Canadian Journal of Cardiology 36 (2020) 1574–1576

Editorial

The Rise of Open-Sourced Machine Learning in Small and Imbalanced Datasets: Predicting In-Stent Restenosis

Robert Avram, MD, MSc,[a] Jeffrey E. Olgin, MD,[a] and Geoffrey H. Tison, MD, MPH[a,b]

[a] *Division of Cardiology and Cardiovascular Research Institute, Cardiology, Department of Medicine, University of California, San Francisco, California, USA*

[b] *Bakar Computational Health Sciences Institute, University of California, San Francisco, California, USA*

See article by Sampedro-Gómez et al., pages 1624–1632 of this issue.

After the successful placement of a coronary stent and revascularization, in-stent restenosis (ISR) occurs in up to 12% of patients and represents the gradual failure of the stent by lumen renarrowing.[1] Clinically, ISR is important to recognize because it is usually associated with recurrent angina symptoms, a higher risk for acute coronary syndrome, and increased mortality.[1] The ability to accurately predict ISR would enable closer monitoring of those patients at higher risk of ISR or the consideration of alternate therapies.[1] Accordingly, various risk prediction models have been developed to predict ISR, such as **P**revention of **Res**tenosis With **T**ranilast and Its **O**utcomes (PRESTO)-1,[2] PRESTO-2,[2] and the and **Ev**aluation of Drug-**E**luting Stents and Ischemic Eve**nt**s (EVENT) risk score[3] (Table 1). However, these prediction models generally suffer from low predictive power: Areas under the receiver operating characteristic curve (AUC-ROCs) are 0.63 for PRESTO-1,[2] 0.63 for PRESTO-2,[2] and 0.68 for EVENT.[3] This presents an opportunity to develop models that more accurately predict the occurrence of ISR.

In recent years, there has been increasing interest in applying machine-learning algorithms toward various tasks in medicine, including risk prediction.[4] Machine-learning algorithms offer several potential advantages over conventional statistical techniques used in risk prediction, such as improved performance as training data increases and the ability to model nonlinear relationships within the data.[4] Such models often can accommodate a far larger number of predictors (also called "features") but may require large dataset sizes to achieve improved predictive power over conventional risk-prediction models.[4] Machine-learning algorithms also have limitations, which include less interpretability than general linear models, and may not offer performance improvements in settings of small dataset sizes or fewer numbers of predictive features.[4]

In this issue of the *Canadian Journal of Cardiology*, Sampedro-Gómez et al. applied various machine-learning algorithms to develop risk models for the prediction of ISR.[5] The authors sought to predict ISR in patients with ST-segment-elevation myocardial infarction (STEMI) undergoing percutaneous coronary intervention (PCI) within the **Gr**upo de **A**nálisis de la **C**ardiopatía **I**squemica **A**guda (GRACIA)-3 study population.[5] The GRACIA-3 trial was a 2×2 randomized, open-label, multicenter trial that compared the efficacy of the paclitaxel-eluting stent with conventional bare-metal stents in patients with STEMI.[6] ISR was assessed by core laboratory–validated coronary angiography at baseline and at 12-month follow-up in all patients.[6] To develop machine learning–based risk prediction models, the authors used the 23 available case subjects with ISR and 240 control subjects in GRACIA-3, obtaining 68 demographic, clinical, and angiographic characteristics from each patient. A notable characteristic of the data used was that there were relatively low numbers of ISR events available (n = 23), and the total dataset size was also not very large (n = 263), both characteristics which might be less favourable to attempting machine-learning approaches. To address the smaller size of their dataset, the authors appropriately used an approach called *k*-fold cross-validation, which serves to provide more generalizable estimates of model performance and limit overfitting by splitting the dataset repeatedly into *k*-sized parts. Although this can be less ideal than maintaining a true hold-out subset of the data for model testing, *k*-fold cross-validation can be useful when the dataset is too small to realistically maintain a hold-out test subset. In addition, because the balance between those with and without ISR was nearly 1:10, this dataset was considered to be "imbalanced," which can have implications regarding both model training and the assessment of model performance. To account for this dataset imbalance, the authors, again appropriately, examined the area under the precision-recall curve (AUC-PR) in addition to the more common AUC-ROC. Unlike, the AUC-ROC, which can be overly optimistic in the setting of

Received for publication January 30, 2020. Accepted February 3, 2020.

Corresponding author: Dr Robert Avram, 555 Mission Bay Blvd South, San Francisco, California 94158, USA.
E-mail: robert.avram@ucsf.edu

See page 1576 for disclosure information.

https://doi.org/10.1016/j.cjca.2020.02.002

Table 1. Risk-score models for predicting in-stent restenosis (ISR) derived from the PRESTO, EVENT and GRACIA-3 cohorts

Characteristic	PRESTO 1[2]	PRESTO 2[2]	EVENT[3]	GRACIA-3[5]
Study population	All comers for PCI with bare-metal stent	All comers for PCI with bare-metal stent	All PCI patients in a "real-world" registry treated with drug-eluting stents, except those who had a PCI or bypass surgery within the previous 4 weeks	STEMI patients randomized to paclitaxel-eluting stent or bare-metal stent
Number of patients	1312	1312	8829	263
Lesion characteristics	601 with ISR; 711 without	601 with ISR; 711 without	321 with target lesion revascularization; 8508 without	23 with ISR; 240 without
Model type	Logistic regression	Logistic regression	Logistic Regression	Extremely randomized trees
Performance (AUC-ROC)	0.63	0.63	0.68	0.77
Performance in the GRACIA-3 cohort (AUC-ROC)	0.67	0.64	0.63	0.77
Performance in the GRACIA-3 cohort (AUC-PR)	0.31	0.27	0.18	0.46
Covariates	Treated diabetes mellitus	Treated diabetes mellitus	Age < 60 y	Diabetes mellitus
	Vessel size	Vessel size	Previous PCI	Two-vessel coronary artery disease
	≤ 2.5 mm	≤ 2.5 mm	Unprotected left main PCI	Post-PCI TIMI flow
	2.5-3 mm	2.5-3 mm	Saphenous vein graft PCI	Post-PCI thrombus
	3-3.5 mm	3-3.5 mm	Minimum stent diameter ≤ 2.5 mm	Abnormal platelets count
	3.5-4 mm	3.5-4 mm	Total stent length ≥ 40 mm	Elevated total cholesterol
	> 4 mm	> 4 mm		
	Lesion length > 20 mm	Lesion length > 20 mm		
	Smoking status	Smoking status		
	Type C lesion	Type C lesion		
	Previous PCI	Previous PCI		
	Female gender	Ostial location		
	Unstable angina			

AUC-PR, area under the precision-recall curve; AUC-ROC, area under the receiver operating characteristic curve; EVENT, Evaluation of Drug-Eluting Stents and Ischemic Events; GARCIA-3, **Gr**upo de **A**nálisis de la **C**ardiopatía **I**squemica **A**guda-3; PCI, percutaneous coronary intervention; PCI, percutaneous coronary intervention; PRESTO, **P**revention of **Res**tenosis With **T**ranilast and Its **O**utcomes; STEMI, ST-segment-elevation myocardial infarction; TIMI, Thrombolysis in Myocardial Infarction.

imbalanced data, the AUC-PR is less affected by imbalanced data and paints a more complete picture of the predictive performance of the model in the minority class, which in this dataset represented the 23 cases of ISR (9% rate).[7] Using both of these techniques, in part to address the limitation of having a lower number of ISR events and a smaller dataset size, the authors then developed risk-prediction models using 6 different machine-learning classifiers. They compared the performance of these machine learning–derived models against existing risk prediction models, including PRESTO-1, PRESTO-2, and EVENT scores applied to the GRACIA-3 population.

The best-performing model was developed using an extremely randomized trees (ERT) classifier, but it was not statistically different from the other machine-learning models used, except for the support vector classifier. The small dataset size may have contributed to the inability to detect differences between the various machine-learning classifiers. Compared with existing ISR prediction models, the ERT classifier achieved a higher AUC-PR of 0.46 than the other models applied to this study cohort (AUC-PRs of 0.31 for PRESTO-1, 0.27 for PRESTO-2, and 0.18 for EVENT). Again, likely because of the small dataset size, the confidence intervals of the AUC-PRs for all of these models were large, and the AUC-PR of the best-performing machine-learning model was not statistically significantly higher than those for PRESTO-1 or PRESTO-2. However, the ERT model appeared to outperform the EVENT risk score, and there was a trend toward improved discrimination compared with both PRESTO scores. It is noteworthy that despite the small dataset size and low numbers of events, the machine-learning models performed as well as they did. In our view, some of the greatest potential may lie in applying similar methods as shown in this work to similar prediction tasks with larger datasets and greater numbers of events.

In addition to performing ISR prediction, the most important predictors out of the 68 available predictorsc were identified during the feature selection process. The GRACIA-3 population looked at STEMI patients undergoing PCI, which is inherently different from the population of the PRESTO and EVENT studies, which recruited a majority of non-STEMI cases.[2,5] Several key predictors identified by Sampedro-Gómez et al., such as ≥ 2 coronary vessel disease, post-PCI Thrombolysis in Myocardial Infarction flow, post-PCI thrombus, abnormal platelets count, and elevated total cholesterol, have not been consistently reported in previous

models that looked at non-STEMI cases. Only diabetes has previously been used for ISR risk prediction in the PRESTO-1, PRESTO-2, and EVENT risk models.[2,5] The other predictors highlighted in this study could represent novel predictors of post-STEMI PCI–associated ISR.[1] It is particularly noteworthy to see that the strongest predictors of ISR in the ERT machine-learning model are supported by literature and consistent with our understanding of ISR pathophysiology, involving both neointimal proliferation and neoatherosclerosis.[1] Future studies with larger numbers of events and larger datasets should examine the value of these features and, if consistent, these features could be considered for inclusion in future ISR risk-prediction models. Furthermore, the feature selection approach used in this study could be valuable in other settings to identify other novel predictors of ISR.[1] Because the strongest predictors in this model are readily available in clinical practice, this model can be easily clinically applied to identify the subset of STEMI patients who are at risk for ISR after PCI,[5] prompting more frequent follow-up to rule out the development of ISR.

The open-source code made available by the authors to accompany this article makes it possible for other institutions and investigators to develop similar machine-learning approaches to data analysis. This study provides an important proof-of-concept demonstration that machine-learning models can be used to develop effective risk-prediction models for ISR. Because the findings from this study were likely limited by the size of the available data, by open-sourcing the code, the authors have laid the groundwork to replicate their methods with the use of larger datasets or possibly datasets combined from multiple institutions or clinical trials. Using datasets of larger size as well as larger numbers of predictors would enable the community to truly leverage the strengths of machine-learning approaches. In addition, open-sourcing machine-learning models enables rapid confirmation, implementation, and iterative improvement, which should be a standard to which we all aspire, especially for machine-learning applications in medicine.

Despite the methodology and findings, there remain important limitations to the work presented by Sampedro-Gómez.[5] The GRACIA-3 trial compared first-generation drug-eluting stents with bare-metal stents. Second-generation drug-eluting stents implanted during STEMI could have a different risk profile for ISR, and the model developed here might not apply to that population. In addition, the event rate was low (9%), and the trial lacked an external validation in an independent cohort. Until the latter is established, individualized risk of ISR predicted by this machine-learning model needs to be taken with a grain of salt because the model may suffer from overfitting and lack external generalizability. However, the authors used approaches such as cross-validation which alleviate some of these concerns.

In summary, Sampedro-Gómez et al. have described a process to apply machine-learning algorithms to develop different models predicting the 12-month risk of ISR in a STEMI population receiving first-generation drug-eluting stents or bare-metal stents. The best-performing model they report provides superior risk prediction for ISR compared with at least 1 previously published model[3] and has the potential to be further improved if more data are available. Other risk-prediction efforts that use similarly small or imbalanced datasets might consider adopting a similar approach of cross-validation and examination of AUC-PRs. Future research, such as prospective validation of this score in other cohorts and evaluation of the reported novel predictors, will be crucial to determine the clinical implications of these findings.

Funding Sources

R.A. is supported by the Fonds de la recherche en santé du Québec (274831). J.E.O. has received research funding from the National Institutes of Health (U2CEB021881). G.H.T. received support from the National Heart, Lung, and Blood Institute, National Institutes of Health (K23HL135274).

Disclosures

J.E.O. has received research funding from Samsung and iBeat. The other authors have no conflicts of interest to disclose.

References

1. Alfonso F, Byrne RA, Rivero F, Kastrati A. Current treatment of in-stent restenosis. J Am Coll Cardiol 2014;63:2659-73.

2. Singh Mandeep, Gersh Bernard J, McClelland Robyn L, et al. Clinical and angiographic predictors of restenosis after percutaneous coronary intervention. Circulation 2004;109:2727-31.

3. Stolker JM, Kennedy KF, Lindsey JB, et al. Predicting restenosis of drug-eluting stents placed in real-world clinical practice: derivation and validation of a risk model from the EVENT registry. Circ Cardiovasc Interv 2010;3:327-34.

4. Deo RC. Machine learning in medicine. Circulation 2015;132:1920-30.

5. Sampedro-Gómez J, Dorado-Díaz PI, Vicente-Palacios V, et al. Machine learning to predict stent restenosis based on daily demographic, clinical, and angiographic characteristics. Can J Cardiol 2020;36:1624-32.

6. Sánchez PL, Gimeno F, Ancillo P, et al. Role of the paclitaxel-eluting stent and tirofiban in patients with ST-elevation myocardial infarction undergoing postfibrinolysis angioplasty: the GRACIA-3 randomized clinical trial. Circ Cardiovasc Interv 2010;3:297-307.

7. Saito S, Isshiki T, Kimura T, et al. Efficacy and safety of adjusted-dose prasugrel compared with clopidogrel in Japanese patients with acute coronary syndrome. Circ J 2014;78:1684-92.