Machine Learning Based Risk Prediction for Major Adverse Cardiovascular Events

Michael SCHREMPF[a,1], Diether KRAMER[a], Stefanie JAUK[a,b], Sai P. K. VEERANKI[a], Werner LEODOLTER[a] and Peter P. RAINER[b]

[a] Steiermärkische Krankenanstaltengesellschaft m. b. H., Graz, Austria
[b] Medical University of Graz, Graz, Austria

Abstract. Background: Patients with major adverse cardiovascular events (MACE) such as myocardial infarction or stroke suffer from frequent hospitalizations and have high mortality rates. By identifying patients at risk at an early stage, MACE can be prevented with the right interventions. Objectives: The aim of this study was to develop machine learning-based models for the 5-year risk prediction of MACE. Methods: The data used for modelling included electronic medical records of more than 128,000 patients including 29,262 patients with MACE. A feature selection based on filter and embedded methods resulted in 826 features for modelling. Different machine learning methods were used for modelling on the training data. Results: A random forest model achieved the best calibration and discriminative performance on a separate test data set with an AUROC of 0.88. Conclusion: The developed risk prediction models achieved an excellent performance in the test data. Future research is needed to determine the performance of these models and their clinical benefit in prospective settings.

Keywords. Major Adverse Cardiovascular Events, Cardiovascular Diseases, Myocardial Infarction, Stroke, Machine Learning, Cardiovascular Risk, Electronic Medical Records, Risk Assessment

1. Introduction

1.1. Problem Statement

Cardiovascular diseases are one of the major death causes. In Austria, almost 40% of deaths are due to major adverse cardiovascular event (MACE) such as myocardial infarction or stroke [1]. During their lifetime, patients with cardiovascular diseases suffer from frequent hospitalizations and treatments, and their outcome depends often on 24-hour care after a severe cardiovascular event [2]. By identifying the risk at an early stage, many of these cases could have been prevented [3].

In many cases, a MACE is caused by advanced atherosclerotic vascular diseases. Atherosclerosis is strongly influenced by smoking, poor diet and lack of exercise. Since atherosclerotic vascular diseases develop over a long period of time, an early risk identification and preventive actions can decrease the risk of MACE [3].

---

[1] Corresponding Author: Michael Schrempf, Steiermärkische Krankenanstaltengesellschaft m.b.H., Graz, Austria, E-Mail: michael.schrempf@kages.at

One way of identifying patients at risk is the use of clinical prediction models. Common risk scores for cardiovascular diseases are the Framingham Risk Score [4], QRISK3 [5], SCORE [6] or ACC/AHA Risk Score [7].

The advantage of these scores is the use of a few predictors for the risk estimation. QRISK3 uses 22 predictors including age, smoking and prevalent diseases to predict the 10-year risk of a cardiovascular event. In a validation cohort, the score achieved a high discriminative performance with a C-statistic of 0.88 [5].

Compared to this, the Framingham Risk Score achieved an Area Under the Receiver Operating Characteristic (AUROC) of 0.79 in women. However, the Framingham Risk Score is using seven parameters to calculate the 10-year risk of cardiovascular disease. Furthermore, it should not applied to patients with diabetes.

The ACC/AHA Risk Score achieved an AUROC of 0.818 for African-American women. The poorest AUROC performance for non-Hispanic white men was 0.713.

Similar to the before mentioned score, the SCORE AUROC ranged from 0.71 to 0.84. It is using similar feature as the Framingham Risk Score and ACC/AHA Risk Score and was developed for European clinical practice.

However, all of the mentioned scores require manual input for risk assessment. The manual scoring often presents an additional burden for healthcare professionals instead of optimizing care. In addition, most risk scores are based on linear methods which are not able to account for non-linear relationships in highly complex data [8].

Therefore, the use of machine learning (ML) techniques is an alternative. ML algorithms can be trained on already available electronic medical records (EMRs) and account for numerous predictions in order to achieve high predictive performance.

Weng et al. [9] demonstrated that machine learning based approaches are able to predict cardiovascular events. However, their best performing model, a neural network, achieved only a performance of an AUROC of 0.764, which was comparable to the ACC/AHA score. Krittanawong et al. [10] demonstrated that the performances of the risk score could be outperformed by machine learning. This meta-analysis of prediction models for cardiovascular diseases showed that machine learning models can achieve AUROCs between 0.88 and 0.93.

1.2. Objectives

The aim of this study was to develop a prediction model based on electronic medical records to predict the 5-year risk of MACE. A machine learning approach was used in order to estimate the risk for each patient without additional efforts by clinicians and to account for various predictors available in the EMR system.

2. Methods

2.1. Data Overview

Data for modelling was provided by the public hospital provider Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes). KAGes covers approximately 90% of acute care beds in the province of Styria, Austria, and provides ambulatory care services. Therefore, longitudinal health records of 2.4 million patients in Styria are available starting from 2006. The extracted EMR data included among others demographic data, ICD-10 coded diagnoses, nursing assessment, LOINC codes for laboratory data,

procedures, transfer data and ATC classes for medication. Available EMR data was included within the time span of 01/01/2009 and 12/31/2019. All patients under the age of 18 were excluded from modeling.

2.2. Label Definition

In order to identify patients with the prediction outcome MACE, transfer data, diagnoses and procedures were used. First, all patients with a coded diagnosis of angina pectoris (ICD-10 code I20) with a manual description associated to an acute coronary syndrome were included. In addition patients with one ICD-10 code of the diagnoses shown in Table 1 were included, if a matching procedure or transfer to a coronary catheterization laboratory was recorded for the hospital stay.

Table 1. ICD-10 Codes for outcome definition

ICD-10 code	Description	Type
I20	Angina Pectoris	
I21	Myocardial Infarction	Coronary heart disease
I24	Other acute ischemic heart disease	
I46	Cardiac arrest	
I63	Cerebral infarction	Cerebrovascular disease
I64	Stroke	
I71	Aortic aneurysm and dissection	Other cardiovascular disease
I74	Arterial embolism and thrombosis	

All patients with the following procedures were included in the cohort of MACE patients: coronary bypass surgeries, percutaneous coronary catheterization and systemic thrombolysis for myocardial infarction or stroke. In addition, magnetic resonance and computed imaging procedures for cerebrovascular diseases were used.

In case of a MACE, there is a high probability that the patient will receive cardiovascular associated procedures in one of three coronary catheterization facilities in the State of Styria. Thus, patients with transfers to the coronary catheterization laboratories in combination with a defined diagnosis or procedure were included as MACE patients.

Furthermore, the death data from Austria's Federal Statistical Office (Statistik Austria) enriched the definition of MACE patients: KAGes patients with death due to myocardial infarction (I21), cerebral infarction (I63) or stroke (I64) were included as MACE patients. Patients with the following ICD 10 codes were excluded, because these disease are mostly not caused by atherosclerosis [11]: intracerebral haemorrhage (I61, I62), aneurysm and dissection (I72).

After applying all inclusion and exclusion criteria, 86,365 admissions of 29,262 patients were included in the MACE group. The overlap of the identification of MACE patients using the four data modalities is shown in Fig. 1.
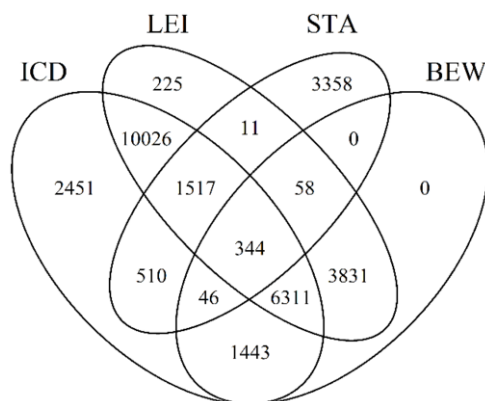
Figure 1. Venn diagram of the final selection of MACE patients using four modalities of identification: Diagnosis (ICD), procedures (LEI), death data (STA), and transfer data (BEW).

Finally, 181,150 admissions of 98,405 patients without MACE identification within 5 years after admission were extracted. For patients with multiple stays in the specified time period, all stays were treated separately. This resulted in 267,515 admissions for modeling.

2.3. Feature Extraction & Modelling

As a reference date for risk prediction, the discharge date of the admission included in the data set was set. Until this reference date, various features were created using already available data from the patient. Out of the data modalities described in section 2.1, 9,270 features were created.

Following, all constant features were deleted, before feature selection was applied. As a method of univariate feature selection, Chi-squared tests and ANOVA were applied, which removed 2,425 features from the data set. For further selection, a Generalized Linear Model with Lasso regularization and a Random Forest model were trained with the remaining features. Using only features with its coefficients greater than zero in or greater than the average feature importance in Random Forest, a feature size of 826 was achieved.

The final data of 267,515 admissions and 826 features was split into a 70% training and 30% test data set. Model training was performed only on the training data using a 5-fold cross-validation. Four different methods were used for training: Generalized Linear Model with Elastic Net regularization (GLM), Random Forest (RF), Gradient Boosting Machines (GBM) and Linear Discriminant Analysis (LDA).

To optimize the hyper-parameters of each method, the GridSearchCV function from scikit-learn [12] was used. GridSearchCV performs a 5-fold cross-validation for every hyper-parameter combination.

Finally, the performances of the models were evaluated on the separate test data set. The discriminative performance was measured using the AUROC value with 95% confidence intervals and ROC plots. The confidence intervals were calculated using the DeLong method [13]. Based on the closest topleft threshold, sensitivity, specificity, and precision were calculated. In order to analyse the calibration of the models, calibration plots were used showing the percentiles of risk probabilities in the x-axis and relative frequency of MACE on the y-axis. Additionally, an ethical vote was issued by the

Medical University of Graz for the development of these predictive models (30-146 ex 17/18). Furthermore, the TRIPOD statement [14] was used for model development, validation and reporting.

3. Results

In Table 2, the results of the evaluation on the test data are presented. Although the discriminative performance of all methods was excellent with AUROC above 0.86, the Gradient Boosting Machines achieved a slightly better performance than the others. Furthermore, a plot with the ROC curves of all models is shown in Fig. 2.

Table 2. 5-Year MACE prediction using four machine learning algorithms.

Method	AUROC [95%-CI]	Sensitivity	Specificity	Precision
GLM	0.867 (0.848 - 0.886)	0.798	0.767	0.619
RF	0.879 (0.860 - 0.897)	0.808	0.776	0.632
GBM	0.896 (0.879 - 0.913)	0.808	0.820	0.681
LDA	0.867 (0.848 - 0.886)	0.785	0.780	0.630

Fig. 3 presents the calibration plots of the models. Results of Gradient Boosting Machines showed the poorest performance among the models, with an underestimation of the risk for MACE at the lower percentiles and an overestimation of the risk for higher percentiles. Random forest and Generalized Linear Model showed a good calibration for all risk percentiles.
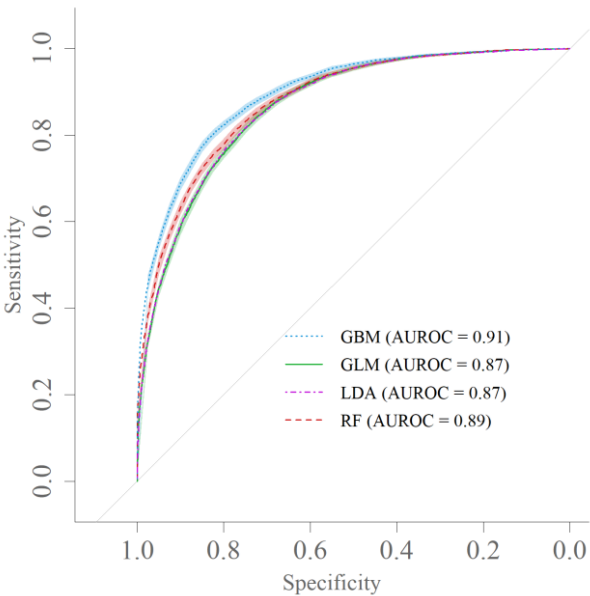

Figure 2. ROC curves for the trained models when applied on the test data (Generalized Linear Model (GLM), Random Forest (RF), Gradient Boosting Machines (GBM), Linear Discriminant Analysis (LDA))
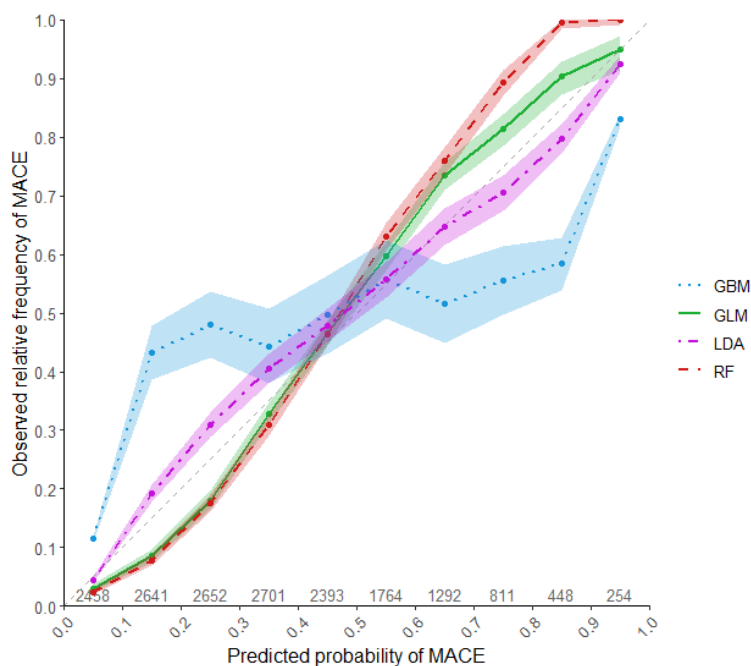
Figure 3. Calibration plots of the models when applied on the test data.

In addition, the 17 most important features of the Random Forest were investigated (see Fig. 4). Similar to QRISK3 or the other cardiovascular diseases related risk scores, age had a very large impact on risk prediction. An increased value for urea in serum or plasma, and potassium in serum or plasma were features with the highest influence, among the others. These features are associated with renal disorders, leading to an increased risk for cardiovascular diseases [5] [15] [16].
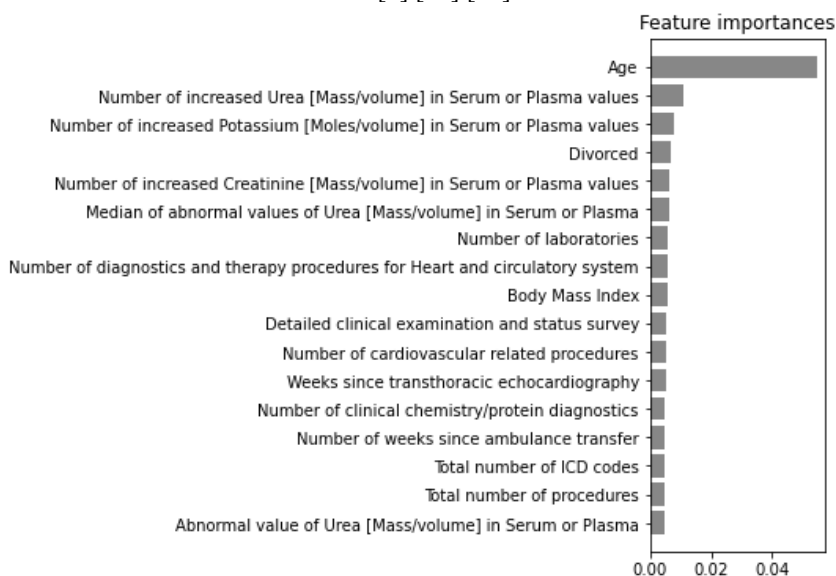

Figure 4. Gini Importance based Feature Importance of the Random Forest model.

4. Discussion

Our study presents the development of various machine learning models predicting the 5-year risk of a major cardiovascular event (MACE) of in-patients. EMRs of more than 127,667 patients were used to train the risk prediction models using different methods. Although GBM proved to be the best discriminator, the results of calibration curve were not satisfying. The three remaining models showed equally good results when considering both discrimination and calibration. This could be due to a comprehensive and effective feature selection.

With an AUROC of 0.879, the Random Forest model outperforms established cardiovascular risk scores such as QRISK3 or the ACC/AHA. However, the comparison has to be considered with care, as the models developed in our study predict the 5-year risk of MACE, while the published scores mostly predict a 10-year risk.

Further limitations need to be discussed. A main advantage of clinical scores over machine learning models is their generalizability. The use of few predictors for a risk prediction is independent of the availability of EMRs. As EMR systems and their records vary between hospitals, machine learning models developed in one hospital might not be applicable to other hospitals. External validation of machine learning-based models is crucial for a broad deployment.

KAGes covers almost 90% of the in-hospital stays in the region, and thus the EMR system provides a comprehensive population-based view of patient histories. However, the identification of patients with MACE can be biased if the event was recorded in another hospital outside the KAGes network. Using different modalities for labelling, including death records from the official Austrian statistical office, presents a major advantage of this study.

Although risk prediction by machine learning models might be highly accurate, more actions are needed in order to achieve a clinical benefit. In a first step, clinicians need to be informed about patient's individual risk factors using transparent visualization methods embedded in the hospital information systems. In a second step, it is essential to inform patients about preventive measures to modify the risk. Thus, future work should focus on how a predicted risk can be changed by staging preventive actions or treatment for patients with high risk of MACE.

The prediction models achieved an excellent performance in this study. In previous work, we demonstrated the high performance of machine learning model predicting delirium in a clinical setting [17]. Future work needs to focus on the prospective evaluation of the risk prediction models for MACE and determine the overall clinical benefit for patients.

References

[1]   STATISTIK AUSTRIA, 'Todesursachen'.
      http://pic.statistik.at/web_de/statistiken/menschen_und_gesellschaft/gesundheit/todesursachen/index.ht
      ml (accessed Jan. 28, 2021).

[2]   L. Bergman, J. H. P. van der Meulen, M. Limburg, and J. D. F. Habbema, 'Costs of Medical Care
      After First-Ever Stroke in the Netherlands', *Stroke*, vol. 26, no. 10, pp. 1830–1836, Oct. 1995, doi:
      10.1161/01.STR.26.10.1830.

[3]   M. F. Piepoli *et al.*, '2016 European Guidelines on cardiovascular disease prevention in clinical
      practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on
      Cardiovascular Disease Prevention in Clinical Practice', *Eur. Heart J.*, vol. 37, no. 29, pp. 2315–2381,
      Aug. 2016, doi: 10.1093/eurheartj/ehw106.

[4]   R. B. D'Agostino *et al.*, 'General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study', *Circulation*, vol. 117, no. 6, pp. 743–753, Feb. 2008, doi: 10.1161/CIRCULATIONAHA.107.699579.

[5]   J. Hippisley-Cox, C. Coupland, and P. Brindle, 'Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study', *BMJ*, p. j2099, May 2017, doi: 10.1136/bmj.j2099.

[6]   R. Conroy, 'Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project', *Eur. Heart J.*, vol. 24, no. 11, pp. 987–1003, Jun. 2003, doi: 10.1016/S0195-668X(03)00114-3.

[7]   D. C. Goff *et al.*, '2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines', *Circulation*, vol. 129, no. 25 suppl 2, pp. S49–S73, Jun. 2014, doi: 10.1161/01.cir.0000437741.48606.98.

[8]   L. Breiman, 'Statistical Modeling: The Two Cultures', Stat. Sci., vol. 16, pp. 199–215, 2001.

[9]   S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, 'Can machine-learning improve cardiovascular risk prediction using routine clinical data?', *PLOS ONE*, vol. 12, no. 4, p. e0174944, Apr. 2017, doi: 10.1371/journal.pone.0174944.

[10]   C. Krittanawong *et al.*, 'Machine learning prediction in cardiovascular diseases: a meta-analysis', *Sci. Rep.*, vol. 10, no. 1, p. 16057, Dec. 2020, doi: 10.1038/s41598-020-72685-1.

[11]   K. Ritz, N. P. Denswil, O. C. G. Stam, J. J. van Lieshout, and M. J. A. P. Daemen, 'Cause and Mechanisms of Intracranial Atherosclerosis', *Circulation*, vol. 130, no. 16, pp. 1407–1414, Oct. 2014, doi: 10.1161/CIRCULATIONAHA.114.011147.

[12]   F. Pedregosa *et al.*, 'Scikit-learn: Machine Learning in Python', *J. Mach. Learn. Res.*, vol. 12, p. 2825−2830, Oct. 2011.

[13]   E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, 'Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach', *Biometrics*, vol. 44, no. 3, p. 837, Sep. 1988, doi: 10.2307/2531595.

[14]   G. S. Collins, J. B. Reitsma, D. G. Altman, and K. Moons, 'Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement', *BMC Med.*, vol. 13, no. 1, p. 1, 2015, doi: 10.1186/s12916-014-0241-z.

[15]   D. Pandya, 'Assessment and Correlation of Urea and Creatinine Levels in Saliva and Serum of Patients with Chronic Kidney Disease, Diabetes and Hypertension– A Research Study', *J. Clin. Diagn. Res.*, 2016, doi: 10.7860/JCDR/2016/20294.8651.

[16]   J. R. Montford and S. Linas, 'How Dangerous Is Hyperkalemia?', *J. Am. Soc. Nephrol.*, vol. 28, no. 11, pp. 3155–3165, Nov. 2017, doi: 10.1681/ASN.2016121344.

[17]   S. Jauk *et al.*, 'Risk prediction of delirium in hospitalized patients using machine learning: An implementation and prospective evaluation study', *J. Am. Med. Inform. Assoc.*, vol. 27, no. 9, pp. 1383–1392, Sep. 2020, doi: 10.1093/jamia/ocaa113.