



Canadian Journal of Cardiology ■ (2020) 1–9

Clinical Research

Machine Learning to Predict Stent Restenosis Based on Daily Demographic, Clinical, and Angiographic Characteristics

Jesús Sampedro-Gómez, Eng,^{a,b,‡} P. Ignacio Dorado-Díaz, Eng,^{a,b,‡}
 Víctor Vicente-Palacios, Eng, PhD,^c Antonio Sánchez-Puente, Phys, PhD,^a
 Manuel Jiménez-Navarro, MD, PhD,^{b,d} J. Alberto San Roman, MD, PhD,^{b,e}
 Purificació Galindo-Villardón, Bstt, PhD,^f Pedro L. Sanchez, MD, PhD,^{a,b} and
 Francisco Fernández-Avilés, MD, PhD^{b,g}

^aDepartment of Cardiology, Hospital Universitario de Salamanca-IBSAL, Universidad de Salamanca, Salamanca, Spain

^bCIBERCV, Instituto de Salud Carlos III, Madrid, Spain

^cPhilips Healthcare, xxx

^dUGC Área del Corazón, Hospital Virgen de la Victoria-IBIMA, Universidad de Málaga, Málaga, Spain

^eInstituto de Ciencias del Corazón, Hospital Clínico Universitario de Valladolid, Valladolid, Spain

^fDepartment of Statistics, Universidad de Salamanca, Salamanca, Spain

^gDepartment of Cardiology, Hospital General Universitario Gregorio Marañón-IiSGM, Universidad Complutense de Madrid, Madrid, Spain

See editorial by Avram et al., pages xxx-xxx of this issue.

ABSTRACT

Background: Machine learning (ML) has arrived in medicine to deliver individually adapted medical care. This study sought to use ML to discriminate stent restenosis (SR) compared with existing predictive scores of SR. To develop an easily applicable model, we performed our predictions without any additional variables other than those obtained in daily practice.

Methods: The dataset, obtained from the Grupo de Análisis de la Cardiopatía Isquémica Aguda (GRACIA)-3 trial, consisted of 263 patients with demographic, clinical, and angiographic characteristics;

Percutaneous coronary intervention (PCI) with stenting has become routine clinical practice for the revascularization of atherosclerotic coronary vessel disease with significant obstruction. Even in the worst clinical scenario, ST-elevation acute myocardial infarction (STEMI), coronary stent

RÉSUMÉ

Contexte : L'avènement de l'apprentissage automatique (AA) en médecine permet de fournir des soins personnalisés. Cette étude portait sur l'utilisation de l'AA comme outil prédictif du risque de resténose après l'implantation d'une endoprothèse comparativement aux scores prédictifs existants. Pour mettre au point un modèle facilement applicable, nous avons effectué nos prévisions sans autres variables que celles obtenues dans la pratique quotidienne.

Méthodologie : L'ensemble de données, issu de l'essai Grupo de Análisis de la Cardiopatía Isquémica Aguda (GRACIA)-3, portait sur

implantation is associated with lower adverse cardiovascular events and longer survival compared with medical treatment.^{1,2}

The long-term success of the procedure has as its main downside the possibility of developing stent restenosis (SR),³ an iatrogenic process caused by excessive neointimal hyperplasia, leading to recurrent lumen narrowing at the site of initial PCI, which may manifest clinically as stable angina but also in the form of acute coronary syndrome, as a new myocardial infarction (MI) and need for new target-vessel revascularization.⁴ Identifying patients with SR is a major challenge, as management of patients with recurrent SR represents a major therapeutic dilemma.²

Received for publication October 8, 2019. Accepted January 14, 2020.

[‡]These authors contributed equally to this study as first authors.

Address for correspondence: Dr Pedro L Sanchez, Cardiology Department, Hospital Universitario de Salamanca-IBSAL, Paseo de San Vicente 58-187, 37007 Salamanca, Spain. Tel.: +34-923291100 (ext 55356).

E-mail: pedrolsanchez@secardiologia.es

See page 8 for disclosure information.

<https://doi.org/10.1016/j.cjca.2020.01.027>

0828-282X/© 2020 Canadian Cardiovascular Society. Published by Elsevier Inc. All rights reserved.

23 (9%) of them presented with SR at 12 months after stent implantation. A methodology to work with small imbalanced datasets, based in cross-validation and the precision/recall (PR) plots, was used, and state-of-the-art ML classifiers were trained.

Results: Our best performing model (0.46, area under the PR curve [AUC-PR]) was developed with an extremely randomized trees classifier, which showed better performance than chance alone (0.09 AUC-PR, corresponding to the 9% of patients presenting SR in our dataset) and 3 existing scores; *Prevention of Restenosis With Tranilast and its Outcomes* (PRESTO)-1 (0.31 AUC-PR), PRESTO-2 (0.27 AUC-PR), and *Evaluation of Drug-Eluting Stents and Ischemic Events* (EVENT) (0.18 AUC-PR). The most important variables ranked according to their contribution to the predictions were diabetes, ≥ 2 vessel-coronary disease, post-percutaneous coronary intervention thrombolysis in myocardial infarction (PCI TIMI)-flow, abnormal platelets, post-PCI thrombus, and abnormal cholesterol. To counteract the lack of external validation for our study, we deployed our ML algorithm in an open source calculator, in which the model would stratify patients of high and low risk as an example tool to determine generalizability of prediction models from small imbalanced sample size.

Conclusions: Applied immediately after stent implantation, a ML model better differentiates those patients who will present with SR over current discriminators.

Machine-learning (ML) techniques are more and more present in medicine, with the aim of automating repetitive tasks and delivering individually adapted medical care. ML classifiers can be applied to clinical databases in which these algorithms can sift through large amounts of data and detect multivariable patterns that reliably predict outcomes.⁵⁻⁸ We postulate that ML classifiers may develop a model that discriminates SR for individual patients better than existing predictive multivariate logistic regression models of SR.^{9,10} We combined the use of daily available demographic, clinical, and angiographic data with ML to test this hypothesis.

Material and Methods

Figure 1 guides through the phases we have followed to build our ML model: preparation of the model, training the model, and evaluating the model.

Preparing to Build the ML Model

Task definition

The aim of our study was to predict 12-month follow-up SR in patients with STEMI undergoing PCI. To develop an easily applicable ML model, we performed our prediction without any additional variables other than those obtained in routine clinical practice.

Input data (features) consist of patient demographic and clinical characteristics, quantitative coronary analysis (QCA)

263 patients dont les caractéristiques démographiques, cliniques et angiographiques étaient consignées; 23 (9 %) d'entre eux avaient présenté une resténose dans les 12 mois suivant l'implantation d'une endoprothèse. Une méthodologie permettant de travailler avec de petits ensembles de données déséquilibrées, fondée sur la validation croisée et la courbe précision-rappel (PR), a été utilisée et des classificateurs d'AA de pointe ont été entraînés.

Résultats : Notre modèle le plus performant (aire sous la courbe PR [ASC-PR] : 0,46) a été mis au point à l'aide d'un classificateur d'arbres extrêmement aléatoires, présentant une meilleure performance que le hasard seul (ASC-PR : 0,09, correspondant aux 9 % de patients présentant une resténose dans notre ensemble de données), et à l'aide de trois scores existants, à savoir : *Prevention of Restenosis With Tranilast and its Outcomes* (PRESTO)-1 (ASC-PR : 0,31), PRESTO-2 (ASC-PR : 0,27) et *Evaluation of Drug-Eluting Stents and Ischemic Events* (EVENT) (ASC-PR : 0,18). Les variables les plus importantes, classées en fonction de leur contribution aux prévisions, étaient le diabète, la coronaropathie multitrunculaire (≥ 2), le flux TIMI (thrombolyse de l'infarctus du myocarde) après une intervention coronarienne percutanée, les anomalies plaquettaires, la thrombose après une intervention coronarienne percutanée et un taux de cholestérol anormal. Pour contrebalancer le manque de validation externe dans le cadre de notre étude, nous avons déployé notre algorithme d'AA dans un calculateur libre et ouvert, où le modèle opérait la stratification des patients très à risque et peu à risque à titre d'outil d'exemplification en vue de déterminer la généralisabilité des modèles prédictifs à partir d'un échantillon déséquilibré de petite taille.

Conclusions : Un modèle d'AA appliqué immédiatement après l'implantation d'une endoprothèse cerne mieux les patients qui présenteront une resténose que les indicateurs actuels.

for assessment of coronary artery dimensions before and after stent implantation, physical examination parameters, ejection fraction, and routine biochemical parameters available at the time of the PCI procedure. As for the corresponding outcome (label), we defined SR of the infarct-related lesion as a $>50\%$ narrowing of the lumen diameter in the target segment, defined as all portions of the vessel that received treatment within the stent zone, including the proximal and distal 5-mm margins.¹¹

Data collection

The data used to train and validate the ML models comes from the Grupo de Análisis de la Cardiopatía Isquémica Aguda (GRACIA)-3 trial.¹² Briefly, the GRACIA-3 trial was a 2 x 2 randomized, open-label, multicentre clinical trial that compared the efficacy of the paclitaxel-eluting stent with the conventional bare-metal stent. Patients with STEMI were enrolled from 20 Spanish hospitals. To determine the incidence of SR, coronary angiography was performed at baseline and after 12 months of follow-up. All QCA for assessment of coronary artery dimensions before and after stent implantation were analyzed at an independent angiography core laboratory (ICICOR, Valladolid, Spain) with a well-validated quantitative computer-based system (Medis, Leesburg, Virginia). The rate of SR was assessed by an experienced reader who was blinded to—and not directly involved in—the stent-implantation project.

The GRACIA-3 consisted of 436 patients, of whom 299 (69%) underwent 12-month angiographic follow-up. The

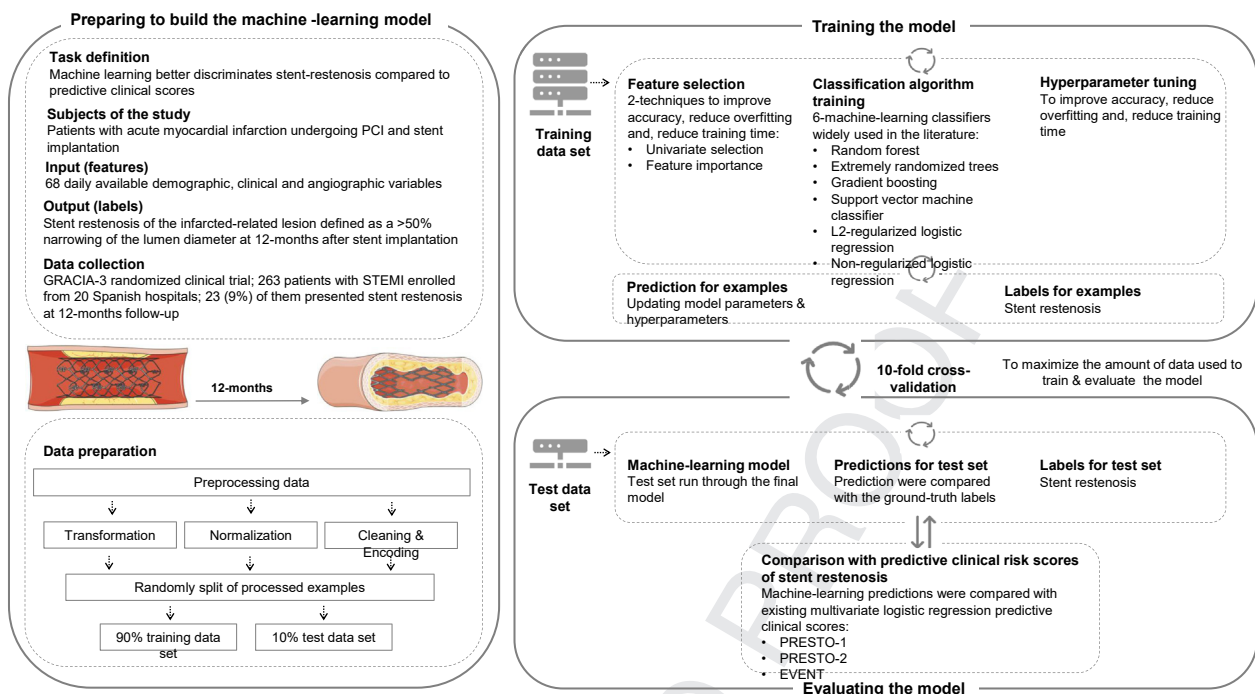


Figure 1. Overview of the phases we have followed to build our machine-learning (ML) model. In preparation of the model, we defined the task, selected the subjects of the study, discriminated input (features) and output (labels), and proceeded to the data collection and preparation. The set of processed examples came from the GRACIA-3 randomized clinical trial¹² and was divided into 2 sets. The first, the training dataset, was used to build the model; the second, the test set, was used to assess how well the model performs. In the model training, the model was developed including a selection of features, the training of different ML classifiers and the application of cross-validation methodology. The parameters of the classification algorithm and the particular feature selection strategy were chosen by means of a hyperparameter-tuning phase to improve the training algorithm further. Finally, in evaluating the model, the test set was run through the final ML model to estimate its performance in a real-world scenario. Comparisons of the development ML model with existing predictive clinical risk scores of stent restenosis were also performed.

dataset used for model development is composed of 263 of 299 (88%) patients for whom 68 features related with demographic, clinical, stent characteristics, and angiographic data were available. This model development cohort was identified before merging input data with outcome, to avoid selection bias.

The GRACIA-3 executive committee and the institutional Ethical Committee of the University Hospital of Salamanca approved the retrospective use of the identified data from the trial for this study (PI201902178).

Data preparation

An important step in preparing any ML model consists in preprocessing the raw data as a set of features to be usable by ML algorithms. For this purpose, multicategory features were 1-hot encoded in binary variables. Missing values were filled with the median and mode of each continuous and categorical feature, respectively.

In addition, different approaches are required in data preparation when the dataset used to train the model is not large and has imbalanced outputs, as in our case. We performed k-fold cross validation methodology¹³ to split the dataset randomly into k-equally sized parts: k-1 parts constituted our training dataset, and the remaining one was used as test dataset for evaluating the model. In our model, we used 10-fold cross-validation with 20 repetitions: a typical setup

that guarantees a minimum number of minority class cases represented in both training and test (Supplemental Fig. S1).

Training the ML Model

ML models and feature-selection techniques

We applied and compared the performance of 6 ML classifiers that are widely used in the literature:⁷ random forest (RF), extremely randomized trees (ERT), gradient boosting (GB), support vector machine classifier (SVC), and, L2-regularized logistic regression (LR). In addition, non-regularized logistic regression (LR_NOREG) was trained as frame of reference. The regularization terms in LR consists of a penalty in the β coefficient's fit added to limit the overfitting of the algorithm. These 6 ML classifiers were trained with 2 different feature-selection techniques, making a total of 12 different model pipelines. The first feature selection method was based in univariate analysis of variance (ANOVA) to select those features that have the strongest relationship with the presence of SR at 12 months after stent implantation. The second feature selection method was based in feature importance in which, using a RF classifier, we obtained a score for each feature of the dataset; the higher the score, the more important is the feature toward the output variable.

Model development codes were developed in Python and for the implementation of k-fold splitting, feature selection,

and ML classifiers, the open source library scikit-Learn was used.¹⁴

Cross-validation scheme and hyperparameter tuning

All models were trained and evaluated with the aforementioned 10-fold cross-validation scheme. Tuning the hyperparameters to improve the performance of ML classifiers was also performed, following a nested cross-validation scheme, performing 9-fold cross-validations with each training subset of the outer cross-validation. This approach is computationally costly because of its highly iterative nature, but it is affordable when dealing with small datasets and allows the use of all the dataset without holding out a part of it to test the generalization error.¹⁵ The optimization of the parameters was performed with a grid search and randomized search algorithms,¹⁶ with the aim of assessing the trade-off between model accuracy and computational efficiency. The fixed values of not optimized hyperparameters and the ranges of optimized ones for each classification and feature selection algorithm can be consulted in [Supplemental Table S1](#).

Evaluating the ML Model

Evaluation scheme

We used the test dataset we set aside on each k-fold train-test split to evaluate our models. Each time we evaluate a new instance of our model using these data, we followed all the steps previously described for training the model, including a new feature selection, hyperparameter tuning, and 10-fold cross-validation.

Comparison of ML models with predictive clinical risk scores of SR

We compared the performance of the developed ML algorithms with standard predictive multivariate logistic regression models: **P**revention of **R**estenosis With **T**ranilast and its **O**utcomes (PRESTO)-1, PRESTO-2, and **E**valuation of **D**rug-**E**luting **S**tents and **I**schemic **E**vents (EVENT) risk scores.

Briefly, PRESTO scores were developed from the PRESTO trial of 1312 patients.⁹ Investigators constructed 2 risk scores: PRESTO-1 used preprocedural variables (female gender, vessel size, lesion length, diabetes, smoking status, type C lesion, any previous PCI, and unstable angina) that are simple to assess and have frequently been reported to be strong predictors of SR; PRESTO-2 considered significant univariate clinical and angiographic predictors of SR identified from the PRESTO dataset (treated diabetes mellitus, nonsmoker, vessel size, lesion length, type C lesion, ostial location, and previous PCI). Both scores achieved an area under the receiver operating characteristic (AUC-ROC) curve of 0.63 on the PRESTO population. EVENT risk score (age < 60, previous PCI, unprotected left main PCI, saphenous vein graft PCI, minimum stent diameter, and total stent length) used data from the EVENT registry of 8829 patients.¹⁰ EVENT achieved a 0.68 AUC-ROC on its independent validation cohort.

For the comparison of ML models with PRESTO-1, PRESTO-2, and EVENT scores, we evaluated the existing

scores directly on our dataset, essentially performing an external validation of the prediction rules. However, as comparing the external performance of these scores with the internal (cross-validation) performance of the ML algorithms could give an unfair advantage to the internal validated ML algorithm, we further evaluated PRESTO and EVENT score performance in 2 additional ways: We used the odds ratios calculated in their respective original cohort studies and the integer scores reported by their authors, following **T**ransparent **R**eporting of a **M**ultivariable **P**rediction **M**odel for **I**ndividual **P**rognosis or **D**iagnosis (TRIPOD) statement,¹⁷ and we refitted the existing scores with β coefficients in the GRACIA population. β coefficients of these refitted models can be consulted in [Supplemental Table S2](#). The evaluation of the existing scores was computed using the same cross-validation scheme and splits described in the ML model development, obtaining 200 randomized test subsamples (10-fold with 20 repetitions).

Evaluation metrics

The differences in patient demographic, clinical characteristics, quantitative coronary analysis, physical examination, ejection fraction, and routine biochemical parameters for each partition were compared using χ^2 tests for categorical variables and ANOVA for continuous variables.

The precision-recall (PR) and the ROC curve analysis were used to assess the predictive capacity of each ML model and predictive clinical risk scores. Although the ROC curves are considered the standard metric for evaluation of ML models, in our case, the PR plots should be preferred as they are more informative when evaluating binary decision problems on imbalanced datasets.^{18,19} We concatenated the predictions of all the test sets and constructed the PR and the ROC curves, and we computed and saved the AUC-PR and the AUC-ROC of each test set of the outer cross-validation scheme. We used the first strategy to plot the PR and the ROC curves for illustration purposes and the second one to compute the mean scores, confidence intervals (CIs), and Student's *t*-test significance between ML models and predictive clinical risk scores.²⁰ We calculated these CIs and Student's *t*-test significance, taking into account that the results over each k-fold are not completely independent.²¹

The classification performance at particular cutoff thresholds based on the PR and the ROC curves were also evaluated according to its sensitivity, specificity, precision, and negative predictive value.

Finally, we computed feature importance of the best model by measuring how AUC decreases when a feature is not available through the method known as permutation importance or mean decrease accuracy (MDA).²² The method consists of replacing each feature on the test dataset with a permutation of its values and measuring the performance for the ML model. The weight of the features with positive impact in the predictive model is scaled to 1.

Open source software

The developed code used to train and evaluate the models can be consulted as open source in <https://github.com/IA-Cardiologia-husa/Restenosis>. Finally, we deployed our ML classifier in 2 open source calculators as an example tool for

external validation of ML models from small imbalanced sample size. The first is a desktop application that can be found in the GitHub repository; the second is an online tool that can be run on a Google Colaboratory [notebook](#).

Results

Characteristics of the study population

The continuous and categorical data input of the patients used for ML model development is shown in the [Supplemental Tables S3 and S4](#), respectively. Twenty-three patients (9%) presented with angiographic binary SR at 12-month follow-up of randomization. In the bivariate analysis, patients who experienced SR were significantly more likely to have the following characteristics: diabetes, abnormal cholesterol, ≥ 2 -vessel coronary disease, total number of stents implanted in the PCI, lower post-PCI minimal luminal diameter, higher post-PCI percent diameter stenosis, higher post-PCI thrombus, and lower post-PCI thrombolysis in myocardial infarction (TIMI) flow.

Comparison of prediction models

[Figure 2](#) shows the PR and the ROC curves for the 6 ML models. The ERT model performed better in both the PR and the ROC curve spaces, with a mean AUC-PR of 0.46 (95% CI: 0.29-0.63) and AUC-ROC of 0.77 (95% CI: 0.66-0.89). Considering that the PR curve in a small size imbalanced dataset like ours gives a more accurate picture of the ERT model's performance, its interpretation is of importance. The PR curve shows a good prediction capacity (0.46) compared with the 0.09 AUC-PR expected under chance alone, which corresponds to the 9% of the patients presenting with SR in our dataset. Of note, no significant differences between the ERT model with other ML models were observed, except for SVC ($P = 0.003$). Thus, other ML classifiers, especially LR and RF, would score similar results as ERT in an external validation. Computation times to fit all the models, using 4 CPU cores, were 1 and 3 days for the randomized and grid

search hyperparameter tuner, respectively, obtaining almost identical results.

[Figure 3](#) shows the PR and the ROC curves for the comparison of the ERT model with the PRESTO-1, PRESTO-2, and EVENT scores. The prediction accuracy of the ERT model outperformed predictive clinical risk scores; PRESTO-1 (AUC-PR 0.27 [0.13-0.40]) and AUC-ROC 0.55 [0.52-0.58]), PRESTO-2 (AUC-PR 0.26 [0.13-0.39], and AUC-ROC 0.58 [0.55-0.62]) and EVENT (AUC-PR 0.17 [0.11-0.22] and AUC-ROC 0.62 [0.59-0.64]), in which the best curve performance was obtained when using refitted scores with β coefficients in the GRACIA population. Differences in P values and plots for comparison between the ERT model and scores when evaluation was directly performed, or using integer scores, are shown in [Supplemental Table S5](#) and [Figure S2](#), respectively.

Variable importance

[Figure 4](#) shows the importance of the first 6 features that were found as main predictors for the ERT model: diabetes, ≥ 2 coronary vessel disease, post-PCI TIMI flow, abnormal platelets, post-PCI thrombus, and abnormal total cholesterol. Diabetes, only included in the PRESTO-2 score, appeared to be more important for SR prediction than other cardiovascular risk factors or habits such as smoking or alcohol consumption. Among the angiographic variables, post-PCI TIMI flow was the most important feature of the top predictors.

This variable importance selection is different from the feature selection used for the computation of the generalized metric score of the model. During that process, the number of features, their particular selection, and the rest of the model hyperparameters are chosen exclusively with the data of the training set in the nested cross-validation scheme. If we calculated the generalization score with a classifier and features optimized with the entire dataset, we would be leaking information from the test set to the model, incurring overfitting, and thus the generalization score would be overly optimistic. This kind of miscalculation is not uncommon in the application of cross-validation, and, as example, in this study, the

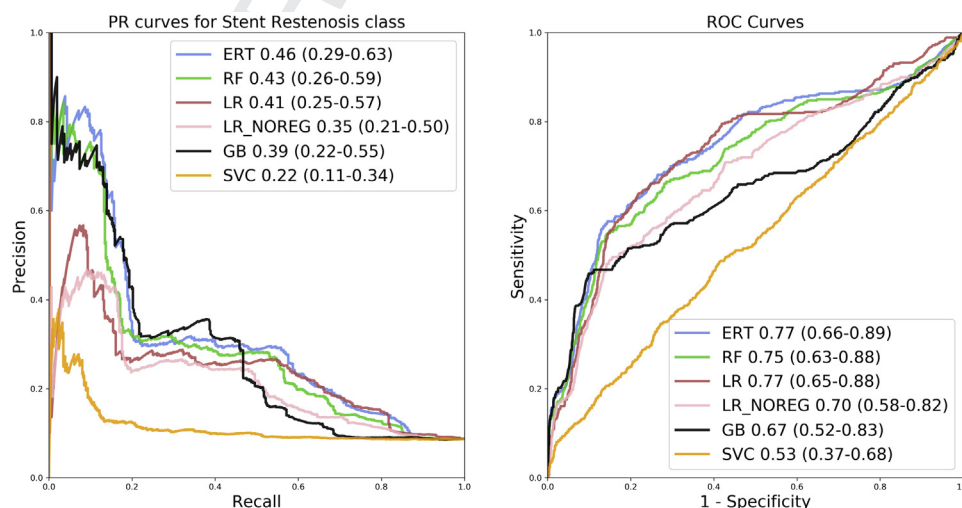


Figure 2. Areas under the precision/recall (PR) and receiver operating characteristic (ROC) curves for machine-learning models. The best final model was obtained with the extremely randomized trees (ERT) classifier. GB, gradient boosting; LR, L2-regularized logistic regression; LR_NOREG, nonregularized logistic regression; RF, random forest; SVC, support vector machine classifier.

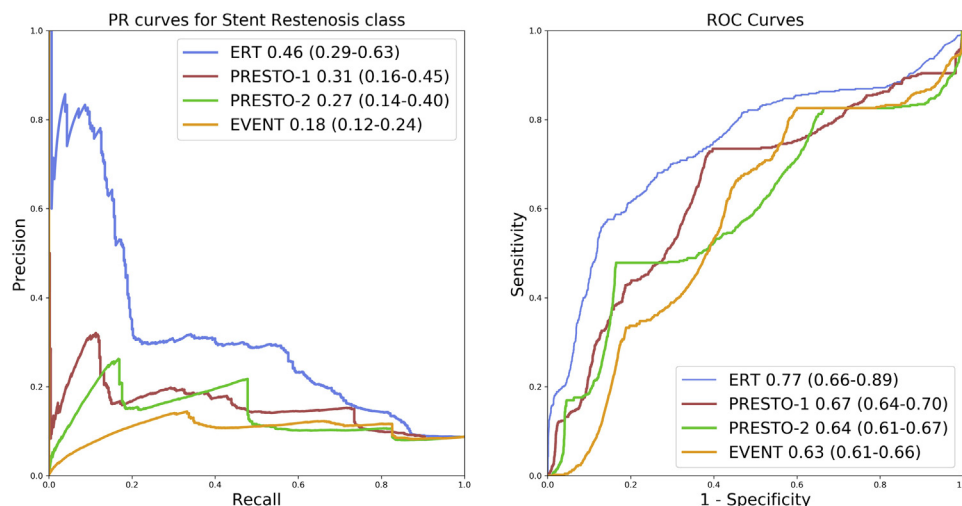


Figure 3. Areas under the precision/recall (PR) and receiver operating characteristic (ROC) curves comparing the best machine-learning classifier with those of the existing risk scores. Our developed extremely randomized trees (ERT) model outperformed predictive clinical discriminators of stent restenosis as the PRESTO-1, PRESTO-2, and EVENT risk scores. Refitted risk scores were calculated with the β coefficients in the GRACIA-3 population.

AUC-PR and AUC-ROC curve for the ERT model would have reached a value of 0.48 and 0.82, respectively.

ML model deployment in a calculator and clinical interpretation

To better assess the clinical significance of our results, we selected—based on the PR and the ROC curves—3 particular cutoff thresholds to discriminate patients at high and low risk of SR for a population similar to the GRACIA-3 cohort. For a threshold considered optimal for high precision (first operating point), our ERT model is able to achieve 89.6% of true negatives and detect a group of 3.2% patients with 50% risk of SR who should be routinely scheduled for 12-month angiographic follow-up. In contrast, for a maximum negative predictive value threshold (third operating point), our model is able to detect a group of 50.4% patients with 3.4% risk of SR avoiding unnecessary follow-up workload and lowering costs.

We deployed our ERT algorithm in a calculator in which one can input the information on the 6 features that were found as main predictors for the model and see the individual risk for SR of a certain patient over the PR and the ROC spaces in which low- and high-risk areas based on the previously described thresholds are plotted. Figure 5 shows 2 examples of the calculator run for high- and low-risk patients.

Discussion

An important goal in current medicine is the incorporation of individualized recommendations for a specific patient about a particular disease. In this regard, our ML model is of relevance, as we could avoid unnecessary follow-up and decrease economic costs for patients with predicted low risk of SR and, by contrast, improve outcomes in high-risk patients, establishing 12-month ischemic tests or catheterization.

Although we are aware of the limitation of working with a small size-imbalanced dataset to develop an ML model, we

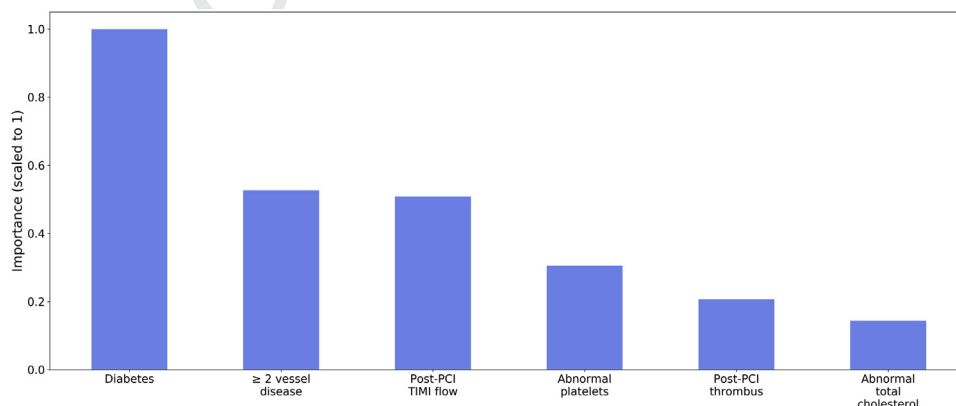


Figure 4. Bar chart of feature importance. The importance of each of the features in predicting 12-month follow-up stent restenosis with the extremely randomized trees classifier. The most important features were diabetes, ≥ 2 coronary vessel disease, and post-percutaneous coronary intervention thrombolysis in myocardial infarction (PCI TIMI) flow.

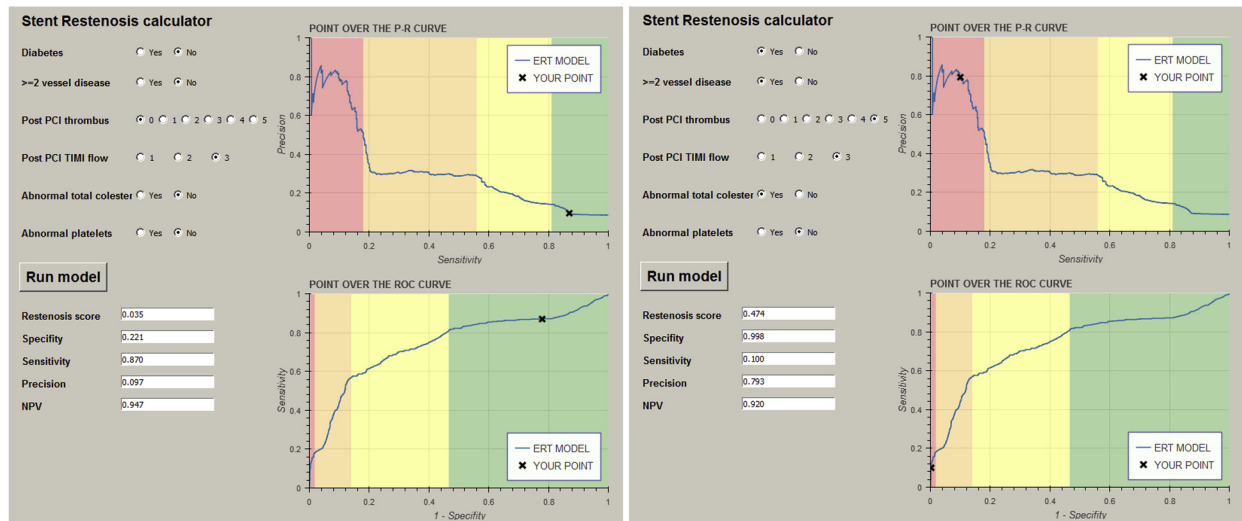


Figure 5. The stent restenosis calculator. We deployed our machine-learning algorithm in an open source calculator, in which one can input the patients' information on the 6 features described in Figure 4 and see the individual risk for stent restenosis over the precision/recall (PR) and the receiver operating characteristic (ROC) spaces. Left and right panels displayed 2 different patients' predictions at low and high risk of stent restenosis, respectively.

consider the GRACIA-3 trial a unique and valuable source of data for several reasons. First, it has allowed us to incorporate as a feature (or variable) input the use of bare-metal vs drug-eluting stents in the STEMI scenario, in which the benefit of drug-eluting stents is still under debate.²³ Our ML model would therefore be specifically useful in this clinical context and did not identify differences between both types of stents regarding predictions of SR. Second, as now it is not usual practice to perform a follow-up scheduled angiography after coronary stent implantation, the number of paired angiographies in our study (60.3%) acquires great relevance and is similar to large-cohort of patient studies in which follow-up catheterization was also scheduled.²⁴ Finally, the dataset comes from a randomized clinical study involving 20 Spanish hospitals with centralized analysis of angiograms at a core laboratory. Thus, the methodology to assess SR was blind to the type of stent, medical treatment, and outcome of the patient and consequently represents an excellent opportunity to develop an ML SR predictive model.

ML algorithms are a growing trend in the analysis of biomedical data through their ability to find complex patterns that typically appear in this field.⁵ This approach has been used to study several cardiac pathologies.^{6,7} In the case of SR, to our knowledge, the only study that has used a ML approach was developed by Cui et al.²⁵ In their study, a set of 6 plasma metabolites was identified to predict SR with an excellent AUC-ROC curve of 0.94. For their study, 2 patient

cohorts were used. The first cohort of 400 patients—in which 36 (9%) presented with SR—was used to develop the model identifying the 6 plasma metabolites. The investigators also used a cross-validation scheme in their methodology. The external validation of the prognostic value of these metabolites was performed in a second cohort of 500 patients, in which 48 (10%) presented with SR. We find their study of great relevance, as it was performed on a clinical cohort and had 86 patients with SR, which is far larger than the 23 subjects in the study presented here. Of interest, the percentage of positive class (SR) was similar in both studies; taking into account that, in the study by Cui et al., all patients underwent second-generation drug-eluting stent implantation, both models could be used irrespective of the type of stent use. Considering that phospholipids and sphingolipids metabolites are not usually measurable in daily cardiac practice, developing a model to predict SR based on variables obtained in routine clinical practice can be a valuable tool for improving the risk stratification of conventional patients, providing individually tailored follow-up and treatment.

ML models work efficiently when they are provided with large number of data and balanced classes. However, scenarios such as the one shown in this study, small datasets with imbalanced target classes are common in the clinical field, in which some pathologies are rare, and the data collection is costly. For this reason, we used a consistent methodology based on cross-validation to evaluate the generalization of our

Table 1. Values of true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), sensitivity, specificity, precision, and negative predictive value (NPV) with 3 different thresholds for the ERT model. Values of TP, TN, FP, FN are scaled for a population of 1000 patients based on the cross-validation performed on the GRACIA-3 study

Model - point	Sensitivity	Specificity	Precision	NPV	TP	TN	FP	FN
ERT - 1	0.18	0.98	0.49	0.93	16	896	16	72
ERT - 2	0.56	0.86	0.28	0.95	49	784	128	38
ERT - 3	0.81	0.53	0.14	0.97	70	487	426	17

ERT, extremely randomized trees

results.²¹ The use of cross-validation and steps such as the fine tuning of the algorithms, the data preprocessing, and, especially, the feature selection, must be handled with care, as they can lead to a leak of information from the test to the training sets and thus an overly optimistic outcome while trying to maximize a metric.

Finally, we agree with the prioritization of the PR curves for the assessment of prediction models in highly skewed datasets.^{18,19} The PR curves are dependent on the ratio between positive and negative classes, and their goal is to be in the upper-right zone. They are more sensitive than the ROC curves to changes in the zones of high specificity, and therefore they can be more intuitive to assess the capacity of the model to predict true positives. Importantly, the value of the AUC-PR for a random classifier is equivalent to the prevalence of the positive class (0.09 for our study), providing our developed ML model a good performance compared with chance alone.

Limitations

The principal limitations of this study are the relative low number of positive class (SR) and the lack of external validation. Although we used cross-validation methodology to address the first drawback, the generalization scores, which—on average—show promising results, have also deviations higher than those desired. In addition, an external validation would have added significant value to the model, but data collection was not possible for us, as clinical practice does not contemplate follow-up angiography at present. This is the main reason why we deployed our algorithm in an open source calculator as an example tool to determine generalizability of prediction models from small imbalanced sample size. Current PCI data, including second-generation drug-eluting stents and wider cohorts, may lead to additional conclusions and better generalization of the predictions.

Conclusions

Using ML, we developed an ETR model to predict SR based on variables obtained in routine clinical practice in a retrospective analysis of the GRACIA-3 trial. Compared with existing predictive multivariate logistic regression models, such as the PRESTO-1, PRESTO-2, and EVENT risk scores, our ETR model increased predictions.

Funding Sources

This study was supported by the Spanish Cardiovascular Network (CIBERCV) and the Fondo de Investigación Sanitaria (FIS: PI040308, PI040235, PI040361, PI042035, PI042686, PI040166, PI040276, PI040306, PI040309, PI040350, PI040478, and PI0402037), both from the Institute of Health Carlos III (Spanish Ministry of Science, Innovation and Universities) and by Philips Healthcare, Spain.

Disclosures

Dr Vicente-Palacios is an employee of Philips Healthcare. All other authors have no conflicts of interest to disclose.

Uncited Table

Table 1

References

- McLellan CS, Le May MR, Labinaz M. Current reperfusion strategies for ST elevation myocardial infarction: a Canadian perspective. *Can J Cardiol* 2004;20:525-33.
- Simard T, Hibbert B, Ramirez FD, Froeschl M, Chen YX, O'Brien ER. The evolution of coronary stents: a brief review. *Can J Cardiol* 2014;30:35-45.
- Califf RM, Fortin DF, Frid DJ, et al. Restenosis after coronary angioplasty: an overview. *J Am Coll Cardiol* 1991;17:2B-13B.
- Dangas GD, Claessen BE, Caixeta A, Sanidas EA, Mintz GS, Mehran R. In-stent restenosis in the drug-eluting stent era. *J Am Coll Cardiol* 2010;56:1897-907.
- Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019;380:1347-58.
- Westcott RJ, Tcheng JE. Artificial intelligence and machine learning in cardiology. *JACC Cardiovasc Interv* 2019;12:1312-4.
- Dorado-Diaz PI, Sampedro-Gomez J, Vicente-Palacios V, Sanchez PL. Applications of artificial intelligence in cardiology: the future is already here. *Rev Esp Cardiol* 2019;72:1065-75.
- Heckman GA, Hirdes JP, McKelvie RS. The role of physicians in the era of big data. *Can J Cardiol* 2019;36:19-21.
- Singh M, Gersh BJ, McClelland RL, et al. Clinical and angiographic predictors of restenosis after percutaneous coronary intervention: insights from the Prevention of Restenosis With Tranilast and Its Outcomes (PRESTO) trial. *Circulation* 2004;109:2727-31.
- 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.
- Kuntz RE, Baim DS. Defining coronary restenosis: newer clinical and angiographic paradigms. *Circulation* 1993;88:1310-23.
- Sanchez 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.
- Rodriguez JD, Perez A, Lozano JA. Sensitivity analysis of kappa-fold cross validation in prediction error estimation. *IEEE Trans Pattern Anal Mach Intell* 2010;32:569-75.
- Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825-30.
- Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 2006;7:91.
- Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Mach Learn Res* 2012;13:281-305.
- Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1-73.
- Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10:e0118432.

19. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning (ICML-06)* 2006;233-40.
20. Airola A, Pahikkala T, Waegeman W, De Baets B, Salakoski T. An experimental comparison of cross-validation techniques for estimating the area under the ROC curve. *Comput Stat Data Anal* 2011;55: 1828-44.
21. Bouckaert RR. Choosing between two learning algorithms based on calibrated tests. *Proceedings of the 20th International Conference on Machine Learning (ICML-03)* 2003:51-8.
22. Breiman L. Random forests. *Mach Learn* 2001;45:5-32.
23. Feinberg J, Nielsen EE, Greenhalgh J, et al. Drug-eluting stents versus bare-metal stents for acute coronary syndrome. *Cochrane Database Syst Rev* 2017;8:CD012481.
24. Mindrescu C, Brener SJ, Guerchicoff A, et al. Impact of scheduled angiographic follow-up in patients treated with primary percutaneous coronary intervention for ST-segment elevation myocardial infarction. *J Interv Cardiol* 2013;26:319-24.
25. Cui S, Li K, Ang L, et al. Plasma phospholipids and sphingolipids identify stent restenosis after percutaneous coronary intervention. *JACC Cardiovasc Interv* 2017;10:1307-16.

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.2020.01.027>.