Article:
Hadanny, A, Shouval, R, Wu, J orcid.org/0000-0001-6093-599X et al. (8 more authors) (2021) Predicting 30-day mortality after ST elevation myocardial infarction: Machine learning- based random forest and its external validation using two independent nationwide datasets. Journal of Cardiology, 78 (5). pp. 439-446. ISSN 0914-5087

https://doi.org/10.1016/j.jjcc.2021.06.002

Predicting 30-day mortality after ST Elevation Myocardial Infarction: machine learning- based random forest and its external validation using two independent nationwide datasets

Amir Hadanny (MD)[a], Roni Shouval MD, PhD[1], Jianhua Wu PhD[2], Nir Shlomo MSc[3], Ron Unger PhD[4], Doron Zahger MD[5], Shlomi Matetzky MD[6], Ilan Goldenberg MD[7], Roy Beigel MD[8], Chris Gale MD[9], Zaza Iakobishvili MD, PhD[10]

[Author: Please indicate affiliations using superscript a,b,c etc. Please list the author contributions separately]

[1] Sheba Medical Center affiliated with the Sackler School of Medicine, Tel-Aviv University, Ramat-Gan, Israel; The Mina and Everard Goodman Faculty of Life Sciences. Adult Bone Marrow Transplantation Service, Memorial Sloan Kettering Cancer Center, New York, NY, USA. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation.

[3] Leeds institute for Data Analytics, Univesity of Leeds, Leeds, United Kingdom. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation.

[3] Israeli Association for Cardiovascular Trials, Sheba Medical Center, Tel Hashomer, Israel. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation

[4] The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation

[5] Department of Cardiology, Soroka University Medical Center, Faculty of Health Sciences, Ben Gurion University of the Negev. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation

[6] The Heart Institute, Sheba Medical Center, Tel Hashomer, Sackler School of Medicine, Tel Aviv University, Israel. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation

[7] The Heart Institute, Sheba Medical Center, Tel Hashomer, Sackler School of Medicine, Tel Aviv University, Israel. Israeli Association for Cardiovascular Trials, Sheba Medical Center, Tel Hashomer, Israel. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation

[8] The Heart Institute, Sheba Medical Center, Tel Hashomer, Sackler School of Medicine, Tel Aviv University, Israel. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation

[9] Leeds institute of Cardiovascular and Metabolic Medicine, School of Medicine, University of Leeds, Leeds, United Kingdom. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation.

[10] Department of Cardiology, Tel Aviv Jaffa district, Clalit Health Services and Samson Assuta Ashdod University Hospital, affiliated to the Faculty of Health Sciences, Ben Gurion University of the Negev. This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation

[a] The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel. Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

*Corresponding author: Amir Hadanny, MD, Sagol Center, Shamir Medical Center, 70300, Israel

Tel: (972) 89779395; Fax: (972) 39542238

E-mail: amir.had@gmail.com

Keywords: ST-segment elevation myocardial infarction; Machine learning; Data mining; Outcome; Mortality

Abstract

Background: Various prognostic models for mortality prediction following ST-segment elevation myocardial infarction (STEMI) have been developed over the past two decades. Our group has previously demonstrated that machine learning (ML)-based models can outperform known risk scores for 30-day mortality post-STEMI. The study aimed to redevelop an ML-based random forest prediction model for 30-day mortality post-STEMI and externally validate it on a large cohort.

Methods: This was a retrospective, supervised learning, data mining study developed on the Acute Coronary Syndrome Israeli Survey (ACSIS) registry and the Myocardial Ischemia National Audit Project (MINAP) for external validation. Patients included received reperfusion therapy for STEMI between 2006 and 2016. Discrimination and calibration performances were assessed for two developed models and compared with the Global Registry of Acute Cardiac Events (GRACE) score.

Results: The ACSIS cohort (2,782 included /15,212 total) and MINAP cohort (22,693 included/735,000 total) were significantly different in most variables, yet similar in 30-day mortality rate (4.3-4.4%). Random forest models were developed on the ACSIS cohort with a full model including all 32 variables and a simple model including the 10 most important ones. Features' importance was calculated using the varImp function measuring how much each feature contributes to the data's homogeneity. Applying the optimized models on the MINAP validation cohort showed high discrimination of area under the curve (AUC)=0.804 (0.786-

0.822) for the full model, and AUC=0.787 (0.748-0.780) using the simple model, compared with the GRACE risk score discrimination of AUC=0.764 (0.748-0.780). All models were not well calibrated for the MINAP data. Following Platt scaling on 20% of the MINAP data, the random forest models calibration improved while the GRACE calibration did not change.

Conclusions: The random forest predictive model for 30-day mortality post STEMI, developed on the ACSIS national registry, has been validated in the MINAP large external cohort and can be applied early at admission for risk stratification. The model performed better than the commonly used GRACE score. Furthermore, to the best of our knowledge, this is the first externally validated ML-based model for STEMI.

1. Introduction

The treatment of acute ST-segment elevation myocardial infarction (STEMI) has undergone significant advances in recent years [1-5]. Rates of 30-day mortality following STEMI range between 2.7% and 8% [6]. Most patients benefit from early hospital discharge. However, some may suffer considerable morbidity and mortality [7-9].

Various prognostic models for mortality prediction following STEMI have been developed over the past two decades. The Thrombolysis in Myocardial Infarction (TIMI) risk score was originally focused on the 30-day mortality after thrombolysis [10]. Subsequently, it was validated for STEMI patients treated by primary PCI [11]. The Zwolle score was created to predict 30-day mortality to identify low-risk patients suitable for an early discharge from hospital [12]. The Controlled Abciximab and Device Investigation to Lower Late Angioplasty Complications (CADILLAC) risk score is used to predict one-year mortality [13]. The Evaluation of Methods of Management of Acute Coronary Events (EMMACE) was developed as a simple model for 30-day mortality [14]. The Global Registry of Acute Cardiac Events (GRACE) score, which is based on a multinational acute coronary syndrome (ACS) registry has been the most widely used [10,15-17] and recommended for routine use in the European Society of Cardiology STEMI guidelines [18]. Such scores, including GRACE and TIMI, are based on a conventional statistical approach that carries inherent limitations, including fixed assumptions on data behavior and the need for variables selection [19,20].

Studying the Acute Coronary Syndrome Israeli Survey (ACSIS) registry, our group has demonstrated that machine learning (ML)-based models performed similarly to the GRACE

score and outperformed the TIMI score in predicting 30-day mortality post STEMI [21]. Among the algorithms studied, random forest (RF) had the highest discrimination [area under the receiver operating characteristic curve (AUC) = 0.91]. In the current study, we sought to redevelop an RF-based model for prediction of 30-day mortality post-STEMI on a cohort from ACSIS and to externally validate it on the British national registry [Myocardial Ischemia National Audit Project (MINAP)].

2. Methods

2.1 Study population

This was a retrospective, supervised learning (i.e. a ML task of inferring a function from labeled training data), data mining study based on the ACSIS registry and the MINAP.

ACSIS is a biennial prospective observational national survey of all patients with acute coronary syndrome (ACS) hospitalized in Israel since 2000 [22,23]. In every survey during a two-month period, data are prospectively collected from all ACS admissions in each of the 25 cardiology wards operating in Israel [22-25]. All patients with STEMI are included in the survey and followed for one month. The number of patients who refused to participate was small (20 individuals in the 2006-2016 period, corresponding to 0.1% of the cohort). The ACSIS was approved by each medical center's institutional review board. Clinical data were recorded on pre-specified forms for all patients. Case report form definitions were centrally determined. The attending physicians determined admission and discharge diagnosis based on clinical, electrocardiographic, and biochemical criteria. Patients' management was at the discretion of the attending physicians.

The MINAP is a national registry of patients admitted to hospitals in England and Wales with ACS [26]. Data collection began in October 2000. MINAP accrues approximately 85,000

episodes of care per year for patients with ACS admitted to all acute National Health Service (NHS) hospitals in England and Wales. The MINAP dataset comprises 123 separate fields under the following groups: patient demographics, admission method, clinical features and investigations, medical history, drug treatment before admission, detail of primary reperfusion treatment, drug treatment in hospital, clinical complications, interventional treatments, hospital outcome, discharge diagnosis, and discharge (secondary prevention) treatment. The dataset is revised every two years to meet the requirements of users and to respond to developments in management of ACS. Care is taken to maintain continuity of fields when new options are added, and redundant fields are archived. Both national cohorts are audited regularly by designated data committees. Both national registries datasets were pulled and underwent preprocessing.

All patients (or surrogates) gave written informed consent for inclusion in the registries. The study follows the TRIPOD (Transparent Reporting of a Multivariable prediction model for Individual Prognosis or Diagnosis) recommendations for the reporting of studies prediction scores [27].

2.2 Study Objectives

The primary objective was to develop a RF-based model for 30-day mortality post-STEMI on the ACSIS registry and validate it on a MINAP cohort. The secondary objective was to compare the newly developed model with the GRACE risk score.

2.3 Study Design

Missing values on both datasets were replaced using the mean and mode values of the ACSIS dataset for numeric and nominal variables, respectively. Prediction models for overall mortality 30 days after STEMI were developed on the ACSIS dataset using RF algorithm (parameters detailed in Table 1) on all 32 variables, for the full model, and an additional RF algorithm

developed on the 10 most important variables, for the simplified model. Models were redeveloped in order to fit MINAP available variables. The predictive performance of the two RF models was evaluated in the MINAP dataset.

2.4 Model development

The current analysis included patients from both datasets (ACSIS and MINAP) who received reperfusion (fibrinolysis, primary PCI, angiography demonstrating spontaneous reperfusion), for STEMI between 2006 and 2016. STEMI was defined as typical symptoms associated with ST segments elevation >2 mm in V2-V3 or 1 mm in all other leads or presumably new left bundle branch block present on the admission electrocardiogram (ECG). Thirty-two variables available on both ACSIS and MINAP databases describing procedural (e.g. reperfusion) and pre-procedural information (e.g. demographics, prior medical history, prior chronic drugs, clinical presentation, basic laboratory data with admission, and angiography) were evaluated. Variables with more than 15% missing values were not included in model development, similarly to our model developed previously on the ACSIS with the same methodology [21]. In order to develop a second, simpler model, the ten most important variables in the full model were included, as known cardiology risk scores include 7-10 features [10, 28]. The 10 most important variables were selected by a feature selection algorithm (using the VIMP package, see Fig. 1). By changing the variables and calculating the difference in prediction error with each change, the variables are ranked according to their impact on prediction performance [29].

The 30-day mortality time frame in ACSIS and MINAP was defined as 30 days from admission day.

RF is a learning method consisting of many individual decision trees that operate as an ensemble. In a classification task, each tree in the RF produces a class prediction and the class with the most

votes (mode) becomes the overall model prediction [30]. Each decision tree node uses a subset of features randomly selected from the entire list of features. The number of trees within the forest are increased aiming to seek out for the number from which there is no more significant performance gain. The number of random features to consider in each tree is called mtrys. Lower values of mtry lead to more different, less correlated trees, yielding better stability. In addition, low mtry also tends to better exploit variables with moderate effect on the response variable. However, lower values of mtry also lead to trees that perform on average worse, since they are built based on suboptimal variables. Lastly, the minimal size of terminal nodes (number of patients classified at that node) is called node size. Setting this number larger causes smaller trees to be grown. All the models' hyperparameters (mtrys, nodesize, number of trees) were tuned using the grid search method, in which all possible combinations were evaluated (Table 2).

2.5 Performance and calibration

The TRIPOD statement [27] provides reporting recommendations for articles that describe the development and external validation of prediction models, aiming to enhance reporting transparency and hence interpretability, reproducibility, and clinical usability of these models. Thus, both performance (discrimination) and calibration analysis were performed.

Discrimination is the ability of a predictive model to separate data into classes (dead/alive). Calibration is a measure of the closeness of model probability to the underlying probability of the population under study [31].

The RF models' performance was evaluated using the means of the area under the receiver operating characteristic (ROC) curve (AUC).

Thus, calibration plots were evaluated for all models. Calibration refers to the agreement between observed outcomes and predictions [32]. The predictions binned to deciles are plotted on x-axis

and the outcome on the y-axis, enabling visual assessment. Perfect predictions would be on the 45 degrees line [33].

Calibration is important, as the model should communicate an accurate absolute risk to physicians. A predicted risk that is too high or too low may result in incorrect treatment decisions. Thus, in case of non-calibrated predictions on the new external data, each model predictions were re-calibrated, i.e. rescaled so their values will better match the distribution of the external data. The MINAP dataset was split to 80% validation set and 20% calibration set. Both RF and GRACE models' calibrations were re-calibrated on 20% of the MINAP data using the Platt scaling method [31]. In short, model's predictions on the 20% calibration data trained a logistic regression classifier with the outcomes on those data. The model's predictions on the 80% MINAP data were transformed with the new classifier to produce new calibrated predictions. The goal of recalibration is to produce accurate predictions suitable for the external dataset (rather than the trained dataset alone). Notably, re-calibration improves calibration but not discrimination performance.

As single imputation may underestimate the population variability and depend on specific assumptions (mean, mode), multiple imputation confirmation was used. Using multiple imputation with the Multiple Imputation by Chained Equations (MICE) package [34], we performed sensitivity analysis on 10 imputed datasets for both full and simple models. The MICE algorithm is broken down into four general steps [34]:

1. A simple imputation, such as imputing the mean, is performed for every missing value in the dataset. These mean imputations can be thought of as “place holders.”
2. The “place holder” mean imputations for one variable (“X”) are reversed back to missing.

3. The observed values from the variable "X" are regressed on the other variables in the imputation model, which may or may not consist of all the variables in the dataset, i.e. that specific feature serves as the dependent variable and all the other variables are independent variables in the regression model
4. The missing values for variable "X" are then replaced with predictions (imputations) from the regression model.

Steps 2–4 are then repeated for each variable that has missing data. At the end of one cycle all the missing values have been replaced with predictions from regressions that reflect the relationships observed in the data.

2.5 Comparison to GRACE score

The GRACE score was calculated, as described previously [10,15] on the entire MINAP dataset without missing values, and its performance (AUC) and calibration were compared with the developed RF models [35,36]. DeLong test was performed to compare models' performance [37].

Sensitivity analysis for the GRACE score was performed similarly on the 10 imputed datasets.

2.6 Statistical analysis

The normal distribution for all variables was tested using the Kolmogorov-Smirnov test. Results were expressed as mean and standard-deviation (SD) for continuous variables and as frequencies/percentages for categorical variables. Univariate analysis was performed using chi-square test to identify significant variables and two-sided independent Student t-test (p<0.05). Confidence intervals of 95% are shown in brackets [].

The analysis was performed using R (v.3.6.1; R Foundation, Vienna, Austria) with packages mlr, imputeMissings, RF, ROCR, vimp, rms, MICE.

3. Results

3.1 Patient characteristics

Of the 15,212 patients in the ACSIS registry, 2,782 fulfilled inclusion criteria (out of 3941 STEMI patients between 2006 and 2016), while in the validation MINAP registry, 22,963 out of 213,880 STEMI patients from the corresponding time frame underwent reperfusion and had 30-day mortality data. The time frame was selected to represent current treatment guidelines and modalities. The two cohorts were significantly different in most variables, yet similar in 30-day mortality rate (4.3-4.4%). Table 1 compares the characteristics of the two cohorts. In general, the validation cohort patients were significantly older, had fewer cardiovascular risk factors, previously underwent less revascularization, less history of myocardial infarction or stroke, and had lower glucose, creatinine, and hemoglobin levels at admission.

3.1 Models development

The full RF model included all variables on the imputed and processed data. The model hyperparameters were tuned using the grid search method until optimized (Table 2).

Variable importance for predictions was evaluated for construction of the simple model (Fig. 1). Creatinine level at admission was the most important variable. In addition to the Killip class at admission, all the continuous parameters were included in the simple model: age, creatinine, heart rate, mean arterial pressure, hemoglobin, glucose, total cholesterol, body mass index, and time from onset of symptoms to PCI. The simple model hyperparameters were tuned using the grid search method (Table 2).

3.2 External validation - Discrimination

The RF models were developed and optimized on the entire ACSIS cohort. Using all variables, the full model had high predictive discrimination of AUC=0.908 (0.889-0.931). Including only the 10 most important variables, the simple model provided similar high discrimination of 0.897 (0.871-0.918) on the ACSIS data.

Applying the optimized models on the MINAP validation cohort showed high discrimination of AUC=0.804 (0.786-0.822) for the full model, and AUC=0.787 (0.771-0.804) using the simple model (Table 2, Fig. 3A[Author: Please cite Figure 2 before Figure 3 or renumber]). The GRACE risk score discrimination with AUC=0.764 (0.748-0.780) was significantly lower than the full model (p=0.008) and the difference from the simple model was not statistically significant (p=0.085) (Table 2, Fig. 4A). Sensitivity analysis demonstrated similar discrimination values on 10 imputed datasets for both full, simple, and GRACE models (see Online Table S1).

3.3 External validation - Calibration

The calibration plots of both full and simple models revealed poor calibration, especially in the high-risk probabilities which were overestimated (Fig. 2B, C). The GRACE risk score showed similarly poor calibration in the high-risk probabilities and goodness of fit (p<0.001, Fig. 4B). Sensitivity analysis demonstrated similar calibration curves for both full, simple, and GRACE models on the 10 imputed datasets (see Online Figs S1,S2, S3).

Due to the poor calibration of all models on the MINAP data, recalibration was performed. The goal of recalibration was to rescale predictions to fit the external dataset. Using a 20% random sample of the MINAP validation set, all models were re-calibrated using Platt's scaling. Tested

on the remaining 80% of the MINAP cohort, both RF models showed better calibration (Fig. 3). However, goodness of fit remained poor (p<0.001). Re-calibration of the GRACE risk score did not change the calibration plot (Fig. 4C).

Sensitivity analysis demonstrated similar re-calibration curves for both full, simple, and GRACE models on the 10 imputed datasets (see Online Figs S4,S5, S6).

4. Discussion

In the current study, we have validated the ML-based RF prediction models for 30-day mortality post-STEMI. Both the full model, including 32 variables, and the simple model, which includes 10 variables, showed high discrimination on an external validation set, comprised a large registry. The full model outperformed the GRACE risk score which was developed using the conventional statistical approach. However, calibration of the models was poor and required re-calibration using a subsample of the external data. Recalibration has provided improved calibration plots in the RF models. Following recalibration, models enable physicians much more accurate predictions (the chance of a STEMI patient to die within 30 days). Sensitivity analysis using multiple imputation confirmed our findings. From a clinical perspective, the RF simple model could be used upon admission of a STEMI patient for risk stratification. From a methodological perspective, we reconfirm the value of applying ML predictions, which can be easily incorporated with the electronic medical records' digitalization.

We have developed several ML-based models for 30-day mortality prediction in our previous work, which performed similarly to the GRACE risk score, and significantly better than the TIMI risk score [21]. The RF model had the highest discrimination, which improved as more

features were included in the model. In the current study, we confirmed the validity of this model for real-world data using the MINAP dataset. As before, the full model performed better than the selected features simple model.

Cardiology literature offers a plethora of proposed models, with little evidence about which are reliable and under what circumstances. The majority of models are developed using classic statistical methods (such as Cox regressions / logistic regressions), without further validation or with a validation study on a small and specific sample (100-1000 patients [13,38,39]). In the past few years, ML-based models have originated however most did not get externally validated [40,41]. External validation uses new patient data, external to those used for model development, to examine whether the model's predictions are reliable in individuals for clinical use. The current study uses one of the largest cohorts in the setting of ACS for external validation. Moreover, the British population's external validation was significantly different to the model's development cohort (ACSIS) of the Israeli population, reaffirming the necessity of these studies as well as the validity of the models.

The GRACE risk score is a known validated tool for risk stratification, developed on one of the largest databases in the field of ACS of over 240 hospitals in 30 countries. The GRACE score was originally developed for predicting mortality at six months for the entire spectrum of ACS [10,15-17] and later validated for 30-day mortality. Mendez-Eirin et al. found a discrimination c-statistic of 0.9 for 30-day mortality in 1500 STEMI patients [42]. Abelin et al. found a c-statistic of 0.84 in a 501-patient cohort [43], and recently Hizoh et al. showed a c-statistic of 0.87 in 5203 STEMI patients [44]. Interestingly, the GRACE score performed significantly worse in the MINAP validation data. In a previous study assessing different risk scores on previous MINAP data (years 2003-2005), the discrimination performance for 30-day mortality was 0.78 [45]. The

GRACE risk score includes 8 predictors: age, medical history, vital signs at admission, creatinine, cardiac markers, and ECG finding. The RF full model includes all of these and the simple model includes 5 out of them. Both RF models outperformed the GRACE risk score (0.78-0.80).

In a recent study on a large prospective cohort, a different ML-based model utilizing deep learning has shown superior discrimination performance over the GRACE model in post STEMI mortality prediction [46]. However, this study has not been validated externally and did not report the model calibration.

Both RF models and the GRACE risk score had poor calibration on the MINAP cohort, especially overestimating high-risk patients. RF models have an inherent difficulty making predictions that are near to 0 and 1,and are biased away from these values. Re-calibration usually mitigates this known problem [47]. Indeed, following re-calibration, the RF models' calibrations improved while we observed no change for the GRACE score. The validated and calibrated simple model was saved and developed as an interactive online calculator app using the Shiny package.

In addition to the clinical value of the validated model, the current study has several methodological implications. First, it reconfirms the feasibility of a data mining approach in predicting outcomes in STEMI patients and cardiology in general. Second, as opposed to Cox/logistic regression-based models, non-parametric ML-based models such as RF can perform better with a higher number of features. With the growing development of electronic medical records, where information on ACS patients is accumulating in unprecedented quantities, models accounting for the very high amount of features are warranted [48].

Contemporary focus in STEMI is reducing the extent of injury in the reperfused STEMI beyond timely reperfusion. The ischemic/reperfusion syndrome following STEMI reperfusion can induce significant morbidity and mortality. There is a great need to select patients who most likely will benefit from cardioprotective interventions [49]. Therefore, our models included patients who underwent reperfusion therapy to represent standard of care. The present study has several limitations. First, it is a retrospective analysis susceptible to data selection and measurement biases. Yet, the large MINAP registry analyzed reflects prospectively collected, real-world data, hence conveying more recent practice [48]. Second, we focus on a short-term outcome, i.e. 30-day mortality. Third, we have adopted a simplistic approach for missing data imputation, subject to potential bias [50]. However, missing values were imputed as constants based on the ACSIS cohort which reduces this bias. In addition, we used multiple imputation for sensitivity analysis to confirm our models and findings. Fourth, variables with a fraction of missing values higher than 15% were dropped, similarly to our original model, to avoid bias in a relatively small data set (~2700) [21]. Notably, cases were not dropped due to missing values. Lastly, the type of reperfusion was not included in the model due to missing values.

Clinical implications

Once classified as having STEMI, the contemporary patient management follows an algorithmic approach without a pre-reperfusion risk assessment. Ideally every eligible STEMI patient should undergo reperfusion therapy as soon as possible with preference to primary PCI. However, at the early stages of ICU stay, risk assessment scores proved to be useful for prediction of ICU length of stay, early discharge safe planning, effective secondary prevention medications prescription, and referral to cardiac rehabilitation (where higher-risk patients can gain significantly more than

lower-risk patients) [46,51-53]. RF score which performs better than the guideline-recommended GRACE score in our derivation and external validation cohorts even with different baseline characteristics may serve as a more precise tool for risk prediction. Although our full model has a significant improvement over the GRACE model, further prospective validation and additional external validations should be performed.

5. Conclusions

The RF predictive model for 30-day mortality post STEMI, developed on the ACSIS national registry, has been validated in the MINAP large external cohort. The model can be applied early at admission for risk stratification. The model performed better than the commonly used GRACE score. Additionally, to the best of our knowledge, this is the first externally validated ML-based model in ACS.

The simple model is available at: https://amirhad.shinyapps.io/30-Days-Mortality-Post-STEMI-Calculator/

5. References

[Author: Please list up to 6 authors before et al.]

1. Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. *Lancet* 2003;361:13-20.
2. Montalescot G, Antoniucci D, Kastrati A, et al. Abciximab in primary coronary stenting of ST-elevation myocardial infarction: a European meta-analysis on individual patients' data with long-term follow-up. *Eur Heart J* 2007;28:443-9.
3. Chua D, Lo C, Babor EM. Addition of clopidogrel to aspirin and fibrinolytic therapy for myocardial infarction. *N Engl J Med* 2005;352:2647-8.
4. Vlaar PJ, Svilaas T, van der Horst IC, et al. Cardiac death and reinfarction after 1 year in the Thrombus Aspiration during Percutaneous coronary intervention in Acute myocardial infarction Study (TAPAS): a 1-year follow-up study. *Lancet* 2008;371:1915-20.
5. Parodi G, Bellandi B, Valenti R, et al. Comparison of double (360 mg) ticagrelor loading dose with standard (60 mg) prasugrel loading dose in ST-elevation myocardial infarction patients: the Rapid Activity of Platelet Inhibitor Drugs (RAPID) primary PCI 2 study. *Am Heart J* 2014;167:909-14.
6. Pedersen F, Butrymovich V, Kelbaek H, et al. Short- and long-term cause of death in patients treated with primary PCI for STEMI. *J Am Coll Cardiol* 2014;64:2101-8.
7. Hasdai D, Topol EJ, Califf RM, Berger PB, Holmes DR, Jr. Cardiogenic shock complicating acute coronary syndromes. *Lancet* 2000;356:749-56.
8. Goldberg RJ, Spencer FA, Gore JM, Lessard D, Yarzebski J. Thirty-year trends (1975 to 2005) in the magnitude of, management of, and hospital death rates associated with cardiogenic shock in patients with acute myocardial infarction: a population-based perspective. *Circulation* 2009;119:1211-9.
9. Yeh RW, Sidney S, Chandra M, Sorel M, Selby JV, Go AS. Population trends in the incidence and outcomes of acute myocardial infarction. *N Engl J Med* 2010;362:2155-65.
10. Morrow DA, Antman EM, Charlesworth A, et al. TIMI risk score for ST-elevation myocardial infarction: A convenient, bedside, clinical score for risk assessment at presentation: An intravenous nPA for treatment of infarcting myocardium early II trial substudy. *Circulation* 2000;102:2031-7.
11. Morrow DA, Antman EM, Parsons L, et al. Application of the TIMI risk score for ST-elevation MI in the National Registry of Myocardial Infarction 3. *JAMA* 2001;286:1356-9.
12. De Luca G, Suryapranata H, van 't Hof AW, et al. Prognostic assessment of patients with acute myocardial infarction treated with primary angioplasty: implications for early discharge. *Circulation* 2004;109:2737-43.
13. Halkin A, Singh M, Nikolsky E, et al. Prediction of mortality after primary percutaneous coronary intervention for acute myocardial infarction: the CADILLAC risk score. *J Am Coll Cardiol* 2005;45:1397-405.
14. Dorsch MF, Lawrance RA, Sapsford RJ, et al. A simple benchmark for evaluating quality of care of patients following acute myocardial infarction. *Heart* 2001;86:150-4.
15. Fox KA, Dabbous OH, Goldberg RJ, et al. Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE). *BMJ* 2006;333:1091.

16. Addala S, Grines CL, Dixon SR, et al. Predicting mortality in patients with ST-elevation myocardial infarction treated with primary percutaneous coronary intervention (PAMI risk score). *Am J Cardiol* 2004;93:629-32.
17. Granger CB, Goldberg RJ, Dabbous O, et al. Predictors of hospital mortality in the global registry of acute coronary events. *Arch Intern Med* 2003;163:2345-53.
18. Ibanez B, James S, Agewall S, et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. *Rev Esp Cardiol (Engl Ed)* 2017;70:1082.
19. Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Stat Sci* 2001;16:199-231.
20. Shouval R, Bondi O, Mishan H, Shimoni A, Unger R, Nagler A. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. *Bone Marrow Transplant* 2014;49:332-7.
21. Shouval R, Hadanny A, Shlomo N, et al. Machine learning for prediction of 30-day mortality after ST elevation myocardial infraction: An Acute Coronary Syndrome Israeli Survey data mining study. *Int J Cardiol* 2017;246:7-13.
22. Arbel Y, Matetzky S, Gavrielov-Yusim N, et al. Temporal trends in all-cause mortality of smokers versus non-smokers hospitalized with ST-segment elevation myocardial infarction. *Int J Cardiol* 2014;176:171-6.
23. Kornowski R. The ACSIS Registry and primary angioplasty following coronary bypass surgery. *Catheter Cardiovasc Interv* 2011;78:537-9.
24. Segev A, Matetzky S, Danenberg H, et al. Contemporary use and outcome of percutaneous coronary interventions in patients with acute coronary syndromes: insights from the 2010 ACSIS and ACSIS-PCI surveys. *EuroIntervention* 2012;8:465-9.
25. Gottlieb S, Behar S, Hod H, et al. Trends in management, hospital and long-term outcomes of elderly patients with acute myocardial infarction. *Am J Med* 2007;120:90-7.
26. Wilkinson C, Weston C, Timmis A, Quinn T, Keys A, Gale CP. The Myocardial Ischaemia National Audit Project (MINAP). *Eur Heart J Qual Care Clin Outcomes* 2020;6:19-22.
27. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55-63.
28. Ray J. The TIMI risk score predicted mortality in patients with ST-elevation myocardial infarction. *ACP J Club* 2002;136:76.
29. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. *BMC Bioinformatics* 2008;9:307.
30. Ishwaran H. The effect of splitting on random forests. *Mach Learn* 2015;99:75-118.
31. Walsh CG, Sharman K, Hripcsak G. Beyond discrimination: A comparison of calibration methods and clinical usefulness of predictive models of readmission risk. *J Biomed Inform* 2017;76:9-18.
32. Rossello X, Dorresteijn JA, Janssen A, et al. Risk prediction tools in cardiovascular disease prevention: A report from the ESC Prevention of CVD Programme led by the European Association of Preventive Cardiology (EAPC) in collaboration with the Acute Cardiovascular Care Association (ACCA) and the Association of Cardiovascular Nursing and Allied Professions (ACNAP). *Eur J Cardiovasc Nurs* 2019;18:534-44.
33. 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.
34. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* 2011;20:40-9.

35. Ian H. Witten EF, Mark A. HAll. Data mining - Practical machine learning tools and techniques 3rd edition. 2011.[Author: Please include full details - See Instructions for Authors]
36. Bengio Y, Nadeau C. *Inference for the generalization error.* CIRANO;1999. [Author: Please include full details - See Instructions for Authors]
37. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-45.
38. Timoteo AT, Aguiar Rosa S, Nogueira MA, Belo A, Cruz Ferreira R. External validation of the ProACS score for risk stratification of patients with acute coronary syndromes. *Rev Port Cardiol* 2016;35:323-8.
39. Cambou JP, Simon T, Mulak G, Bataille V, Danchin N. The French registry of Acute ST elevation or non-ST-elevation Myocardial Infarction (FAST-MI): study design and baseline characteristics. *Arch Mal Coeur Vaiss* 2007;100:524-34.
40. Sanchez-Cabo F, Rossello X, Fuster V, et al. Machine learning improves cardiovascular risk definition for young, asymptomatic individuals. *J Am Coll Cardiol* 2020;76:1674-85.
41. Al-Zaiti S, Besomi L, Bouzid Z, et al. Machine learning-based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. *Nat Commun* 2020;11:3966.
42. Mendez-Eirin E, Flores-Rios X, Garcia-Lopez F, et al. Comparison of the prognostic predictive value of the TIMI, PAMI, CADILLAC, and GRACE risk scores in STEACS undergoing primary or rescue PCI. *Rev Esp Cardiol* 2012;65:227-33.
43. Abelin AP, David RB, Gottschall CA, Quadros AS. Accuracy of dedicated risk scores in patients undergoing primary percutaneous coronary intervention in daily clinical practice. *Can J Cardiol* 2014;30:125-31.
44. Hizoh I, Domokos D, Banhegyi G, Becker D, Merkely B, Ruzsa Z. Mortality prediction algorithms for patients undergoing primary percutaneous coronary intervention. *J Thorac Dis* 2020;12:1706-20.
45. Gale CP, Manda SO, Weston CF, Birkhead JS, Batin PD, Hall AS. Evaluation of risk scores for risk stratification of acute coronary syndromes in the Myocardial Infarction National Audit Project (MINAP) database. *Heart* 2009;95:221-7.
46. Kwon JM, Jeon KH, Kim HM, et al. Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction. *PLoS One* 2019;14:e0224502.
47. Dankowski T, Ziegler A. Calibrating random forests for probability estimation. *Stat Med* 2016;35:3949-60.
48. Lauer MS, D'Agostino RB, Sr. The randomized registry trial--the next disruptive technology in clinical research? *N Engl J Med* 2013;369:1579-81.
49. Rossello X, Lobo-Gonzalez M, Ibanez B. Editor's Choice- Pathophysiology and therapy of myocardial ischaemia/reperfusion syndrome. *Eur Heart J Acute Cardiovasc Care* 2019;8:443-56.
50. 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.
51. Shavadia JS, Chen AY, Fanaroff AC, de Lemos JA, Kontos MC, Wang TY. Intensive care utilization in stable patients with ST-segment elevation myocardial infarction treated with rapid reperfusion. *JACC Cardiovasc Interv* 2019;12:709-17.
52. Brogan RA, Malkin CJ, Batin PD, Simms AD, McLenachan JM, Gale CP. Risk stratification for ST segment elevation myocardial infarction in the era of primary percutaneous coronary intervention. *World J Cardiol* 2014;6:865-73.
53. Gibson WJ, Nafee T, Travis R, et al. Machine learning versus traditional risk stratification methods in acute coronary syndrome: a pooled randomized clinical trial analysis. *J Thromb Thrombolysis* 2020;49:1-9.

Acknowledgments

We would like to thank Mr Raz Tuval for proofreading the manuscript.

Source of funding

The authors had no funding source or grant for this study.

Disclosures

The authors had no conflict of interests

Figure Legends

Figure 1 Variable importance grading for RF model. Variable importance is the difference between prediction error when feature values are altered randomly (adding noise) to the prediction error when using its original values. The higher the difference, the importance of the feature is higher. The 10 features with the highest feature importance value were selected for the simple models.
[Author: Please define all abbreviations]
Figure 2 Full and simple models' performance and calibration on the MINAP data. (A) ROC curves. (B) Full model calibration plot on the MINAP data. (C) Simple model calibration plot on the MINAP data.
[Author: Please define all abbreviations]

Figure 3 Full and simple re-calibration plots. Following calibration on 20% of the MINAP data, calibration plots on the other 80% of the MINAP data.
[Author: Please define all abbreviations]

Figure 4 GRACE risk score predictive performance and calibration. (A) ROC curve, AUC=0.76 on both 80% and 100% of the MINAP data. (B) Calibration plot of GRACE on the entire MINAP data. (C) Re-calibration plot of the GRACE risk score on 80% of MINAP following calibration on the remainder 20%.[Author: Please define all abbreviations]

Tables

Table 1 Comparison of ACSIS and MINAP cohorts

	ACSIS	Imputed	MINAP	Imputed	Sig.
N	2476		22963		
Age (years)	60.8±12.7	0.0%	63.5±13.1	0.0%	<0.001
Males	82.30%	0.0%	73.1%	0.0%	<0.001
Heart Rate (bmp)	78.4±19.7	1.4%	76.7 ±18.3	8.2%	<0.001
MAP (mmHg)	98.9±20.5	1.2%	100.4±19.1	8.2%	<0.001
BMI	27.5±4.3	8.9%	27.5±4.2	12.5%	0.5
Total Cholesterol (mg/dL)	179.2±41.9	10.0%	192.9±47.2	15.8%	<0.001
Glucose (mmol/L)	156.4±82.1	4.6%	150.9±70.0	8.7%	0.001
Hemoglobin (g/dL)	14.4±6.6	1.4%	13.7±1.8	7.2%	0.001

Creatinine (mg/dL)		1.11±0.77	2.9%	1.01±0.47	10.9%	<0.001
Time from Onset to PCI (minutes)		292.8±401	13.7%	308.5±440.8	8.5%	0.033
Past MI		526 (21.2%)	0.2%	2417 (10.5%)	0.2%	<0.001
Chronic heart failure		68 (2.7%)	0.2%	298 (1.3%)	0.3%	<0.001
Past Revascularization		560 (22.6%)	0.0%	1890 (8.2%)	0.2%	<0.001
Past Stroke		148 (6%)	0.1%	1039 (4.5%)	0.2%	0.002
Family history of MI		730 (29.5%)	8.1%	7626 (33.2%)	7.7%	0.007
Chronic renal failure		135 (5.5%)	0.2%	462 (2.0%)	0.1%	<0.001
PVD		125 (5%)	0.2%	653 (2.8%)	0.3%	<0.001
Diabetes Mellitus		712 (28.8%)	0.3%	3129 (13.6%)	0.4%	<0.001
Hypertension		1273 (51.4%)	0.5%	9538 (41.5%)	0.4%	<0.001
Killip class at admission			1.3%		0.0%	0.098
	I	2166 (87.5%)		19767 (86.1%)		
	II	162 (6.5%)		1772 (7.7%)		
	III	68 (2.7%)		732 (3.2%)		
	IV	80 (3.2%)		692 (3.0%)		
Chronic beta blockers		559 (22.6%)	2.9%	4589 (20.0%)	0.0%	0.684
Chronic antiplatelet drugs		881 (35.6%)	0.0%	2681 (11.7%)	0.0%	<0.001
Any smoking		1680 (67.8%)	0.1%	15325 (66.7%)	0.1%	0.631
Chronic ACE-I/ARBB		697 (28.2%)	0.2%	6340 (27.6%)	1.7%	0.683
Typical angina presenting symptom		2191 (88.5%)	0.0%	21962 (95.6%)	0.0%	0.012

Normal Sinus Rhythm	2133 (86.1%)	0.0%	19704 (85.8%)	0.0%	0.911
CPR/DCS	129 (5.2%)	17.8%	859 (3.7%)	0.4%	0.001
Aspirin before admission	2088 (84.3%)	0.0%	17273 (75.2%)	0.0%	<0.001
Clopidogrel before admission	411 (11.6%)	0.0%	3406 (14.8%)	0.9%	0.049
Chronic Hyperlipidemia drugs	910 (36.8%)	0.0%	6950 (30.3%)	0.0%	<0.001
Anterior MI	1169 (46.9%)	0.0%	8663 (37.7%)	5.7%	<0.001
30-day mortality	110 (4.4%)	0.0%	989 (4.3%)	0.0%	0.803

MAP, mean arterial pressure; PCI, percutaneous coronary intervention; MI, myocardial infarction; PVD, peripheral vascular disease; ACE-I, angiotensin-converting enzyme inhibitors; ARBB, angiotensin II receptor blockers; CPR/DCS, resuscitation or electric shock. [Author: Please define all abbreviations]

Table 2 Models' parameters and predictive performance on both ACSIS and MINAP cohorts.

	mtry	nodesize	trees	ACSIS AUC	MINAP AUC	Significance
1.RandomForest	3	4	1000	0.908 [0.889-0.931]	0.804 [0.786-0.822]	0.008
2.RandomForest-Simple	1	9	1000	0.897 [0.871-0.918]	0.787 [0.771-0.804]	0.085
3.GRACE	-	-	-	0.881 [0.872-0.899]	0.764 [0.748-0.780]	Reference

Confidence intervals of 95% are shown in brackets [].

[Author: Please define all abbreviations]