Canadian Journal of Cardiology ■ (2019) 1–8

Clinical Research

Evaluation of Machine Learning Algorithms for Predicting Readmission After Acute Myocardial Infarction Using Routinely Collected Clinical Data

Shagun Gupta, HBSc, MASc,[a] Dennis T. Ko, MD, MSc,[b,c,d] Paymon Azizi, MSc,[c,d] Mohamed Reda Bouadjenek, PhD,[a] Maria Koh, MSc,[c] Alice Chong, MSc,[c] Peter C. Austin, PhD,[c,d] and Scott Sanner, PhD[a]

[a] *Department of Mechanical and Industrial Engineering, University of Toronto, Ontario, Canada*

[b] *Schulich Heart Centre, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada*

[c] *Institute for Clinical Evaluative Service (ICES), Toronto, Ontario, Canada*

[d] *Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada*

ABSTRACT

Background: The ability to predict readmission accurately after hospitalization for acute myocardial infarction (AMI) is limited in current statistical models. Machine-learning (ML) methods have shown improved predictive ability in various clinical contexts, but their utility in predicting readmission after hospitalization for AMI is unknown.

Methods: Using detailed clinical information collected from patients hospitalized with AMI, we evaluated 6 ML algorithms (logistic regression, naïve Bayes, support vector machines, random forest, gradient boosting, and deep neural networks) to predict readmission within 30 days and 1 year of discharge. A nested cross-validation approach was used to develop and test models. We used C-statistics to compare discriminatory capacity, whereas the Brier score was used to indicate overall model performance. Model calibration was assessed using calibration plots.

RÉSUMÉ

Contexte : Les modèles statistiques actuels ne permettent pas de prédire avec exactitude la réadmission après une hospitalisation pour cause d'infarctus aigu du myocarde (IAM). Les méthodes de prédiction faisant appel à l'apprentissage automatique ont été associées à une amélioration de la capacité de prédiction dans divers contextes cliniques, mais leur utilité pour prédire la réadmission après une hospitalisation pour cause d'IAM demeure inconnue.

Méthodologie : À l'aide de données cliniques détaillées recueillies auprès de patients hospitalisés pour un IAM, nous avons évalué six algorithmes d'apprentissage automatique (régression logistique, classification naïve bayésienne, machine à vecteurs de support, forêt aléatoire, boosting par descente de gradient fonctionnelle et réseaux neuronaux d'apprentissage profond) pour prédire la réadmission dans les 30 jours et dans l'année suivant la sortie de l'hôpital. Les modèles

Acute myocardial infarction (AMI) has a high readmission burden, as approximately 1 in 6 patients are readmitted within 30 days of discharge.[1] With the introduction of the Hospital Readmission Reduction Program in 2012 in the United States, hospitals have been financially incentivized to reduce readmission rates.[2] Although this has led to the development of various intervention strategies aimed at reducing readmissions, they tend to be costly and resource intensive.[3] In this context, predictive models would allow hospitals to identify patients at high risk for readmission and target the delivery of these interventions towards them, thereby reducing unplanned readmissions.[4,5]

Several models exist for predicting patients' risk of readmission. The LACE index was developed in 2010 to estimate risk of all-cause readmission in medical and surgical patients based on 4 criteria: length of stay (L), acuity of admission (A), comorbidity of the patient (C), and emergency department use in the 6 months before admission (E). Although this criteria-based system is easy to apply, it is not very accurate (C-statistic range, 0.51-0.72).[6-8] AMI-specific logistic and Cox models have also been developed to model and predict readmission in an AMI population more accurately.[9-11] However, despite the more focused approach, these models have only improved discrimination capabilities moderately (median C-statistic 0.65; range 0.53-0.79). Moreover, none of these models exclusively use clinical data, thus limiting their applicability for clinical use.[12]

In many clinical disease contexts, machine-learning (ML) models have shown improved discriminatory power at predicting outcomes compared with traditional approaches.[13,14] ML methods can automatically identify patterns in data that

Received for publication June 14, 2019. Accepted October 21, 2019.

Corresponding author: Dr Dennis T. Ko, ICES G1-06, 2075 Bayview Ave, Toronto, Ontario M4N3M5, Canada. Tel.: +1-416-480-4055, ext 7835; fax: +1-416-480-6048.

E-mail: dennis.ko@ices.on.ca

See page 7 for disclosure information.

https://doi.org/10.1016/j.cjca.2019.10.023

Results: The 30-day readmission rate was 16.3%, whereas the 1-year readmission rate was 45.1%. For 30-day readmission, the discriminative ability for the ML models was modest (C-statistic 0.641; 95% confidence interval (CI), 0.621-0.662 for gradient boosting) and did not outperform previously reported methods. For 1-year readmission, different ML models showed moderate performance, with C-statistics around 0.72. Despite modest discriminatory capabilities, the observed readmission rates were markedly higher in the tenth decile of predicted risk compared with the first decile of predicted risk for both 30-day and 1-year readmission.

Conclusions: Despite including detailed clinical information and evaluating various ML methods, these models did not have better discriminatory ability to predict readmission outcomes compared with previously reported methods.

ont été mis au point et testés à l'aide d'une approche de validation croisée imbriquée. Nous avons utilisé la statistique C pour comparer la capacité de discrimination des différents modèles, et le score de Brier pour en chiffrer le rendement global. Le calage des modèles a été évalué au moyen de courbes d'étalonnage.

Résultats : Le taux de réadmission à 30 jours était de 16,3 %, tandis que le taux de réadmission à 1 an était de 45,1 %. Dans le cas de la réadmission à 30 jours, la capacité de discrimination des modèles d'apprentissage automatique était modeste (statistique C : 0,641; intervalle de confiance [IC] à 95 % : 0,621-0,662 pour le boosting par descente de gradient fonctionnelle) et n'était pas supérieure à celle des méthodes déjà utilisées. Dans le cas de la réadmission à 1 an, différents modèles d'apprentissage automatique se sont révélés modérément efficaces, la statistique C se chiffrant à environ 0,72. En dépit des modestes capacités de discrimination des différentes méthodes, les taux de réadmission observés étaient nettement plus élevés dans le dixième décile du risque prédit comparativement à ceux du premier décile, pour la réadmission à 30 jours comme pour la réadmission à 1 an.

Conclusions : Malgré le recours à des données cliniques détaillées et à différentes méthodes d'apprentissage automatique, les modèles évalués n'ont pas montré une capacité de discrimination supérieure à celle des méthodes déjà utilisées pour prédire la réadmission.

are predictive of relevant health outcomes and include various algorithms that can model complex nonlinear interactions among variables. With respect to AMI readmissions, it has not yet been established whether these methods can produce more accurate predictions compared with the existing methods described here. The goal of this study was to conduct a comprehensive analysis of ML methods for the task of predicting readmission in patients with AMI at the time of discharge, using clinical data collected during the course of care. To this end, we compared the performance of commonly used supervised ML algorithms on 2 different readmission tasks: readmission within 30 days and readmission within 1 year.

Methods

Data sources

The study sample was derived from the **E**nhanced **F**eedback for E**ff**ective **C**ardiac **T**reatment (EFFECT) study,[15] a cluster randomized trial to evaluate the effectiveness of public reporting. Detailed clinical data were collected from 81 hospital corporations in Ontario via retrospective chart review performed by trained nurse abstractors. Additional databases that were used for this study include the Ontario Registered Persons Database, a registry of all Ontario residents with health insurance coverage, and the Canadian Institute for Health Information Discharge Abstract Database (DAD), which contains information on all admissions to acute care hospitals in Ontario. This database was used to identify any subsequent hospital readmissions. These datasets were linked using unique encoded identifiers and analyzed at Institute for Clinical Evaluative Services (ICES). The use of data in this project was authorized under section 45 of Ontario's Personal Health Information Protection Act, which does not require review by a Research Ethics Board.

Study sample

Patients between the ages of 20 to 105 years with a most responsible diagnosis of AMI (identified by International Classification Disease-9 code 410 in DAD) and a valid Ontario Health Insurance Plan number were included in the EFFECT phase 2 study. However, patients were excluded if they died before the end of their respective analysis periods of 30 days (n = 91, 1.28%) or 1 year (n = 211, 2.96%) without a readmission.

Predictors

A total of 204 routinely available clinical variables were initially selected from the EFFECT study using *a priori* clinical knowledge, level of completeness, and applicability to the entire cohort (ie, nonconditional questions). These included demographic variables (age, sex, race/ethnicity, and employment status), past medical history, vital signs, characterization of AMI (including admission symptoms, complications and severity scores), cardiac procedures, select laboratory tests and medications, medical imaging procedures, and patient counselling. To account for predictors with missing data, variables with more than 30% missing values were dropped from each readmission cohort. This resulted in a total of 192 variables. The final list of variables used for input is provided in Supplemental Table S1. The continuous variables were normalized by z-scoring across all patients so that each continuous variable had zero mean and unit variance. Finally, dummy variables were created for all categorical variables.

Outcome

We developed models to predict all-cause hospital readmission within 2 timeframes: readmission within 30 days of discharge from the index hospitalization and readmission within 1 year of discharge from the index hospitalization.

Analysis

We developed prediction models using 6 commonly used ML algorithms: (1) logistic regression (LR); (2) naïve Bayes (NB); (3) support vector machines (SVMs); (4) random forests (RF); (5) gradient boosting (GB); and (6) deep neural networks (DNN). LR is a simple linear model that provides a good baseline comparison for more complex models. We considered both L1 (lasso) and L2 (ridge regression) regularization for LR. NB is a probabilistic classifier that makes a "naïve" assumption that the input variables are independent, given the outcome. Although this assumption does not usually hold, the resulting model often performs well owing to its simplicity and relative immunity to overfitting.[16] SVM attempts to find an optimal hyperplane that maximizes the margin (distance) between the data points of the 2 classes. With instances in which data are not linearly separable, the assumption of linearity is relaxed by allowing outlier data points to fall on the wrong side of the decision margin.[17] In contrast to the aforementioned methods, RF, GB, and DNN are nonlinear algorithms; they do not make previous assumptions about the functional form of the input and are therefore able to model nonlinear interactions in the data. RF is an ensemble method that fits a number of decision trees on subsamples of the dataset and then uses model averaging to improve accuracy and control for overfitting.[18] GB is another ensemble learning method; however, unlike RF, the decision trees are grown sequentially using gradient descent, with each subsequent tree aiming to reduce the errors of the previous tree.[19] A DNN consists of a series of fully connected hidden layers, which transform an input vector into a probability distribution estimating the output.[20] All analyses were done using Python v2.7. LR, NB, SVM and RF were implemented using scikit-learn v0.18.1.[21] GB was implemented using and XGBoost v0.6.[22] The DNN was implemented using Keras.[23] A detailed discussion of the ML algorithms is provided in Supplemental Text S1.

Each model was assessed using a nested k-fold cross-validation approach consisting of 2 loops: an outer loop for model evaluation and an inner loop for tuning hyperparameters. We used 10 nonoverlapping training and test sets in the outer loop. Performance estimates were calculated by averaging test set scores across the 10 dataset splits of the outer loop. Each training fold of the outer loop was further split into 5 nonoverlapping training and validation sets (the inner loop), which were used to tune hyperparameters for each model using grid searches. A grid search builds a model for every combination of the specified hyperparameters, using the training set, and then evaluates each model on the validation set to identify the optimal combination of hyperparameter values. Further details including a list of hyperparameters are provided in the Supplemental Text S2 and Supplemental Table S2).

To address the class imbalance in the cohort for 30-day readmission, we experimented with 3 strategies to create a balanced dataset of readmitted and not-readmitted patients during model training: upsampling, creating duplicates of the minority class; downsampling, randomly subsampling subjects of the majority class; and weighting algorithms by providing a weight for each class. The weight selected for each class (readmitted and not readmitted) was inversely proportional to class frequencies in the training dataset. In each of these cases, we applied the modification only to the training dataset. We also considered an approach in which no weighting, and no up- or downsampling was applied. For each algorithm, we report the results for the technique that gave the best prediction.

We chose the C-statistic (equivalent to the area under the receiver operating characteristic [ROC] curve) as our primary performance measure for all models. The C-statistic is the probability that a randomly selected readmitted patient will have a higher predicted probability of readmission than a randomly selected patient who was not readmitted.[24] We report the Brier score as a measure of accuracy of probabilistic predictions. It can take on values between 0 and 1, with a score of 0 indicating perfect prediction.[24] We assessed model confidence using the logarithmic-loss (log-loss) function. The log-loss function calculates a score by penalizing the difference between the predicted probabilities and expected values; as the penalty is logarithmic in nature, log-loss heavily penalizes models that are confident about an incorrect prediction. A perfectly accurate model has a log-loss score of 0, whereas less accurate models have increasingly larger scores. Model calibration was assessed by stratifying the test sample into 10 risk strata using the deciles of the predicted probability of the outcome. We then compared the observed probability of the outcome within each decile with the mean predicted probability of the outcome within each decile.

Results

Baseline characteristics of patients in 30-day and 1-year readmission cohorts are detailed in Table 1. Overall, there were 7018 patients in the 30-day readmission cohort, with 1146 readmissions, resulting in a readmission rate of 16.3%. For 1-year readmission, there were a total of 6898 patients in the cohort, of whom 3113 were readmitted, resulting in a readmission rate of 45.1%. On average, patients who were readmitted were older than those not readmitted. Comorbid conditions and AMI-specific hospital complications were also higher in readmitted patients. Finally, patients who were readmitted were less likely to have undergone cardiac procedures during the hospital stay than those were not readmitted.

Thirty-day readmission

Overall, discrimination for the various models was modest (Table 2). The discriminative ability of different models, as represented by ROC curves, is shown in Figure 1A. The GB model had the highest discrimination (C-statistic = 0.641; 95% confidence interval [CI], 0.621-0.662), although predictive accuracy was average (Brier score, 0.151; 95% CI, 0.144-0.158, log-loss 0.484; 95% CI, 0.468-0.500). The RF and DNN models displayed similar discrimination (RF: C-statistic, 0.639; 95% CI, 0.617-0.662 and DNN: C-statistic = 0.637; 95% CI, 0.617-657), with improved predictive accuracy and confidence compared with the GB model (Brier score ~0.132, log-loss ~0.430 for both RF

Table 1. Characteristics of patients in 30-day and 1-year readmission cohorts

Characteristic	30-day readmission		1-year readmission	
	Readmitted n (%)	Not readmitted n (%)	Readmitted n (%)	Not Readmitted n (%)
N	1146	5872	3113	3785
Median age, years (SD)	69.8 (13.7)	66.7 (14.0)	70.1 (13.5)	64.4 (13.8)
Women	435 (38.0)	2040 (34.7)	1198 (38.5)	1220 (32.2)
Past medical history (cardiovascular comorbidities)				
Stroke/transient ischemic attack	181 (15.8)	621 (10.6)	478 (15.4)	284 (7.5)
Diabetes	395 (34.5)	1500 (25.5)	1046 (33.6)	804 (21.2)
Congestive heart failure	101 (8.8)	243 (4.1)	254 (8.2)	70 (1.8)
Hypertension	748 (65.3)	3284 (55.9)	1971 (63.3)	1981 (52.3)
Hyperlipidemia	529 (46.2)	2665 (44.4)	1455 (46.7)	1692 (44.7)
Myocardial infarction	335 (29.2)	1286 (21.9)	892 (28.7)	678 (17.9)
Atrial fibrillation	131 (11.4)	409 (7.0)	346 (11.1)	170 (4.5)
Peripheral arterial disease	122 (10.6)	432 (7.4)	334 (10.7)	202 (5.3)
Angina	397 (34.6)	1612 (27.5)	1042 (33.5)	914 (24.1)
Past medical history (medical comorbidities)				
Renal disease	18 (1.6)	45 (0.8)	51 (1.6)	12 (0.3)
Cancer	25 (2.2)	88 (1.5)	69 (2.2)	40 (1.1)
Chronic obstructive pulmonary disease	152 (13.3)	528 (9.0)	427 (13.7)	229 (6.1)
Chronic liver disease	9 (0.8)	26 (0.4)	23 (0.7)	12 (0.3)
Peptic ulcer disease	67 (5.8)	253 (4.3)	169 (5.4)	143 (3.8)
Past cardiac procedures				
Coronary artery bypass grafting	94 (8.2)	464 (7.9)	309 (9.9)	242 (6.4)
Percutaneous coronary intervention	85 (7.4)	355 (6.0)	225 (7.2)	210 (5.5)
Types of myocardial infarction				
STEMI	370 (32.3)	2111 (36.0)	941 (30.2)	1526 (40.3)
Non-STEMI	745 (65.0)	3640 (62.0)	2090 (67.1)	2194 (58.0)
In-hospital treatments and procedures				
Cardiac catheterization	430 (37.5)	2884 (49.1)	1148 (36.9)	2152 (56.9)
Percutaneous coronary intervention in hospital	176 (15.4)	1403 (23.9)	474 (15.2)	1102 (29.1)
Coronary artery bypass surgery in hospital	43 (3.8)	263 (4.5)	103 (3.3)	201 (5.3)
AMI-specific hospital complications (>24 hours after arrival)				
Heart failure in hospital	122 (10.6)	353 (6.0)	276 (8.9)	184 (4.9)
Shock	16 (1.4)	46 (0.8)	30 (1.0)	31 (0.8)
Recurrent angina/ischemia	222 (19.4)	883 (15.0)	518 (16.6)	575 (15.2)
Hemorrhage requiring intervention	28 (2.4)	123 (2.1)	69 (2.2)	82 (2.2)
Lab values, mean (SD)				
Blood urea nitrogen	4.5 (1.2)	7.4 (4.5)	8.7 (5.9)	6.7 (3.5)
Cholesterol	4.5 (1.2)	4.6 (1.2)	4.5 (1.2)	4.7 (1.1)
Hemoglobin (first)	133.2 (21.7)	138.9 (19.5)	133.8 (21.3)	141.8 (18.0)
Hemoglobin (last)	123.1 (18.6)	128.9 (17.9)	124.5 (18.3)	131.1 (17.5)

SD, standard deviation; STEMI, ST-elevated myocardial infarction.

and DNN). SVM (with class weighting) and LR performed similarly, both with slightly worse discrimination and predictive accuracy. Finally, although NB also showed similar discrimination (C-statistic 0.627; 95% CI, 0.602-0.652), it had the lowest overall prediction score (Brier score, 0.189; 95% CI, 0.176-0.201, log-loss 1.141; 95% CI, 1.018-1.263).

The LR, SVM, DNN, and RF models demonstrated good concordance between the observed and predicted probabilities of readmission (Fig. 2A), although the range of predicted probabilities was limited between 0.042 and 0.489. The calibration curves show poor alignment for the GB model, despite having a C-statistic value similar to RF. The NB model had the worst model calibration, despite a large predictive range of predicted probabilities (0.00 to 1.00), with no change in observed readmission rate, regardless of predicted probability. Finally, despite the small range of predicted probabilities for the LR, DNN, and RF models, the mean observed readmission rates were markedly higher in the highest predicted risk group (30.1%, 31.6%, and 30.8%, respectively) than the lowest predicted risk group (6.8%, 7.1%, and 6.4%) (Fig. 2A, with detailed values in Supplemental Table S3).

One-year readmission

The LR, SVM, DNN, RF, and GB models resulted in similar scores for both measures of predictive accuracy (C-statistic ~0.72, Brier Score ~0.21, and log-loss ~0.61), with no model showing superior performance compared with the other (Table 2 and Fig. 1B). As in the 30-day readmission sample, NB had worse discrimination (C-statistic 0.692; 95% CI, 0.679-0.705) and worse predictive accuracy and confidence (Brier score, 0.321; 95% CI, 0.306-336, log-loss 2.485; 95% CI, 2.317-2.653).

The LR, SVM, RF, and DNN models all resulted in predictions with good concordance between observed and predicted probabilities for all ranges of predicted probability (Fig. 2B). The range of predicted probabilities were also much larger compared with the models for 30-day readmission. The GB model also showed good calibration, unlike for 30-day readmission in which calibration was poor. NB had the

Table 2. Measures of predictive accuracy for various machine learning models

Outcome and model	C-statistic (95% CI)	Brier score (95% CI)	Log loss (95% CI)
Readmission within 30 days			
Logistic regression	0.631 (0.611-0.651)	0.132 (0.124-0.141)	0.430 (0.410-0.451)
Naïve Bayes	0.627 (0.602-0.652)	0.189 (0.176-0.201)	1.141 (1.018-1.263)
Support vector machine	0.627 (0.608-0.645)	0.133 (0.128-0.137)	0.431 (0.420-0.443)
Deep neural network	0.637 (0.617-0.657)	0.132 (0.124-0.141)	0.430 (0.410-0.450)
Random forest	0.639 (0.616-0.661)	0.132 (0.124-0.141)	0.429 (0.408-0.450)
Gradient boosting	0.641 (0.621-0.662)	0.151 (0.144-0.158)	0.484 (0.468-0.500)
Readmission within 1 year			
Logistic regression	0.719 (0.706-0.732)	0.212 (0.207-0.217)	0.614 (0.604-0.624)
Naïve Bayes	0.692 (0.679-0.705)	0.321 (0.306-0.336)	2.485 (2.317-2.653)
Support vector machine	0.718 (0.706-0.731)	0.212 (0.208-0.216)	0.613 (0.605-0.621)
Deep neural network	0.716 (0.702-0.731)	0.213 (0.208-0.218)	0.615 (0.604-0.626)
Random forest	0.716 (0.704-0.729)	0.213 (0.209-0.217)	0.616 (0.607-0.625)
Gradient boosting	0.720 (0.708-0.732)	0.212 (0.207-0.216)	0.612 (0.603-0.621)

worst model calibration, with low concordance between observed and predicted probabilities for all ranges, similar to the 30-day readmission sample.

The LR, SVM, DNN, RF, and GB models were able to risk stratify patients adequately, with observed mean readmission rates ranging between ~17% and 18% among the lowest-risk persons to ~80.0% among the highest-risk persons. The range was smaller for NB, with an observed mean readmission rate of 21.5% in the first risk decile and 76.7% in the highest-risk decile (Fig. 2B, with detailed values in Supplemental Table S4).

Discussion

In this study, we examined the ability of various ML algorithms to predict readmission, using detailed clinical data in patients who were hospitalized with AMI. We found that the use of ML algorithms did not lead to substantial improvements in prediction for either 30-day or 1-year readmission when compared with previously reported statistical methods. All models developed for 30-day readmission had similar performance and showed modest discrimination, with C-statistics in the range of 0.63 to 0.64. These values are consistent with the median C-statistic (0.65) reported by a recent review of risk-prediction models for 30-day AMI readmission,[12] although it should be noted that this number is likely inflated, as not all models were validated. The models for 1-year readmission showed moderate performance, with C-statistics consistently around 0.72 (with the exception of the NB model), whic was also consistent with results from the literature.[9]

Given the considerable burden placed on the health care system by readmissions after AMI,[1,25] predicting readmission risk would enable hospitals to target readmission reduction

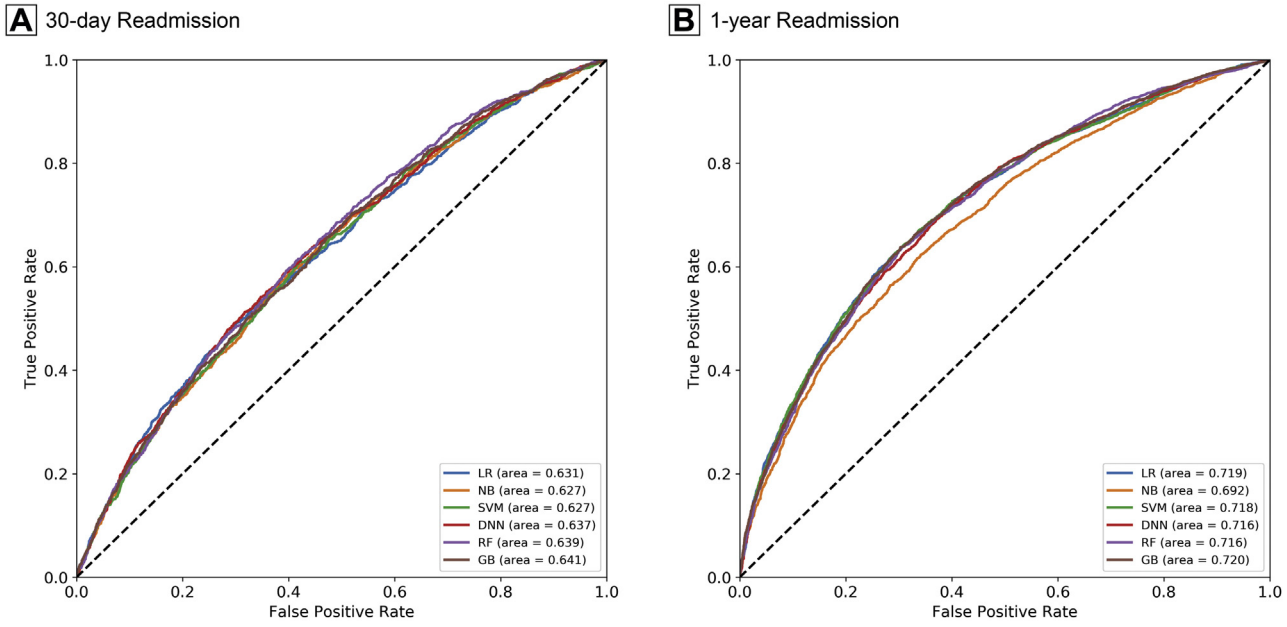

Figure 1. Receiver operating characteristic curves for (**A**) 30-day readmission and (**B**) 1-year readmission. The corresponding values of the area under the curve for each model (ie, C-statistics) are presented in Table 2. LR (**blue line**) indicates logistic regression; NB (**orange line**), na ve Bayes; SVM (**green line**), support vector machine; DNN (**red line**), deep neural network; RF (**purple line**), random forest; and GB (**brown line**), gradient boosting.

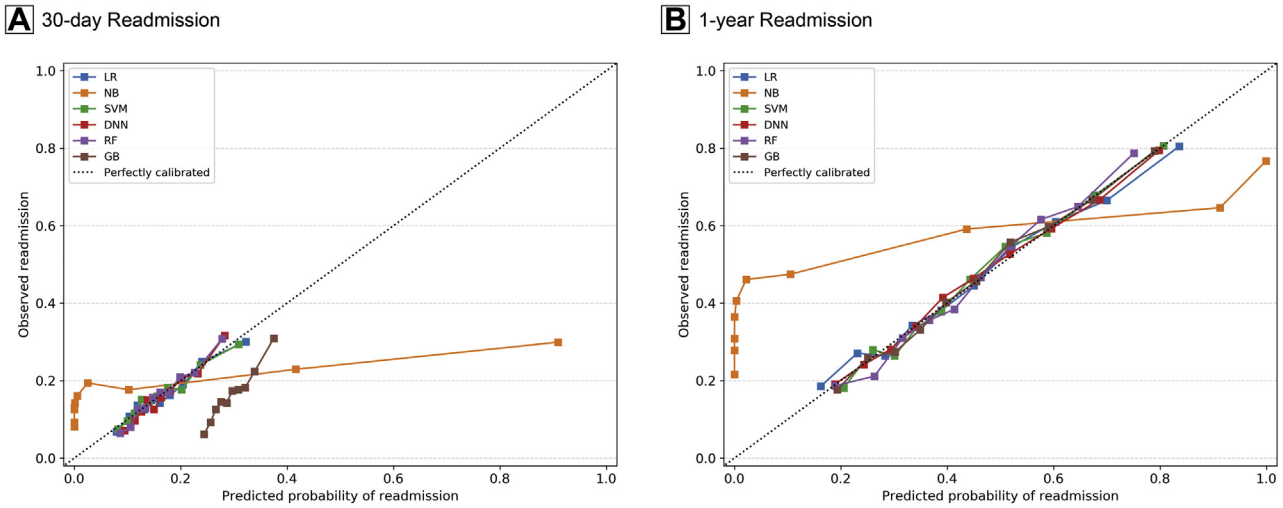

Figure 2. Calibration plots for machine learning models for predicting (**A**) 30-day readmission and (**B**) 1-year readmission. LR (**blue line**) indicates logistic regression; NB (**orange line**), na ve Bayes; SVM (**green line**), support vector machine; DNN (**red line**), deep neural network; RF (**purple line**), random forest; and GB (**brown line**), gradient boosting.

interventions toward patients more likely to benefit from them, thereby improving patient health and reducing costs. As such, several models have been developed to estimate a patient's risk of readmission after hospitalization for AMI. However, many of the existing prediction models use data that are not available until well after discharge (eg, administrative claims or registry data),[12] are developed using data derived from a single centre, or lack any validation.[26-28] We overcame these limitations by relying solely on data collected during a patient's hospital stay, from across 81 hospitals in Ontario, Canada. To our knowledge, this is the first study that uses detailed clinical data that are routinely available to clinicians to evaluate ML algorithms for the task of predicting AMI readmission. Moreover, using data that have been collected from multiple locations ensures that the variables included in the models represent standard care of practice, rather than esoteric tests and measurements. These factors help increase the applicability of the models to clinical settings.

Our study found that ML methods do not provide an adequate increase in accuracy for predicting readmission for AMI to warrant clinical use, when compared with conventional approaches reported in literature. The limited number of previously published papers for prediction of readmission for AMI rely mostly on points-based risk scores or LR models,[9,10,29] which tend to underperform when there are complex decision boundaries present in the data. Employing ML algorithms that are able to leverage complex higher-level interactions among input variables could potentially overcome this drawback. We found minor improvement in discrimination for 30-day readmission when using models such as RF, GB, and DNN (0.64) as opposed to LR (0.63), which is encouraging but ultimately insufficient. This suggests that the inclusion of complex interactions does not improve predictive accuracy in this context. Furthermore, calibration plots showed that no model was as—or more—accurate in predicting readmissions than LR. The outcome was similar in the case of 1-year readmission; LR performed just as well as more complex models, such as DNN or RF, when comparing C-statistic and model calibration. Despite the overall modest discrimination for both 30-day and 1-year readmission, we found that the range of observed readmission events among deciles of predicted risk varies sufficiently to stratify patients into low-, medium-, and high-risk groups to target readmission intervention prevention.

We attribute the modest predictive ability of the various ML algorithms to 2 factors. First, predicting readmission is a difficult problem, especially when compared with other[30] prediction tasks such as mortality.[6,12,31] Unlike mortality, which is largely dependent upon severity of disease and comorbidity burden,[32-36] readmission is likely a result of more complex interactions among patients' clinical conditions, sociodemographic factors, and psychosocial environments.[33-36] Such complex factors are unlikely to be captured within the data from a single hospital stay. Second, our study adds to the growing trend of literature, which confirms that, although more complex ML methods have potential, they do not necessarily confer an advantage when the data are tabular, contain a relatively small number of features, and are of a modest sample size. This is particularly true of deep learning (deep neural network) models; although they are capable of using raw data as input, they require enormous numbers of data to train and achieve sufficiently high performance. For example, researchers recently demonstrated a deep learning model capable of achieving high accuracy on various tasks including in-hospital mortality (C-statistic 0.93 to 0.94), 30-day readmission (C-statistic 0.75 to 0.76), and prolonged length of stay (C-statistic 0.85 to 0.86).[37] This deep learning model was trained using raw time-ordered electronic health records data consisting of 46 billion data points (including clinical notes) from 216,221 patients. It is because of these large clinical datasets and computational resources that complex ML algorithms have become practical and useful. However, they cannot replace the use of classifiers and regressors on smaller tabular datasets such as in this study.

Limitations

This study has several potential limitations. First, the ML models developed use data from the patient's entire stay to make a prediction at the time of discharge. Although this would enable hospitals to implement postdischarge interventions, there is some evidence to suggest that certain interventions are more effective at reducing readmissions if implemented well before discharge.[12,38] Second, certain AMI-severity measures, such as troponin value and frailty, were not included in the dataset, although it is unlikely that the addition of further severity scores will improve accuracy drastically, given that alternate measures of disease severity are included. Third, we did not include postdischarge factors as predictors because our intention was to predict rehospitalization before discharge and concerns for bias because the overlap in ascertainment period because the predictor and the outcome. Finally, although we used a robust nested cross-validation approach to verify our models, if these models were to be implemented, they would require external validation on more recent data than used in the study to reflect any changes in care practices for AMI.[39]

Conclusions

In this analysis of readmission prediction after hospitalization for AMI using clinical data, we found that ML methods do not improve discrimination when compared with previously reported approaches. Future work needs to focus on further improvement of predictive ability through the use of larger datasets that contain both clinical time-series data and information on sociodemographic factors to facilitate earlier and better targeting of interventions to patients at high risk for readmission.

Acknowledgements

This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). The opinions, results, and conclusions reported in this article are those of the authors and are independent from the funding sources. No endorsement by ICES or the MOHLTC is intended or should be inferred. Parts of this material are based on data and information compiled and provided by CIHI. The analyses, conclusions, opinions, and statements expressed herein are those of the author and not necessarily those of CIHI.

Funding Sources

This study was funded by a University of Toronto EMH Seed grant and a Foundation grant (FDN-154333) from the Canadian Institutes of Health Research.

Disclosures

Dr Ko and Dr Austin are supported by a Mid-Career Investigator Award from the HSF, Ontario Provincial Office. The other authors have no conflicts of interest to disclose.

References

1. Krumholz HM, Lin Z, Keenan PS, et al. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. JAMA 2013;309:587-93.
2. Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. JAMA 2013;309:342-3.
3. Kripalani S, Theobald CN, Anctil B, Vasilevskis EE. Reducing hospital readmission rates: current strategies and future directions. Annu Rev Med 2014;65:471-85.
4. Dunlay SM, Pack QR, Thomas RJ, Killian JM, Roger VL. Participation in cardiac rehabilitation, readmissions, and death after acute myocardial infarction. Am J Med 2014;127:538-46.
5. Feltner C, Jones CD, Cene CW, et al. Transitional care interventions to prevent readmissions for persons with heart failure: a systematic review and meta-analysis. Ann Intern Med 2014;160:774-84.
6. van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ 2010;182:551-7.
7. Zhou H, Della PR, Roberts P, Goh L, Dhaliwal SS. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review. BMJ Open 2016;6:e011060.
8. Wang H, Robinson RD, Johnson C, et al. Using the LACE index to predict hospital readmissions in congestive heart failure patients. BMC Cardiovasc Disord 2014;14:97.
9. Rana S, Tran T, Luo W, Phung D, Kennedy RL, Venkatesh S. Predicting unplanned readmission after myocardial infarction from routinely collected administrative hospital data. Aust Health Rev 2014;38:377-82.
10. Krumholz HM, Lin Z, Drye EE, et al. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2011;4:243-52.
11. Yu S, Farooq F, van Esbroeck A, Fung G, Anand V, Krishnapuram B. Predicting readmission risk with institution-specific prediction models. Artif Intell Med 2015;65:89-96.
12. Smith LN, Makam AN, Darden D, et al. Acute myocardial infarction readmission risk prediction models: a systematic review of model performance. Circ Cardiovasc Qual Outcomes 2018;11:e003885.
13. Crown WH. Potential application of machine learning in health outcomes research and some statistical cautions. Value Health 2015;18:137-40.
14. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One 2017;12:e0174944.
15. Tu JV, Donovan LR, Lee DS, et al. Effectiveness of public report cards for improving the quality of cardiac care: the EFFECT study: a randomized trial. JAMA 2009;302:2330-7.
16. Murphy KP. Machine Learning: A Probabilistic Perspective. Cambridge, MA: MIT Press;2012:82-3.
17. Cortes C, Vapnik V. Support-sector networks. Machine Learning 1995;20:273-97.
18. Breiman L. Random forests. Machine Learning 2001;45:5-32.
19. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001;29:1189-232.

20. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323:533-6.

21. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in python. J Machine Learning Res 2011;23:2825-30.

22. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016; San Francisco, California.

23. Chollet F. Keras, https://keras.io. Updated 2015.

24. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128-38.

25. Khera R, Jain S, Pandey A, et al. Comparison of readmission rates after acute myocardial infarction in 3 patient age groups (18 to 44, 45 to 64, and ≥65 years) in the United States. Am J Cardiol 2017;120:1761-7.

26. Meddings J, Reichert H, Smith SN, et al. The impact of disability and social determinants of health on condition-specific readmissions beyond Medicare risk adjustments: a cohort study. J Gen Intern Med 2017;32:71-80.

27. Brown JR, Conley SM, Niles NW, II. Predicting readmission or death after acute ST-elevation myocardial infarction. Clin Cardiol 2013;36:570-5.

28. McManus DD, Saczynski JS, Lessard D, et al. Reliability of predicting early hospital readmission after discharge for an acute coronary syndrome using claims-based data. Am J Cardiol 2016;117:501-7.

29. Burke RE, Schnipper JL, Williams MV, et al. The HOSPITAL score predicts potentially preventable 30-day readmissions in conditions targeted by the hospital readmissions reduction program. Med Care 2017;55:285-90.

30. Joynt KE, Orav EJ, Jha AK. Thirty-day readmission rates for Medicare beneficiaries by race and site of care. JAMA 2011;305:675-81.

31. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. JAMA 2011;306:1688-98.

32. Amarasingham R, Moore BJ, Tabak YP, et al. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. Med Care 2010;48:981-8.

33. Foraker RE, Rose KM, Suchindran CM, Chang PP, McNeill AM, Rosamond WD. Socioeconomic status, Medicaid coverage, clinical comorbidity, and rehospitalization or death after an incident heart failure hospitalization: atherosclerosis risk in communities cohort (1987 to 2004). Circ Heart Fail 2011;4:308-16.

34. Herrin J, St Andre J, Kenward K, Joshi MS, Audet AM, Hines SC. Community factors and hospital readmission rates. Health Serv Res 2015;50:20-39.

35. Alter DA, Franklin B, Ko DT, et al. socioeconomic status, functional recovery, and long-term mortality among patients surviving acute myocardial infarction. PLoS One 2013;8:e65130.

36. Singh S, Lin YL, Kuo YF, Nattinger AB, Goodwin JS. Variation in the risk of readmission among hospitals: the relative contribution of patient, hospital and inpatient provider characteristics. J Gen Intern Med 2014;29:572-8.

37. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. NPJ Digital Med 2018;1:18.

38. Rennke S, Nguyen OK, Shoeb MH, Magan Y, Wachter RM, Ranji SR. Hospital-initiated transitional care interventions as a patient safety strategy: a systematic review. Ann Intern Med 2013;158:433-40.

39. Wong GC, Welsford M, Ainsworth C, et al. 2019 Canadian Cardiovascular Society/Canadian Association of Interventional Cardiology Guidelines on the Acute Management of ST-Elevation Myocardial Infarction: focused update on regionalization and reperfusion. Can J Cardiol 2019;35:107-32.

Supplementary Material

To access the supplementary material accompanying this article, visit the online version of the *Canadian Journal of Cardiology* at www.onlinecjc.ca and at https://doi.org/10.1016/j.cjca.2019.10.023.