

# 1 Prediction of peak oxygen consumption using cardiorespiratory parameters 2 from warm-up and submaximal stage of treadmill cardiopulmonary 3 exercise test

4

5 Maciej Rosol<sup>1\*</sup>, Monika Petelczyc<sup>2</sup>, Jakub S. Gąsior<sup>3</sup>, Marcel Młyńczak<sup>1</sup>

6

7 <sup>1</sup> Warsaw University of Technology, Faculty of Mechatronics, Institute of Metrology and  
8 Biomedical Engineering, Boboli 8 Street, Warsaw, Poland

9 <sup>2</sup> Warsaw University of Technology, Faculty of Physics, Warsaw, Poland

10 <sup>3</sup> Medical University of Warsaw, Department of Pediatric Cardiology and General Pediatrics,  
11 Warsaw, Poland

12 \*maciej.rosol.dokt@pw.edu.pl

13

## 14 Abstract

15

16 This study investigates the quality of peak oxygen consumption ( $VO_{2peak}$ ) prediction based on  
17 cardiac and respiratory parameters calculated from warmup and submaximal stages of treadmill  
18 cardiopulmonary exercise test (CPET) using machine learning (ML) techniques and assesses  
19 the importance of respiratory parameters for the prediction outcome. The database consists of  
20 the following parameters: heart rate (HR), respiratory rate (RespRate), pulmonary ventilation  
21 (VE), oxygen consumption ( $VO_2$ ) and carbon dioxide production ( $VCO_2$ ) obtained from 369  
22 treadmill CPETs. Combinations of features calculated based on the HR, VE and RespRate time-  
23 series from different stages of CPET were used to create 11 datasets for  $VO_{2peak}$  prediction.  
24 Thirteen ML algorithms were employed, and model performances were evaluated using cross-  
25 validation with mean absolute percentage error (MAPE),  $R^2$  score, mean absolute error (MAE),  
26 and root mean squared error (RMSE) calculated after each iteration of the validation. The  
27 results demonstrated that incorporating respiratory-based features improves the prediction of  
28  $VO_{2peak}$ . The best results in terms of  $R^2$  score (0.47) and RMSE (5.78) were obtained for the  
29 dataset which included both cardiac- and respiratory-based features from CPET up to 85% of  
30 age-predicted  $HR_{max}$ , while the best results in terms of MAPE (10.5%) and MAE (4.63) were  
31 obtained for the dataset containing cardiorespiratory features from the last 30 seconds of  
32 warmup. The study showed the potential of using ML models based on cardiorespiratory  
33 features from submaximal tests for prediction of  $VO_{2peak}$  and highlights the importance of the  
34 monitoring of respiratory signals, enabling to include respiratory parameters into the analysis.  
35 Presented approach offers a feasible alternative to direct  $VO_{2peak}$  measurement, especially when  
36 specialized equipment is limited or unavailable.

37

38

39 **Keywords:** cardiorespiratory fitness, peak oxygen consumption, cardiorespiratory signals,  
40 submaximal exercise test, respiratory rate, machine learning, time-series modeling

## 41 1. Introduction

42  
43 Peak oxygen consumption ( $VO_{2peak}$ ) obtained through cardiopulmonary exercise test (CPET) is  
44 the gold standard measure of cardiorespiratory fitness (1). It is a reliable predictor of cardiac  
45 events, as well as lung cancer and liver transplantation survival and risk of postoperative  
46 complications (2–6). Moreover,  $VO_{2peak}$  is a predictor of sport performance (7–9) and planetary  
47 mission task performance during spaceflight (10). Although CPET is the most reliable form of  
48 test, it is costly, requires specialized personnel and advanced equipment (11).

49 While conducting CPET, heart rate (HR) data are usually obtained through  
50 electrocardiography (ECG), while respiratory rate (RespRate) and pulmonary ventilation (VE)  
51 are gathered using tight-fitting masks. Nevertheless, this data can be acquired with relative ease,  
52 using heart rate monitors or smartwatches in case of HR, and impedance pneumography (IP) in  
53 case of RespRate and VE (12,13). Moreover, CPET is physically demanding as assumes the  
54 participants' exhaustion and thus it is contraindicated for patients with acute myocardial  
55 infarction, unstable angina, uncontrolled arrhythmia causing symptoms or hemodynamic  
56 compromise, uncontrolled asthma, and other pathological conditions (11). Maximal  
57 cardiopulmonary exercise test might also interfere with an athletes training program (14).

58 Actually, thanks to: a) the growing development of aforementioned measurement  
59 devices, b) availability of simply field-based tests such as incremental shuttle walk test (15,16)  
60 and c) new statistical prediction models and equations, clinicians and/or researchers are able to  
61 estimate  $VO_{2peak}$ , and/or  $VO_{2max}$ , based on selected parameters without performing maximal  
62 CPET (17–22). Unfortunately, estimated  $VO_{2peak}$  using, e.g., only 6-min Walk Test distance  
63 demonstrated poor agreement with measured  $VO_{2peak}$  from a CPET (23). Addition of other data  
64 such as demographic, anthropometric, and functional characteristics improved the accuracy of  
65  $VO_{2peak}$  estimate based on walking tests at least in elderly patients with stable coronary artery  
66 disease (model with all variables explained 73% of  $VO_{2peak}$  variance) (24). Thus, estimation of  
67 peak oxygen consumption based on combination of demographic factors and cardiac  
68 parameters obtained during submaximal (or even not) physical effort is possible, however, may  
69 be biased.

70 Reliable and accurate estimation of  $VO_{2peak}$  without performing maximal CPET may  
71 require more input physiological data to perform more sophisticated analyses. Thus, the  
72 development of new prediction models or equations, which will be able to accurately estimate  
73  $VO_{2peak}$ , and/or  $VO_{2max}$ , and will not relies on performing maximal CPET, is still ongoing  
74 (18,25). In recent years with the growth of the popularity of machine learning tools (ML)  
75 incorporated during the data analysis phase, those techniques were also utilized for the  
76 prediction of  $VO_2$  kinetics and  $VO_{2max}$  (26,27). ML models were also used by *Szjarto et al.* for  
77 prediction of  $VO_{2peak}$  based on the anthropometric data and 2D echocardiography (2DE) (28).  
78 This approach was more accurate than a model based on anthropometric factors, however, it  
79 required performing a 2DE examination with sophisticated equipment and a trained physician.  
80 Importantly, not only the model or prediction algorithm might be important in terms of the  
81 prediction accuracy, but also the features used for the training. There are existing studies  
82 utilizing respiratory rate and ML for prediction of oxygen uptake dynamics during CPET (29–  
83 31). However, to the best of our knowledge, there have been no previous studies utilizing

84 features from cardiorespiratory time-series obtained from submaximal CPET, for the prediction  
85 of  $VO_{2peak}$  using ML models.

86 The aim of this paper was hence to investigate the quality of  $VO_{2peak}$  prediction by  
87 models based on cardiac and respiratory features obtained from different stages of CPET.  
88 Additionally, we assessed the importance of respiratory-based features included in the models  
89 for  $VO_{2peak}$  prediction.

90

## 91 2. Materials and Methods

92

### 93 2.1. Data and study population

94

95 The database of cardiorespiratory time-series acquired during treadmill maximal  
96 cardiopulmonary exercise tests presented by *Mongin et al.* was used (32,33). The database  
97 comprises 992 recordings from experiments undertaken among amateur and professional  
98 athletes in the Exercise Physiology and Human Performance Lab of the University of Málaga  
99 between 2008 and 2018 with two types of protocols: a continuous increase of treadmill speed  
100 and a graded approach. The test itself was preceded by a warmup at 5 km/h. The study was  
101 conducted according to the principles of the Declaration of Helsinki, the study protocol was  
102 approved by the Research Ethics Committee of the University of Málaga, written informed  
103 consent was obtained from the participants and all the data were analyzed anonymously.

104 During each test, the following cardiorespiratory time-series were acquired: heart rate  
105 (HR), respiratory rate (RespRate), pulmonary ventilation (VE), oxygen uptake ( $VO_2$ ) and  
106 carbon dioxide production ( $VCO_2$ ). Data were acquired on a breath-to-breath basis. HR was  
107 monitored via a 12-lead ECG (Mortara Instrument, Inc., USA), while respiratory signals were  
108 obtained using the CPX MedGraphics gas analyzer system (Medical Graphics Corporation,  
109 USA) (32).

110 Participants between 18 and 40 years old were chosen for the analysis reducing the  
111 sample size to 692. Tests only with ramp speed increments were selected in order to obtain  
112 more consistent conditions along the study population. In result, 485 recordings have left. Next,  
113 subjects who were determined as outliers based on the 1.5 interquartile range method in terms  
114 of weight, height, and  $VO_{2peak}$ , with respect to the given sex, were excluded from the study,  
115 limiting to 462 recordings. Furthermore, the obtained data was visually evaluated in order to  
116 discard measurements during which there were visible artefacts in HR acquisition (e.g., sudden  
117 drop of over 30 bpm or lack of continuity of HR time-series during CPET probably due to  
118 electrode detachment); ultimately 369 recordings became background for the analysis. The final  
119 recordings belong to 327 unique subjects (42 subjects had more than one test) including 275  
120 men and 52 women. The demographic summary of the final group is presented in Table 1.

121

122 Table 1. Descriptive statistics of the study population.

	Age [years]	Height [cm]	Weight [kg]	BMI	$VO_{2peak}$ [ml/min/kg]
Men	27.3 ± 5.8	177.4 ± 6.3	76.6 ± 8.3	24.3 ± 2.2	47.7 ± 7.5
Woman	26.9 ± 6.3	165.2 ± 6.1	62.2 ± 8.2	22.8 ± 2.3	38.1 ± 6.3

All	$27.3 \pm 5.9$	$175.5 \pm 7.6$	$74.5 \pm 9.7$	$24.1 \pm 2.3$	$46.3 \pm 8.1$
-----	----------------	-----------------	----------------	----------------	----------------

123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164

## 2.2. Modeling

Based on the aforementioned dataset, we decided to investigate the quality of  $VO_{2peak}$  prediction from different stages of CPET based on cardiac and respiratory parameters, and to assess the importance of respiratory-based features included in the modeling of  $VO_{2peak}$ . For this purpose, we utilized recorded time-series of HR, RespRate, and VE.  $VO_{2peak}$  was determined as the maximal value of the signal obtained after a 15-breath  $VO_2$  moving average window according to the recommendation presented by *Robergs et al.* (34).

As features for ML models, basic statistics such as mean, standard deviation, maximal and minimal value, median, 25<sup>th</sup> and 75<sup>th</sup> quantile, skewness, kurtosis, coefficient from linear regression, impulse and shape factors were calculated for HR, RespRate, and VE, for a given stage of the maximal CPET. On this basis, 11 datasets were created based on different combinations of parameters and CPET stages, as presented in Table 2. Our research is focused on the submaximal stage from the cardiopulmonary exercise test, which equals 85% of the maximal measured and age-predicted  $HR_{max}$  as a threshold. Studied value of HR termination is commonly used in submaximal testing (35–37). We also used both actual  $HR_{max}$  obtained during the treadmill cardiopulmonary exercise test, and age-predicted  $HR_{max}$  (220-age) in order to provide insights about the utility of the prediction of  $VO_{2peak}$  in submaximal tests without prior knowledge about the value of  $HR_{max}$  for a given subject. The example plot of the signals, alongside the threshold for all the stages of the CPET for which the features were calculated, is presented in Figure 2.

The 10-fold cross-validation (CV) was used to assess the accuracy of the prediction. In each iteration, standardization of the non-categorical features based on the mean and standard deviation from the training dataset was performed. The only feature that was not standardized was participants' sex: -1 was assigned to male, and 1 to female subjects. Different ML algorithms were utilized: Linear, Lasso and Ridge Regression, Random Forest, XGBoost, Multilayer perceptron, Epsilon-Support Vector Regression, Bayesian Ridge Regression, Bayesian Automatic Relevance Determination (ARD) Regression, Gaussian Process Regression, Gradient Boosting for Regression, Huber Regression and Theil-Sen Estimator (38–40). The hyperparameter tuning was performed for each algorithm using the grid-search technique. In each iteration of the validation, metrics like mean absolute percentage error (MAPE),  $R^2$  score, mean absolute error (MAE) and root mean squared error (RMSE) were calculated. The best model for each dataset was determined based on the lowest MAPE score (which was chosen arbitrarily) obtained from the cross-validation. For the best model, Lin concordance correlation coefficient was calculated and results were visualized as the dependency between predicted and actual values of  $VO_{2peak}$  and as Bland-Altman plot.

Metrics obtained from all datasets were pairwise compared using the Wilcoxon signed-rank test. The significance level was set to 0.05. For the calculations, Python 3.9.13 was used. The whole modeling pipeline is presented in Figure 2.

165 Table 2. Characteristics of all datasets with an indication of features belonging to individual datasets.

Dataset	subjects' demography	HR features from the last 30 s of warmup	RespRate and VE features from the last 30 s of warmup	HR features from CPET up to 85% of $HR_{max}$	RespRate and VE features from CPET up to 85% of $HR_{max}$	HR features from CPET up to 85% of age-predicted $HR_{max}$	RespRate and VE features from CPET up to 85% of age-predicted $HR_{max}$
D1	+						
D2	+	+					
D3	+	+	+				
D4	+			+			
D5	+			+	+		
D6	+	+		+			
D7	+	+	+	+	+		
D8	+					+	
D9	+					+	+
D10	+	+				+	
D11	+	+	+			+	+

167  
168  
169

170 Figure 1. Typical representation of the time-series for participants with selected fragments used  
171 in the analysis. Part A presents the linearly increasing treadmill speed, part B heart rate  
172 fluctuations, part C respiratory rate and part D pulmonary ventilation kinetics. The segment  
173 between the blue and orange dashed lines is the last 30 seconds of warmup. The segment  
174 between the orange and green lines corresponds to the section of CPET up to 85% of the age-  
175 predicted  $HR_{max}$ . Finally, the segment between the orange and red lines corresponds to the  
176 increasing workload in CPET up to 85% of the measured  $HR_{max}$ , which is marked with a red  
177 circle on the HR plot.

178  
179

180 Figure 2. Modeling pipeline applied for each dataset and algorithm.

181

### 182 2.3. Explainable AI

183

184 In order to investigate the importance of the individual features used for ML modeling,  
185 explainable artificial intelligence (XAI) tools were applied. For this purpose, the Dalex Python  
186 package was used (41). During each iteration of the cross-validation, Shapley values and model-  
187 level variable importance based on drop-out loss values were calculated on the test set. After  
188 the whole cross-validation, all Shapley values for each sample and feature, as well as mean  
189 variable importance values were visualized. For the variable importance, *model\_parts* function  
190 of *dalex.Explainer* class was used. 30 permutation rounds were performed on each variable  
191 with MAE as a loss function and no data sampling (argument *N* was equal to *None*) due to the  
192 small number of samples.

193

## 194 3. Results

195

196 The metrics obtained for the best algorithm in terms of the lowest MAPE from the cross-  
197 validation for each dataset are presented in Table 3 alongside the model names. The violin-plots  
198 of the obtained metrics for each dataset were visualized in Figure 3. The p-values from the  
199 Wilcoxon signed-rank test from a pairwise comparison of the metrics are presented in Figure  
200 3.

201

202 The lowest MAPE and MAE - 10.51% and 4.63, respectively - were obtained for dataset  
203 D11 (demographic data along with cardiac and respiratory features from the last 30 seconds of  
204 warmup and CPET up to 85% of age-predicted  $HR_{max}$ ), while the lowest RMSE and highest  $R^2$   
205 score (5.78 and 0.47, respectively) were obtained for D9 (demographic data along with cardiac  
206 and respiratory features from CPET up to 85% of age-predicted  $HR_{max}$ ). The worst prediction  
207 of  $VO_{2peak}$  in terms of all metrics was achieved by using the D1 (demographic data) dataset.  
208 Results obtained for D11 were statistically significantly better in terms of all metrics than results  
209 for all the rest of the datasets excluding D9 as presented in Figure 3. Regarding  $R^2$  score and  
210 RMSE metrics, datasets that included respiratory-based features from the part of CPET  
(irrespective of  $HR_{max}$  determination, whether measured or estimated) showed statistically

211 significant superiority over datasets lacking features based on VE and respiratory rate during  
 212 the corresponding period as presented in Figure 3. Similarly, for MAPE and MAE, datasets  
 213 containing respiratory-based features calculated up to 85% of age-predicted  $HR_{max}$   
 214 demonstrated significantly better metrics than datasets without such features.

215  
 216 Table 3. Mean and standard deviation of metrics from cross-validation for each dataset for the  
 217 model which resulted in the lowest MAPE for the given dataset. The highest metric values were  
 218 highlighted.

	MAPE [%]	$R^2$	MAE [ml/min/kg]	RMSE [ml/min/kg]	Model
D1	12.52 ± 2.11	0.26 ± 0.09	5.50 ± 0.80	6.84 ± 0.81	Ridge regression
D2	11.95 ± 1.84	0.31 ± 0.07	5.24 ± 0.84	6.61 ± 0.71	Huber regression
D3	11.63 ± 1.84	0.34 ± 0.05	5.13 ± 0.75	6.45 ± 0.65	Bayesian ARD regression
D4	11.51 ± 1.72	0.36 ± 0.06	5.07 ± 0.68	6.36 ± 0.64	Bayesian ARD regression
D5	10.86 ± 1.23	0.44 ± 0.06	4.78 ± 0.52	5.95 ± 0.51	Bayesian ARD regression
D6	11.67 ± 1.72	0.34 ± 0.07	5.15 ± 0.69	6.46 ± 0.64	Lasso regression
D7	11.10 ± 1.26	0.42 ± 0.08	4.90 ± 0.53	6.07 ± 0.50	Bayesian ARD regression
D8	11.36 ± 1.49	0.38 ± 0.06	4.99 ± 0.61	6.29 ± 0.59	Bayesian ARD regression
D9	10.54 ± 1.20	<b>0.47 ± 0.06</b>	4.64 ± 0.49	<b>5.78 ± 0.50</b>	Bayesian ARD regression
D10	11.50 ± 1.49	0.36 ± 0.07	5.06 ± 0.62	6.37 ± 0.59	Bayesian ARD regression
D11	<b>10.51 ± 1.24</b>	0.47 ± 0.07	<b>4.63 ± 0.52</b>	5.78 ± 0.52	Bayesian ARD regression

219  
 220  
 221  
 222 Figure 3. Violin-plots of the calculated metrics for each dataset with the visualization of the  
 223 metrics obtained in each iteration of the 10-fold cross-validation. Black dots represent metrics  
 224 obtained from datasets without respiratory-based features, while red dots represent these that  
 225 include such features.

226  
 227  
 228

229 Figure 4. The p-values from Wilcoxon signed-rank test from pairwise comparison of the metrics  
230 obtained from different datasets. P-values smaller than 0.05 are marked with a black  
231 background.

232  
233 The measured values of  $VO_{2peak}$  and values predicted for the dataset that obtained the lowest  
234 MAPE score (D11) were visualized in Figure 5. The Lin concordance correlation coefficient  
235 between predicted and measured  $VO_{2peak}$  values was 0.66. The Bland-Altman plot for this  
236 dataset is presented in Figure 6.

237  
238  
239 Figure 5. The plot of measured and predicted  $VO_{2peak}$  values for dataset D11. The solid black  
240 line represents the function where predicted value is equal to the measured one.

241  
242 Figure 6. Bland-Altman plot of measured (gold standard from CPET) and predicted  $VO_{2peak}$   
243 values based on the results for dataset D11.

244  
245 As the smallest mean MAPE was obtained for D11, Shapley values and feature importance  
246 were visualized for this dataset in Figures 7 and 8, respectively. The discussion of the XAI  
247 results can be found in the next section.

248  
249 Figure 7. Shapley values obtained for dataset D11. Feature names are explained in the Appendix  
250 1.

251  
252  
253 Figure 8. Variable importance for dataset D11. Feature names are explained in the Appendix 1.

#### 254 255 256 **4. Discussion**

257  
258 Considering the features calculated from HR, VE, and RespRate time-series (attainable without  
259 the specialized equipment used in CPET), it is possible to predict  $VO_{2peak}$  from a submaximal  
260 test relying on age-predicted  $HR_{max}$ , achieving a mean absolute percentage error of 10.51% (for  
261 D11), using Bayesian ARD regression method. The addition of respiratory-based parameters  
262 resulted in an improvement of prediction compared to datasets based solely on the  
263 corresponding stage of the treadmill cardiopulmonary exercise test in 4 out of 5 cases in terms  
264 of  $R^2$  score and RMSE, and 2 out of 5 cases in terms of MAPE and MAE. When limiting  
265 treadmill cardiopulmonary exercise test to 85% of age-based  $HR_{max}$ , the inclusion of features  
266 based on VE and RespRate improved the prediction in terms of all the specified metrics. The  
267 fact that the best results were achieved for the dataset considering 85% aged-based  $HR_{max}$  and  
268 parameters obtained from easily accessible time-series indicates the possibility of using the  
269 presented method in clinical practice to determine  $VO_{2peak}$  without the prior knowledge of the  
270 actual  $HR_{max}$  value and the necessity to perform a maximal treadmill cardiopulmonary exercise  
271 test.



272 Obtaining  $VO_{2peak}$  from maximal CPET might be costly, time-consuming and in some  
273 cases impossible or contraindicated to carried out due to observed cardiac or pulmonary  
274 dysfunction, musculoskeletal diseases, or strict training programs. Therefore, there is a growing  
275 interest in the prediction of  $VO_{2peak}$  and/or  $VO_{2max}$  from submaximal tests (14,42–48). Our study  
276 focused on investigating ML algorithms to predict  $VO_{2peak}$  with the set of features, which could  
277 be obtained using simpler techniques than commonly used spirometry, and the significance of  
278 incorporating respiration into the prediction process. The presented results are similar or  
279 superior compared to some other presented  $VO_{2peak}$  prediction methods like WFI  $VO_{2peak}$   
280 prediction equation, deep-learning model based on 2DE, or regression models from PACER  
281 20-m shuttle run (19,28,49–52). However, in the existing literature, there are also techniques,  
282 which managed to obtain better performance like regression models based on submaximal  
283 exercise test protocol using a total body recumbent stepper (53–55). Nonetheless, in those  
284 studies more heterogeneous groups of patients were present in terms of age or health status  
285 (patients after heart failure or individuals with low to moderate risk of cardiovascular diseases).  
286 Further improvement of the prediction of  $VO_{2peak}$  might be achieved by increasing sample size,  
287 and inclusion of other parameters based on the raw signals (especially ECG) like HRV and  
288 parameters from information and causal domain (56–59).

289 Another notable aspect of the study was the utilization of XAI tools, specifically Shapley  
290 values and model-level variable importance, to obtain insights into the feature importance for  
291 prediction. For most datasets (including D9 and D11, which produced the best results) Bayesian  
292 ARD Regression model was used, which has an ability to automatically determine the relevance  
293 of each feature, effectively pruning irrelevant or redundant information, while accentuating the  
294 impactful variables (60). In our analysis, we found that the top five most influential features  
295 were consistent between Shapley values and variable importance. The most impactful feature  
296 of the prediction was the maximal value of VE during the test, up to 85% of age-predicted  
297  $HR_{max}$ . Additionally, subjects' weight and sex influenced the prediction results, with higher  
298  $VO_{2peak}$  observed in lighter individuals and males compared to females. Notably, 13 out of the  
299 20 features with the highest Shapley values and 10 out of the 15 features with the highest  
300 variable importance score were related to respiratory signals. Those findings seem to be in line  
301 with results presented in other studies, where the importance of respiratory signals in the context  
302 of oxygen consumption was presented (31,61,62). The presented configuration offers the  
303 benefit of avoiding monitoring  $O_2$  consumption and  $CO_2$  production through laboratory device,  
304 instead allowing for the application of less sophisticated respiratory monitoring techniques,  
305 such as IP. Simultaneous acquisition of both ECG and IP can be performed using e.g.,  
306 Pneumonitor device, which is a recently developed device, designed for research in the fields  
307 of physiology and sports medicine (12,13,63). Thus, all the cardiorespiratory features under  
308 current study could be obtained using Pneumonitor without any additional equipment.

309 There are several limitations of the study. First of all, the raw ECG/RR-intervals signals  
310 and raw respiratory curves were unavailable, and thus more sophisticated parameters and  
311 parameters from information and causal domains, which could provide additional insights into  
312 the predictive models could have not been calculated. Moreover, the sample size in this study  
313 was limited, as only 369 recordings from the initial database of 992 CPET recordings were used  
314 for analysis after applying exclusion criteria based on outlier detection methods and visual  
315 inspection of the signals. Furthermore, the dataset was imbalanced in terms of patients' sex as

316 there were 275 men and 52 women. A larger and more balanced dataset could prove beneficial  
317 for ML model training. There was also lack of information about the amount of sport activity  
318 undertaken by the participants, which might introduce inconsistency in the study population.  
319 Additionally, one approach of age-predicted  $HR_{max}$  calculation and one threshold of  $HR_{max}$  were  
320 introduced. Some of these limitations could be overcome by the usage of the Pneumonitor  
321 device, which allows for the simultaneous acquisition of raw ECG and IP signals (63). Thanks  
322 to this, the pulmonary activity (including RespRate and VE) can be monitored without the usage  
323 of sophisticated apparatus for gas analysis and tight-fitting masks may stress some groups of  
324 patients (e.g., children). Future studies may explore the optimal percentage of  $HR_{max}$  as well as  
325 other than treadmill forms of cardiopulmonary exercise tests in order to determine the optimal  
326 settings for the prediction of  $VO_{2peak}$  for clinical practice.

327 This study expands the discussion on predicting cardiorespiratory fitness by  
328 highlighting the important role of submaximal testing and incorporating respiratory signals in  
329 the prediction process. The presented analysis indicates that the inclusion of respiratory  
330 parameters might improve the quality of the  $VO_{2peak}$  prediction. The use of a submaximal test  
331 based on age-predicted  $HR_{max}$  and the utilization of cardiological and respiratory parameters  
332 that can be obtained without specialized CPET equipment is an advantage of the presented  
333 approach and facilitates its potential application in clinical practice.

334

## 335 **5. Supporting information**

336

337 Feature names presented in Figures 7 and 8 are explained in the S1 Appendix.

338

## 339 **6. Author Contributions**

340

341 **Conceptualization:** Maciej Rosoł, Monika Petelczyc

342 **Data Curation:** Maciej Rosoł, Monika Petelczyc

343 **Formal Analysis:** Maciej Rosoł

344 **Methodology:** Maciej Rosoł, Monika Petelczyc, Jakub S. Gąsior, Marcel Młyńczak

345 **Software:** Maciej Rosoł

346 **Visualization:** Maciej Rosoł

347 **Writing – Original Draft Preparation:** Maciej Rosoł

348 **Writing – Review & Editing:** Maciej Rosoł, Monika Petelczyc, Jakub S. Gąsior, Marcel  
349 Młyńczak

350

## 351 **7. Acknowledgments**

352

353 Research was funded by POB Biotechnology and biomedical engineering of Warsaw  
354 University of Technology within the Excellence Initiative: Research University (IDUB)  
355 programme.

## 356 8. References

- 357
- 358 1. Lee J, Zhang XL. Physiological determinants of VO<sub>2</sub>max and the methods to evaluate  
359 it: A critical review. Vol. 36, Science and Sports. 2021.
  - 360 2. Laukkanen JA, Rauramaa R, Salonen JT, Kurl S. The predictive value of  
361 cardiorespiratory fitness combined with coronary risk evaluation and the risk of  
362 cardiovascular and all-cause death. *J Intern Med.* 2007;262(2).
  - 363 3. Laukkanen JA, Kurl S, Salonen R, Rauramaa R, Salonen JT. The predictive value of  
364 cardiorespiratory fitness for cardiovascular events in men with various risk profiles: A  
365 prospective population-based cohort study. *Eur Heart J.* 2004;25(16).
  - 366 4. Jones LW, Watson D, Herndon JE, Eves ND, Haithcock BE, Loewen G, et al. Peak  
367 oxygen consumption and long-term all-cause mortality in nonsmall cell lung cancer.  
368 *Cancer.* 2010;116(20).
  - 369 5. Bernal W, Martin-Mateos R, Lipcsey M, Tallis C, Woodsford K, McPhail MJ, et al.  
370 Aerobic capacity during cardiopulmonary exercise testing and survival with and  
371 without liver transplantation for patients with chronic liver disease. *Liver*  
372 *Transplantation.* 2014;20(1).
  - 373 6. Snowden CP, Prentis JM, Anderson HL, Roberts DR, Randles D, Renton M, et al.  
374 Submaximal cardiopulmonary exercise testing predicts complications and hospital  
375 length of stay in patients undergoing major elective surgery. *Ann Surg.* 2010;251(3).
  - 376 7. Schabert EJ, Killian SC, St Clair Gibson A, Hawley JA, Noakes TD. Prediction of  
377 triathlon race time from laboratory testing in national triathletes. *Med Sci Sports Exerc.*  
378 2000;32(4).
  - 379 8. Billat VL, Demarle A, Slawinski J, Paiva M, Koralsztein JP. Physical and training  
380 characteristics of top-class marathon runners. *Med Sci Sports Exerc.* 2001;33(12).
  - 381 9. Staib JL, Im J, Caldwell Z, Rundell KW. Cross-Country Ski Racing Performance  
382 Predicted by Aerobic and Anaerobic Double Poling Power. *J Strength Cond Res.*  
383 2000;14(3).
  - 384 10. Sutterfield SL, Alexander AM, Hammer SM, Didier KD, Caldwell JT, Barstow TJ, et  
385 al. Prediction of Planetary Mission Task Performance for Long-Duration Spaceflight.  
386 *Med Sci Sports Exerc.* 2019;51(8).
  - 387 11. Levett DZH, Jack S, Swart M, Carlisle J, Wilson J, Snowden C, et al. Perioperative  
388 cardiopulmonary exercise testing (CPET): consensus clinical guidelines on indications,  
389 organization, conduct, and physiological interpretation. *Br J Anaesth.* 2018;120(3).
  - 390 12. Młynczak M, Zylinski M, Niewiadomski W, Cybulski G. Ambulatory Devices  
391 Measuring Cardiorespiratory Activity with Motion. In: *BIODEVICES 2017 - 10th*  
392 *International Conference on Biomedical Electronics and Devices, Proceedings; Part of*  
393 *10th International Joint Conference on Biomedical Engineering Systems and*  
394 *Technologies, BIOSTEC 2017.* 2017.
  - 395 13. Młynczak MC, Niewiadomski W, Zyliński M, Cybulski GP. Ambulatory impedance  
396 pneumography device for quantitative monitoring of volumetric parameters in  
397 respiratory and cardiac applications. In: *Computing in Cardiology.* 2014.
  - 398 14. Petelczyc M, Kotlewski M, Bruhn S, Weippert M. Maximal oxygen uptake prediction  
399 from submaximal bicycle ergometry using a differential model. *Sci Rep [Internet].*  
400 2023;13(1):11289. Available from: <https://doi.org/10.1038/s41598-023-38089-7>
  - 401 15. Marsico A, Corso SD, de Carvalho EF, Arakelian V, Phillips S, Stirbulov R, et al. A  
402 more effective alternative to the 6-minute walk test for the assessment of functional  
403 capacity in patients with pulmonary hypertension. *Eur J Phys Rehabil Med.*  
404 2021;57(4).

- 405 16. Alves R, Lima MM, Fonseca C, Dos Reis R, Figueiredo PH, Costa H, et al. Peak  
406 oxygen uptake during the incremental shuttle walk test in a predominantly female  
407 population with Chagas heart disease. *Eur J Phys Rehabil Med.* 2016;52(1).
- 408 17. Buttar KK, Saboo N, Kacker S. A Review: Maximal Oxygen Uptake (VO<sub>2</sub> Max) and  
409 Its Estimation Methods. *International Journal of Physical Education, Sports and Health.*  
410 2019;6(6).
- 411 18. Lima LP, Leite HR, de Matos MA, Neves CDC, da Silva Lage VK, da Silva GP, et al.  
412 Cardiorespiratory fitness assessment and prediction of peak oxygen consumption by  
413 Incremental Shuttle Walking Test in healthy women. *PLoS One.* 2019;14(2).
- 414 19. Selland CA, Kelly J, Gums K, Meendering JR, Vukovich M. A Generalized Equation  
415 for Prediction of VO<sub>2</sub>peak from a Step Test. *Int J Sports Med.* 2021;42(9).
- 416 20. Cureton KJ, Sloniger MA, O'Bannon JP, Black DM, McCormack WP. A generalized  
417 equation for prediction of VO<sub>2</sub>peak from 1-mile run/walk performance. *Med Sci*  
418 *Sports Exerc.* 1995;27(3).
- 419 21. Cooper KD, Shafer AB. Validity and Reliability of the Polar A300's Fitness Test  
420 Feature to Predict VO<sub>2</sub>max. *Int J Exerc Sci.* 2019;12(4).
- 421 22. Hernandez B, Roberts B, Kodidhi A, Roelle L, Miller N, Littell LM, et al. Evaluating  
422 accuracy of estimated VO<sub>2</sub>max with wrist worn Polar Ignite compared to peak VO<sub>2</sub> on  
423 formal cardiopulmonary exercise testing in healthy and fontan pediatric patients. *J Am*  
424 *Coll Cardiol.* 2023;81(8).
- 425 23. Chirico D, Davidson TW, Terada T, Scott K, Keast ML, Reid RD, et al. Using the 6-  
426 min walk test to monitor peak oxygen uptake response to cardiac rehabilitation in  
427 patients with heart failure. *J Cardiopulm Rehabil Prev.* 2020;40(6).
- 428 24. Mandic S, Walker R, Stevens E, Nye ER, Body D, Barclay L, et al. Estimating exercise  
429 capacity from walking tests in elderly individuals with stable coronary artery disease.  
430 *Disabil Rehabil.* 2013;35(22).
- 431 25. Wiecha S, Kasiak PS, Szwed P, Kowalski T, Cieśliński I, Postuła M, et al. VO<sub>2</sub>max  
432 prediction based on submaximal cardiorespiratory relationships and body composition  
433 in male runners and cyclists: a population study. Löllgen H, Barton M, Löllgen H,  
434 Knechtle B, editors. *Elife [Internet].* 2023;12:e86291. Available from:  
435 <https://doi.org/10.7554/eLife.86291>
- 436 26. Ashfaq A, Cronin N, Müller P. Recent advances in machine learning for maximal  
437 oxygen uptake (VO<sub>2</sub> max) prediction: A review. Vol. 28, *Informatics in Medicine*  
438 *Unlocked.* 2022.
- 439 27. Hedge ET, Amelard R, Hughson RL. Prediction of oxygen uptake kinetics during  
440 heavy-intensity cycling exercise by machine learning analysis. *J Appl Physiol.*  
441 2023;134(6).
- 442 28. Szijarto A, Tokodi M, Fabian A, Lakatos BK, Shiida K, Tolvaj M, et al. Deep-learning  
443 based prediction of peak oxygen uptake in athletes using 2D echocardiographic videos.  
444 *Eur Heart J Cardiovasc Imaging [Internet].* 2023 Jun 1;24(Supplement\_1):jead119.244.  
445 Available from: <https://doi.org/10.1093/ehjci/jead119.244>
- 446 29. Zignoli A, Fornasiero A, Ragni M, Pellegrini B, Schena F, Biral F, et al. Estimating an  
447 individual's oxygen uptake during cycling exercise with a recurrent neural network  
448 trained from easy-to-obtain inputs: A pilot study. *PLoS One.* 2020;15(3).
- 449 30. Amelard R, Hedge ET, Hughson RL. Temporal convolutional networks predict  
450 dynamic oxygen uptake response from wearable sensors across exercise intensities.  
451 *NPJ Digit Med.* 2021;4(1).
- 452 31. Wang Z, Zhang Q, Lan K, Yang Z, Gao X, Wu A, et al. Enhancing instantaneous  
453 oxygen uptake estimation by non-linear model using cardio-pulmonary physiological  
454 and motion signals. *Front Physiol.* 2022;13.

- 455 32. Mongin D, Chabert C, Courvoisier DS, García-Romero J, Alvero-Cruz JR. Heart rate  
456 recovery to assess fitness: comparison of different calculation methods in a large cross-  
457 sectional study. *Research in Sports Medicine*. 2023;31(2).
- 458 33. Mongin D, García-Romero J, Alvero-Cruz JR. [https://physionet.org/content/treadmill-](https://physionet.org/content/treadmill-exercise-cardioresp/1.0.1/)  
459 [exercise-cardioresp/1.0.1/](https://physionet.org/content/treadmill-exercise-cardioresp/1.0.1/). 2021. Treadmill Maximal Exercise Tests from the Exercise  
460 Physiology and Human Performance Lab of the University of Malaga (version 1.0.1)  
461 PhysioNet.
- 462 34. Robergs RA, Dwyer D, Astorino T. Recommendations for improved data processing  
463 from expired gas analysis indirect calorimetry. *Sports Medicine*. 2010;40(2).
- 464 35. Noonan V, Dean E. Submaximal exercise testing: Clinical application and  
465 interpretation. Vol. 80, *Physical Therapy*. 2000.
- 466 36. Shushan T, Lovell R, Buchheit M, Scott TJ, Barrett S, Norris D, et al. Submaximal  
467 Fitness Test in Team Sports: A Systematic Review and Meta-Analysis of Exercise  
468 Heart Rate Measurement Properties. Vol. 9, *Sports Medicine - Open*. 2023.
- 469 37. Leopold E, Tuller T, Scheinowitz M. A computational predictor of the anaerobic  
470 mechanical power outputs from a clinical exercise stress test. *PLoS One*. 2023;18(5  
471 MAY).
- 472 38. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the*  
473 *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.  
474 2016.
- 475 39. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-  
476 learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011;12.
- 477 40. Chollet F. Keras (2015). URL <http://keras.io>. 2017;
- 478 41. Baniecki H, Kretowicz W, Piatyszek P, Wisniewski J, Biecek P. dalex: Responsible  
479 machine learning with interactive explainability and fairness in python. *Journal of*  
480 *Machine Learning Research*. 2021;22.
- 481 42. Mart MF, Ely EW, Tolle JJ, Patel MB, Brummel NE. Physiologic responses to exercise  
482 in survivors of critical illness: an exploratory pilot study. *Intensive Care Medicine*  
483 *Experimental* . 2022;10(1).
- 484 43. Baldasseroni S, Silverii MV, Pratesi A, Burgisser C, Orso F, Lucarelli G, et al. Cardiac  
485 Rehabilitation in Advanced aGE after PCI for acute coronary syndromes: predictors of  
486 exercise capacity improvement in the CR-AGE ACS study. *Aging Clin Exp Res*. 2022;
- 487 44. Izquierdo MC, Lopes S, Teixeira M, Polónia J, Alves AJ, Mesquita-Bastos J, et al. The  
488 Chester step test is a valid tool to assess cardiorespiratory fitness in adults with  
489 hypertension: reducing the gap between clinical practice and fitness assessments. Vol.  
490 42, *Hypertension Research*. 2019.
- 491 45. Garcia-Tabar I, Iturricastillo A, Castellano J, Cadore EL, Izquierdo M, Setuain I.  
492 Predicting Cardiorespiratory Fitness in Female Soccer Players: The Basque Female  
493 Football Cohort Study. *Int J Sports Physiol Perform*. 2022;17(1).
- 494 46. Altmann S, Neumann R, Härtel S, Woll A, Buchheit M. Using submaximal exercise  
495 heart rate for monitoring cardiorespiratory fitness changes in professional soccer  
496 players: A replication study. *Int J Sports Physiol Perform*. 2021;16(8).
- 497 47. Sartor F, Vernillo G, De Morree HM, Bonomi AG, La Torre A, Kubis HP, et al.  
498 Estimation of maximal oxygen uptake via submaximal exercise testing in sports,  
499 clinical, and home settings. Vol. 43, *Sports Medicine*. 2013.
- 500 48. Albouaini K, Egred M, Alahmar A, Wright DJ. Cardiopulmonary exercise testing and  
501 its application. Vol. 83, *Postgraduate Medical Journal*. 2007.
- 502 49. Klaren RE, Horn GP, Fernhall B, Motl RW. Accuracy of the VO<sub>2</sub>peak prediction  
503 equation in firefighters. *Journal of Occupational Medicine and Toxicology*. 2014;9(1).

- 504 50. Mahar MT, Welk GJ, Rowe DA, Crotts DJ, McIver KL. Development and Validation  
505 of a Regression Model to Estimate VO<sub>2</sub>peak from PACER 20-m Shuttle Run  
506 Performance. *J Phys Act Health*. 2016;3(s2).
- 507 51. Loe H, Nes BM, Wisløff U. Predicting VO<sub>2</sub>peak from submaximal-and peak exercise  
508 models: The HUNT 3 fitness study, Norway. *PLoS One*. 2016;11(1).
- 509 52. Harmsen WJ, Ribbers GM, Slaman J, Heijnenbrok-Kal MH, Khajeh L, van Kooten F, et  
510 al. The six-minute walk test predicts cardiorespiratory fitness in individuals with  
511 aneurysmal subarachnoid hemorrhage. *Top Stroke Rehabil*. 2017;24(4).
- 512 53. BILLINGER SA, VAN SWEARINGEN E, MCCLAIN M, LENTZ AA, GOOD MB.  
513 Recumbent Stepper Submaximal Exercise Test to Predict Peak Oxygen Uptake. *Med*  
514 *Sci Sports Exerc* [Internet]. 2012;44(8). Available from:  
515 [https://journals.lww.com/acsm-](https://journals.lww.com/acsm-msse/Fulltext/2012/08000/Recumbent_Stepper_Submaximal_Exercise_Test_to.17.aspx)  
516 [msse/Fulltext/2012/08000/Recumbent\\_Stepper\\_Submaximal\\_Exercise\\_Test\\_to.17.as](https://journals.lww.com/acsm-msse/Fulltext/2012/08000/Recumbent_Stepper_Submaximal_Exercise_Test_to.17.aspx)  
517 [x](https://journals.lww.com/acsm-msse/Fulltext/2012/08000/Recumbent_Stepper_Submaximal_Exercise_Test_to.17.aspx)
- 518 54. Herda AA, Lentz AA, Matlaga AE, Sisante JF, Billinger SA. Cross-validation of the  
519 recumbent stepper submaximal exercise test to predict peak oxygen uptake in older  
520 adults. *Phys Ther*. 2014;94(5).
- 521 55. Deka P, Pozehl BJ, Pathak D, Williams M, Norman JF, Alonso WW, et al. Predicting  
522 maximal oxygen uptake from the 6 min walk test in patients with heart failure. *ESC*  
523 *Heart Fail*. 2021;8(1).
- 524 56. Mlyńczak M. Temporal orders and causal vector for physiological data analysis. In:  
525 *Proceedings of the Annual International Conference of the IEEE Engineering in*  
526 *Medicine and Biology Society, EMBS*. 2020.
- 527 57. Rosoł M, Młyńczak M, Cybulski G. Granger causality test with nonlinear neural-  
528 network-based methods: Python package and simulation study. *Comput Methods*  
529 *Programs Biomed*. 2022;216.
- 530 58. Schulz S, Adochiei FC, Edu IR, Schroeder R, Costin H, Bär KJ, et al. Cardiovascular  
531 and cardiorespiratory coupling analyses: A review. Vol. 371, *Philosophical*  
532 *Transactions of the Royal Society A: Mathematical, Physical and Engineering*  
533 *Sciences*. 2013.
- 534 59. Rosol M, Gasior JS, Walecka I, Werner B, Cybulski G, Mlyńczak M. Causality in  
535 cardiorespiratory signals in pediatric cardiac patients. *Annu Int Conf IEEE Eng Med*  
536 *Biol Soc*. 2022;2022.
- 537 60. Wipf D, Nagarajan S. A new view of automatic relevance determination. In: *Advances*  
538 *in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*.  
539 2008.
- 540 61. Gastinger S, Sorel A, Nicolas G, Gratas-Delamarche A, Prioux J. A comparison  
541 between ventilation and heart rate as indicator of oxygen uptake during different  
542 intensities of exercise. *J Sports Sci Med*. 2010;9(1).
- 543 62. Beltrame T, Amelard R, Wong A, Hughson RL. Prediction of oxygen uptake dynamics  
544 by machine learning analysis of wearable sensors during activities of daily living. *Sci*  
545 *Rep*. 2017;7.
- 546 63. Gašior JS, Młyńczak M, Rosoł M, Wieniawski P, Walecka I, Cybulski G, et al.  
547 Validity of the Pneumonitor for RR intervals acquisition for short-term heart rate  
548 variability analysis extended with respiratory data in pediatric cardiac patients. *Kardiol*  
549 *Pol*. 2023;81(5).

550

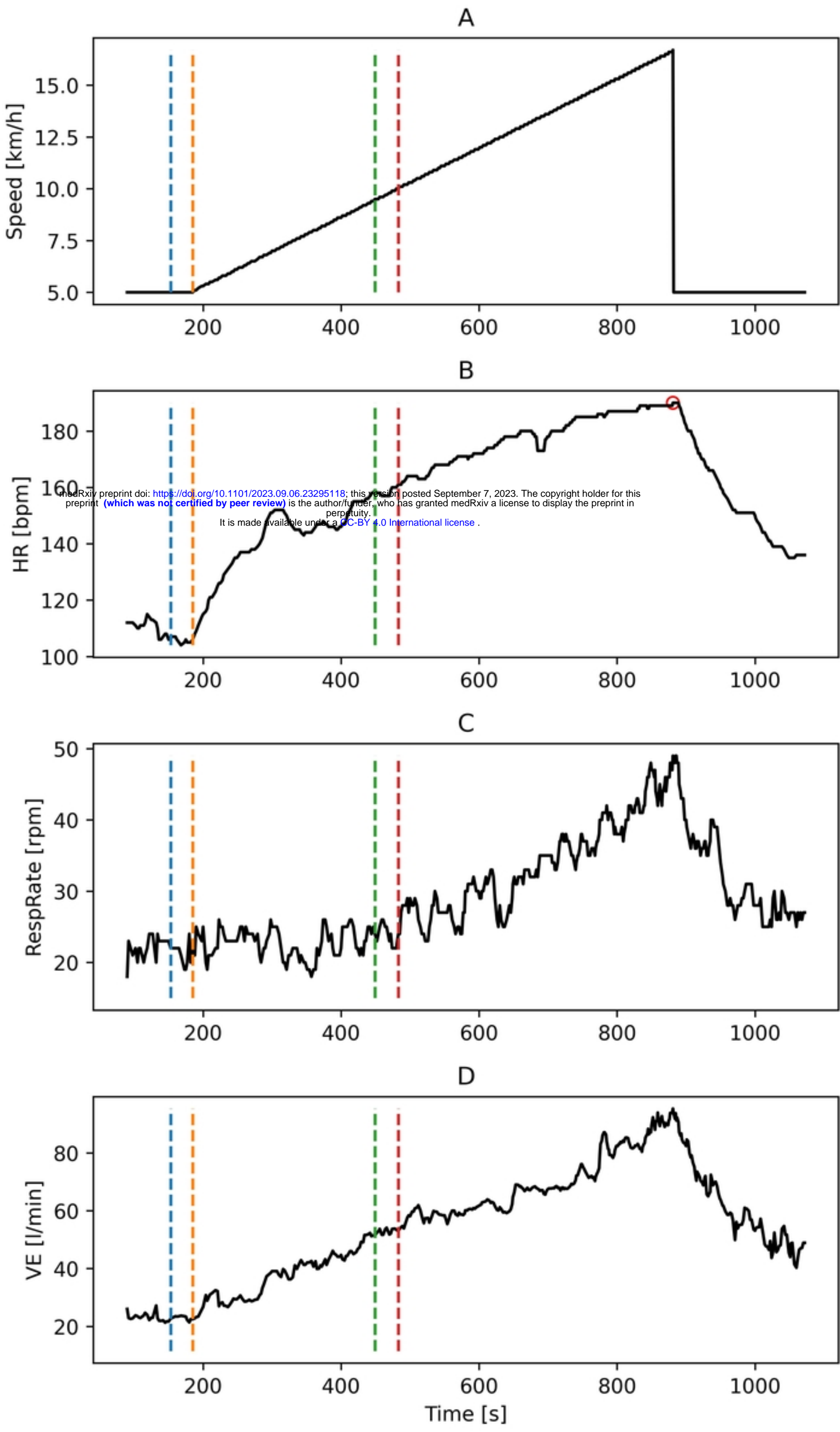


Figure 1

Dataset

10-fold  
cross-validation

Test set

Train set

<https://doi.org/10.1101/2023.09.06.23295118>  
(which was not certified by peer review)

CC-BY 4.0 International license

Standardization  
coefficients

Data standardization

Data standardization

Standardized test set

Standardized train set

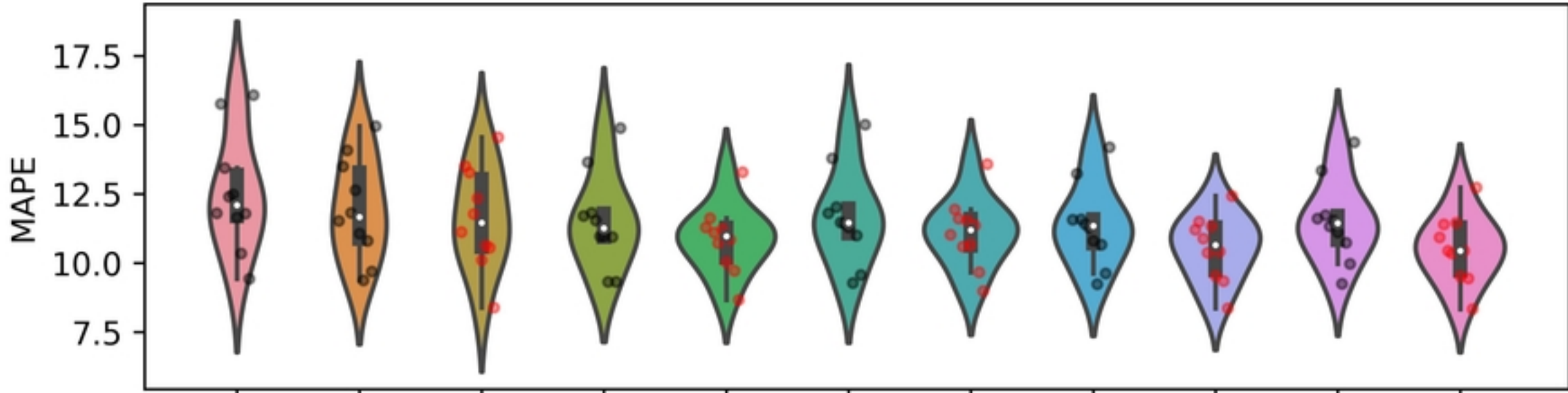
Model training

Metrics calculation

Final metrics  
calculation and  
results visualization

Figure 2





medRxiv preprint doi: <https://doi.org/10.1101/2023.09.06.23295118>; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

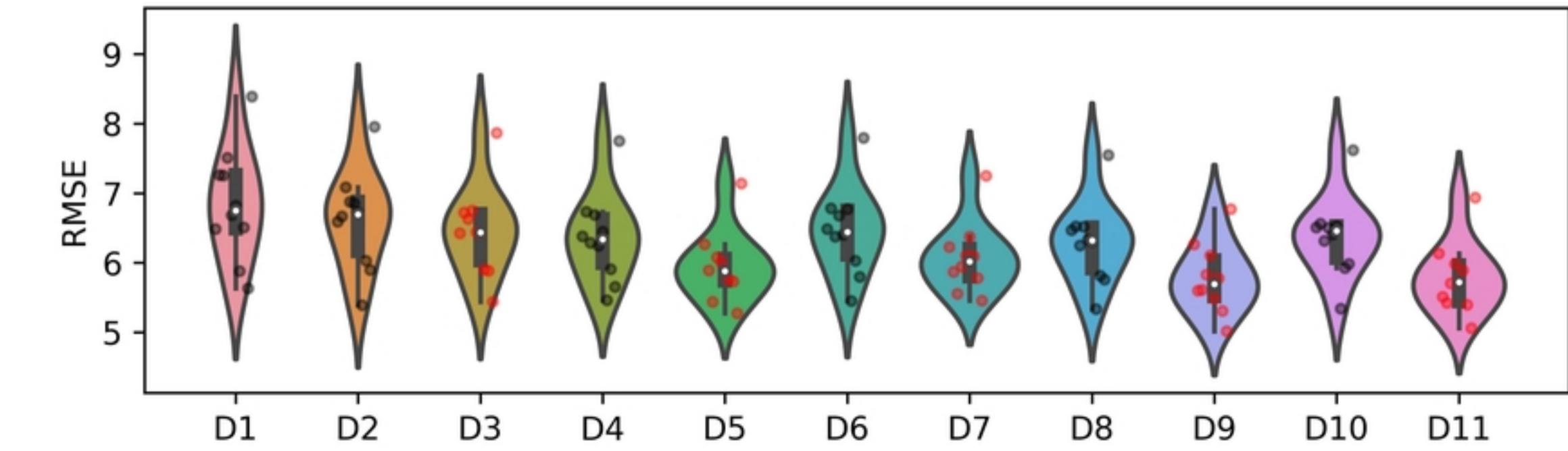
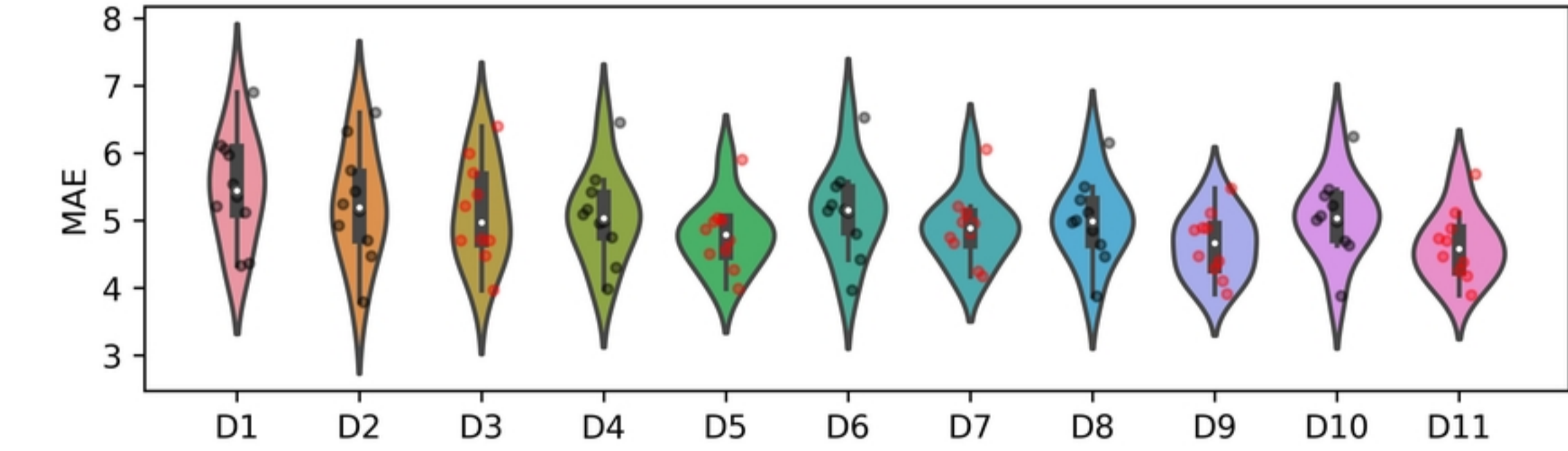
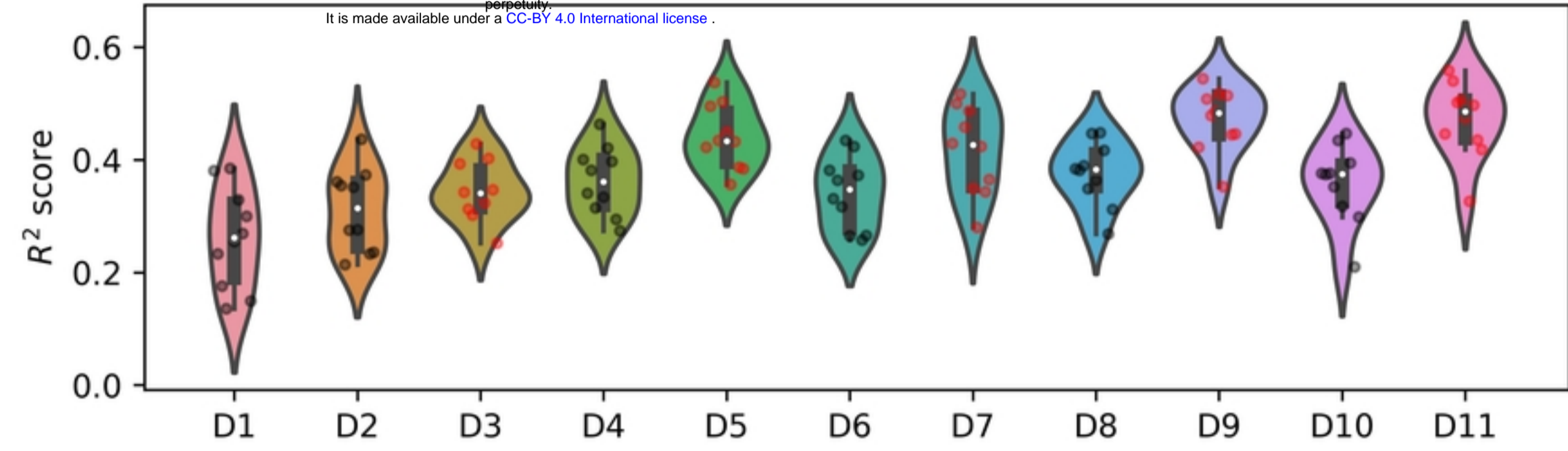
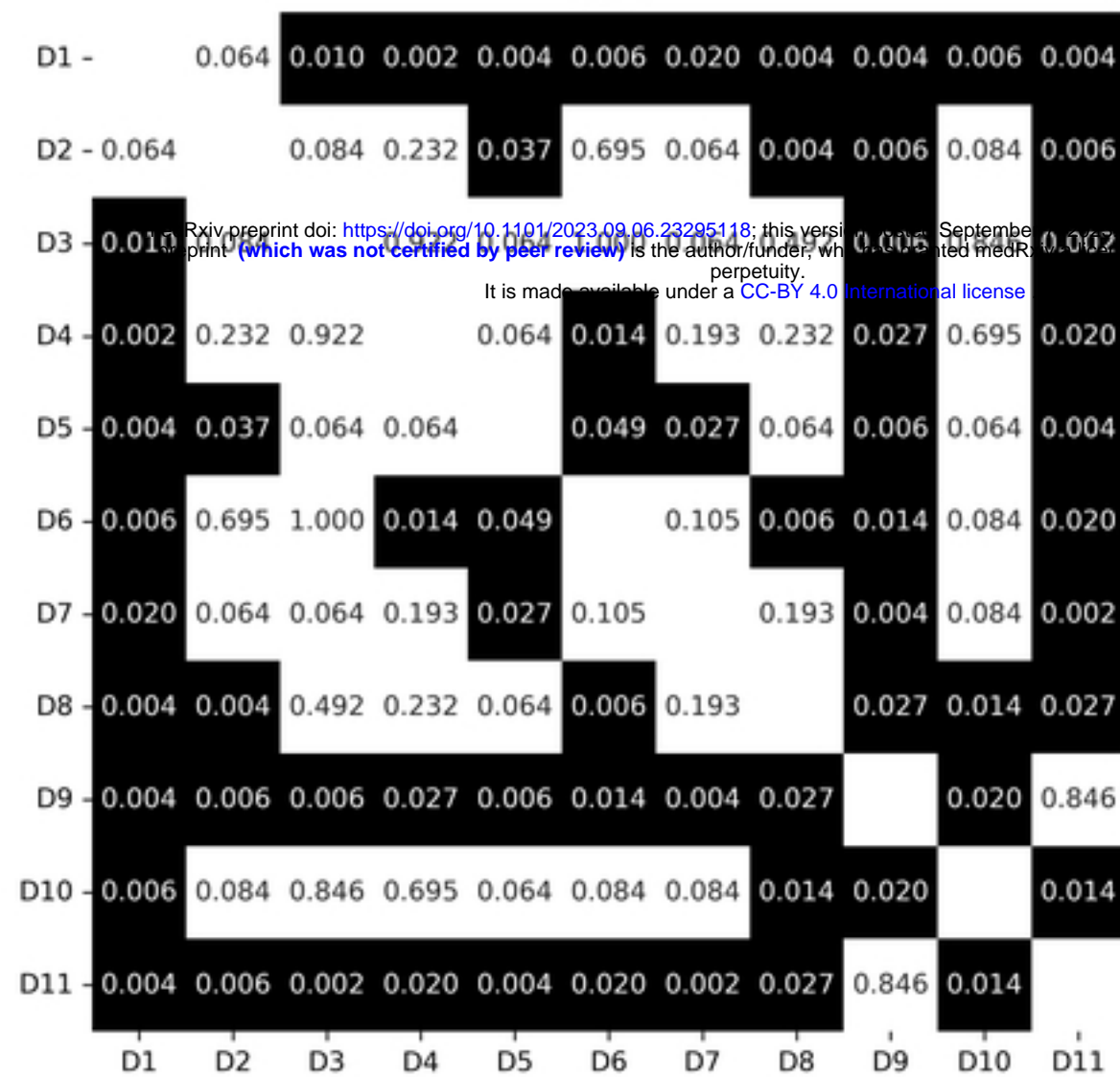
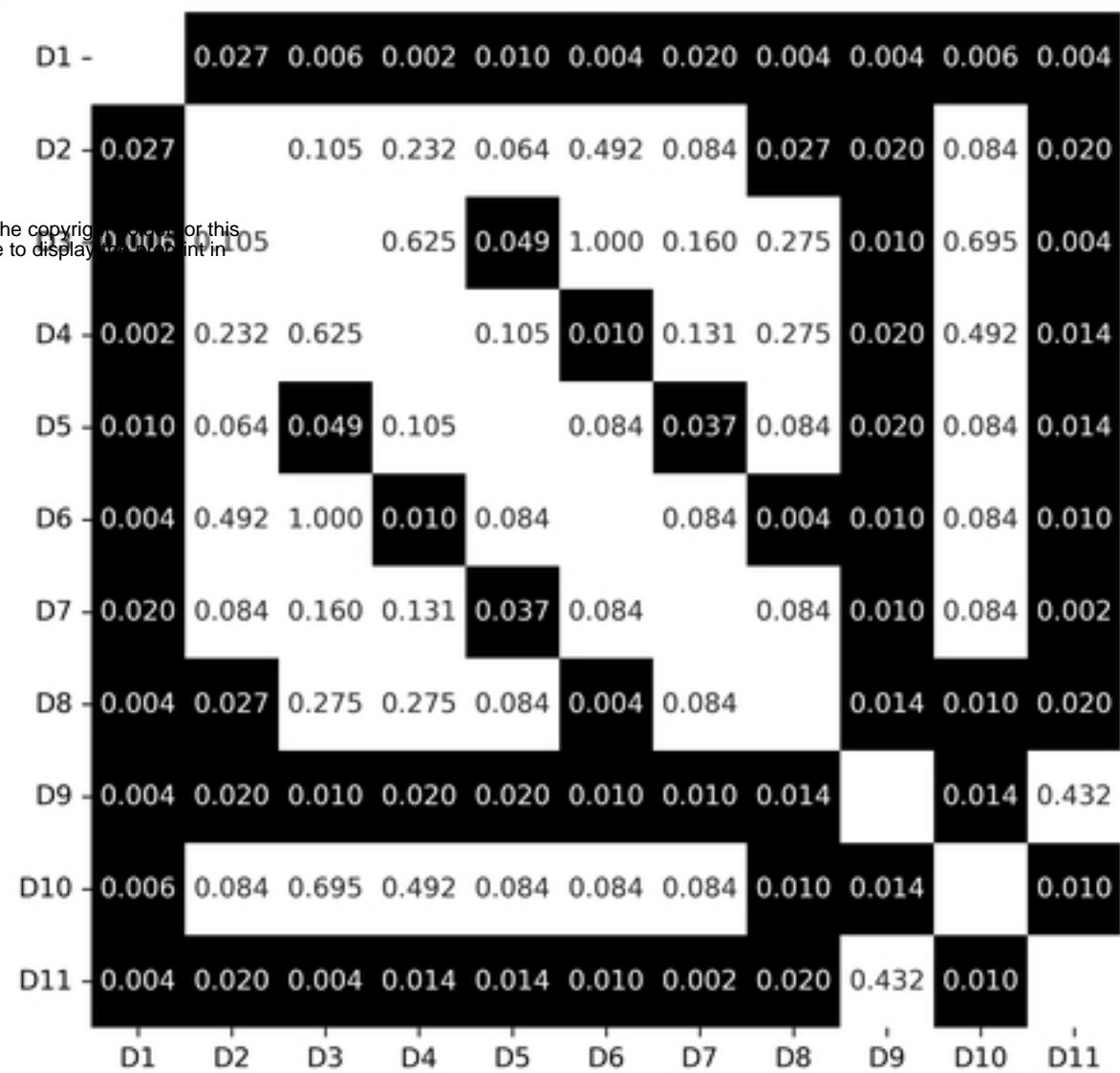
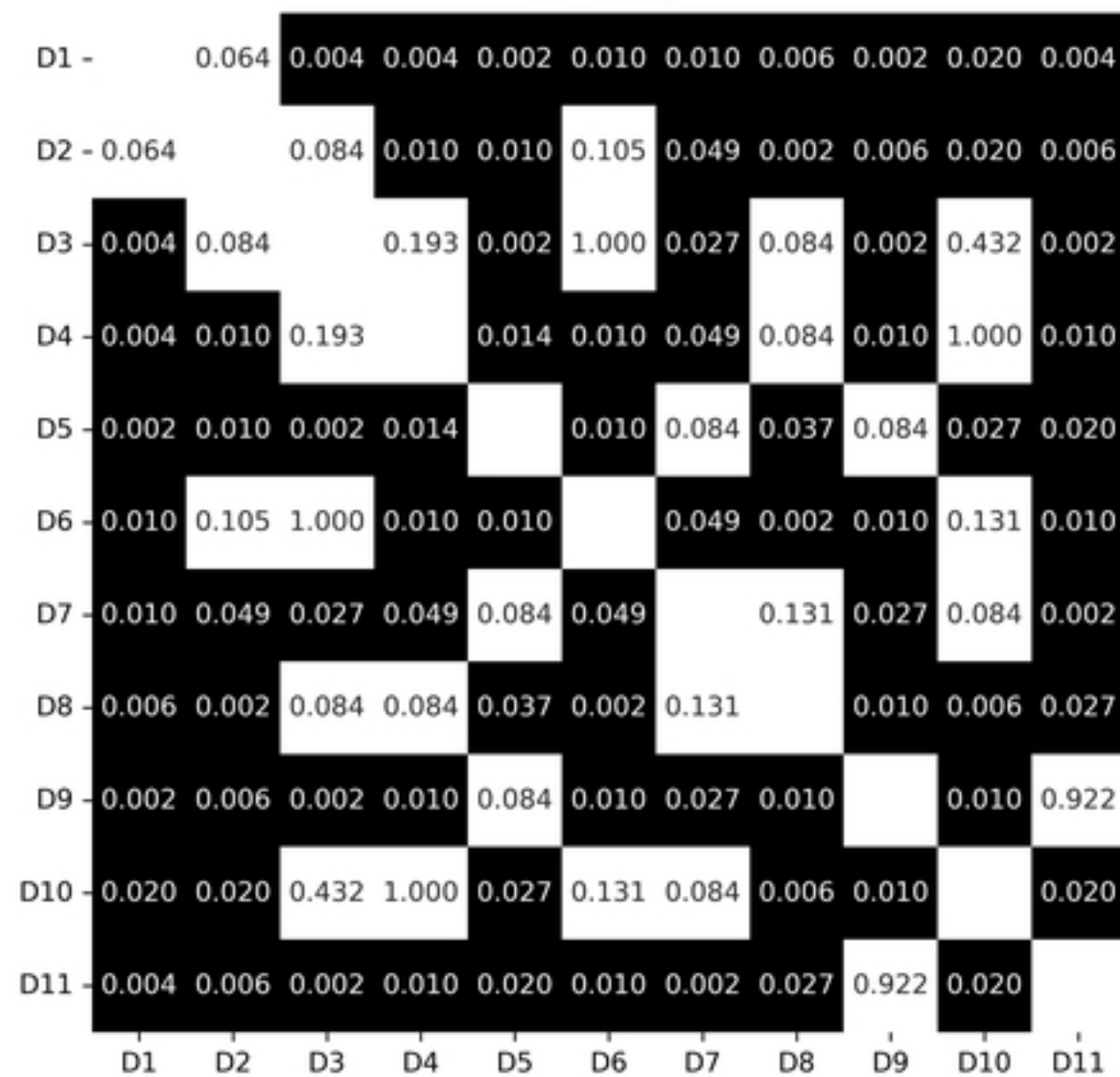


Figure 3

MAPE



MAE

 $R^2$  score

RMSE

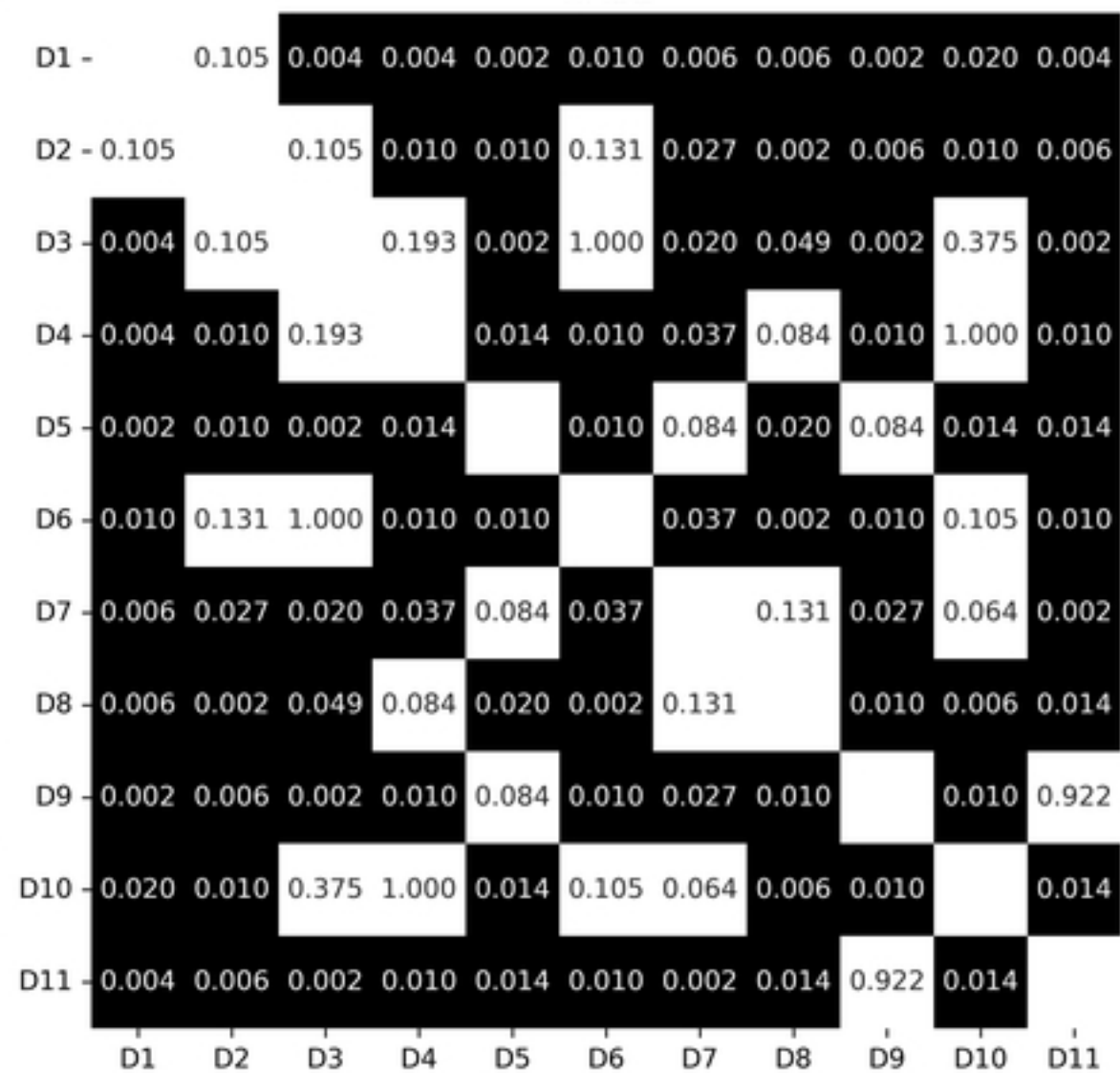


Figure 4

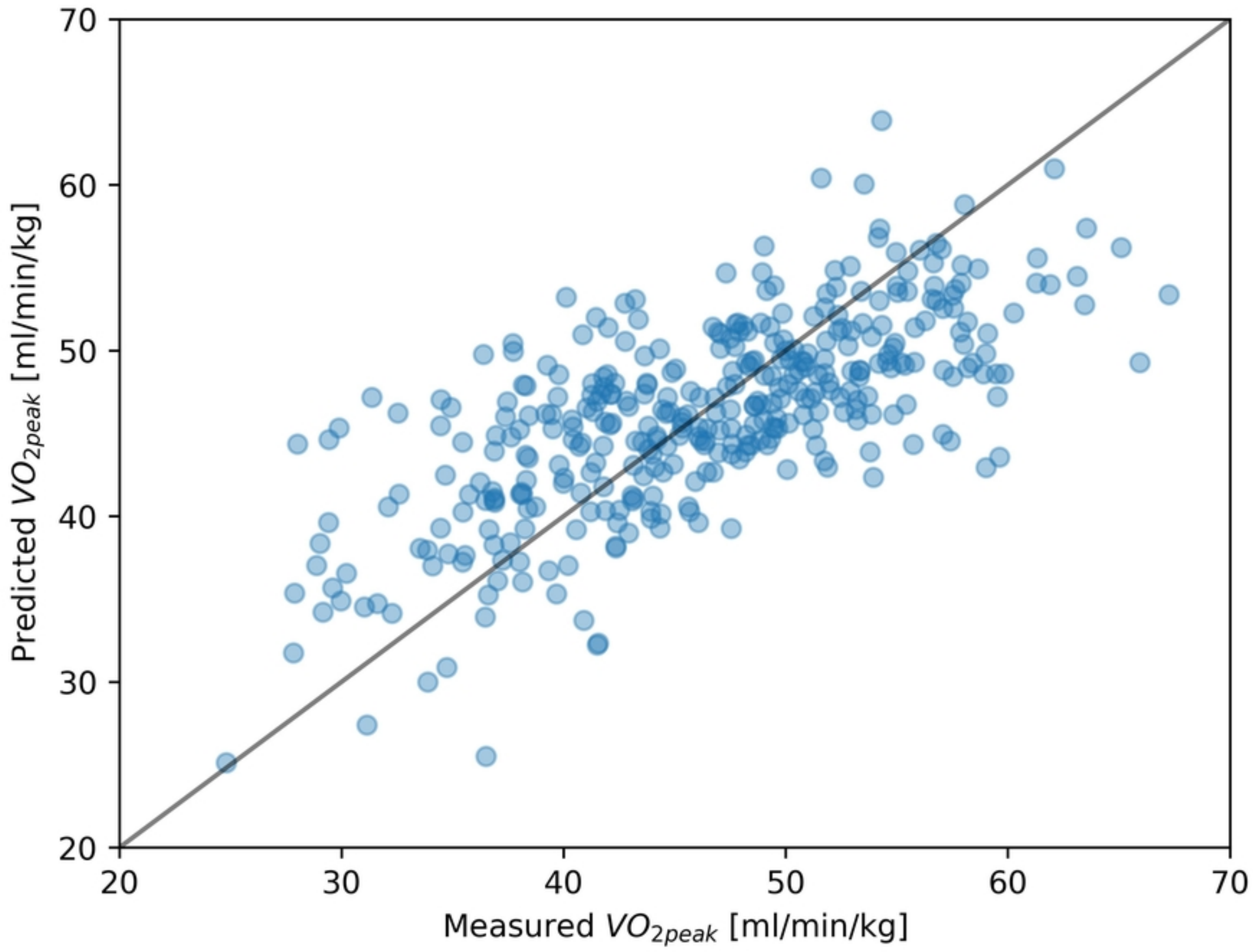


Figure 5

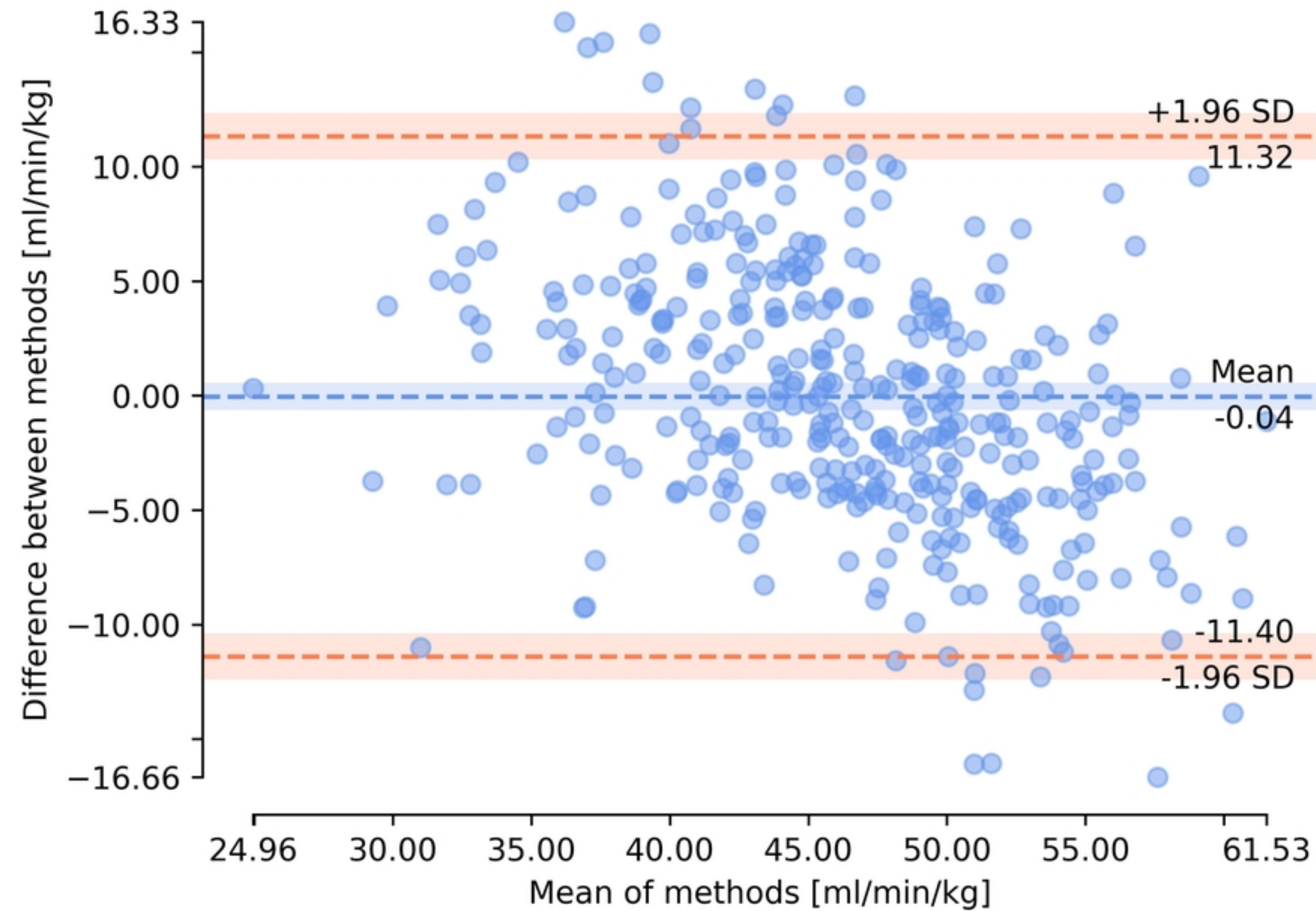


Figure 6

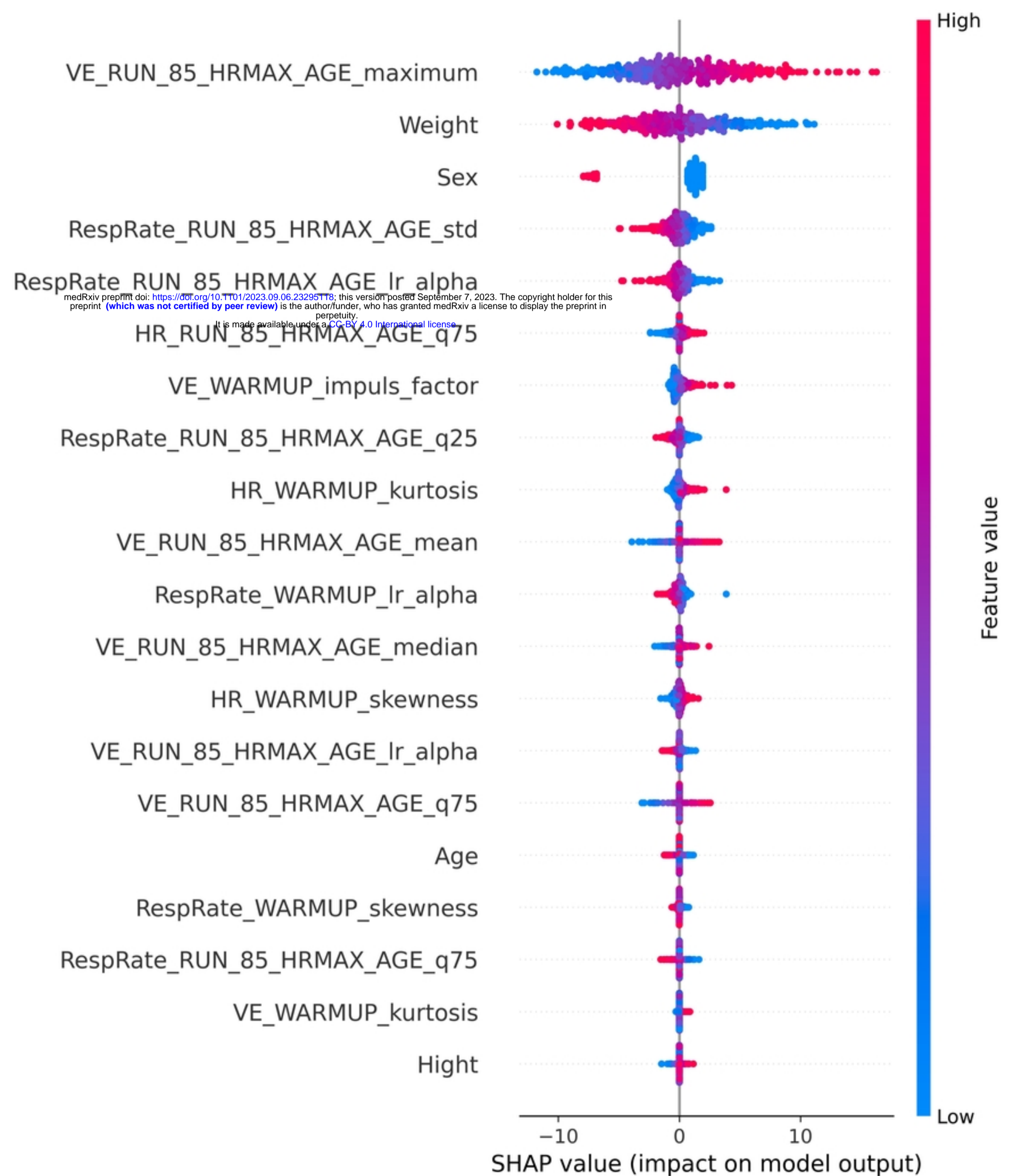


Figure 7

# Variable Importance

VO2\_peak

medRxiv preprint doi: <https://doi.org/10.1101/2023.09.06.23295118>; this version posted September 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](#).

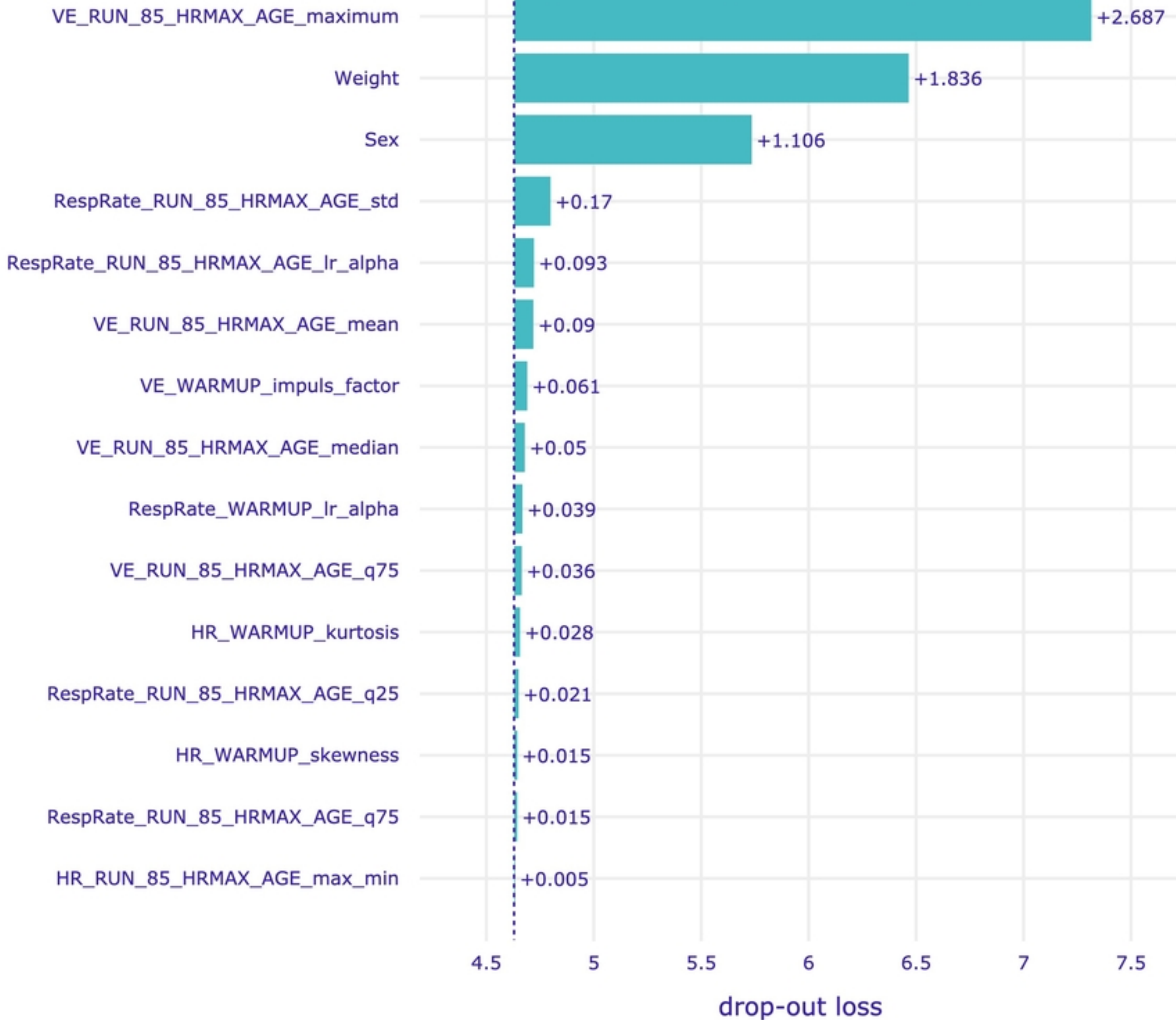


Figure 8