

1 **Precision phenotyping from routine laboratory parameters for machine**
2 **learning out-of-hospital survival prediction using 4D time-dependent SHAP**
3 **plots in an all-comers prospective PCI registry**

4
5 Paul-Adrian Călborean, MD, PhD ^{a,b,*}, Anda-Cristina Scurtu, MSc ^{a,b}, Paul Grebenișan, MD
6 ^a, Ioana-Andreea Nistor ^a, Victor Vacariu, MD ^a, Reka-Katalin Drincal, MD ^a, Ioana Paula
7 Șulea, MD ^a, Tiberiu Oltean ^a, László Hadadi, MD, PhD ^a

8
9 ^a Emergency Institute for Cardiovascular Diseases and Transplantation Târgu Mureș, Târgu
10 Mureș, Romania

11 ^b University of Medicine, Pharmacy, Science and Technology “George Emil Palade” of Târgu
12 Mureș, Târgu Mureș, Romania

13
14 ***[Corresponding author]:** Paul-Adrian Călborean, University of Medicine, Pharmacy,
15 Science and Technology “George Emil Palade” of Târgu Mureș, Gheorghe Marinescu Str.,
16 no. 38, Târgu Mureș, Romania. **E-mail:** calboreanpaul@gmail.com

17
18 **Total word count:** 3347 (excluding abstract, references, tables, and figures)

19 **Funding:** This research received no funding.

20 **Abstract**

21 **Introduction**

22 Out-of-hospital mortality in coronary artery disease (CAD) is particularly high and
23 established adverse event prediction tools are yet to be available. Our study aimed to
24 investigate whether precision phenotyping can be performed using routine laboratory
25 parameters for the prediction of out-of-hospital survival in a CAD population treated by
26 percutaneous coronary intervention (PCI).

27 **Materials and methods**

28 All patients treated by PCI and discharged alive in a tertiary center between January 2016 –
29 December 2022 that have been included prospectively in the local registry were analyzed.
30 115 parameters from the PCI registry and 266 parameters derived from routine laboratory
31 testing were used. An extreme gradient-boosted decision tree machine learning (ML)
32 algorithm was trained and used to predict all-cause and cardiovascular-cause survival.

33 **Results**

34 A total of 7186 PCI hospitalizations for 5797 patients were included with more than 610.000
35 laboratory values. All-cause and cardiovascular cause mortality was 17.5% and 12.2%,
36 respectively, during a median follow-up time of 1454 (687 – 2072) days. The integrated area
37 under the receiver operator characteristic curve for prediction of all-cause and cardiovascular
38 cause mortality by the ML on the validation dataset was 0.844 and 0.837, respectively (all
39 $p < 0.001$). The integrated area under the precision-recall curve for prediction of all-cause and
40 cardiovascular cause mortality by the ML on the validation dataset was 0.647 and 0.589,
41 respectively (all $p < 0.001$).

42 **Conclusion**

43 Precise survival prediction in CAD can be achieved using routine laboratory parameters. ML
44 outperformed clinical risk scores in predicting out-of-hospital mortality in a prospective all-
45 comers PCI population.

46 **Keywords**

47 Machine learning, coronary artery disease, percutaneous coronary intervention, survival
48 analysis

49 **Introduction**

50 Extensive research has been dedicated to the prediction of adverse events in the coronary
51 artery disease (CAD) population. Numerous clinical, angiographic, and combined risk scores
52 were developed to estimate survival, such as Global Registry of Acute Coronary Events
53 (GRACE) score, Age, Creatinine, and Ejection fraction (ACEF) score, thrombolysis in
54 myocardial infarction (TIMI) score and Synergy Between PCI with TAXUS and Cardiac
55 Surgery (SYNTAX) score.¹⁻⁴ However, in many available risk scores the prediction
56 performance is mainly driven by high accuracy in predicting in-hospital death, since the
57 presence of very high-risk clinical features almost invariably leads to in-hospital death (e.g.,
58 cardiogenic shock, acute pulmonary edema, resuscitated cardiac arrest, mechanical
59 complications of acute myocardial infarction, severe anemia or acute kidney injury), while
60 out-of-hospital death is more difficult to predict. Moreover, long-term survival in CAD is less
61 determined by the complexity of coronary anatomy.⁵ Thus, more research is needed to find
62 robust predictors of impaired survival in the CAD population, especially after hospital
63 discharge.

64 Phenotype, initially defined as “the observable traits of an organism”, nowadays
65 includes clinical and paraclinical characteristics of a patient (e.g., diagnostic imaging,
66 laboratory parameters, electrocardiographic changes).⁶ According to the patient’s phenotypic
67 abnormalities, both diagnostic and prognostic assessments can be made.⁶ Precision
68 phenotyping is a concept that implies a comprehensive analysis of a wide panel of clinical
69 parameters usually using artificial intelligence (AI) techniques which can quantify even slight
70 changes in a patient’s phenotype and offer a precise prognostic estimate. Precision
71 phenotyping has been previously proposed using clinical, ultrasound, computed tomography,
72 or big “-omic” data.⁷⁻⁹ Whether a precise survival estimate can be given using a spectrum of
73 laboratory parameters routinely measured in clinical practice is yet to be proven.

74 The aim of the present study was (1) to train an ML model to predict out-of-hospital
75 all-cause and cardiovascular cause mortality using a wide panel of laboratory parameters, (2)
76 to identify predictors of out-of-hospital adverse events, and (3) to compare the predictive
77 performance of ML model with clinical risk scores relevant for CAD population, in an all-
78 comers patient population treated by PCI in a Romanian tertiary cardiovascular center.

79

80 **Materials and methods**

81 **Study population**

82 The study design is illustrated in Figure 1. All patients treated by PCI in the Emergency
83 Institute for Cardiovascular Diseases and Transplantation of Târgu Mureş have been
84 prospectively included at discharge in the local PCI Registry of the Institute since January
85 2016. For the current analysis, all patients treated between January 2016 – December 2022
86 were included, in order to have a minimum of 6 months of follow-up. Subjects were analyzed
87 by PCI hospitalization (hospitalization in which PCI was performed). Exclusion criteria
88 consisted of (1) age less than 18 years old, (2) presence of in-hospital death, or (2) lack of
89 available survival data (e.g., foreign patients). The Registry is accessible online at the website
90 <http://pci.cardio.ro/> and is based on the criteria of Cardiology Audit and Registration Data
91 Standards (CARDS) developed by the Department of Health and Children, European Society
92 of Cardiology, Irish Cardiac Society, and the European Commission.¹⁰ Briefly, the CARDS
93 recommendations address data regarding demographics, relevant medical history, and
94 comorbid conditions, clinical status at hospital admission, PCI indication, affected and treated
95 coronary artery segments, usage of invasive diagnostic or therapeutic devices, procedural
96 complications, medical treatment during hospitalization and at discharge and in-hospital
97 evolution. All the information available regarding all the variables proposed in that document
98 was collected for each included patient, at every PCI. A number of 115 parameters were

99 available from the PCI registry. Moreover, to complete and update the original CARDS
100 recommendations regarding PCI procedures, an additional 266 laboratory parameters were
101 also collected (Supplemental Table 1). Noteworthy, all laboratory data during the PCI
102 hospitalization was acquired from the electronic health record (EHR). Since the timing and
103 number of determinations of laboratory parameters differ between patients and
104 hospitalizations, each laboratory parameter was considered separately first value in hospital,
105 last value before discharge, average value during hospitalization, absolute maximum and
106 minimum value, and number of each separate determination of that certain laboratory
107 parameter (e.g., first creatinine, last creatinine, average creatinine, maximum and minimum
108 creatinine, number of creatinine determinations, Supplemental Table 1). A total of 381
109 parameters were used for the final analysis. Furthermore, clinical, and angiographic risk
110 scores such as ACEF, GRACE, and SYNTAX score were also calculated for comparison with
111 the ML model's survival prediction performance.

112 All patients or their legal representatives (e.g., for critical patients) provided written
113 signed informed consent regarding the PCI procedure and their participation in the study. The
114 study was approved by the ethical committee of our institution (decision number 8646 from
115 22 December 2015 approved by the Ethical Committee of the Emergency Institute for
116 Cardiovascular Diseases and Transplantation of Târgu Mureş). The protocol was carried out
117 in accordance with the ethical principles for medical research involving human subjects
118 established by the Declaration of Helsinki, protecting the confidentiality of personal
119 information of the patients.

120 **Follow-up and clinical outcomes**

121 The clinical endpoint of this study was the incidence of out-of-hospital cardiovascular and
122 all-cause mortality. Out-of-hospital mortality was defined as death that occurred after
123 discharge, but not occurring during the initial hospital admission when the PCI procedure was

124 performed. The Romanian National Health Insurance System database supplied mortality
125 rates as of July 2023 for all the patients. For patients who had died during follow-up, the
126 Regional Statistics Office of the Romanian National Institute of Statistics supplied the exact
127 date and cause of death according to the tenth revision of the International Classification of
128 Diseases (ICD-10). If the cause of death belonged to diseases of the circulatory system, then
129 death was considered to be of cardiovascular cause.

130 **Machine learning**

131 All patients had a minimum of 6 months survival status available and up to 6 years of follow-
132 up. Patients were dichotomized every 6 months into alive/deceased groups and censored
133 events were removed. Afterward, for each timeframe, the dataset was randomly divided into
134 70% training and 30% validation datasets. To prevent data leaks since a patient could have
135 multiple hospitalizations, each patient was assigned to either a training or validation set. On
136 the 70% training dataset, a 5-fold cross-validation was performed and to prevent data leak
137 between folds, each patient was only assigned to one-fold. An extreme gradient boosted
138 decision tree (XGBoost) algorithm¹¹ was evaluated as a binary classifier for predicting both
139 all-cause- and cardiovascular-cause mortality occurred out of hospital, for each timeframe.
140 XGBoost algorithm was implemented in Python version 3.9.13. Hyperparameter optimization
141 was performed by using Bayesian search on 6 years survival status, being the least
142 imbalanced dataset, and those best training parameters were used for all timeframes. The
143 accessibility of underlying supporting scripts is detailed in the data availability section. In
144 order to explain the ML decision process, the open-source Shapley additive explanations
145 (SHAP) framework was used.¹² Most important features were obtained as follows: (1) SHAP
146 values were obtained for each of the database values at each timeframe, for both all-cause and
147 cardiovascular cause events; (2) at each timeframe, absolute SHAP values were added, and
148 parameters' importance was considered by the highest sum of absolute SHAP value (e.g., the

149 parameter with the highest absolute SHAP values sum at 5-years cardiovascular survival was
150 the most important); (3) overall most important predictors were obtained by ascending
151 classification of averaged positions at all timeframes for all-cause and cardiovascular cause
152 survival, separately (see supporting code). Illustration of time-dependent SHAP values for a
153 certain parameter on a specified survival (e.g., time-dependent SHAP values for age impact
154 on all-cause survival) was performed as follows: (1) SHAP values were obtained for every
155 value of that parameter at each timeframe on the specified survival; (2) for the specified
156 parameter, ten percentile intervals were obtained according to the minimum and maximum
157 value; (3) SHAP values were averaged for each percentile at each timeframe; (4) 3D mesh
158 plot was illustrated using the ten percentiles of that parameter, all investigated timeframes
159 and averaged SHAP values for that percentile; (5) in order to correlate averaged SHAP values
160 with actual observed survival, the now 4D mesh plot was colored by averaged survival for
161 each parameter percentile at each timeframe (see supporting code).

162 **Statistical analysis**

163 A significance level α of 0.05 and a 95% confidence interval (CI) were considered.
164 Continuous variables were evaluated for normal distribution using the Shapiro-Wilk test.
165 Continuous variables with parametric distributions were reported as mean \pm standard
166 deviation and compared using non-paired Student's t-test, while continuous variables with
167 non-parametric distributions and discrete variables were reported as median (interquartile
168 range) and compared using Mann Whitney test. Categorical variables were reported as
169 absolute and relative frequencies and compared using Fisher's exact test for clinical
170 parameters with frequencies less than 5 and the Chi² test otherwise. For baseline clinical
171 characteristics, univariate Cox proportional hazard was used to predict the association in the
172 form of hazard ratio (HR) between observed survival and an independent categorical
173 variable. All continuous variables were dichotomized related to the median to facilitate

174 comparability of HR and CI between variables. Statistical analysis was performed using R
175 version 4.1.1 and R Studio version 1.4.17. Performance on survival prediction was measured
176 by integrating the time-dependent area under the curve of receiver-operator characteristic
177 (AUC-ROC), the area under the curve of precision-recall (AUC-PR), Matthews correlation
178 coefficient (MCC), F1 score, and Brier score for a specified survival. While for AUC-ROC
179 the baseline is constant at 0.5, in the case of AUC-PR, the baseline is determined by positive
180 cases per total cases ratio.¹³ Thus, a baseline-corrected AUC-PR was calculated by
181 subtracting the baseline value from the AUC-PR value. MCC and F1 score calculation require
182 dichotomization from probability for ML model and score value for clinical risk scores into
183 predicted deceased or alive, which was performed using Youden's method for cut-off
184 calculation.

185

186 **Results**

187 **Study population**

188 A total of 7186 PCI hospitalizations for 5797 patients were included in the present study, with
189 a total of 611309 laboratory values. Of those patients, 4062 (70.07%) were male sex, the
190 median age was 65.0 (57.2-71.8) years and the median BMI was 28.3 (25.9-31.8) kg/m².
191 Complete clinical characteristics of the studied patients are reported in Table 1 and
192 Supplemental Table 1. Complete clinical characteristics of the studied PCI hospitalizations
193 are reported in Table 2 and Supplemental Table 2. A total of 1017 (17.5%) patients died of all
194 causes during a median follow-up time of 1454 (687 – 2072) days. When censoring events of
195 non-cardiovascular cause, a total of 710 (12.2%) patients died of cardiovascular causes
196 during a median follow-up time of 1544 (748– 2155) days. Cumulative per-patient mortality
197 incidence is illustrated in Figure 2A and cumulative per-PCI mortality incidence is illustrated
198 in Figure 2B. A total of 115 clinical parameters were available from the PCI registry. A

199 median of 114 (110-115) parameters were available for each patient as some parameters were
200 missing or not applicable (e.g., time from symptom onset to PCI, which is measured in acute
201 but not in chronic coronary syndrome). A total of 611309 laboratory values were analyzed for
202 all 7186 hospitalizations. Per hospitalization, a median of 78 (65-116) laboratory values were
203 available. For each laboratory parameter, the first and last in-hospital value, maximum and
204 minimum value during hospitalization, averaged value during hospitalization, and number of
205 determinations for that parameter during hospitalization were calculated, therefore 266
206 additional parameters were analyzed.

207 **Machine learning analysis**

208 Machine learning models accurately predicted out-of-hospital survival at all timeframes on
209 all time-dependent performance metrics (Figure 3). While AUC-ROC was consistently high
210 (Figure 3A, 3C), AUC-PR gradually increased as the deceased-to-alive ratio became less
211 imbalanced (Figure 3B, 3D). Minor differences were between testing dataset predictions and
212 validations dataset predictions, revealing that overfitting did not occur (Figure 3). Tuned
213 hyperparameters found by Bayesian search included a total of 5000 aggregated decision trees,
214 with a maximum tree depth of 12 levels and a learning rate of 0.01 (see supporting code).
215 Most important survival predictors were obtained by classifying the averaged importance
216 position at all timeframes. For all-cause mortality, left ventricular ejection fraction (LVEF),
217 age, hospitalization cost, heart rate at presentation, renal function reflected by maximum
218 creatinine during hospitalization, standard deviation of red cell distribution width and last
219 determined before discharge lymphocyte per monocyte ratio were among the most important
220 parameters. For cardiovascular cause mortality, left ventricular ejection fraction (LVEF), age,
221 systolic and diastolic blood pressure at presentation, heart rate at presentation, hospitalization
222 cost, renal function reflected by maximum creatinine during hospitalization, the standard
223 deviation of red cell distribution width and last determined before discharge lymphocyte per

224 monocyte ratio was among the most important parameters. Furthermore, the most important
225 factors for the occurrence of adverse events and most important factors against the
226 occurrence of adverse events of any cause and cardiovascular cause are reported in
227 Supplemental Figure 1. Time-dependent SHAP values are reported in Figure 5 and reveal
228 how the predictor's values (e.g., advanced age or reduced LVEF) predispose to adverse
229 events. Patient individualized survival prediction can be illustrated using ML probability
230 output at each timeframe with event occurrence estimated at the point where probability is
231 less than the cut-off for that specific timeframe (Figure 6).

232 **Comparison between machine learning and clinical scores**

233 To illustrate the predictive performance of ML models, some important clinical risk scores,
234 such as ACEF, GRACE, or SYNTAX scores, were also calculated and compared with ML
235 models. ML consistently and significantly outperformed clinical risk scores on all timeframes
236 (Figure 3) and all performance metrics (Table 3). Among clinical scores, the ACEF score had
237 the highest performance metrics, while the SYNTAX score had the lowest performance
238 metrics. The integrated AUC-ROC for prediction of all-cause mortality by the ML model
239 versus ACEF score was 0.844 and 0.735, respectively (Table 3). The integrated AUC-ROC
240 for prediction of cardiovascular cause mortality by the ML model versus ACEF score was
241 0.837 and 0.761, respectively (Table 3). A more striking increase in prediction performance is
242 reflected by the integrated AUC-PR and integrated baseline-corrected AUC-PR. The
243 integrated AUC-PR for prediction of cardiovascular and all-cause mortality by the ML model
244 was 0.647 and 0.589, respectively, while for ACEF score was 0.380 and 0.410, respectively
245 (Table 3). The integrated baseline corrected AUC-PR for prediction of cardiovascular and all-
246 cause mortality by the ML model was 0.407 and 0.428, respectively, while for ACEF score
247 was 0.199 and 0.190, respectively (Table 3). Additional details regarding performance

248 metrics are reported in Supplemental Table 3. Other metrics, such as integrated MCC, F1, and
249 Brier score were also in favor of ML models.

250

251 **Discussions**

252 The main findings of our study are: (1) In a prospective all-comers PCI registry, both all-
253 cause and cardiovascular cause out-of-hospital mortality is high, with 6-years mortality of up
254 to 42%; (2) ML models can accurately predict long-term out-of-hospital survival in a CAD
255 population treated by PCI using routine laboratory parameters; (3) Predictive performance of
256 ML models was better than relevant clinical scores, such as GRACE, ACEF or SYNTAX
257 score on all performance metrics, including integrated AUC-ROC, AUC-PR, MCC, F1 and
258 Brier score; (3) Important factors for survival prediction includes age, LVEF, blood pressure
259 and heart rate on presentation. However, most important factors were those derived from
260 laboratory parameter analysis; (4) Time-dependent SHAP values reveal how the values of a
261 parameter impact predictions for each timeframe.

262 Ischemic heart disease is the leading cause of death¹⁴, although precise long-term
263 follow-up from the moment of CAD diagnosis, especially out-of-hospital, is scarce. In our
264 study 5-year and 6-year all-cause, out-of-hospital mortality was 21.7% and 33.2%,
265 respectively, relatively lower than the 5-year all-cause mortality of 37.3% from the SYNTAX
266 trial although it included only three-vessel CAD disease.¹⁵ In patients with established CAD,
267 clinical risk prediction tools for secondary prevention are useful, considering the high
268 mortality associated with CAD. For this purpose, certain clinical scores such as ACEF,
269 SYNTAX, and GRACE scores were developed.^{1,2,4} Refinements of clinical scores were
270 attempted to improve predictive performance, particularly for the original SYNTAX score,
271 from which several SYNTAX-derived scores were reported.¹⁶ In our analysis, ML models
272 outperformed clinical scores on all timeframes and all statistical metrics (Table 3).

273 Accordingly, the use of machine learning instead of continuously refined traditional risk
274 scores seems to be more efficient in long-term mortality risk prediction in CAD patients,
275 especially if clinical data can reliably be obtained from EHR. In the present study, over
276 600,000 laboratory values were automatically obtained from the EHR, making future
277 integration in clinical practice possible.

278 In our analysis, LVEF, age, blood pressure, and heart rate were among the most
279 important predictors, which are also partially included in clinical scores such as ACEF and
280 GRACE scores. High hospitalization cost and increased contrast volume were also predictors
281 of events, reflecting the need for advanced therapies and long or complex PCI procedures,
282 respectively. Interestingly, numerous important clinical features were obtained from
283 laboratory values, reflecting renal function (first and maximum serum creatinine and urea
284 levels), hematologic function (red cell distribution width, mean corpuscular hemoglobin
285 concentration, platelet distribution width), inflammatory status (lymphocyte per monocyte
286 ratio, maximum neutrophil count) and glycemic status (Figure 4). Most of the mentioned
287 parameters were previously reported as predictors of adverse events. The association between
288 lymphocyte-to-monocyte ratio and impaired prognosis in the CAD population has been
289 documented in a recent meta-analysis.¹⁷ Other studies showed that red cell distribution width
290 is an inexpensive prognostic marker in both myocardial infarction and heart failure.^{18,19} This
291 illustrates that precision phenotyping can be achieved using laboratory parameters. Moreover,
292 while the clinician works with dichotomized outcomes from continuous variables (e.g.,
293 patient has heart failure with reduced ejection fraction if LVEF \leq 40%) leading to an
294 information loss²⁰, ML models are particularly powerful in analyzing continuous data without
295 the need for fitting into categories – for this reason, the most important parameters were
296 laboratory values instead of disease labels. Interestingly, the severity of coronary lesions was

297 not particularly important for long-term prognosis. Similarly, SYNTAX score performance
298 decreased over time, revealing that coronary anatomy is less important in the long term.

299 The emergence of AI brought numerous tools for prediction making, which can be
300 simplistically categorized as supervised learning using machine learning techniques (e.g.,
301 support vector machines or gradient-boosted decision trees) or unsupervised learning using
302 deep learning techniques (e.g., neural networks). Depending on the underlying task, certain
303 algorithms could be more suitable. Indeed, for tabular data, deep learning seldom
304 outperforms machine learning.²¹ Regarding survival analysis, Kaplan–Meier and Cox
305 proportional hazards models have successfully been used for decades.²² However, both
306 techniques consider a linear relationship between the risk factor and the log hazard function²³.
307 This constant effect of the risk factor during the follow-up assumed by the Kaplan–Meier and
308 Cox models could be an oversimplification.²³ Recently, ML methods were proposed to
309 analyze survival data.²⁴ More specifically, standard ML techniques use binary classification
310 to predict the outcome at a certain timeframe. This requires dichotomization with removal of
311 censored data, does not consider hazards a function of time, but has a high interpretability of
312 the model. On the other hand, modified ML techniques for survival data are adjusted to
313 consider hazards as a function of time, to handle censored and time-to-event data, but have
314 lower interpretability.²⁴ Moreover, numerous performance metrics to evaluate survival
315 prediction models are time-dependent and require the removal of censored events (e.g., Brier
316 score or time-dependent area under receiver operator characteristic).²⁵ Indeed, there are a
317 variety of ML and deep learning techniques to analyze survival data, and the best technique
318 to perform survival analysis is yet to be determined.

319 A reproducibility crisis is increasingly recognized in the area of AI, driven mainly by
320 data leakage and lack of transparency by unpublished code.^{26,27} Indeed, data leakage – the
321 ability of the AI model to see the outcome of the training dataset – is an important problem

322 that may remain undetected.²⁸ In our previous study, an unexpected data leak was observed –
323 the ML model was making an unrealistic nearly perfect in-hospital mortality prediction by
324 looking at treatment at discharge, recognizing that if a patient did not have medical treatment
325 at hospital discharge it was because the patient was deceased.²⁹ In the current study, we
326 consider data leak to be highly improbable, since (1) only in-hospital data was analyzed, (2)
327 only out-of-hospital death was considered, and (3) survival data was acquired from a different
328 institution. Moreover, survival data was not merged with clinical data during the analysis.
329 Regarding underlying code, a report revealed that 6% of AI papers are accompanied by
330 complete code, while half of the AI papers contained “pseudocode”, a brief description of the
331 algorithm.³⁰ In line with current recommendations, the underlying algorithm code for the
332 current study is publicly disclosed.

333

334 **Study limitations and future research directions**

335 The main limitation of our study is the lack of external validation. In addition, the study
336 population is typical for Eastern Europe, thus extrapolating to other populations could be
337 limited. Although not a limitation per se, we are the first to illustrate a time-dependent SHAP
338 plot and certain time-integrated statistical indicators (e.g., integrated AUC-PR and baseline
339 corrected AUC-PR) which warrants further investigation.

340

341 **Conclusions**

342 Machine learning can accurately predict out-of-hospital all-cause and cardiovascular cause
343 death using routinely performed laboratory parameters and outperforms classic clinical
344 parameter-based risk scores in a prospectively followed CAD population treated by PCI.

345

346 **Acknowledgments**

347 The authors would like to thank all the resident colleagues who continuously introduced the
348 data regarding PCIs performed in the Emergency Institute for Cardiovascular Diseases and
349 Transplantation since 2016. We also would like to thank Mr. Ioan Matei for his help in
350 achieving long-term mortality data.

351 **Author contributions**

352 P.-A.C. and L.H. were involved in the study design. P.-A.C. and L.H. performed the literature
353 search. L.H. supervised and coordinated the data collection in the registry. P.-A.C. was
354 involved in the development of machine learning models and explainable artificial
355 intelligence. All authors were involved in data collection. T.O. was responsible for the
356 development and maintenance of the online platform. P.-A.C. and L.H. analyzed the data. P.-
357 A.C. and L.H. were involved in the writing of the initial draft. All authors were involved in
358 the writing of the final draft and approved the manuscript.

359 **Conflict of interest**

360 The authors have nothing to disclose.

361 **Data availability**

362 Due to national and EU regulations, particularly the General Data Protection Regulation, the
363 data used in this study cannot be made publicly available and shared with the wider research
364 community. The data can be shared by the corresponding author for use in secure
365 environments upon request. However, all Python scripts underlying this study are publicly
366 available at <https://github.com/paul-md/Survival-ML>.

367 **References**

368 1. Tang, E. W., Wong, C.-K. & Herbison, P. Global Registry of Acute Coronary Events
369 (GRACE) hospital discharge risk score accurately predicts long-term mortality post acute
370 coronary syndrome. *American Heart Journal* **153**, 29–35 (2007).

- 371 2. Chichareon, P. *et al.* Predictive ability of ACEF and ACEF II score in patients
372 undergoing percutaneous coronary intervention in the GLOBAL LEADERS study.
373 *International Journal of Cardiology* **286**, 43–50 (2019).
- 374 3. Mok, Y. *et al.* International Validation of the Thrombolysis in Myocardial Infarction
375 (TIMI) Risk Score for Secondary Prevention in Post-MI Patients: A Collaborative Analysis
376 of the Chronic Kidney Disease Prognosis Consortium and the Risk Validation Scientific
377 Committee. *Journal of the American Heart Association* **7**, e008426 (2018).
- 378 4. Sianos, G. *et al.* The SYNTAX Score: an angiographic tool grading the complexity of
379 coronary artery disease. *EuroIntervention* [https://eurointervention.pconline.com/article/the-](https://eurointervention.pconline.com/article/the-syntax-score-an-angiographic-tool-grading-the-complexity-of-coronary-artery-disease)
380 [syntax-score-an-angiographic-tool-grading-the-complexity-of-coronary-artery-disease](https://eurointervention.pconline.com/article/the-syntax-score-an-angiographic-tool-grading-the-complexity-of-coronary-artery-disease).
- 381 5. Thuijs, D. J. F. M. *et al.* Percutaneous coronary intervention versus coronary artery
382 bypass grafting in patients with three-vessel or left main coronary artery disease: 10-year
383 follow-up of the multicentre randomised controlled SYNTAX trial. *The Lancet* **394**, 1325–
384 1334 (2019).
- 385 6. Robinson, P. N. Deep phenotyping for precision medicine. *Human Mutation* **33**, 777–
386 780 (2012).
- 387 7. Precision Phenotyping in Heart Failure and Pattern Clustering of Ultrasound Data for
388 the Assessment of Diastolic Dysfunction.
389 <https://www.jacc.org/doi/epdf/10.1016/j.jcmg.2016.10.012> doi:10.1016/j.jcmg.2016.10.012.
- 390 8. Oikonomou, E. K., Siddique, M. & Antoniadou, C. Artificial intelligence in medical
391 imaging: A radiomic guide to precision phenotyping of cardiovascular disease.
392 *Cardiovascular Research* **116**, 2040–2054 (2020).
- 393 9. Delude, C. M. Deep phenotyping: The details of disease. *Nature* **527**, S14–S15
394 (2015).
- 395 10. Flynn, M. R. *et al.* The Cardiology Audit and Registration Data Standards (CARDS),
396 European data standards for clinical cardiology practice. *European Heart Journal* **26**, 308–
397 313 (2005).
- 398 11. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings*
399 *of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data*
400 *Mining* 785–794 (2016). doi:10.1145/2939672.2939785.
- 401 12. Lundberg, S. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions.
402 *arXiv:1705.07874 [cs, stat]* (2017).
- 403 13. Saito, T. & Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the
404 ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE* **10**,
405 e0118432 (2015).
- 406 14. Nowbar, A. N., Howard, J. P., Finegold, J. A., Asaria, P. & Francis, D. P. 2014 Global
407 geographic analysis of mortality from ischaemic heart disease by country, age and income:
408 Statistics from World Health Organisation and United Nations. *International Journal of*
409 *Cardiology* **174**, 293–298 (2014).

- 410 15. Mohr, F. W. *et al.* Coronary artery bypass graft surgery versus percutaneous coronary
411 intervention in patients with three-vessel disease and left main coronary disease: 5-year
412 follow-up of the randomised, clinical SYNTAX trial. *The Lancet* **381**, 629–638 (2013).
- 413 16. Yadav, M. *et al.* Prediction of Coronary Risk by SYNTAX and Derived Scores.
414 *Journal of the American College of Cardiology* **62**, 1219–1230 (2013).
- 415 17. Quan, X.-Q., Wang, R.-C., Zhang, Q., Zhang, C.-T. & Sun, L. The predictive value of
416 lymphocyte-to-monocyte ratio in the prognosis of acute coronary syndrome patients: a
417 systematic review and meta-analysis. *BMC Cardiovasc Disord* **20**, 338 (2020).
- 418 18. Al-Najjar, Y., Goode, K. M., Zhang, J., Cleland, J. G. F. & Clark, A. L. Red cell
419 distribution width: an inexpensive and powerful prognostic marker in heart failure. *European*
420 *Journal of Heart Failure* **11**, 1155–1162 (2009).
- 421 19. Dabbah, S., Hammerman, H., Markiewicz, W. & Aronson, D. Relation Between Red
422 Cell Distribution Width and Clinical Outcomes After Acute Myocardial Infarction. *The*
423 *American Journal of Cardiology* **105**, 312–317 (2010).
- 424 20. Fedorov, V., Mannino, F. & Zhang, R. Consequences of dichotomization.
425 *Pharmaceutical Statistics* **8**, 50–61 (2009).
- 426 21. Shwartz-Ziv, R. & Armon, A. Tabular Data: Deep Learning is Not All You Need.
427 Preprint at <https://doi.org/10.48550/arXiv.2106.03253> (2021).
- 428 22. Cox, D. R. Regression models and life tables. *Journal of the Royal Statistical*
429 *Society: Series B (Methodological)* **34**, 187–202 (1972).
- 430 23. Harrell, F. E. *Regression modeling strategies: with applications to linear models,*
431 *logistic regression, and survival analysis.* vol. 608 (Springer, 2001).
- 432 24. Cygu, S., Seow, H., Dushoff, J. & Bolker, B. M. Comparing machine learning
433 approaches to incorporate time-varying covariates in predicting cancer survival time. *Sci Rep*
434 **13**, 1370 (2023).
- 435 25. Park, S. Y., Park, J. E., Kim, H. & Park, S. H. Review of Statistical Methods for
436 Evaluating the Performance of Survival or Other Time-to-Event Prediction Models (from
437 Conventional to Deep Learning Approaches). *Korean J Radiol* **22**, 1697–1707 (2021).
- 438 26. Hutson, M. Artificial intelligence faces reproducibility crisis. *Science* **359**, 725–726
439 (2018).
- 440 27. Gibney, E. Could machine learning fuel a reproducibility crisis in science? *Nature*
441 **608**, 250–251 (2022).
- 442 28. Kapoor, S. & Narayanan, A. Leakage and the Reproducibility Crisis in ML-based
443 Science. Preprint at <https://doi.org/10.48550/arXiv.2207.07048> (2022).
- 444 29. Călborean, P.-A. *et al.* Prediction of 3-year all-cause and cardiovascular cause
445 mortality in a prospective percutaneous coronary intervention registry: Machine learning
446 model outperforms conventional clinical risk scores. *Atherosclerosis* **350**, 33–40 (2022).

447 30. Gundersen, O. E. & Kjensmo, S. State of the art: Reproducibility in artificial
448 intelligence. in *Proceedings of the AAAI Conference on Artificial Intelligence* vol. 32 (2018).
449

450 **Tables and figures**

Figure 1 – Illustration of the study design.

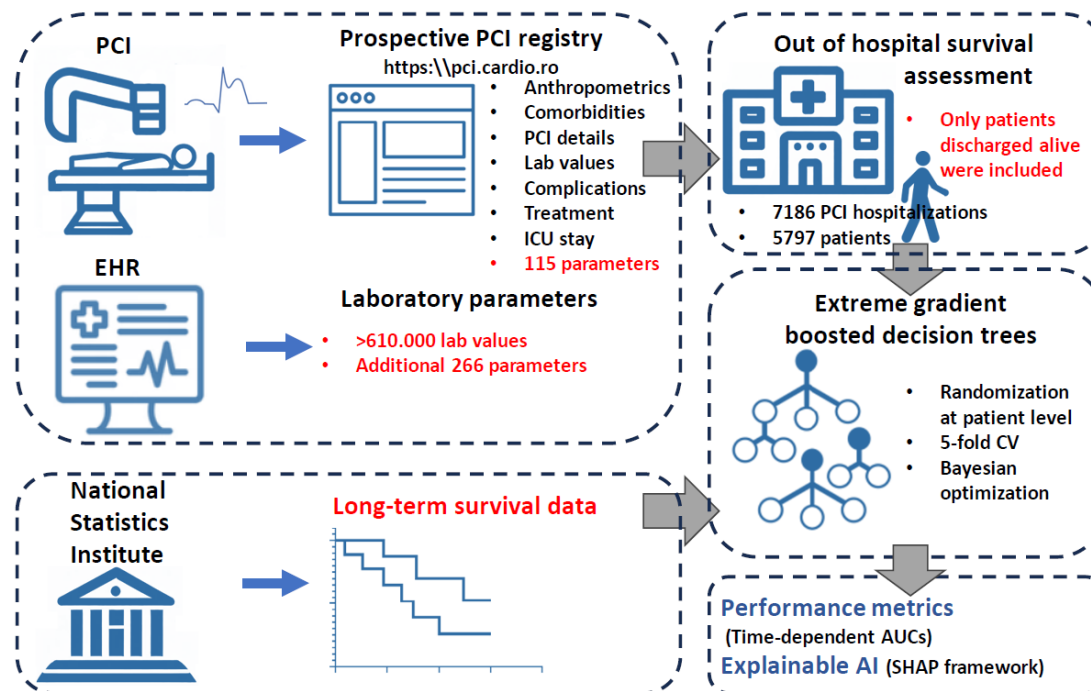
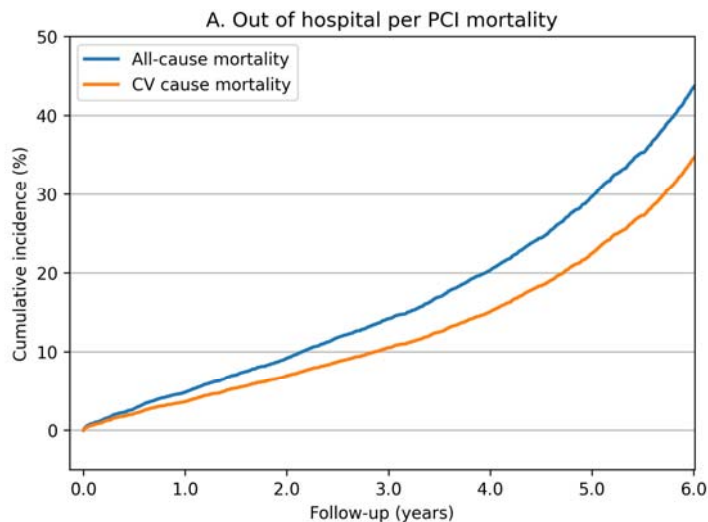
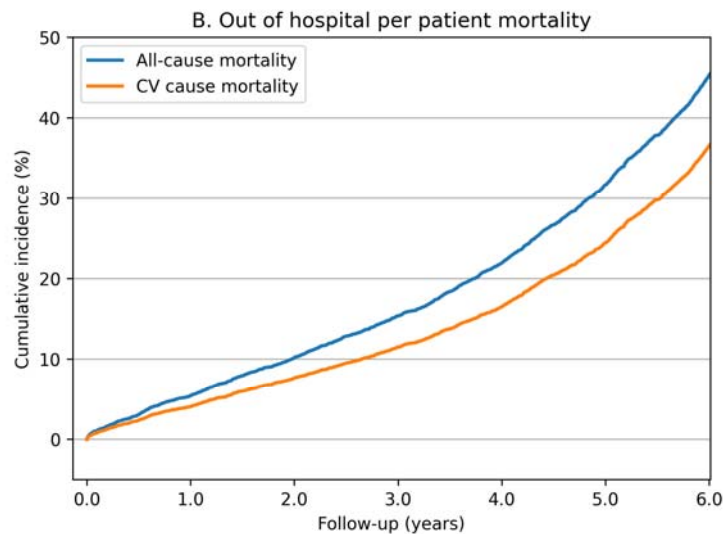


Figure 2 – Observed survival in the studied population.



All-cause mortality							
At risk	7184	6421	5414	4667	3789	2679	1562
Censored	0	438	1225	1747	2428	3376	4417
Events	2	327	547	772	969	1131	1207
CV cause mortality							
At risk	7184	6421	5414	4667	3789	2679	1562
Censored	1	522	1369	1970	2723	3731	4798
Events	1	243	403	549	674	776	826



All-cause mortality							
At risk	5795	5119	4241	3586	2882	2005	1191
Censored	0	384	1076	1560	2105	2862	3620
Events	2	294	480	651	810	930	986
CV cause mortality							
At risk	5795	5119	4241	3586	2882	2005	1191
Censored	1	460	1205	1746	2346	3143	3920
Events	1	218	351	465	569	649	686

CV – cardiovascular; PCI – percutaneous coronary intervention.

Figure 3 – Time-dependent AUC-ROC and AUC-PR for cardiovascular cause and all-cause death prediction.

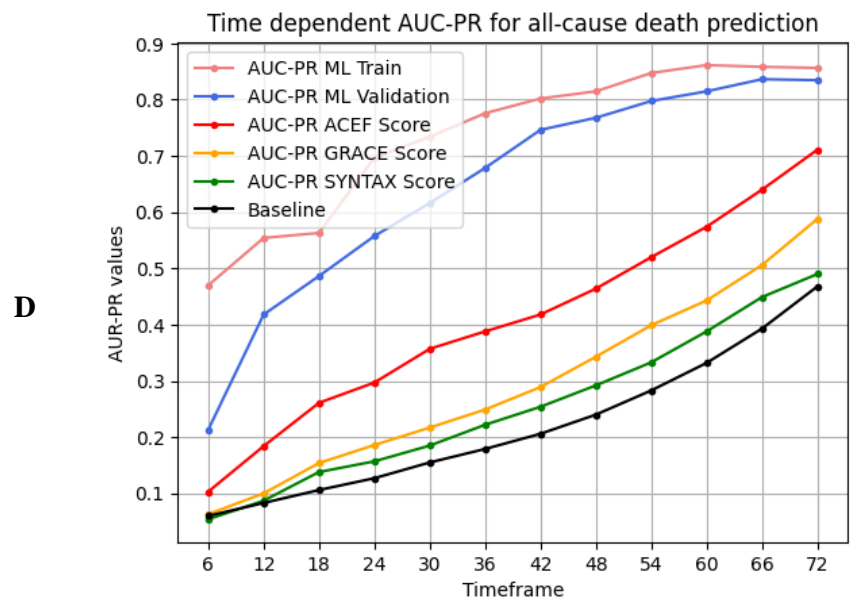
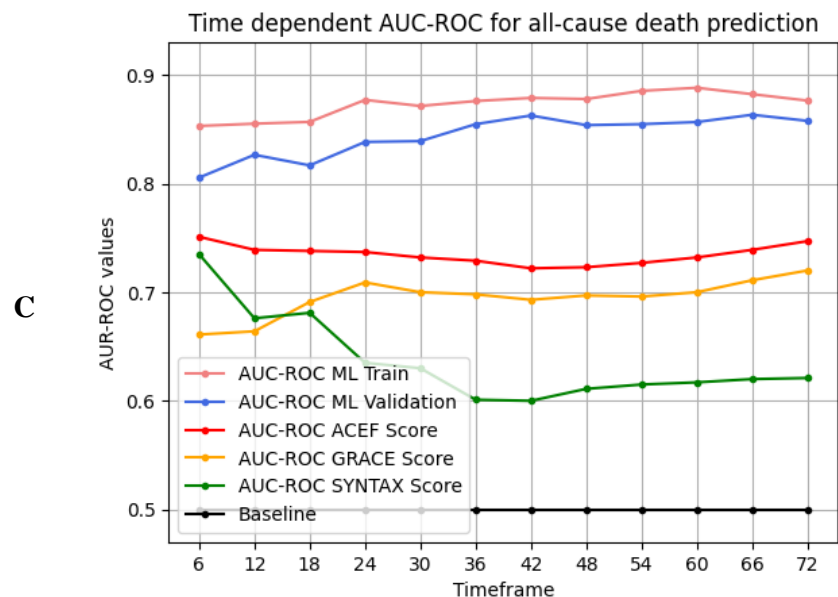
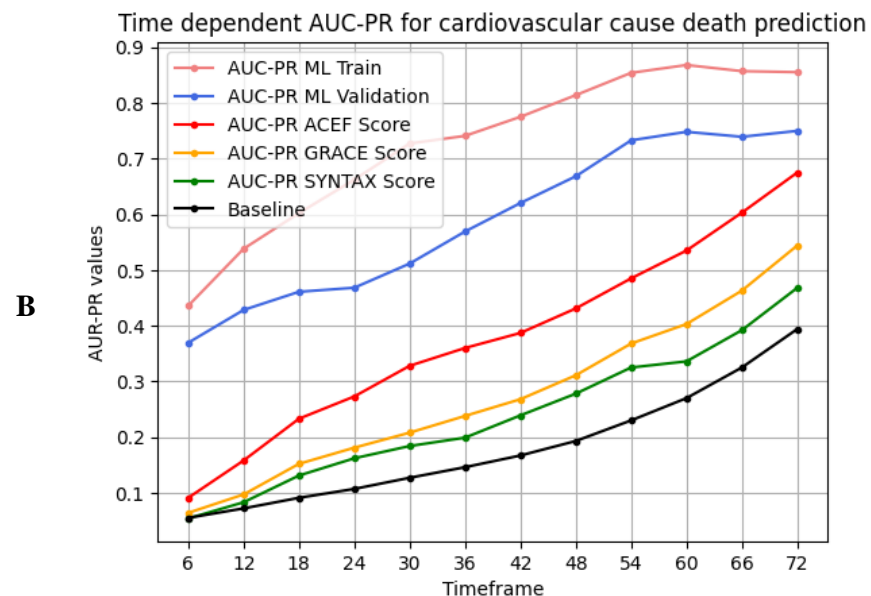
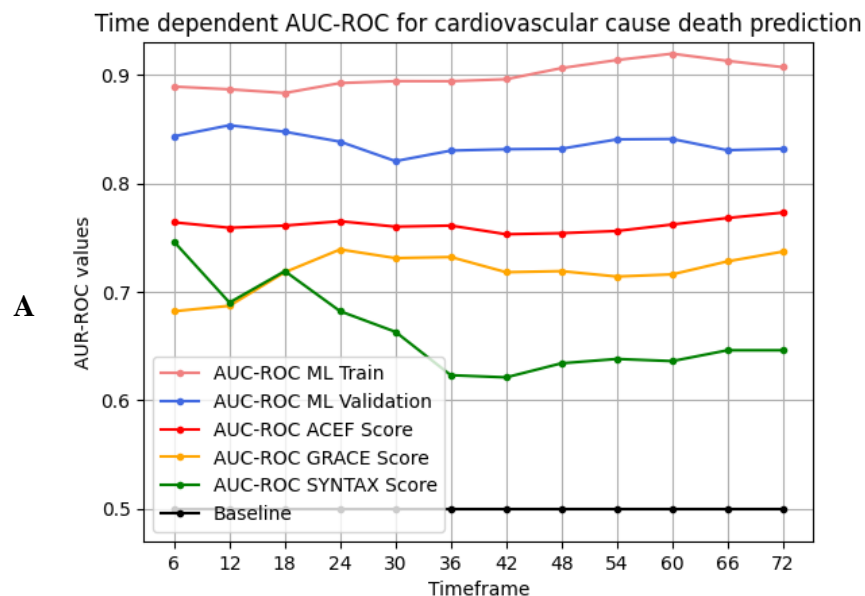
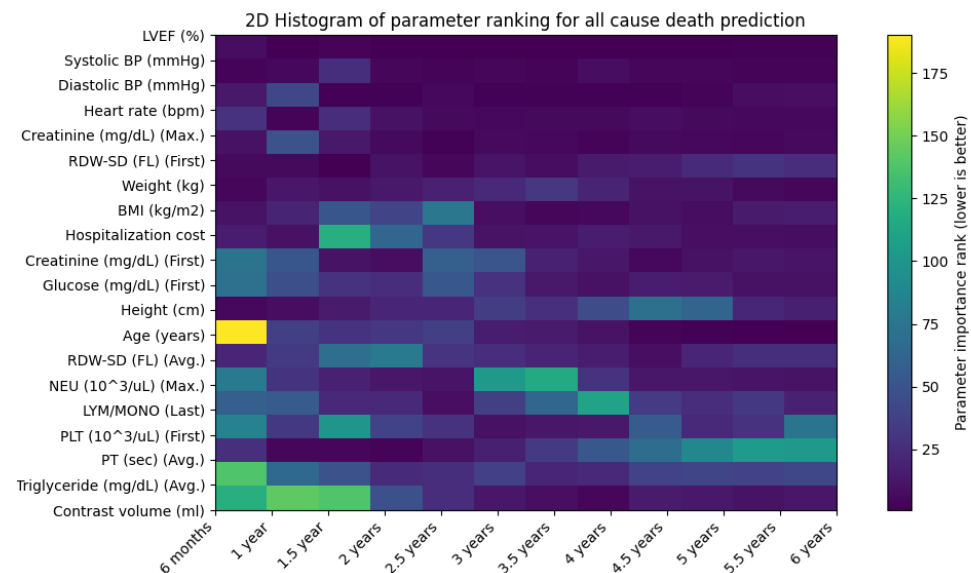
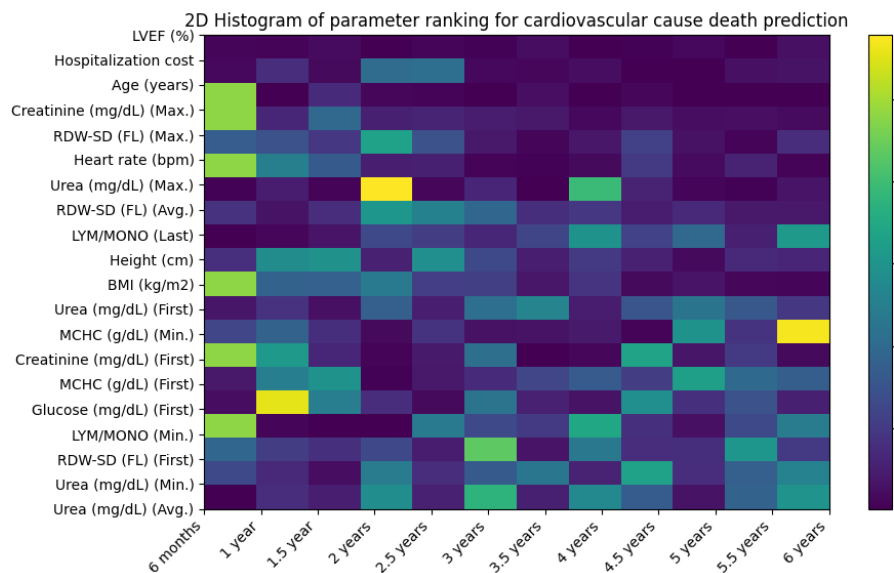


Figure 4 – Most important parameters in the prediction of cardiovascular cause and all-cause mortality (lower is better).



3

4

Figure 5 – 4D time-dependent SHAP mesh plot for various predictors of all-cause mortality.

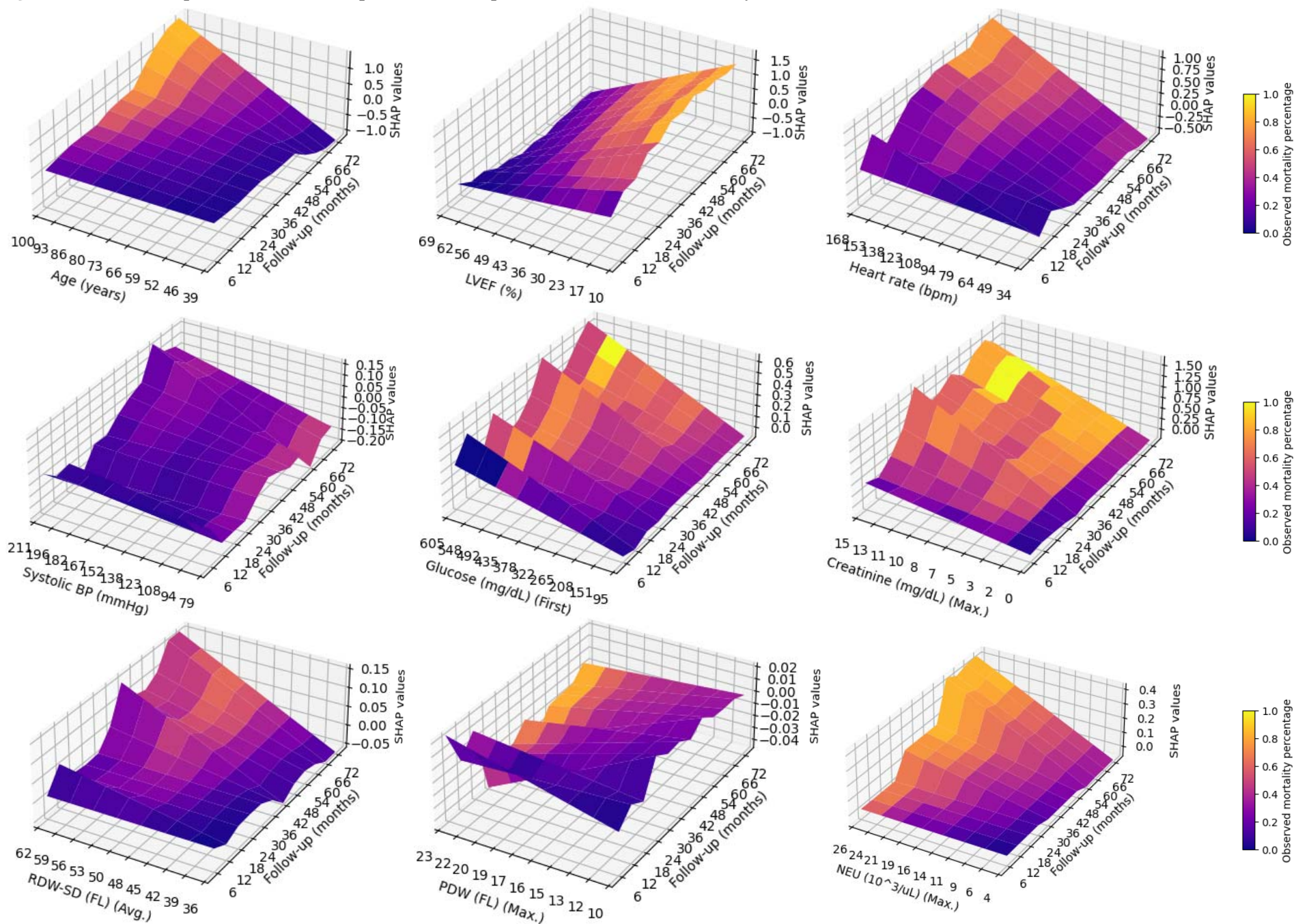


Figure 6 – An example of a patient specific estimated survival curve.

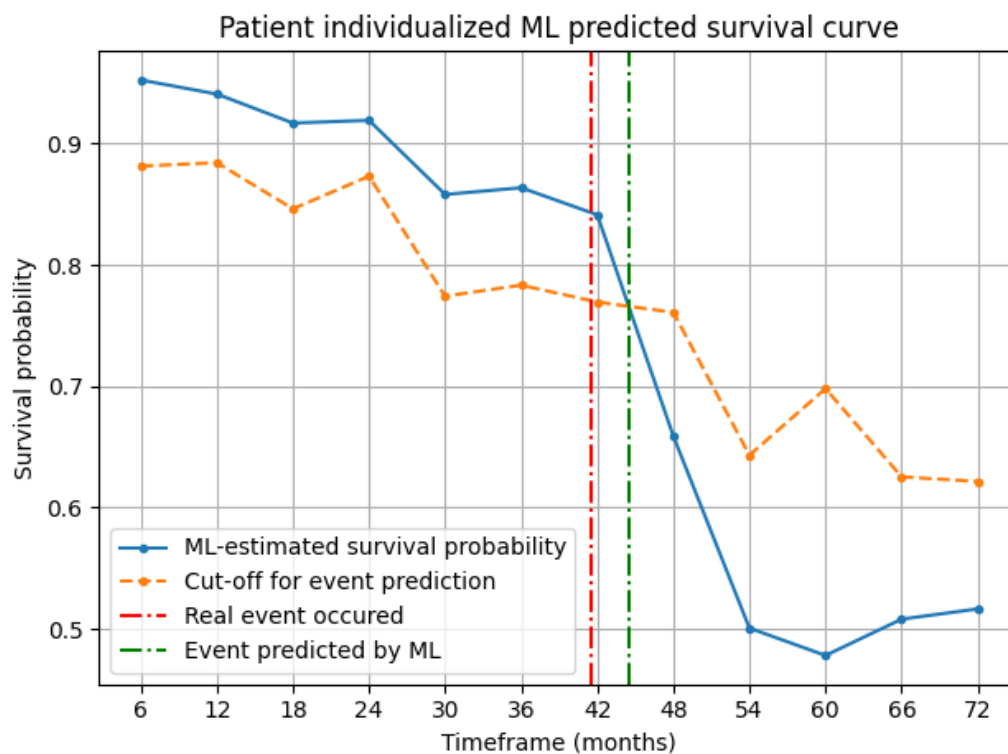


Table 1 – Baseline characteristics of the studied parameters on a per patient analysis.

Parameter	All patients (n=5797)	All-cause survival			CV-cause survival		
		HR	95%CI	p	HR	95%CI	P
Male sex	4062 (70.07%)	1.04	0.91-1.19	0.6	0.99	0.85-1.17	0.93
Age (years)	65 (57-71)	2.56	2.24-2.92	<10 ⁻⁴²	2.64	2.25-3.1	<10 ⁻³¹
BMI (kg/m ²)	28 (25-31)	0.89	0.77-1.02	0.1	0.87	0.74-1.04	0.13
History of MI	1140 (19.67%)	1.16	1.01-1.33	0.04	1.24	1.05-1.47	<10 ⁻²
CABG	154 (2.66%)	1.06	0.75-1.5	0.73	1.13	0.76-1.68	0.56
Previous PCI	1284 (22.15%)	0.89	0.77-1.03	0.13	0.89	0.75-1.06	0.18
Stroke/TIA	386 (6.66%)	2.13	1.76-2.57	<10 ⁻¹⁴	2.15	1.71-2.69	<10 ⁻¹⁰
Atrial fibrillation	695 (11.99%)	2.14	1.84-2.5	<10 ⁻²²	2.33	1.95-2.78	<10 ⁻¹⁹
PAD	607 (10.47%)	2.02	1.73-2.35	<10 ⁻¹⁸	1.83	1.52-2.22	<10 ⁻⁹
CKD	415 (7.16%)	3.46	2.93-4.09	<10 ⁻⁴⁷	3.68	3.02-4.47	<10 ⁻³⁸
VHD	332 (5.73%)	2.16	1.78-2.62	<10 ⁻¹⁴	2.35	1.88-2.94	<10 ⁻¹³
DCM	481 (8.30%)	1.88	1.58-2.25	<10 ⁻¹¹	2.23	1.82-2.73	<10 ⁻¹⁴
COPD	531 (9.16%)	2.4	2.05-2.82	<10 ⁻²⁶	2.58	2.14-3.11	<10 ⁻²²
Diabetes Mellitus	1072 (18.49%)	1.54	1.35-1.77	<10 ⁻⁹	1.51	1.29-1.78	<10 ⁻⁶
HBP	2873 (49.56%)	1.19	1.05-1.36	<10 ⁻²	1.21	1.04-1.42	0.01
Dyslipidaemia	2236 (38.57%)	0.97	0.86-1.1	0.67	0.93	0.8-1.08	0.32
LVEF (%)	50 (40-55)	0.46	0.4-0.54	<10 ⁻²¹	0.41	0.34-0.5	<10 ⁻¹⁸
NTproBNP(pg/ml)	934(212-2831)	2.94	1.66-5.21	<10 ⁻³	1.82	0.98-3.36	0.06
Creatinine(mg/dL)	0.91 (0.79-1.1)	1.81	1.59-2.05	<10 ⁻¹⁹	1.97	1.69-2.3	<10 ⁻¹⁷
HGB (g/dL)	13.8 (12.7-14)	0.55	0.49-0.63	<10 ⁻¹⁸	0.55	0.47-0.64	<10 ⁻¹³
PLT (×10 ³ /μL)	227 (188-272)	0.99	0.88-1.12	0.89	0.95	0.82-1.1	0.5
WBC (×10 ³ /μL)	8.3(6.7-10.4)	1.25	1.1-1.42	<10 ⁻³	1.31	1.13-1.52	<10 ⁻³
STEMI	1574 (27.15%)	1.26	1.1-1.45	<10 ⁻³	1.31	1.12-1.55	<10 ⁻²
NSTE-MI	517 (8.92%)	1.64	1.35-1.98	<10 ⁻⁶	1.73	1.39-2.16	<10 ⁻⁵
UA	1064 (18.35%)	0.9	0.76-1.05	0.18	0.81	0.67-0.99	0.04
CCS	2642 (45.58%)	0.75	0.66-0.85	<10 ⁻⁵	0.75	0.64-0.87	<10 ⁻³

BMI – body mass index; CABG – coronary artery bypass graft; CCS – chronic coronary syndrome; CI – confidence interval; CKD – chronic kidney disease; COPD – chronic obstructive pulmonary disease; CV – cardiovascular; DCM – dilated cardiomyopathy; HGB – haemoglobin; HBP – high blood pressure; LVEF – left ventricular ejection fraction; MI – myocardial infarction; PAD – peripheral artery disease; PCI – percutaneous coronary intervention; PLT – platelet count; NSTE-MI – non-ST segment elevation acute MI; STEMI – ST segment elevation acute MI; TIA – transient ischemic attack; UA – unstable angina; VHD – valvular heart disease; WBC – white blood cells count.

Table 2 – Baseline characteristics of the studied parameters on a per PCI analysis.

Parameter	All PCIs (n=7186)	All-cause survival			CV-cause survival		
		HR	95%CI	p	HR	95%CI	p
Male sex	5068 (70.53%)	1.01	0.89-1.14	0.93	0.91	0.79-1.05	0.2
Age (years)	64 (57-71)	2.52	2.23-2.84	<10 ⁻⁵⁰	2.63	2.27-3.04	<10 ⁻³⁷
BMI (kg/m ²)	28.3 (25.8-31)	0.91	0.8-1.03	0.15	0.85	0.73-1	0.05
History of MI	1549 (21.56%)	1.16	1.02-1.31	0.02	1.23	1.06-1.42	<10 ⁻²
CABG	199 (2.77%)	1	0.73-1.37	0.99	1.04	0.72-1.51	0.83
Previous PCI	1883 (26.20%)	0.91	0.81-1.03	0.15	0.87	0.75-1.01	0.07
Stroke/TIA	490 (6.82%)	2.01	1.69-2.38	<10 ⁻¹⁴	2.04	1.66-2.51	<10 ⁻¹⁰
Atrial fibrillation	820 (11.41%)	2.07	1.79-2.38	<10 ⁻²²	2.25	1.91-2.66	<10 ⁻²⁰
PAD	780 (10.85%)	2	1.74-2.3	<10 ⁻²¹	1.91	1.61-2.26	<10 ⁻¹²
CKD	509 (7.08%)	3.52	3.03-4.1	<10 ⁻⁵⁹	3.87	3.25-4.62	<10 ⁻⁵⁰
VHD	438 (6.10%)	2.19	1.84-2.6	<10 ⁻¹⁸	2.33	1.9-2.85	<10 ⁻¹⁵
DCM	600 (8.35%)	1.95	1.66-2.29	<10 ⁻¹⁵	2.22	1.85-2.67	<10 ⁻¹⁶
COPD	633 (8.81%)	2.32	2.01-2.69	<10 ⁻²⁸	2.43	2.04-2.89	<10 ⁻²²
Diabetes Mellitus	1441 (20.05%)	1.49	1.32-1.68	<10 ⁻¹⁰	1.44	1.25-1.67	<10 ⁻⁵
HBP	3788 (52.71%)	1.17	1.03-1.31	0.01	1.15	1-1.33	0.05
Dyslipidaemia	2973 (41.37%)	0.98	0.87-1.09	0.7	0.91	0.8-1.04	0.19
LVEF (%)	50 (40-55)	0.46	0.4-0.52	<10 ⁻²⁷	0.39	0.33-0.46	<10 ⁻²⁵
NTproBNP(pg/ml)	932(206-2635)	3.33	1.93-5.76	<10 ⁻⁴	2.2	1.23-3.94	<10 ⁻²
Creatinine(mg/dL)	0.92 (0.8-1.11)	1.86	1.66-2.09	<10 ⁻²⁵	2.02	1.76-2.33	<10 ⁻²²
HGB (g/dL)	13.8 (12.6-14)	0.55	0.49-0.62	<10 ⁻²¹	0.54	0.47-0.62	<10 ⁻¹⁶
PLT (×10 ³ /μL)	226 (188-271)	1.01	0.9-1.13	0.84	0.97	0.85-1.11	0.68
WBC (×10 ³ /μL)	8.1 (6.6-10.1)	1.27	1.13-1.42	<10 ⁻⁴	1.32	1.15-1.51	<10 ⁻⁴
STEMI	1721 (23.95%)	1.29	1.13-1.46	<10 ⁻³	1.38	1.19-1.61	<10 ⁻⁴
NSTE-MI	593 (8.25%)	1.73	1.46-2.07	<10 ⁻⁹	1.91	1.56-2.34	<10 ⁻⁹
UA	1377 (19.16%)	0.96	0.83-1.1	0.52	0.9	0.76-1.07	0.22
CCS	3495 (48.64%)	0.72	0.64-0.8	<10 ⁻⁸	0.68	0.59-0.77	<10 ⁻⁷

BMI – body mass index; CABG – coronary artery bypass graft; CCS – chronic coronary syndrome; CI – confidence interval; CKD – chronic kidney disease; COPD – chronic obstructive pulmonary disease; CV – cardiovascular; DCM – dilated cardiomyopathy; HGB – haemoglobin; HBP – high blood pressure; LVEF – left ventricular ejection fraction; MI – myocardial infarction; PAD – peripheral artery disease; PCI – percutaneous coronary intervention; PLT – platelet count; NSTE-MI – non-ST segment elevation acute MI; STEMI – ST segment elevation acute MI; TIA – transient ischemic attack; UA – unstable angina; VHD – valvular heart disease; WBC – white blood cells count.

Table 3 – Performance metrics for ML models and clinical risk scores.

Parameter	Integrated AUC-ROC*		Integrated AUC-PR*		Integrated cAUC-PR	Integrated MCC	Integrated F1 score	Integrated Brier score
Cardiovascular cause death prediction								
ML on training dataset	0.900	-	0.727	-	0.546	0.583	0.664	0.094
ML on validation dataset	0.837	-	0.589	-	0.407	0.438	0.555	0.117
ACEF score	0.761	$p < 10^{-10}$	0.380	$p < 10^{-10}$	0.199	0.282	0.351	0.124
GRACE score	0.718	$p < 10^{-10}$	0.275	$p < 10^{-10}$	0.093	0.231	0.301	0.159
SYNTAX score	0.662	$p < 10^{-10}$	0.238	$p < 10^{-10}$	0.238	0.141	0.216	0.173
All-cause death prediction								
ML on training dataset	0.873	-	0.736	-	0.517	0.543	0.650	0.107
ML on validation dataset	0.844	-	0.647	-	0.428	0.485	0.612	0.125
ACEF score	0.735	$p < 10^{-10}$	0.410	$p < 10^{-10}$	0.190	0.278	0.387	0.150
GRACE score	0.695	$p < 10^{-10}$	0.295	$p < 10^{-10}$	0.075	0.220	0.328	0.175
SYNTAX score	0.637	$p < 10^{-10}$	0.255	$p < 10^{-10}$	0.036	0.137	0.241	0.186

ACEF – Age, Creatinine and Ejection Fraction score; AUC-ROC – area under the curve of receiver-operator characteristic; AUC-PR – area under the curve of precision-recall; cAUC-PR – baseline corrected area under the curve of precision-recall; GRACE – Global Registry of Acute Coronary Events; MCC – Matthews correlation coefficient; ML – machine learning; SYNTAX – Synergy Between PCI with TAXUS and Cardiac Surgery. *p values were obtained by comparing each clinical score with ML on validation dataset at each timeframe.