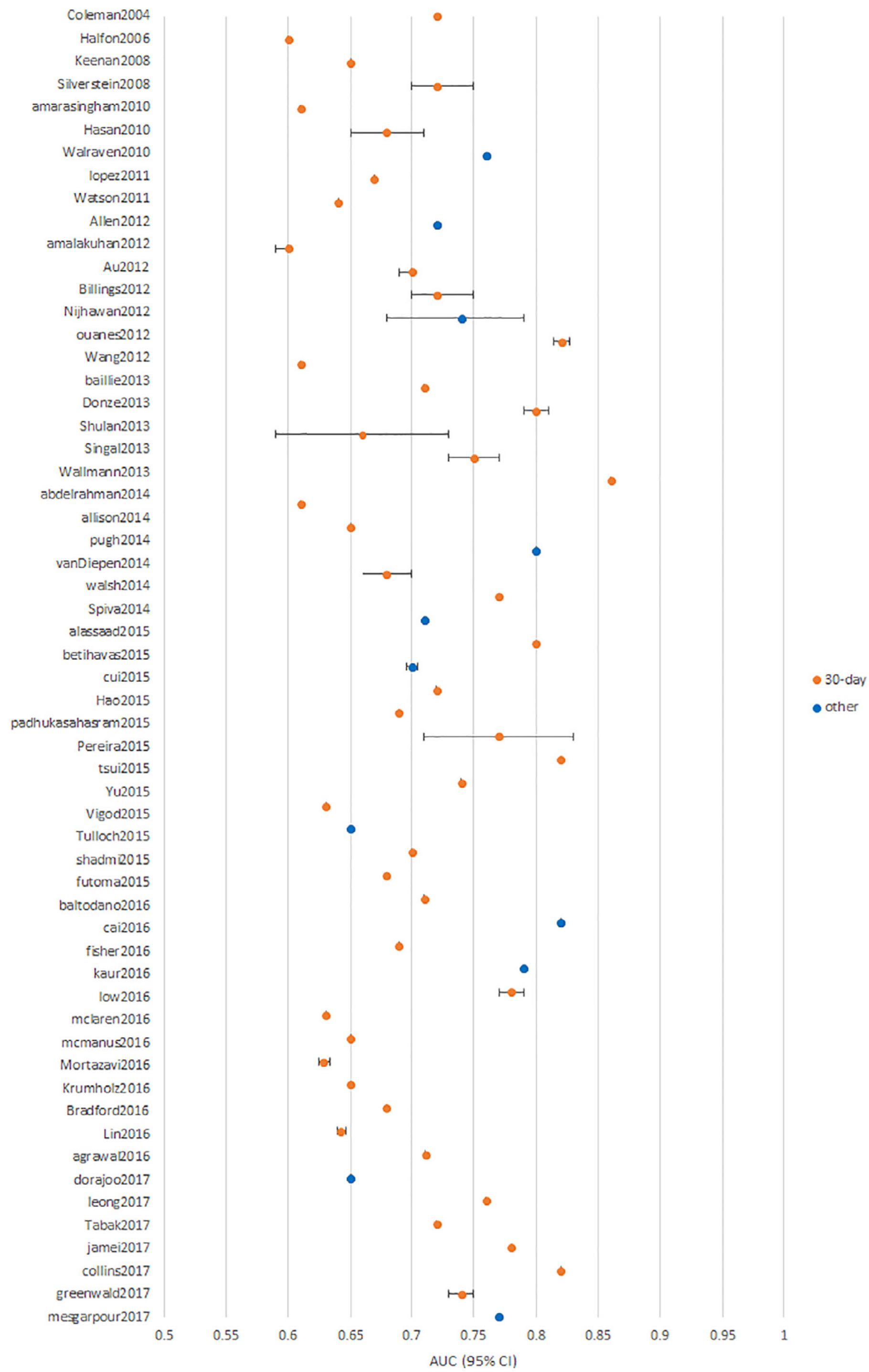

Fig. 6. AUC of the prediction models (with 95% confidence intervals when reported). Thirty-day readmission threshold includes studies using 28-day readmission. Only includes studies reporting AUC.

minority of studies report sensitivity and specificity scores along with PPV instead of AUC [51,25,84,83,79].

The average AUC among studies is 0.71, even though there is a great variability among them (standard deviation is 0.07). Below we summarize the relationship between the predictive performance of the models and different factors.

4.1.1. Readmission threshold

According to Table 5, 30-day is the most widely used readmission threshold, as it is the time span used by more than 76% of the reviewed papers, although we did find time spans ranging from 48 hours to 1 year. Readmission rates also vary depending on this threshold. Longer readmission thresholds are related to higher readmission rates and vice versa. This is not surprising as we would expect a greater probability of being readmitted within a year than in the first 48 hours after discharge. Other factors, such as the population subject to study or the type of clinical study, can greatly influence this indicator.

The observed variability of the readmission rates, and subsequent imbalance ratios (IRs), is caused mainly due to two factors, namely: population or target outcome and readmission thresh-

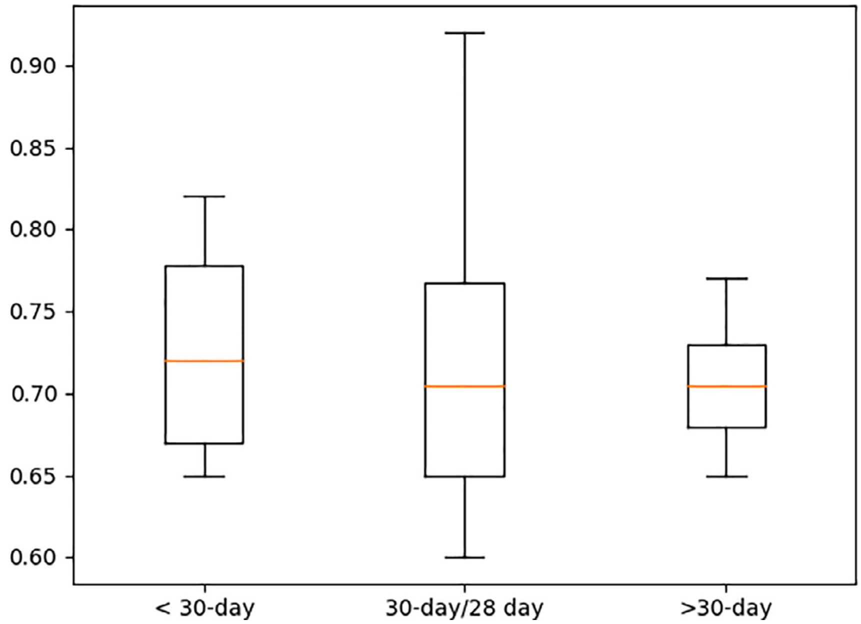

Fig. 7. AUC performance of the models according to the readmission threshold.

Table 6
Thirty-day readmission ratio of target populations.

Target	# of studies	Mean	Max	Min	Std. dev
All-cause	23	15.46	30.4	4.7	6.77
HF	17	16.26	28	4.2	6.57
65<	8	18.88	50	6.1	13.83
Other	8	19.44	27	9.2	6.36

old. Results show (see Table 6) that the all-cause 30-day readmission rate is 15.46% on average, which agrees with average ratios reported in the literature [32]. Some studies introduce variability, such as the one by Halfon et al. [28], which considers only "potentially avoidable" readmissions, causing a much lower rate or [30] which excludes "normal delivery" admissions. The elderly population tends to have higher readmission rates, although great variability is observed among studies. Studies considering only emergency department admissions such those by Cui et al. [31] report considerably lower readmission ratios than the average (18.9%) for elderly population (6.1% and 7.8% respectively).

Table 6 synthesizes the prediction performance of models (those reporting AUC) based on the selected readmission threshold. We can observe from the results that models using shorter thresholds (7 days or lower) perform better than average, whereas longer thresholds perform similar to models using 30-day readmissions. Fig. 7 summarizes the performance difference across readmission thresholds, divided into less than thirty days, thirty days (including 28 days for simplification) and more than thirty days. We observe that there exists a great variability among models within 30-day threshold group.

4.1.2. Effect of the targeting reason

Comparison of models' predictive performance across different studies may be challenging due to differences in the target population. Each target population has its own inherent specificities, so when comparing models targeting different populations (e.g. heart failure and pneumonia) may not be very informative. Some studies can be found where separate models are built for all-cause and specific reasons, so that their performance can be compared. Fig. 8 shows the AUC reported in [19,64,67,70] for models targeting all-cause and Heart Failure (HF) for 30-day readmission.

According to Fig. 8, three studies reported better performance ($7.7\% \pm 6.5$) for models predicting all-cause admissions than HF admissions. Walsh et al. [30] on the contrary, reported significantly better values (35% more) for HF readmission prediction. Nonetheless, the authors hypothesized that those results would be highly sensitive to the definition of the cohort.

4.1.3. Mortality

Results reported in this systematic review target only the readmission prediction, excluding mortality. Although mortality models are not covered in this review, we observed that a minority (9%) of the included studies included death outcomes [22,63,23,52,65,59,29]. Those studies build separate models for readmission and death outcomes, while at times they present results of a combination of both, i.e., hospitalization or death [63,23,65].

According to the reported predictive performance of the models (see Fig. 9) death was more predictable than readmission ($9.9\% \pm 7.7$) in all the studies except in [65], where the readmission prediction models perform slightly better than mortality and hospitalization or mortality for both 30-day and 1-year outcomes (0.82, 0.80, 0.80 and 0.82, 0.76, 0.77 respectively).

As pointed out in [70], in HF it is considered that readmissions are more difficult to predict than mortality outcomes. This difference was shown in a meta-analysis of 117 models by Ouwerkerk et al. [93] where "the average C statistic to predict mortality was 0.71 and hospitalization was 0.63".

4.2. Machine learning techniques

Table 7 various machine learning (ML) techniques were used in the selected studies. They can be classified into the three main categories that were most frequently used, namely tree-based, neural networks and support vector machines. Tree-based methods, which include random forests (RF) and decision trees, are the most frequently used ML classification techniques (13% of all selected studies and 77% of those using ML techniques). In fact, RF is the most utilized ML algorithm (54%), which overall achieved

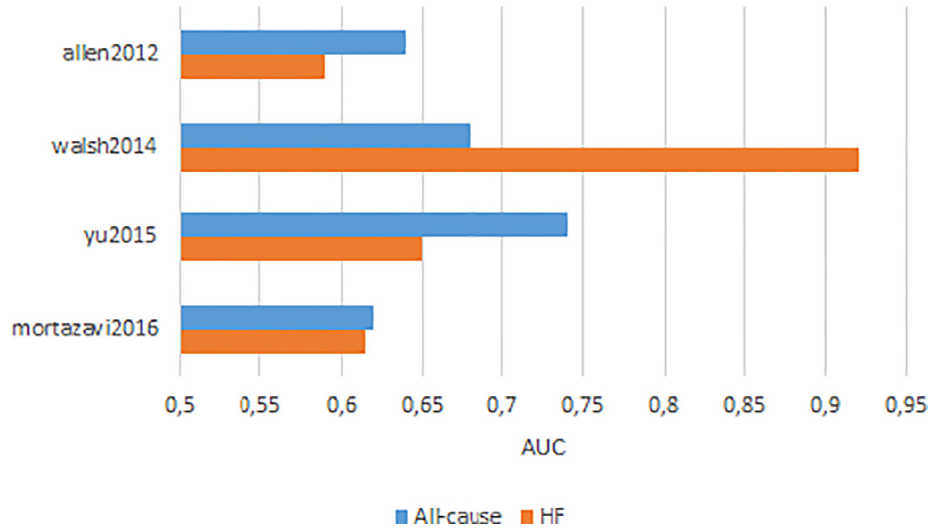

Fig. 8. Comparison of AUC performance of models for all-cause and HF readmission prediction (30-day).

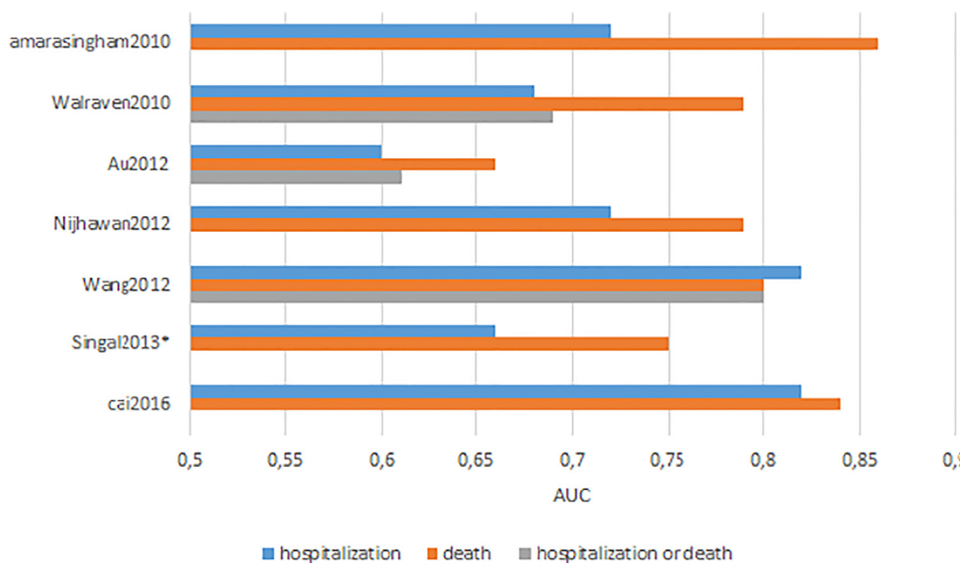

Fig. 9. AUC performance of the models according to the outcome type (hospitalization, death and hospitalization and death). *90-day threshold is used for mortality instead of 30-day in [59].

similar or slightly better performance rates than other prediction techniques such as SVM or SGD. Five of the selected studies (7%) used NN, of which those following deep learning approaches [88,80] outperformed other ML methods. SVM were used in five studies, in which linear kernel was demonstrated to perform better than other kernels, except one study using a radial kernel [69]. The remaining models used either stand-alone classifiers (e.g. Naïve Bayes, Gradient Boosting Method, Stochastic Gradient Descent) or hybrid approaches that combine various classifiers [87,61].

The different approaches reported in different studies cannot be directly compared since each study has its own particular characteristics. According to the results reported by studies where different models are compared under the same conditions, overall we found that more complex models outperform traditional simple ones. Mortazavi et al. [70] concluded that ML methods "can improve both discrimination and range of prediction over traditional statistical techniques", having compared the effectiveness of several ML algorithms against traditional linear regression (LR). Bergese et al. [25] justify the use of decision trees (DT) and artificial neural networks (ANN) "because usually, even if logistic regression achieved greater accuracy, it is outperformed by DT and ANN when comparing other performance measures like sensitivity and specificity" . When comparing both approaches, they claim that DT is the best model since it outperforms ANN in sensitivity (79.8% vs 6.9%). Zheng et al. [69] compared several methods such as radial basis function neural networks (RBFNN), LR, RF and particle swarm optimization (PSO)-SVM to find that the latter outperforms in terms of accuracy and sensitivity. In [61] the authors built a mixed-ensemble that combined a boosted DT and SVM that performed better than standalone classifiers.

Some other studies do not find significant differences among different methods. For instance, Agrawal et al. [87] stated that stacked generalization (stacking) performed marginally better than individual models (Gradient Boosting Method, LR, and RF) which gave similar AUC. Similarly, Walsh and Hripcsak [64] compared regularized regression (LASSO) and SVM concluding that both perform equally. Yu et al. [67] compared models using SVM (various kernels) as a classification approach and Cox regression as a prognosis approach to find that linear SVM performed marginally better. The authors conclude that "more sophisticated learning

Table 7
Reported performance of methods in studies using ML for classification.

	RF	SVM			DT		NN		SGD	Naive Bayes	Bayesian network	BPM	Boosting	GBM
		radial	linear	polynomial	CART	C5.0*		deep learning						
Amalakuhan et al. [21], 2012	0.72													
Au et al. [23], 2012	0.6													
Walsh and Hripcsak [64], 2014		0.68												
Yu et al. [67], 2015			0.74	0.7										
Zheng et al. [69], 2015	0.87*	0.97*	0.03*	0.52*			0.56*							
Futoma et al. [88], 2015	0.684		0.671	0.588				0.638 –0.734	0.672					
Cai et al. [29], 2016											0.82			
Fisher et al. [36], 2016					0.67									
Turgeman et al. [61], 2016			0.643			0.693	0.639			0.676				
Mortazavi et al. [70], 2016	0.628												0.615	
Agrawal et al. [87], 2016	0.714													0.719
Bergese et al. [25], 2017					0.79*		0.07*							
Jamei et al. [80], 2017	0.77							0.78						
Mesgarpour et al. [98], 2017												0.77		

* Sensitivity.

algorithms often may not yield better results due to lack of relevant data". Futoma et al. [88] compared different variations of LR (Maximum-likelihood LR, Penalized LR and LR with multi-step LR) and some ML methods (RF, SVM, Stochastic Gradient Descent and deep NN). Results show that the accuracy of the best performing models, namely RF, PLR and SGD, was statistically indistinguishable. However, the authors reported that deep NN outperforms previous approaches.

Although more complicated modelling approaches such as deep learning are considered to have "the greatest potential to boost predictive accuracy in statistical approaches" [88], they are also the most difficult to manage (due to the large number of parameters) and more challenging to interpret. According to Zheng et al. [69], although their PSO-SVM approach outperforms other prediction models, the difficulty of parameter tuning introduces computational and time-related problems as well as a risk of overfitting in the training process.

4.3. Limitations of this review

This review used a systematic search strategy that aimed to find an appropriate trade-off between coverage and size. There are limitations to any search strategy and it is possible that related articles may have been overlooked. The inclusion of additional terms in the search string as well as the inclusion of additional databases could provide a more comprehensive understanding of the methods used for building readmission risk prediction models.

5. Conclusions

Our literature review included 77 studies that described prediction models for hospital readmission risk. Although statistical modelling techniques have prevailed and are still popular techniques, machine learning approaches have emerged in recent years as a promising technique that can improve the predictive ability of readmission risk prediction models.

Clinical trials tend to follow established procedures when it comes to statistical data analysis. Logistic regression and Cox regression (or proportional hazards regression) are the dominant modelling methods. In the most recent studies, however, machine learning algorithms such as random forests, neural networks or SVMs, have been increasingly used (from none to 38% of yearly publications in the last five years). Within the studies that use machine learning techniques, we found that class imbalance is only addressed in a reduced subset of them (30%), despite being a well-known problem which affects model training and correct evaluation. A comparative study on class-balancing techniques may also be useful for future researchers in this field. In respect of feature selection techniques, we observed that established and widely used univariate approaches are the most common, albeit possibly the simplest. Stepwise regression is also an extended feature reduction procedure intended to produce parsimonious models.

The results obtained show that the performance of the models varies greatly depending on the target population and readmission threshold. Short-time readmission thresholds (7 days or less) tend to achieve better results in terms of AUC. Overall, more complex ML models outperform traditional statistical models such as LR. Nevertheless, emerging techniques (e.g. as deep learning) are often difficult to tune and less interpretable. In addition, the lack of relevant data may limit the performance of any model, so even more complex models may not yield better results.

Recent studies introducing machine learning techniques report promising results and anticipate advantages over classical methods. Nevertheless, further comparative studies are needed to assess the real impact of these techniques in the domain of readmission risk prediction. In this regard, the lack of a benchmark dataset makes a comparison of the models' performance difficult across different studies.

Appendix A

Table 8
Studies including death as outcome and performance of the models (at 30-days threshold) depending on the outcome type.

Study	AUC		
	Hospitalization	Death	Hospitalization or death
amarasingham2010	0.72	0.86	–
Walraven2010	0.68	0.79	0.69
Au2012	0.6	0.66	0.61
Nijhawan2012	0.72	0.79	–
Wang2012	0.82	0.8	0.8
Singal2013*	0.66	0.75*	–
cai2016	0.82	0.84	–

*90-day

Table 9
AUC values of models for different target population.

Study	All-cause	HF	AMI	Pneumonia
allen2012	0.64	0.59		
walsh2014	0.68	0.92		
yu2015	0.74	0.65	0.66	0.63
mortazavi2016	0.62	0.615		

References

[1] R. Wallmann, J. Llorca, I. Gómez-Acebo, Ortega ÁC, F.R. Roldan, T. Dierssen-Sotos, Prediction of 30-day cardiac-related-emergency-readmissions using simple administrative hospital data, Int. J. Cardiol. 164 (2) (2013) 193–200.
[2] K. Dharmarajan, A.F. Hsieh, Z. Lin, H. Bueno, J.S. Ross, L.I. Horwitz, J.A. Barreto-Filho, N. Kim, S.M. Bernheim, L.G. Suter, E.E. Drye, Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia, JAMA 309 (4) (2013) 355–363.
[3] C.A. Baillie, C. VanZandbergen, G. Tait, A. Hanish, B. Leas, B. French, C. William Hanson, M. Behta, C.A. Umscheid, The readmission risk flag: using the electronic health record to automatically identify patients at risk for 30-day readmission, J. Hosp. Med. 8 (12) (2013) 689–695.
[4] Centers for Medicare and Medicaid Services (CMS), HHS. Medicare program; hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system and Fiscal Year 2014 rates; quality reporting requirements for specific providers; hospital conditions of participation; payment policies related to patient status, Final rules. Fed. Regist. 78 (160) (2013) 50495.
[5] Z. Kmietowicz, Hospitals will be fined for emergency readmissions, says Lansley, BMJ: Br. Med. J. (Online) 340 (2010) 340.
[6] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, S. Kripalani, Risk prediction models for hospital readmission: a systematic review, JAMA 306 (15) (2011) 1688–1698.
[7] M.J. Swain, H. Kharrazi, Feasibility of 30-day hospital readmission prediction modeling based on health information exchange data, Int. J. Med. Inf. 84 (12) (2015) 1048–1056.
[8] J.S. Ross, G.K. Mulvey, B. Stauffer, V. Patlolla, S.M. Bernheim, P.S. Keenan, H.M. Krumholz, Statistical models and patient predictors of readmission for heart failure: a systematic review, Arch. Intern. Med. 168 (13) (2008) 1371–1386.
[9] A.L. Leppin, M.R. Gionfriddo, M. Kessler, J.P. Brito, F.S. Mair, K. Gallacher, Z. Wang, P.J. Erwin, T. Sylvester, K. Boehmer, H.H. Ting, Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials, JAMA Intern Med. 174 (7) (2014) 1095–1107.
[10] I. Kadi, A. Idri, J.L. Fernandez-Aleman, Knowledge discovery in cardiology: a systematic literature review, Int. J. Med. Inf. 97 (2017) 12–32.
[11] A. Artetxe, B. Ayerdi, M. Graña, S. Rios, Using anticipative hybrid extreme rotation forest to predict emergency service readmission risk, J. Comput. Sci. 20 (2017) 154–161.
[12] A. Artetxe, M. Graña, A. Beristain, S. Ríos, Emergency department readmission risk prediction: a case study in Chile, in: International Work-Conference on the Interplay between Natural and Artificial Computation, Springer, 2017, pp. 11–20.
[13] D. Tranfield, D. Denyer, P. Smart, Towards a methodology for developing evidence-informed management knowledge by means of systematic review, Br. J. Manag. 14 (3) (2003) 207–222.
[14] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, Prisma Group, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement, PLoS Med. 6 (7) (2009) e1000097.
[15] S. Greenland, Modeling and variable selection in epidemiologic analysis, Am. J. Public Health 79 (3) (1989) 340–349.
[16] S.E. AbdelRahman, M. Zhang, B.E. Bray, K. Kawamoto, A three-step approach for the derivation and validation of high-performing predictive models using an operational dataset: congestive heart failure readmission case study, BMC Med. Inf. Decis. Making 14 (1) (2014) 41.
[17] A. Alassaad, H. Melhus, M. Hammarlund-Udenaes, M. Bertilsson, U. Gillespie, J. Sundström, A tool for prediction of risk of rehospitalisation and mortality in the hospitalised elderly: secondary analysis of clinical trial data, BMJ Open 5 (2) (2015) e007259.
[18] N. Allaudeen, A. Vidyarthi, J. Maselli, A. Auerbach, Redefining readmission risk factors for general medicine patients, J. Hosp. Med. 6 (2) (2011) 54–60.
[19] L.A. Allen, K.E. Tomic, D.M. Smith, K.L. Wilson, I. Agodoa, Rates and predictors of 30-day readmission among commercially insured and medicaid-enrolled patients hospitalized with systolic heart failure, Circ: Heart Fail. 5 (6) (2012) 672–679 CIRCHEARTFAILURE-112.
[20] G.M. Allison, E.G. Muldoon, D.M. Kent, J.K. Paulus, R. Ruthazer, A. Ren, D.R. Snydman, Prediction model for 30-day hospital readmissions among patients discharged receiving outpatient parenteral antibiotic therapy, Clin. Infect. Dis. 58 (6) (2013) 812–819.
[21] B. Amalakuhan, L. Kiljanek, A. Parvathaneni, M. Hester, P. Cheriyath, D. Fischman, A prediction model for COPD readmissions: catching up, catching our breath, and improving a national problem, J. Community Hosp. Intern. Med. Perspect. 2 (1) (2012) 9915.
[22] R. Amarasingham, B.J. Moore, Y.P. Tabak, M.H. Drazner, C.A. Clark, S. Zhang, W.G. Reed, T.S. Swanson, Y. Ma, E.A. Halm, An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data, Med. Care 48 (11) (2010) 981–988.
[23] A.G. Au, F.A. McAlister, J.A. Bakal, J. Ezekowitz, P. Kaul, C. van Walraven, Predicting the risk of unplanned readmission or death within 30 days of discharge after a heart failure hospitalization, Am. Heart J. 164 (3) (2012) 365–372.
[24] P.A. Baltodano, Y. Webb-Vargas, K.C. Soares, C.W. Hicks, C.M. Cooney, P. Cornell, K.K. Burce, T.M. Pawlik, F.E. Eckhauser, A validated, risk assessment tool for predicting readmission after open ventral hernia repair, Hernia 20 (1) (2016) 119–129.
[25] I. Bergese, S. Frigerio, M. Clari, E. Castagno, PonticelliE De Clemente A, E. Scavino, P. Berchialla, An innovative model to predict pediatric emergency department return visits, Pediatr. Emerg. Care (2017), doi:10.1097/PEC.0000000000000910.
[26] K. Berman, S. Tandra, K. Forssell, R. Vuppalanchi, J.R. Burton, J. Nguyen, D. Mullis, P. Kwo, N. Chalasani, Incidence and predictors of 30-day readmission among patients hospitalized for advanced liver disease, Clin. Gastroenterol. Hepatol. 9 (3) (2011) 254–259.
[27] V. Betihavas, S.A. Frost, P.J. Newton, P. Macdonald, S. Stewart, M.J. Carrington, Y.K. Chan, P.M. Davidson, An absolute risk prediction model to determine unplanned cardiovascular readmissions for adults with chronic heart failure, Heart, Lung and Circ. 24 (11) (2015) 1068–1073.
[28] J. Billings, I. Blunt, A. Steventon, T. Georghiou, G. Lewis, M. Bardsley, Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30), BMJ Open 2 (4) (2012) e001667.
[29] X. Cai, O. Perez-Concha, E. Coiera, F. Martin-Sanchez, R. Day, D. Roffe, B. Gallego, Real-time prediction of mortality, readmission, and length of stay using electronic health record data, J. Am. Med. Inform. Assoc. 23 (3) (2015) 553–561.
[30] E.A. Coleman, S.J. Min, A. Chomiak, A.M. Kramer, Posthospital care transitions: patterns, complications, and risk identification, Health Serv. Res. 39 (5) (2004) 1449–1466.
[31] Y. Cui, C. Metge, X. Ye, M. Moffatt, L. Oppenheimer, E.L. Forget, Development and validation of a predictive model for all-cause hospital readmissions in Winnipeg, Canada, J. Health Serv. Res. Policy 20 (2) (2015) 83–91.
[32] M. Deschodt, E. Devriendt, M. Sabbe, D. Knockaert, P. Deboutte, S. Boonen, J. Flamaing, K. Milisen, Characteristics of older adults admitted to the emergency department (ED) and their risk factors for ED readmission based on comprehensive geriatric assessment: a prospective cohort study, BMC Geriatr 15 (1) (2015) 54.
[33] 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.
[34] S.R. Dorajoo, V. See, C.T. Chan, J.Z. Tan, D.S. Tan, S.M. Abdul Razak, T.T. Ong, N. Koomanan, C.W. Yap, A. Chan, Identifying potentially avoidable readmissions: a medication-based 15-day readmission risk stratification algorithm, Pharmacotherapy 37 (3) (2017) 268–277.
[35] A.M. Epstein, A.K. Jha, E.J. Orav, The relationship between hospital admission rates and rehospitalizations, N. Engl. J. Med. 365 (24) (2011) 2287–2295.
[36] S.R. Fisher, J.E. Graham, S. Krishnan, K.J. Ottenbacher, Predictors of 30-day readmission following inpatient rehabilitation for patients at high risk for hospital readmission, Phys. Ther. 96 (1) (2016) 62–70.
[37] G.M. Garrison, M.P. Mansukhani, B. Bohn, Predictors of thirty-day readmission among hospitalized family medicine patients, J. Am. Board Family Med. 26 (1) (2013) 71–77.

[38] P. Halfon, Y. Eggli, G. van Melle, J. Chevalier, J.B. Wasserfallen, B. Burnand, Measuring potentially avoidable hospital readmissions, J. Clin. Epidemiol. 55 (6) (2002) 573–587.
[39] S. Hao, Y. Wang, B. Jin, A.Y. Shin, C. Zhu, M. Huang, L. Zheng, J. Luo, Z. Hu, C. Fu, D. Dai, Development, validation and deployment of a real time 30 day hospital readmission risk assessment tool in the Maine Healthcare Information Exchange, PLoS One 10 (10) (2015) e0140271.
[40] O. Hasan, D.O. Meltzer, S.A. Shaykevich, C.M. Bell, P.J. Kaboli, A.D. Auerbach, T.B. Wetterneck, V.M. Arora, J. Zhang, J.L. Schnipper, Hospital readmission in general medicine patients: a prediction model, J. Gen. Intern. Med. 25 (3) (2010) 211–219.
[41] S.F. Jencks, M.V. Williams, E.A. Coleman, Rehospitalizations among patients in the Medicare fee-for-service program, N. Engl. J. Med. 360 (14) (2009) 1418–1428.
[42] H. Kaur, J.M. Naessens, A.C. Hanson, K. Fryer, M.E. Nemergut, S. Tripathi, PROPER: Development of an early pediatric intensive care unit readmission risk prediction tool, J. Intensive Care Med. 33 (1) (2016) 29–36 0885066616665806.
[43] P.S. Keenan, S.L. Normand, Z. Lin, E.E. Drye, K.R. Bhat, J.S. Ross, J.D. Schuur, B.D. Stauffer, S.M. Bernheim, A.J. Epstein, Y. Wang, An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure., Circ. Cardiovasc. Qual. Outcomes. 1 (1) (2008) 29–37.
[44] K.T. Leong, L.Y. Wong, K.C. Aung, M. Macdonald, Y. Cao, S. Lee, W.L. Chow, S. Doddamani, A.M. Richards, Risk stratification model for 30-day heart failure readmission in a multiethnic south east asian community, Am. J. Cardiol. 119 (9) (2017) 1428–1432.
[45] S. López-Aguilà, J.C. Contel, J. Farre, J.L. Campuzano, L. Rajmil, Predictive model for emergency hospital admission and 6-month readmission, Am. J. Manag. Care 17 (9) (2011) e348–e357.
[46] L.L. Low, N. Liu, S. Wang, J. Thumboo, M.E. Ong, K.H. Lee, Predicting 30-day readmissions in an Asian population: building a predictive model by incorporating markers of hospitalization severity, PLoS One 11 (12) (2016) e0167413.
[47] E.R. Marcantonio, S. McKean, M. Goldfinger, S. Kleefield, M. Yurkofsky, T.A. Brennan, Factors associated with unplanned hospital readmission among patients 65 years of age and older in a Medicare managed care plan, Am. J. Med. 107 (1) (1999) 13–17.
[48] D.P. McLaren, R. Jones, R. Plotnik, W. Zareba, S. McIntosh, J. Alexis, L. Chen, R. Block, C.J. Lowenstein, V. Kutyifa, Prior hospital admission predicts thirty–day hospital readmission for heart failure patients, Cardiol. J. 23 (2) (2016) 155–162.
[49] D.D. McManus, J.S. Saczynski, D. Lessard, M.E. Waring, J. Allison, D.C. Parish, R.J. Goldberg, A. Ash, C.I. Kiefe, TRACE-CORE. Investigators, Reliability of predicting early hospital readmission after discharge for an acute coronary syndrome using claims-based data, Am. J. Cardiol. 117 (4) (2016) 501–507.
[50] J.N. Morris, E.P. Howard, K. Steel, R. Schreiber, B.E. Fries, L.A. Lipsitz, B. Goldman, Predicting risk of hospital and emergency department use for home care elderly persons through a secondary analysis of cross-national data, BMC Health Serv. Res. 14 (1) (2014) 519.
[51] H.Q. Nguyen, L. Chu, I.L. Amy Liu, J.S. Lee, D. Suh, B. Korotzer, G. Yuen, S. Desai, K.J. Coleman, A.H. Xiang, M.K. Gould, Associations between physical activity and 30-day readmission risk in chronic obstructive pulmonary disease, Ann. Am. Thorac. Soc. 11 (5) (2014) 695–705.
[52] A.E. Nijhawan, C. Clark, R. Kaplan, B. Moore, E.A. Halm, R. Amarasingham, An electronic medical record-based model to predict 30-day risk of readmission and death among HIV-infected inpatients, JAIDS J. Acquir. Immune Defic. Syndr 61 (3) (2012) 349–358.
[53] I. Ouanes, C. Schwebel, A. Français, C. Bruel, F. Philippart, A. Vesin, L. Soufir, C. Adrie, M. Garrouste-Orgeas, J.F. Timsit, B. Misset, A model to predict short--term death or readmission after intensive care unit discharge, J. Crit. Care 27 (4) (2012) 422–4e1.
[54] B. Padhukasahasram, C.K. Reddy, Y. Li, D.E. Lanfear, Joint impact of clinical and behavioral variables on the risk of unplanned readmission and death after a heart failure hospitalization, PLoS One 10 (6) (2015) e0129553.
[55] L. Pereira, C. Choquet, A. Perozziello, M. Wargon, G. Juillien, L. Colosi, R. Hellmann, M. Ranaivoson, E. Casalino, Unscheduled-return-visits after an emergency department (ED) attendance and clinical link between both visits in patients aged 75 years and over: a prospective observational study, PLoS One 10 (4) (2015) e0123803.
[56] J.A. Pugh, C.P. Wang, S.E. Espinoza, P.H. Noël, M. Bollinger, M. Amuan, E. Finley, M.J. Pugh, Influence of frailty-related diagnoses, high-risk prescribing in elderly adults, and primary care use on readmissions in fewer than 30 days for veterans aged 65 and older, J. Am. Geriatr. Soc. 62 (2) (2014) 291–298.
[57] M. Shulan, K. Gao, C.D. Moore, Predicting 30-day all-cause hospital readmissions, Health Care Manag. Sci. 16 (2) (2013) 167–175.
[58] M.D. Silverstein, H. Qin, S.Q. Mercer, J. Fong, Z. Haydar, Risk factors for 30-day hospital readmission in patients≥ 65 years of age, Proc. (Baylor Univ. Med. Center). 21 (4) (2008) 363.
[59] A.G. Singal, R.S. Rahimi, C. Clark, Y. Ma, J.A. Cuthbert, D.C. Rockey, R. Amarasingham, An automated model using electronic medical record data identifies patients with cirrhosis at high risk for readmission, Clin. Gastroenterol. Hepatol. 11 (10) (2013) 1335–1341.
[60] E. Tsui, S.Y. Au, C.P. Wong, A. Cheung, P. Lam, Development of an automated model to predict the risk of elderly emergency medical admissions within a month following an index hospital visit: a Hong Kong experience, Health Inf. J. 21 (1) (2015) 46–56.
[61] L. Turgeman, J.H. May, A mixed-ensemble model for hospital readmission, Artif. Intell. Med. 72 (2016) 72–82.
[62] S. van Diepen, M.M. Graham, J. Nagendran, C.M. Norris, Predicting cardiovascular intensive care unit readmission after cardiac surgery: derivation and validation of the Alberta Provincial Project for Outcomes Assessment in Coronary Heart Disease (APPROACH) cardiovascular intensive care unit clinical prediction model from a registry cohort of 10,799 surgical cases, Critical Care 18 (6) (2014) 651.
[63] 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.
[64] C. Walsh, G. Hripcsak, The effects of data sources, cohort selection, and outcome definition on a predictive model of risk of thirty-day hospital readmissions, J. Biomed. Inform. 52 (2014) 418–426.
[65] L. Wang, B. Porter, C. Maynard, C. Bryson, H. Sun, E. Lowy, M. McDonell, K. Frisbee, C. Nielson, S.D. Fihn, Predicting risk of hospitalization or death among patients with heart failure in the veterans health administration, Am. J. Cardiol. 110 (9) (2012) 1342–1349.
[66] A.J. Watson, J. O'Rourke, K. Jethwani, A. Cami, T.A. Stern, J.C. Kvedar, H.C. Chueh, A.H. Zai, Linking electronic health record-extracted psychosocial data in real-time to risk of readmission for heart failure, Psychosomatics 52 (4) (2011) 319–327.
[67] S. Yu, F. Farooq, A. Van Esbroeck, G. Fung, V. Anand, B. Krishnapuram, Predicting readmission risk with institution-specific prediction models, Artif. Intell. Med. 65 (2) (2015) 89–96.
[68] A. Zapatero, R. Barba, J. Marco, J. Hinojosa, S. Plaza, J.E. Losa, J. Canora, Predictive model of readmission to internal medicine wards, Eur. J. Intern. Med 23 (5) (2012) 451–456.
[69] B. Zheng, J. Zhang, S.W. Yoon, S.S. Lam, M. Khasawneh, S. Poranki, Predictive modeling of hospital readmissions using metaheuristics and data mining, Expert Syst. Appl. 42 (20) (2015) 7110–7120.
[70] B.J. Mortazavi, N.S. Downing, E.M. Bucholz, K. Dharmarajan, A. Manhapra, S.X. Li, S.N. Negahban, H.M. Krumholz, Analysis of machine learning techniques for heart failure readmissions, Circ.: Cardiovasc. Qual. Outcomes (2016) CIRCOUTCOMES–CIRCOUTCO116.
[71] H.M. Krumholz, S.I. Chaudhry, J.A. Spertus, J.A. Mattera, B. Hodshon, J. Herrin, Do non-clinical factors improve prediction of readmission risk?: Results from the Tele-Hf study, JACC: Heart Fail. 4 (1) (2016) 12–20.
[72] C. Bradford, B.M. Shah, P. Shane, N. Wachi, K. Sahota, Patient and clinical characteristics that heighten risk for heart failure readmission, Res. Soc. Admin. Pharm. 13 (6) (2016) 1070–1081.
[73] K.P. Lin, P.-C. Chen, L.-Y. Huang, H.-C. Mao, D.-C. (Derrick). Chan, Predicting inpatient readmission and outpatient admission in elderly: a population-based cohort study, Med. (Baltimore). 95 (16) (2016) e3484, doi:10.1097/MD.0000000000003484.
[74] J.M. Corrigan, J.B. Martin, Identification of factors associated with hospital readmission and development of a predictive model, Health Serv. Res. 27 (1) (1992) 81.
[75] S.N. Vigod, P.A. Kurdyak, D. Seitz, N. Herrmann, K. Fung, E. Lin, C. Perlman, V.H. Taylor, P.A. Rochon, A. Gruneir, READMIT: a clinical risk index to predict 30-day readmission after discharge from acute psychiatric units, J. Psychiatr. Res. 61 (2015) 205–213.
[76] A.D. Tulloch, A.S. David, G. Thornicroft, Exploring the predictors of early readmission to psychiatric hospital, Epidemiol. Psychiatr. Sci. 25 (2) (2016) 181–193.
[77] Y.P. Tabak, X. Sun, C.M. Nunez, V. Gupta, R.S. Johannes, Predicting readmission at early hospitalization using electronic clinical data: an early readmission risk score, Med Care 55 (3) (2017) 267–275 PubMed PMID: 27755391, doi:10.1097/MLR.0000000000000654.
[78] L. Spiva, M. Hand, L. VanBrackle, F. McVay, Validation of a predictive model to identify patients at high risk for hospital readmission, J. Healthc. Qual (2014), doi:10.1111/jhq.12070.
[79] F. Casalini, S. Salvetti, S. Memmini, E. Lucaccini, G. Massimetti, P.L. Lopalco, G.P. Privitera, Unplanned readmissions within 30 days after discharge: improving quality through easy prediction, Int. J. Qual. Health Care 29 (2) (2017) 256–261 PubMed PMID: 28453826, doi:10.1093/intqhc/mzx011.
[80] M. Jamei, A. Nisnevich, E. Wetchler, S. Sudat, E. Liu, Predicting all-cause risk of 30-day hospital readmission using artificial neural networks, PLoS One 12 (7) (2017) e0181173 eCollection 2017. PubMed PMID: 28708848; PubMed Central PMCID: PMC5510858, doi:10.1371/journal.pone.0181173.
[81] J. Collins, I.M. Abbass, R. Harvey, B. Suehs, C. Uribe, J. Bouchard, T. Prewitt, T. DeLuzio, E. Allen, Predictors of all-cause 30 day readmission among Medicare patients with type 2 diabetes, Curr. Med. Res. Opin. 33 (8) (2017) 1517–1523 Epub 2017 Jun 9. PubMed PMID: 28498094, doi:10.1080/03007995.2017.1330258.
[82] L. Fernandez-Gasso, L. Hernando-Arizaleta, J.A. Palomar-Rodríguez, M.V. Abellán-Pérez, D.A. Pascual-Figal, Trends, causes and timing of 30-day readmissions after hospitalization for heart failure: 11-year population-based analysis with linked data, Int. J. Cardiol. 248 (2017) 246–251 Epub 2017 Aug 4. PubMed PMID: 28801153, doi:10.1016/j.ijcard.2017.07.094.
[83] N.L. Novotny, M.A. Anderson, Prediction of early readmission in medical inpatients using the Probability of Repeated Admission instrument, Nurs Res. 57 (6) (2008) 406–415 PubMed PMID: 19018215, doi:10.1097/NNR.0b013e31818c3e06.
[84] S. Sfoungaristos, G. Hidas, O.N. Gofrit, S. Rosenberg, V. Yutkin, E.H. Landau, D. Pode, M. Duvdevani, A novel model to predict the risk of readmission in

patients with renal colic, J. Endourol. 28 (8) (2014) 1011–1015 Epub 2014 May 15. PubMed PMID: 24742211, doi:10.1089/end.2014.0082.
[85] E. Shadmi, N. Flaks-Manov, M. Hoshen, O. Goldman, H. Bitterman, R.D. Balicer, Predicting 30-day readmissions with preadmission electronic health record data, Med Care 53 (3) (2015) 283–289 PubMed PMID: 25634089, doi:10.1097/MLR.0000000000000315.
[86] J.L. Greenwald, P.R. Cronin, V. Carballo, G. Danaei, G. Choy, A novel model for predicting rehospitalization risk incorporating physical function, cognitive status, and psychosocial support using natural language processing, Med Care 55 (3) (2017) 261–266 PubMed PMID: 27632767, doi:10.1097/MLR.0000000000000651.
[87] D. Agrawal, CB. Chen, R.W. Dravenstott, C.T. Strömblad, J.A. Schmid, J.D. Darer, P. Devapriya, S. Kumara, Predicting patients at risk for 3-day postdischarge readmissions, ED visits, and deaths, Med Care 54 (11) (2016) 1017–1023 PubMed PMID: 27213544.
[88] J. Futoma, J. Morris, J. Lucas, A comparison of models for predicting early hospital readmissions, J. Biomed. Inform. 56 (2015) 229–238.
[89] S.F. Crone, S. Finlay, Instance sampling in credit scoring: An empirical study of sample size and balancing, Int. J. Forecast 28 (1) (2012) 224–238.
[90] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[91] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognit. 30 (7) (1997) 1145–1159.
[92] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.
[93] W. Ouwerkerk, A.A. Voors, A.H. Zwinderman, Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure, JACC: Heart Fail. 2 (5) (2014) 429–436.
[94] W.J. Lin, J.J. Chen, Class-imbalanced classifiers for high-dimensional data, Brief. Bioinform. 14 (1) (2012) 13–26.
[95] T. Oommen, L.G. Baise, R.M. Vogel, Sampling bias and class imbalance in maximum-likelihood logistic regression, Math. Geosci. 43 (1) (2011) 99–120.
[96] Y. Sun, A.K. Wong, M.S. Kamel, Classification of imbalanced data: a review, Int. J. Pattern Recognit Artif Intell. 23 (04) (2009) 687–719.
[97] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[98] M. Mesgarpour, T. Chaussalet, S. Chahed, Ensemble risk model of emergency admissions (ERMER), Int. J. Med. Inf. 103 (2017) 65–77.