

# A comparison of general and disease-specific machine learning models for the prediction of unplanned hospital readmissions

**Journal Article****Author(s):**

Sutter, Thomas; Roth, Jan A.; Chin-Cheong, Kieran; Hug, Balthasar; Vogt, Julia E.

**Publication date:**

2020-04

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000458934>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

**Originally published in:**

Journal of the American Medical Informatics Association 28(4), <https://doi.org/10.1093/jamia/ocaa299>

## Brief Communications

# A comparison of general and disease-specific machine learning models for the prediction of unplanned hospital readmissions

Thomas Sutter,<sup>1,2</sup> Jan A. Roth,<sup>3,4,5</sup> Kieran Chin-Cheong,<sup>1,2</sup> Balthasar L. Hug <sup>3,6,\*</sup> and Julia E. Vogt<sup>1,2,\*</sup>

<sup>1</sup>Medical Data Science, Department of Computer Science, ETH Zurich, Zurich, Switzerland, <sup>2</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland, <sup>3</sup>University of Basel, Basel, Switzerland, <sup>4</sup>Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, Basel, Switzerland, <sup>5</sup>Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Switzerland, Basel, and <sup>6</sup>Department of Internal Medicine, Kantonsspital Luzern, Lucerne, Switzerland

\*Authors contributed equally.

Corresponding Author: Balthasar L. Hug, MD, MBA, MPH, Department of Internal Medicine, Kantonsspital Luzern, Spitalstrasse, 6000 Luzern 16, Switzerland; balthasar.hug@luks.ch

Received 11 June 2020; Editorial Decision 29 October 2020; Revised 14 October 2020; Accepted 17 November 2020

## ABSTRACT

Unplanned hospital readmissions are a burden to patients and increase healthcare costs. A wide variety of machine learning (ML) models have been suggested to predict unplanned hospital readmissions. These ML models were often specifically trained on patient populations with certain diseases. However, it is unclear whether these specialized ML models—trained on patient subpopulations with certain diseases or defined by other clinical characteristics—are more accurate than a general ML model trained on an unrestricted hospital cohort. In this study based on an electronic health record cohort of consecutive inpatient cases of a single tertiary care center, we demonstrate that accurate prediction of hospital readmissions may be obtained by general, disease-independent, ML models. This general approach may substantially decrease the cost of development and deployment of respective ML models in daily clinical routine, as all predictions are obtained by the use of a single model.

**Key words:** Digital epidemiology, disease-specific model, hospital readmission, machine learning, prediction

## INTRODUCTION

Unplanned hospital readmissions are a burden to patients, increase healthcare costs, and serve as important quality indicators for hospitals.<sup>1</sup> A multitude of specialized machine learning (ML) approaches and models using different types of available data types have been utilized in recent years to optimize predictions of unplanned readmissions based on real-world data.<sup>2</sup> However, it is unclear whether specialized, disease-specific ML models are more accurate than general ML models in predicting unplanned readmissions when using the same set of features. ML models trained on separate disease

cohorts may account for different characteristics of the underlying cohort but complicates the model development process.

Certain disease entities are more prone to readmissions than others. Patients undergoing surgery have a long history of readmission research with readmission reduction programs.<sup>3</sup> Furthermore, patients with medical conditions such as heart failure and chronic obstructive lung disease or patients discharged from an oncology service area are especially prone to readmissions due to disease exacerbations.<sup>4–6</sup> Additionally, certain subgroups of patients in gynecology and obstetrics show a higher tendency toward readmissions.<sup>7</sup>

**Table 1.** Description of datasets

Dataset	Overall Inpatient Cases	Early Readmission Proportion (%)	Training Set (Inpatient Cases)	Validation Set (Inpatient Cases)	Balanced Test Set (Inpatient Cases)
All	180 318	2.4	4788	844	3152
Discharge discipline					
Medicine	78 854	2.5	2084	368	1518
Surgery/orthopedics	72 518	3.0	2510	444	1468
Gynecology/obstetrics	21 821	0.5	126	22	78
Disease category					
Cardiovascular diseases <sup>a</sup>	34 806	3.5	1356	240	868
Oncological diseases <sup>b</sup>	19 760	1.5	370	66	184
Chronic lung diseases <sup>c</sup>	6214	7.4	290	52	184

<sup>a</sup>Codes: I00-I99.

<sup>b</sup>Codes: C00-C97, D00-D09.

<sup>c</sup>Codes: J30-J39, J40-J47, J60-J70, J80-J84, J85-J86, J90-J94, J95-J99.

In this study, we investigate and compare the predictive performance of specialized and general ML models in predicting unplanned readmissions based on routine healthcare data in general and in patients discharged from medical, surgical, and gynecological departments.

## MATERIALS AND METHODS

### Study design and population

We conducted a cohort-based ML study at the University Hospital Basel—a tertiary care center in Switzerland with more than 37 000 inpatient cases per year—by using the routinely collected administrative, nongenetic health data from inpatient cases, as described previously.<sup>2</sup> In summary, administrative data such as sex, date of birth and home address are collected on admission of the patient, clinical data such as laboratory values were collected during hospitalization and diagnoses according to International Classification of Diseases—Tenth Revision—German Modification are determined by professional coders and collected shortly after discharge from hospital. From January 2012 to December 2017, all inpatients  $\geq 18$  years of age on hospital admission were eligible for study inclusion. The Ethics Committee of Northwestern and Central Switzerland approved this study project with a waiver of informed consent (project number 2016-02128).

### Outcomes and modeling

Unplanned readmission, our primary outcome, is routinely ascertained by medical coders as unplanned hospital readmissions according to Swiss health law within 18 days after discharge due to the same diagnosis as the previous hospitalization. In order to predict unplanned readmissions, we trained 3 ML classifiers: (1) a random forest (RF) classifier—a state-of-the-art ML method for structured data based on decision trees<sup>8</sup>; (2) a logistic regression with L1-regularization (LASSO) that is well suited for sparse and structured data<sup>9</sup>; and (3) a neural network (NN).<sup>10</sup> Deep NNs are powerful and flexible models with multiple layers of connected nodes. However, their capacity comes at the price of a need for a larger amount of data for model training. Furthermore, there is no simple method to infer the underlying mechanisms of neural network decisions.

Models were trained and validated to predict unplanned hospital readmissions at day of discharge with a prediction horizon of 18 days; therefore, the index (prediction) date is the date of discharge.

We take the last 2 years of the data as test set and the first 4 years as training set. From the latter, 85% of the early readmission samples are randomly chosen for training, with 15% for validation. The cases without unplanned readmission are randomly subsampled to match the number of cases with unplanned readmission in the respective set. The hyperparameters for all models were selected according to the predictive performance on the validation set. Similar to 10-fold cross-validation, this procedure is repeated 10 times for a more stable assessment of the models' performance.

Last, we evaluated the performance and generalizability of the final ML models on the held-out test set. For each model, we report the mean performance of all classifiers and their 95% confidence interval (CI). The final model architectures are as follows. First, for all experiments, the RF classifier consists of 500 trees with a maximum depth of 50. For decision trees and forests, a measure of importance for each variable used during training can be retrieved. It is based on the decrease of Gini impurity when a variable is chosen to split a node.<sup>11</sup> We used the variable importance for the analysis of what the classifiers have learned. Second, for the LASSO model, the L1 regularization strength, which controls the induced sparsity of the weight vector, is set to 2.0. The L1 penalty can be used for variable selection as follows: the components of the weight vector  $\beta$  are assigned high absolute values if the corresponding variable has a high impact on the prediction task. Owing to the L1 penalty, redundant variables are set to zero. Hence, the resulting weight vector  $\beta$  is sparse, and the nonzero components can be used for interpretability. Third, we used an NN with 4 layers. For all categorical variables, we used an additional input embedding layer, which helps to reduce the number of parameters of the NN. The number of hidden units per layer is 60, 40, 40, and 20. Every hidden layer is followed by a ReLU activation function<sup>12</sup> and a dropout layer with a ratio of 0.5.<sup>13</sup> We use an ADAM optimizer with an initial learning rate of 0.001.<sup>14</sup>

To compare the different ML models' predictive performance, we used the area under the receiver-operating characteristic curve (AUC) and the area under the precision-recall curve (AUPR). For disease-specific classifiers, we trained the ML models on subsets of inpatients with their primary diagnosis belonging to 1 of the following major disease categories according to the International Classification of Diseases—Tenth Revision codes (Table 1): (1) chronic lung diseases, (2) oncological diseases, and (3) cardiovascular diseases. Furthermore, we performed a subset selection on inpatients according to the healthcare discipline. We stratified healthcare disciplines into 4 categories based on the unit from which the inpatient was

discharged, that is, medicine, surgery and orthopedics, gynecology and obstetrics, and other disciplines (eg, ear, nose, and throat clinic; eye clinic). For discipline-specific predictions, we trained ML models on subsets of inpatients belonging to 1 of the 3 most frequented discharge disciplines (1) medicine, (2) surgery, or (3) gynecology.

**RESULTS**

During the study period and after excluding 8135 inpatients who declined the general research consent, 180 318 cases were included in the final analysis with an overall unplanned readmission proportion of 2.4% (n = 4392 of 180 318). We observed 6214 cases with a chronic lung disease as primary diagnosis (unplanned readmission proportion 7.4%). Furthermore, there were 19 760 and 34 806 cases with an oncological disease and a cardiovascular disease as primary diagnosis (unplanned readmission proportion, 1.5% and 3.5%), respectively.

Overall, 78 854 cases were discharged from medicine departments (unplanned readmission proportion, 2.5%). Moreover, 72 518 and 21 821 inpatient cases were discharged from surgical/orthopedic and gynecological/obstetric departments (unplanned readmission proportion, 3.0% and 0.5%), respectively.

The ML models were trained using 5561 clinical and administrative variables. The 3 different ML methods achieved comparable performance levels. The prediction performances for all subsets are shown in Table 2 (AUC; AUPR data are shown in the Supplementary Appendix). The RF and LASSO classifier slightly outperformed the NN algorithm (Figure 1A), with the general and disease-specific models having similar predictive performance. For all inpatient cases, the RF model reached a performance of 0.79 (AUC), the LASSO reached an AUC of 0.79 and the NN reached an AUC Of 0.77. For the subsets regarding the discharge discipline, we reached the best performance for cases in the gynecology department with an AUC of 0.90. For the disease categories, we reached the best performance on the subset of oncological diseases (AUC = 0.82). Calibration curves for the models are provided in the Supplementary Appendix.

The length of stay (LOS) category was the feature with the most predictive weight across all models—especially the value “inlier” as opposed to “low outliers” and “high outliers” with less than expected and longer than expected LOS, respectively (Figure 2). Apart from LOS, most variable values were homogeneously distributed between cases with and without an unplanned readmission.

For the LASSO and the RF classifiers, the importance of variables was also compared between the different subsets of the data. A detailed comparison of important variables for the 2 models may be found in the Supplementary Appendix. For the best-performing model, the RF classifier, similar variables are important for all but 1 dataset. For patients discharged from the gynecological/obstetrics department (Figure 3C), variables that describe the medical reason of the patient’s visit to the hospital were learned. For inpatient cases discharged from surgery (Figure 3B) and the medicine department (Figure 3A), demographic and organizational variables were learned. The same applies for the dataset with all inpatient cases (Figure 1A).

**DISCUSSION**

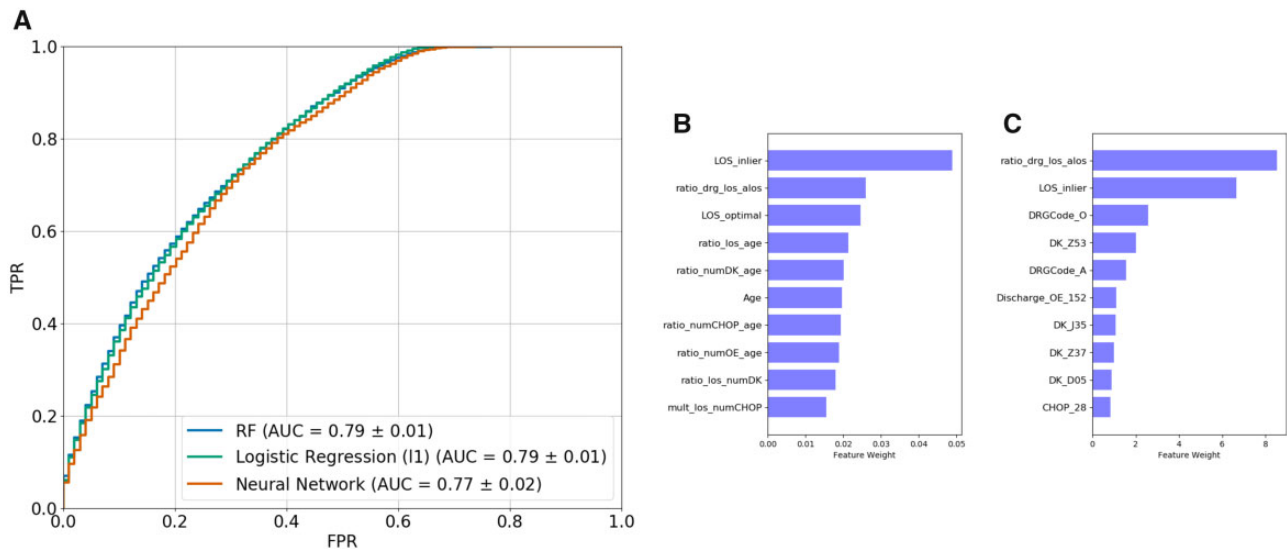
In the current study, we either outperformed existing work such as Brünger et al<sup>15</sup> or Allam et al<sup>16</sup> or reached state-of-the-art perfor-

**Table 2.** The prediction of unplanned hospital readmission within 18 days (test set)

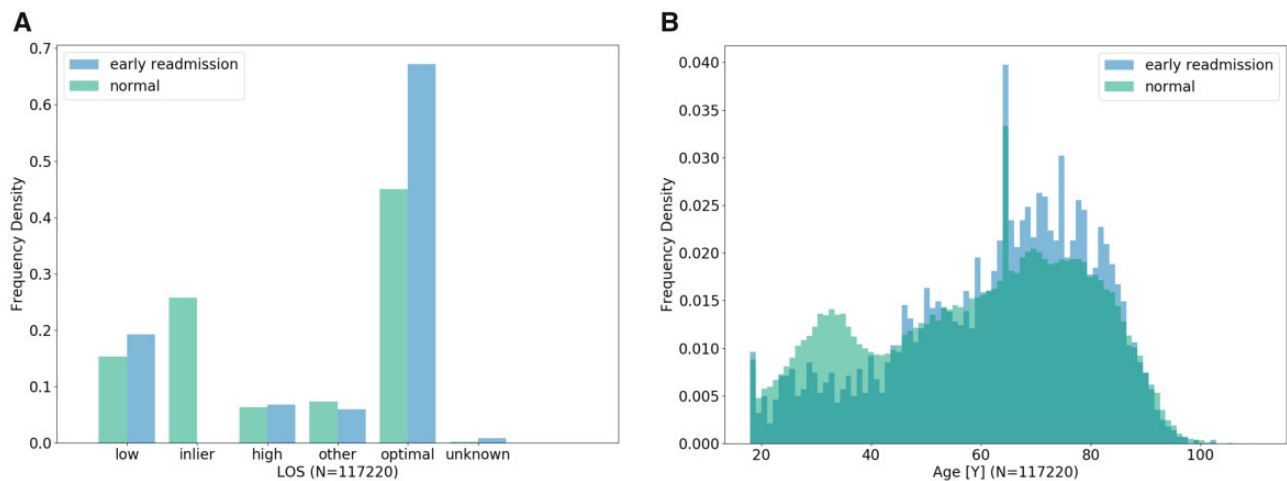
Classifier	Model	All	Discharge Discipline			Disease Category		
			Medicine	Surgery/Orthopedics	Gynecology/Obstetrics	Cardiovascular Diseases	Oncological Diseases	Chronic Lung Diseases
Balanced test set	Logistic regression	0.79 (0.77-0.81) <sup>a</sup>	0.74 (0.72-0.76)	0.78 (0.77-0.79)	0.93 (0.89-0.97) <sup>a</sup>	0.75 (0.72-0.78) <sup>a</sup>	0.82 (0.78-0.86) <sup>a</sup>	0.74 (0.71-0.77) <sup>a</sup>
	Specialized		0.74 (0.72-0.76)	0.78 (0.76-0.80)	0.89 (0.82-0.96)	0.73 (0.70-0.76)	0.82 (0.73-0.91) <sup>a</sup>	0.69 (0.62-0.76)
	Random forest	0.79 (0.78-0.80) <sup>a</sup>	0.75 (0.73-0.77) <sup>a</sup>	0.79 (0.77-0.81)	0.93 (0.87-0.99) <sup>a</sup>	0.73 (0.70-0.76)	0.81 (0.76-0.86)	0.70 (0.65-0.75)
Neural network	Specialized		0.75 (0.73-0.77) <sup>a</sup>	0.80 (0.79-0.81) <sup>a</sup>	0.90 (0.83-0.99)	0.73 (0.71-0.75)	0.80 (0.71-0.89)	0.68 (0.62-0.74)
	General	0.77 (0.75-0.79)	0.73 (0.71-0.75)	0.75 (0.72-0.78)	0.92 (0.87-0.97)	0.74 (0.71-0.77)	0.79 (0.72-0.86)	0.73 (0.69-0.77)
	Specialized		0.72 (0.71-0.75)	0.72 (0.70-0.74)	0.90 (0.85-0.95)	0.73 (0.70-0.76)	0.76 (0.70-0.82)	0.73 (0.67-0.79)
Unbalanced test set	Logistic regression	0.79 (0.78-0.80) <sup>a</sup>	0.75 (0.74-0.76) <sup>a</sup>	0.78 (0.77-0.79)	0.95 (0.94-0.96) <sup>a</sup>	0.75 (0.74-0.76) <sup>a</sup>	0.83 (0.81-0.86) <sup>a</sup>	0.73 (0.72-0.74) <sup>a</sup>
	Specialized		0.73 (0.72-0.74)	0.78 (0.77-0.79)	0.91 (0.90-0.92)	0.74 (0.73-0.75)	0.81 (0.77-0.85)	0.69 (0.65-0.73)
	Random forest	0.79 (0.78-0.80) <sup>a</sup>	0.75 (0.74-0.76) <sup>a</sup>	0.79 (0.78-0.80)	0.92 (0.91-0.93)	0.74 (0.73-0.75)	0.80 (0.78-0.82)	0.71 (0.69-0.73)
Neural network	Specialized		0.74 (0.73-0.75)	0.80 (0.79-0.81) <sup>a</sup>	0.89 (0.87-0.91)	0.73 (0.71-0.75)	0.81 (0.78-0.84)	0.69 (0.65-0.73)
	General	0.77 (0.76-0.78)	0.74 (0.73-0.75)	0.75 (0.74-0.76)	0.92 (0.91-0.93)	0.74 (0.73-0.75)	0.78 (0.77-0.79)	0.73 (0.72-0.74) <sup>a</sup>
	Specialized		0.73 (0.72-0.74)	0.73 (0.72-0.74)	0.89 (0.88-0.90)	0.73 (0.72-0.74)	0.75 (0.73-0.77)	0.72 (0.70-0.74)

Values are area under the receiver-operating characteristic curve (95% confidence interval). General models were trained on all inpatients and evaluated on the disease-specific test set, whereas Specialized models were trained on disease-specific subsets of inpatients cases only.

<sup>a</sup>Best-performing model per stratum.



**Figure 1.** (A) Performance of different machine learning (ML) models (test set;  $n = 3152$  inpatient cases) and variable importance for random forest (RF) and LASSO. (B) The learned variable importance for the RF model trained on all inpatient cases. (C) The learned variable importance for the LASSO model trained on all inpatient cases. A list with explanations of variable names can be found in the [Supplementary Appendix](#). AUC: area under the receiver-operating characteristic curve; FPR: false positive rate (specificity); TPR: true positive rate (sensitivity).

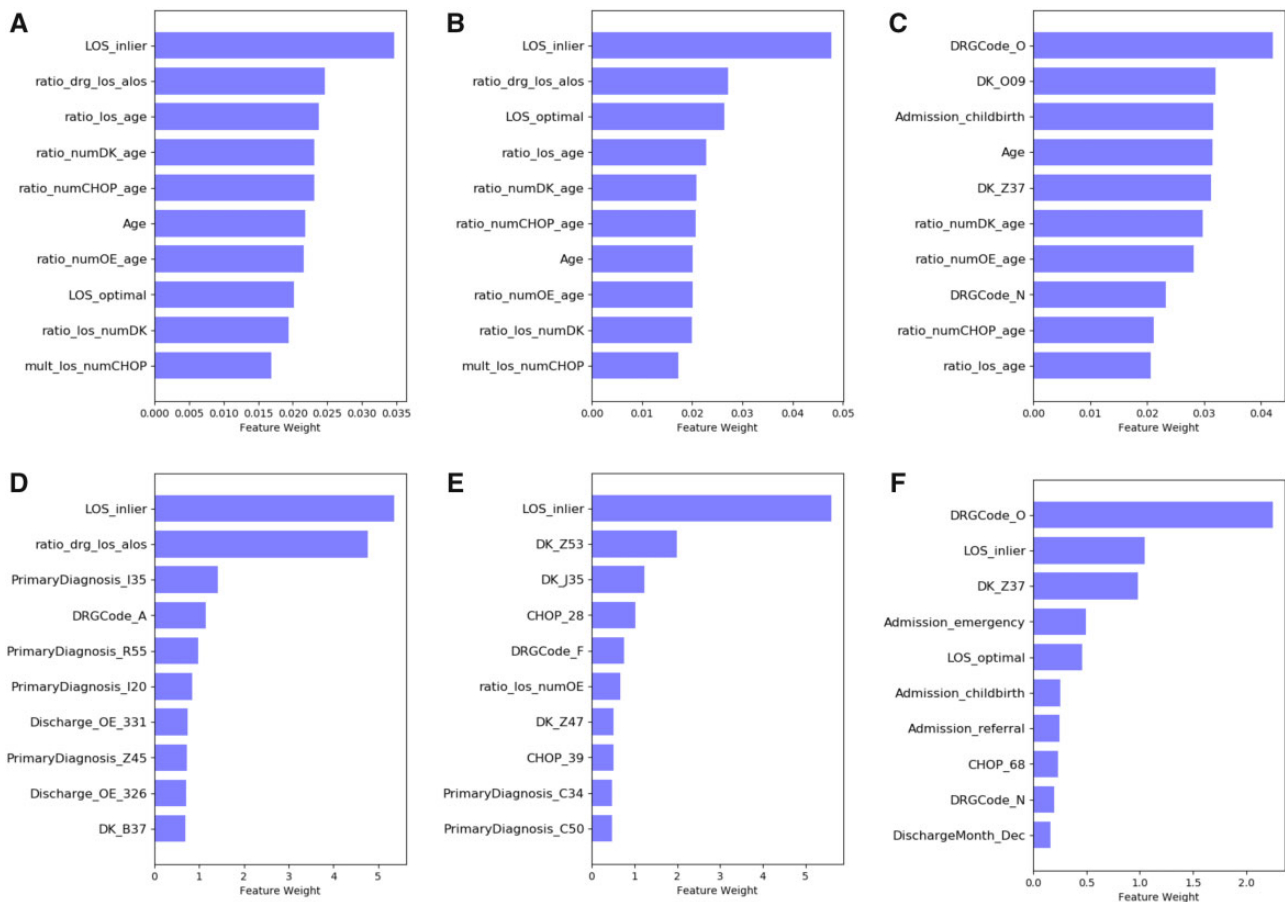


**Figure 2.** Distribution of (A) hospital length of stay (LOS) and (B) age for inpatient cases with and without an unplanned hospital readmission (training set;  $n = 117\,220$  inpatient cases). The pink color shows the percentage of normal patients and the blue color shows the percentage of early readmission patients. (A) The category value “inlier” defines the range of normal length of stays for a given diagnosis related group. “Low” and “high” category values correspond to values out of this range. The value “optimal” is set for length of stays, which are close enough to the lower end of the “inlier” category. No early readmissions have been recorded for the category value “inlier.”

mance in predicting unplanned readmissions for the entire study population as well as for disease-specific subsets.<sup>17</sup> Our work is best comparable to the study by Brünger et al,<sup>15</sup> which was performed in a similar setting with a 30-day prediction horizon. Their model reached an AUC performance of 0.60 for the entire study population of 138 222 inpatient cases. For surgical inpatients, they reached an AUC of 0.62, and for medical inpatients, a performance of 0.58 (AUC). The proposed model in Allam et al<sup>16</sup> reached a predictive performance of 0.64 in terms of AUC for the entire study population of 272 778 inpatient cases. Our RF and LASSO models reach a predictive performance of 0.79 in terms of AUC for all inpatients. For surgical inpatients, our model achieves a performance of 0.80 in terms of AUC and for medical inpatients reached an AUC of 0.75.

Interestingly, the training of specialized, disease-specific ML models did not result in a better predictive performance compared with general models. Our findings challenge current ML prediction projects, which are often restricted to patient populations with specific disease categories—leading to a highly fragmented and complex prediction landscape.<sup>15–17</sup> While acknowledging, for certain outcomes, that ML models trained on separate patient cohorts (eg, defined by different disease characteristics) may improve the resulting predictions through different predictive weighting and inclusion of interactions, our findings indicate that such a disease-specific approach may not always be necessary—especially when the same data source is used.

Our study also has limitations. First, our results may not be generalizable to other healthcare settings with different patient



**Figure 3.** Variable importance of random forest (RF) and LASSO models trained on different discharge discipline subsets based on their organizational exit entity of inpatients and a general model. (A–C) The variable importance for RF models. (D–F) The variable importance for LASSO models. (A, D) The importance for models trained on the subset of inpatients with medicine as their organizational exit entity; (B, E) the variable importance for the surgery/orthopedics subset; and (C, F) the variable importance for the gynecology subset. A list with explanations of variable names can be found in the [Supplementary Appendix](#).

populations and data structures. Second, we focused our analyses on unplanned hospital readmissions: there may be a difference in the predictive performance between general and disease-specific ML models for other health-related outcomes. Third, the comparable predictive performances of our models in terms of AUC may be explained by the difficulty of the prediction task, as a variety of unmeasured factors may affect the risk for unplanned readmissions (eg, socioeconomic status). Last, the limited sample size of the study dataset might have led to similar predictive performances of our ML models. Although there is a large number of inpatient cases, the readmission proportion as outcome parameter was low, which leads to smaller training sets for evaluating the models. In particular, neural networks tend to perform best with access to larger amounts of data. Sample size is further reduced for specialized models, as sample size for subgroups of the data naturally decreases. In future work, we would like to address the open question of the trade-off between patient cohort size and specialized models.

## CONCLUSION

In conclusion, different ML algorithms reached or outperformed state-of-the-art performance in predicting unplanned hospital readmissions. Interestingly, disease-specific and general ML models showed similar performances in predicting unplanned readmissions.

Hence, our study findings suggest that accurate ML prediction of hospital readmissions may be obtained by general, disease-independent models. This general approach may substantially decrease the cost of development and introduction of respective ML models in daily clinical routine, as only one prediction model has to be deployed.

## AUTHOR CONTRIBUTIONS

TS, JAR, BLH, and JEV designed the study and wrote the first draft. TS is responsible for the data analytics together with KC-C. All authors revised and finalized the manuscript.

## ACKNOWLEDGMENTS

We thank Brian Duggan from the University Hospital Basel for his help with the manuscript.

## DATA AVAILABILITY STATEMENT

The data underlying this article cannot be shared publicly due to the privacy of individuals that participated in the study. The data will be shared on reasonable request to the corresponding author.



## CONFLICT OF INTEREST STATEMENT

None declared.

## REFERENCES

- McIlvennan CK, Eapen ZJ, Allen LA. Hospital readmissions reduction program. *Circulation* 2015; 131 (20): 1796–803.
- Roth JA, Goebel N, Sakoparnig T, *et al.*; the PATREC Study Group. Secondary use of routine data in hospitals: description of a scalable analytical platform based on a business intelligence system. *JAMIA Open* 2018; 1 (2): 172–7.
- Ibrahim AM, Nathan H, Thumma J, *et al.* Impact of the Hospital Readmission Reduction Program on surgical readmissions among Medicare beneficiaries. *Ann Surg* 2017; 266: 617–24.
- Bahadori K, FitzGerald JM. Risk factors of hospitalization and readmission of patients with COPD exacerbation—systematic review. *Int J Chron Obstruct Pulmon Dis* 2007; 2 (3): 241–51.
- Gupta A, Allen LA, Bhatt DL, *et al.* Association of the hospital readmissions reduction program implementation with readmission and mortality outcomes in heart failure. *JAMA Cardiol* 2018; 3 (1): 44–53.
- Donzé J, Aujesky D, Williams D, *et al.* Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA Intern Med* 2013; 173 (8): 632–8.
- Wilbur MB, Mannschreck DB, Tanner E, *et al.* 18: Unplanned thirty-day readmission rates as a quality measure: risk factors and costs of readmission on a gynecologic oncology service. *Am J Obstet Gynecol* 2016; 214 (4): S465.
- Breiman L. Random forests. *Mach Learn* 2001; 45 (1): 5–32.
- Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 1996; 58 (1): 267–88.
- Haykin S. *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall PTR; 1994.
- Louppe G, Wehenkel L, Sutter A, *et al.* Understanding variable importance in forests of randomized trees. *Adv Neural Inf Process Syst* 2013; 2013: 431–9.
- Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*; 2010: 807–14.
- Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014; 15: 1929–58.
- Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv*: 1412.6980; 2014.
- Brünger B, Blozik E. Hospital readmission risk prediction based on claims data available at admission: a pilot study in Switzerland. *BMJ Open* 2019; 9 (6): e028409.
- Allam A, Nagy M, Thoma G, *et al.* Neural networks versus Logistic regression for 30 days all-cause readmission prediction. *Sci Rep* 2019; 9 (1): 9277.
- Futoma J, Morris J, Lucas J. A comparison of models for predicting early hospital readmissions. *J Biomed Inform* 2015; 56: 229–38.