

1 A machine learning approach to identifying important features for achieving step  
2 thresholds in individuals with chronic stroke

3  
4 Allison E. Miller<sup>1¶\*</sup>, Emily Russell<sup>2¶</sup>, Darcy S. Reisman<sup>1,3</sup>, Hyosub E. Kim<sup>1,3&</sup>, Vu Dinh<sup>2&</sup>

5  
6 <sup>1</sup>Department of Biomechanics and Movement Science Program, University of Delaware,  
7 Newark, Delaware, United States of America

8 <sup>2</sup>Department of Mathematical Sciences, University of Delaware, Newark, Delaware,  
9 United States of America

10 <sup>3</sup>Department of Physical Therapy, University of Delaware, Newark, Delaware, United  
11 States of America

12  
13 \*Corresponding author

14 Email: [amillers@udel.edu](mailto:amillers@udel.edu) (AM)

15  
16 ¶These authors contributed equally to this work (AM and ER are co first authors)

17 &These authors contributed equally to this work (HEK and VD are co senior authors)

18

19

## 20 **Abstract**

21 **Background:** While many factors are associated with stepping activity after stroke,  
22 there is significant variability across studies. One potential reason for this variability is  
23 that there are some characteristics necessary to achieve greater stepping activity that  
24 differ from others that may need to be targeted to improve stepping activity.

25 **Objective:** Using two step thresholds (2500 steps/day, corresponding to home vs.  
26 community ambulation and 5500 steps/day, corresponding to achieving physical activity  
27 guidelines through walking), we applied 3 different algorithms to determine which  
28 predictors are most important to achieve these thresholds.

29 **Methods:** We analyzed data from 268 participants with stroke that included 25  
30 demographic, performance-based and self-report variables. Step 1 of our analysis  
31 involved dimensionality reduction using lasso regularization. Step 2 applied drop column  
32 feature importance to compute the mean importance of each variable. We then  
33 assessed which predictors were important to all 3 mathematically unique algorithms.

34 **Results:** The number of relevant predictors was reduced from 25 to 7 for home vs.  
35 community and from 25 to 16 for aerobic thresholds. Drop column feature importance  
36 revealed that 6 Minute Walk Test and speed modulation were the only variables found  
37 to be important to all 3 algorithms (*primary characteristics*) for each respective  
38 threshold. Other variables related to readiness to change activity behavior and physical

39 health, among others, were found to be important to one or two algorithms (*ancillary*  
40 *characteristics*).

41 **Conclusions:** Addressing physical capacity is *necessary but not sufficient* to  
42 achieve important step thresholds, as *ancillary characteristics*, such as readiness to  
43 change activity behavior and physical health may also need to be targeted. This  
44 delineation may explain heterogeneity across studies examining predictors of stepping  
45 activity in stroke.

46

47

48

49

50

51

52

53

54

55

56

57

## 58 Introduction

59 Stroke is a leading cause of disability world-wide and results in numerous  
60 sequelae, including reduced walking ability and aerobic deconditioning [1, 2]. This is  
61 problematic because reduced walking ability and aerobic deconditioning are associated  
62 with deficits in physical function [3, 4], depression [5, 6], and reduced self-efficacy [7].  
63 As a result, many individuals with stroke are inactive [8] and not meeting physical  
64 activity recommendations to maximize health benefits [9, 10]. In parallel, individuals with  
65 stroke often report improving their walking ability as a primary goal for rehabilitation [11]  
66 and clinicians spend considerable time on interventions to improve their walking [12].  
67 Thus, two areas of particular relevance for the rehabilitation community are determining  
68 predictors of daily stepping activity that may inform whether an individual with stroke will  
69 be able to walk in the community or if they will be primarily home bound [3, 13, 14] and  
70 whether they will meet aerobic activity guidelines through walking [10, 15]. The latter is  
71 salient for clinicians as reaching physical activity recommendations may have  
72 implications for future health outcomes [16-19].

73 Previous work has suggested that ~2500 daily steps distinguishes between home  
74 versus community ambulators in individuals with stroke [3] and that ~5500 daily steps is  
75 a reasonable target for individuals with disabilities to meet physical activity guidelines  
76 [10]. However, there has been significant heterogeneity in predictors of daily stepping  
77 activity after stroke. A recent meta-analysis including 26 studies and over 30 predictors  
78 of stepping activity post stroke found inconsistencies in the relevance of certain  
79 predictors [20]. This finding, in conjunction with the limited efficacy of interventions

80 targeting daily stepping activity post stroke [21], suggests a need to better understand  
81 which predictors are most important for improving daily stepping activity after stroke.

82 To this end, the large number of variables that may influence walking activity  
83 after stroke requires analytical techniques with the ability to handle large,  
84 heterogeneous datasets. Machine learning techniques have this capacity as well as  
85 other advantages, including requiring fewer assumptions about the distributions of the  
86 data, numerous options for non-parametric models and dimensionality reduction  
87 techniques, and most notably their strong predictive capabilities [22-24]. Recent work  
88 has utilized machine learning to predict recovery of upper limb functioning [23, 24] and  
89 functional outcomes after stroke with high accuracy [25]. In particular, one approach to  
90 determining which predictors may be most relevant is to utilize multiple different  
91 machine learning algorithms and compare predictors across algorithms [24]. This  
92 approach can help gain a better understanding of a target population and increase  
93 predictive power, as features that are shown to be important across mathematically  
94 unique algorithms likely represent fundamental characteristics of that population and are  
95 therefore important in predicting the outcome. Said differently, comparing relevant  
96 predictors across mathematically unique algorithms may help differentiate predictors  
97 that are *most important* for predicting the outcome from predictors that *might* be  
98 important for predicting the outcome in some individuals.

99 Despite these clear advantages, principled application of ML algorithms has not  
100 been used to understand the most important characteristics of stroke survivors who  
101 achieve stepping thresholds for community mobility status [3] or physical activity  
102 recommendations [10]. Thus, we had two objectives for this study. First, we aimed to

103 apply three mathematically unique algorithms to a large dataset to predict achievement  
104 of community ambulation status (2500 steps/day) [3] and/or meeting physical activity  
105 recommendations (5500 steps/day) [10] and compare predictors across algorithms. We  
106 were specifically interested in identifying predictors that were found to be important to all  
107 three algorithms. Based on previous studies, we hypothesized that measures of  
108 physical capacity, specifically gait endurance [3, 26] (6 Minute Walk Test) and gait  
109 speed [14, 27] (10-Meter Walk Test), would be important predictors to all three  
110 algorithms and that balance self-efficacy [28, 29] (Activities Specific Balance  
111 Confidence Scale) and environmental factors [30-33] (Area Deprivation Index), would  
112 be important predictors to one or two algorithms but not all three. Our second objective  
113 was to assess the prediction accuracy of these three different machine learning  
114 algorithms for each threshold.

## 115 **Methods**

### 116 **Participants:**

117 Data was obtained from the baseline timepoint of a randomized clinical trial  
118 comparing the efficacy of specific interventions for improving daily walking activity in  
119 individuals with stroke [34]. Table 1 lists the eligibility criteria for this study. All  
120 participants signed informed consent approved by the Human Subjects Review Board at  
121 the University of Delaware prior to study participation (IRB #878153-50). This work has  
122 been conducted according to the principles expressed in the *Declaration of Helsinki*.

123

124 **Table 1. Eligibility criteria**

| Inclusion Criteria                                          | Exclusion Criteria                                                                                                                                                                                                                             |
|-------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ages 21 – 85                                                | Cerebellar stroke                                                                                                                                                                                                                              |
| ≥6 months post stroke                                       | Other neurologic conditions in addition to stroke                                                                                                                                                                                              |
| Able to walk without assistance at a gait speed of ≥0.3 m/s | Lower extremity Botulinum toxin injection < 4 months earlier                                                                                                                                                                                   |
|                                                             | Current participation in physical therapy                                                                                                                                                                                                      |
|                                                             | Inability to walk outside the home before their stroke                                                                                                                                                                                         |
|                                                             | Coronary artery bypass graft, stent placement, or myocardial infarction within the past 3 months                                                                                                                                               |
|                                                             | Musculoskeletal pain limiting activity                                                                                                                                                                                                         |
|                                                             | Inability to communicate with the investigators                                                                                                                                                                                                |
|                                                             | Inability to answer at least one orientation question correctly (item 1b of the National Institutes of Health Stroke Scale) and inability to follow at least one, two-step command (item 1c of the National Institutes of Health Stroke Scale) |

## 126 **Measures:**

127           The following continuous measures were included in the statistical analysis: 6-  
128 Minute Walk Test (6MWT) [35, 36], self-selected walking speed (SSWS) [36], fastest  
129 walking speed (FWS) [36], speed modulation (calculated as FWS – SSWS), Montreal  
130 Cognitive Assessment (MoCA) [37], Charlson Comorbidity Index (CCI, age-adjusted)  
131 [38], Patient Health Questionnaire-9 (PHQ-9) [39], Activities Specific Balance  
132 Confidence Scale (ABC) [40], body mass index (BMI), age, time since initial stroke,  
133 number of strokes, number of medications (including supplements), Area Deprivation  
134 Index state decile (ADI\_S) [41, 42], and Area Deprivation Index national percentile  
135 (ADI\_N) [41, 42].

136           The following categorical measures were included in the analysis: usual orthotic  
137 device [0= no orthotic device, 1=orthotic device], usual assistive device [0= no assistive  
138 device, 1= assistive device], living situation [0= living alone, 1= living with a family  
139 member/significant other, 2= living alone but has outside assistance daily, 3= other],  
140 marital status [0= married, 1= not married], work status [0= employed full-time, 1=  
141 employed part-time, 2= retired, 3= unemployed (includes being on disability)], years of  
142 education [0= high school ( $\leq 15$  years), college graduate (16 years), above graduate  
143 ( $>16$  years)], gender [0= male, 1= female], side of hemiparesis [0= left, 1= right, 2=  
144 bilateral], readiness to change activity behavior [1= currently not active and do not  
145 intend on becoming active in the next 6 months, 2= currently not active but thinking  
146 about starting to become active in the next 6 months, 3= currently active sometimes but  
147 not regularly, 4= currently active regularly but have only begun doing so within the last 6  
148 months, 5= currently active regularly and have done so for longer than 6 months], and



149 relapse in activity behavior [1= experienced a relapse in activity levels, 2= no relapse in  
150 activity levels] [43, 44].

### 151 **Daily stepping activity:**

152 To measure daily stepping activity, participants were provided with a FitBit™ at  
153 the baseline visit of the clinical trial. The FitBit™ has acceptable accuracy in detecting  
154 stepping activity in individuals with stroke [45-48]. Participants wore the device on their  
155 non-paretic ankle and were instructed to wear it for 7 days to reliably estimate daily  
156 stepping activity [49] and continue with their usual activity. Average steps per day  
157 (ASPD) was calculated by summing the total number of steps taken over all valid  
158 recording days and dividing this sum by the number of valid recording days.

### 159 **Statistical analysis:**

#### 160 **Data processing:**

161 Figure 1 displays a data pipeline that describes the data and analysis  
162 procedures. First, the data was exported from the electronic database, REDCap [50],  
163 and initially comprised 283 individuals and 32 clinical and demographic variables. ASPD  
164 was used to determine stepping thresholds (home vs. community threshold of 2500  
165 steps/day and the aerobic threshold of 5500 steps/day). After removing variables with  
166 more than 5% missing entries and variables that were irrelevant to the outcome (e.g.,  
167 “patient ID”), 25 of the remaining 32 variables were used in our analyses. Of the 283  
168 participants, 15 were excluded for having missing data in one or more of the 25

169 variables selected, leaving a total of 268 participants included in our analyses. Two  
170 individuals inspected the data for accuracy prior to analysis.

171

172 **Fig 1. Data pipeline.** *Abbreviations: ASPD- Average Steps/Day, LR- Logistic*  
173 *Regression, SVM- Support Vector Machine, RF- Random Forest, CI- Confidence*  
174 *Interval*

175

176 All analyses were conducted using custom-written code in the Python  
177 programming language and compiled with Spyder4, using the standard machine  
178 learning library sklearn [51]. The same procedures were repeated for both the home vs.  
179 community and aerobic thresholds. Briefly, the preprocessed data set from 268  
180 participants was imported into a data frame. Our design matrix was composed of all 25  
181 variables except for ASPD, which was used to compute our binary outcome variable,  
182 step threshold category. All data in our design matrix was normalized for stability using  
183 a min-max scaler, which uses the minimum value and range of the distribution to shift  
184 and scale the distribution, respectively, translating to the interval [0,1] while preserving  
185 the shape of the original distribution.

186 For the home versus community threshold, participants were assigned a label of  
187 1 for home ambulator ( $ASPD < 2500$ ) or 0 for community ambulator ( $ASPD \geq 2500$ ).  
188 This resulted in a distribution of 58 (21.64%) home ambulators and 210 (78.36%)  
189 community ambulators. For the aerobic threshold, those whose ASPD were below the  
190 threshold of 5500 daily steps were given a label of 1, and those who met or exceeded

191 the minimum aerobic threshold were labeled with a 0. The distribution in this case was  
192 185 (69.03%) below the aerobic threshold and 83 (30.97%) above the threshold.

### 193 **Drop-column procedure for feature importance**

194 To address the first objective of this study, a two-stage procedure was used. The  
195 first stage was dimensionality reduction using lasso regularization. The purpose of this  
196 stage was to reduce redundancy and noise in our set of variables, thus avoiding  
197 potential bias or conflation in measures of feature importance for strongly correlated  
198 variables in the second stage. The second stage involved computing a measure of  
199 feature importance for the remaining features following dimensionality reduction (Fig 1)  
200 [52]. Throughout the analysis, the performance metrics were assessed using Monte  
201 Carlo Cross-Validation (MCCV) which in each instance consisted of 100 randomly  
202 generated train-test-splits of the data where for each split, 70% of the data was used as  
203 a training set and the remaining 30% was used as the test set.

204 For the first stage, we used logistic regression (LR) and linear support vector  
205 machine (SVM) both with lasso regularization with optimized regularization parameters.  
206 To optimize these parameters, a grid search was used with two refinements, each  
207 around the parameter value that was found to produce the best average model  
208 performance using MCCV. Once the optimized regularization parameters were found,  
209 the following procedure was executed for both regularized models: first, the model was  
210 fit over 100 random 70-30 train-test-splits of the data and the 100 sets of weights (i.e.,  
211 coefficients, see Supplementary Material for more detail) for each variable were  
212 recorded. Then, for each variable, 1000 bootstrap samples of size 100 were generated

213 from the sample of 100 coefficients, and an empirical 95% confidence interval for the  
214 median of the coefficients was computed. Variables for which 0 was included in the 95%  
215 confidence interval for both regularized models were then dropped, and only the  
216 remaining variables would be used in the second stage.

217 In the second stage, a measure of feature importance was computed for three  
218 different machine learning models: LR, SVM with a radial basis function (RBF) kernel,  
219 and random forest (RF). Our criteria for choosing these three models were that they had  
220 to be commonly-used machine learning algorithms, mathematically distinct, and  
221 previously shown to perform well with clinical and biological data (see Supplementary  
222 Material for additional details) [53]. While all three of these algorithms fit this criterion, it  
223 is important to note that, unlike parametric models like LR and SVM whose models can  
224 be written down in function form, nonparametric models like RF are known for having  
225 strong predictive power, while lacking interpretability. In this way, RF could be thought  
226 of as having insight that is more complex yet could be difficult to quantify. For this stage,  
227 we aimed to use a measure of feature importance that is both easily interpretable as  
228 well as uniformly applicable across the three algorithms. We therefore chose the drop  
229 column feature importance process.

230 The drop column importance measure uses a chosen metric of model  
231 performance to quantify how much a given variable contributes to a model's  
232 performance, i.e., whether the variable helps, hurts, or has no effect on the performance  
233 of the model [54]. Once the metric for performance is chosen, the drop column feature  
234 importance can be computed using any model. The idea of this procedure is to fit the  
235 model using all variables first (i.e., the *benchmark model*) and take a measure of the

236 model's performance using the chosen metric (i.e., the *benchmark performance*). To  
237 measure the influence of an individual feature on this metric of model performance, that  
238 feature will then be dropped from the training set and the model is then refitted. A  
239 measure of this new model's performance, the *dropped column performance* is then  
240 taken, and that feature's overall *importance* is taken to be the difference between the  
241 model's performance with and without that feature, i.e.:

$$242 \quad \textit{importance} = \textit{benchmark performance} - \textit{dropped column performance}$$

243 This same procedure is then followed for each feature. Thus, if a particular feature  
244 improves the model's performance, this *importance* measure will be positive because  
245 the *dropped column performance* would be lower than the *benchmark performance* (in  
246 other words, the model performed worse when we removed that particular feature) and  
247 vice versa for features whose importance is negative. The intuition underlying this  
248 procedure is that positive features hold pertinent information about the dependent  
249 variable, as they contributed the most to correctly identifying those in the target class.

250 The drop column procedure was run with MCCV for all three algorithms using  
251 only the features that were retained after regularization. For consistency, a single list of  
252 100 train-test-splits was randomly generated and used for all three algorithms in this  
253 step. For each algorithm, every feature was given a drop column importance for every  
254 train test-split, resulting in 100 importance measures associated with each algorithm for  
255 each feature. From these samples of 100 importance scores for each feature, 1000  
256 bootstrap samples of size 100 were taken and an empirical 95% confidence interval for  
257 the mean importance score was computed for each feature. A feature was considered  
258 important to an algorithm if the 95% confidence interval for the mean importance of that

259 feature was positive. To address our first objective, we compared predictors whose  
260 mean importance was positive across all three algorithms with the framework that  
261 predictors that met this criterion are likely critically important for predicting the outcome.

## 262 **Model performance**

263 For our second objective, we compared the performance of these models by  
264 examining their prediction accuracy for each step threshold. For the aerobic threshold,  
265 the metric of prediction accuracy used was standard accuracy. Due to the nature and  
266 severity of the class imbalance in the home versus community case, particularly that the  
267 target class was the minority [55], the metric of balanced accuracy, which takes the  
268 average of the recall (sensitivity) and specificity, was used to more accurately reflect the  
269 models' performance on the target class (see Supplementary Material). With this class  
270 imbalance, balanced accuracy is better suited than standard accuracy to assess model  
271 performance because taking the average of recall and specificity takes the accurate  
272 classification of both classes into account. This avoids the case where a constant model  
273 (i.e., labeling all points as community ambulators) would result in a conflated standard  
274 accuracy score, while mis-identifying the entire target class. These models were fit  
275 using the algorithms' default parameters from the sklearn documentation for the aerobic  
276 threshold. Due to the class imbalance in the home versus community threshold, a grid  
277 search was performed to optimize only the class weights parameter, *class\_weight*, to  
278 either have the default setting of no weights or to have balanced class weights, which  
279 imposes a weight on each class during fitting that is inversely proportional to the class  
280 frequency.

281 To assess model performance, as well as validate the regularization step in the  
282 feature importance procedure, we computed the appropriate measure of prediction  
283 accuracy, again using MCCV over a single, newly generated set of 100 train-test-splits.  
284 This was done first using all the features, then using the set of features retained in the  
285 regularization phase. The goal of regularization is to eliminate redundant or  
286 uninformative variables, resulting in the use of fewer variables without the loss of any  
287 critical information. Achieving this would be evidenced by either no significant decrease,  
288 or even an improvement in the model's performance when using the variables retained  
289 after regularization versus the full set of variables. The data and code associated with  
290 this work are available on Open Science Framework at <https://osf.io/tgzpb/>

## 291 Results

292 Table 2 displays the demographic characteristics and summary step data of our  
293 full sample (n=268).

294 **Table 2. Characteristics of study sample (n = 268)<sup>a</sup>**

| Characteristic      | Participant                                                  |
|---------------------|--------------------------------------------------------------|
| Age (years)         | 65 (IQR 17)                                                  |
| Gender              | Male: 139 (51.9%)<br>Female: 129 (48.1%)                     |
| Side of Hemiparesis | Left: 142 (53%)<br>Right: 120 (44.8%)<br>Bilateral: 6 (2.2%) |

|                                     |                                     |
|-------------------------------------|-------------------------------------|
| Time Since Initial Stroke (months)  | 24 (IQR 42)                         |
| Assistive Device (yes/no)           | Yes: 126 (47%)<br>No: 142 (53%)     |
| Self-selected Gait Speed (m/s)      | 0.75 (IQR 0.32)                     |
| Average Steps per Day               | 4175 (IQR 3061.5, Range 76 – 18166) |
| Total Number of Valid Stepping Days | 8 (IQR 6, Range 3 - 27)             |

295 <sup>a</sup>Continuous variables presented as median (IQR- Interquartile range)

## 296 **Results for home versus community threshold**

297       After the regularization process, LR and linear SVM dropped all but the same 6  
298 features (6MWT, PHQ-9, readiness to change relapse score, usual assistive device,  
299 years of education, and ADI\_N) with linear SVM also keeping 1 additional feature  
300 (SSWS). This resulted in 7 of the original 25 variables proceeding to the drop column  
301 phase.

302       For both LR and SVM, all 7 variables resulting from the regularization step were  
303 found to be important, with each feature contributing at least 6% improvement to the  
304 balanced accuracy score on average for both algorithms. For RF, only 1 of the 7  
305 variables was found to be important (6MWT). Thus, the only feature that was important  
306 to all three algorithms for the home versus community threshold was 6MWT, suggesting  
307 that walking endurance is critically important for predicting community mobility status in  
308 stroke. Figure 2 displays the results of the drop column phase for this threshold.

309



310 **Fig 2. Drop column feature importance for home vs. community threshold (2500**  
311 **steps/day).** Red markers show mean feature importance with 95% bootstrapped  
312 confidence interval. 6MWT was the only feature found to be important across all three  
313 algorithms. *Abbreviations: ADI\_N- Area Deprivation Index (national percentile), PHQ-9-*  
314 *Patient Health Questionnaire-9, Readiness\_Relapse- Readiness to change relapse*  
315 *score, SSWS- self-selected walking speed, 6MWT- 6-Minute Walk Test, LR- Logistic*  
316 *regression, SVM- Support vector machine, RF- Random forest.*

317

318 Figure 3A displays the model performances for the home versus community  
319 threshold. All models demonstrated a moderate level of test-set accuracy after  
320 optimizing for class weights, where LR and SVM were fit with balanced class weights  
321 and RF was fit with none. Note that for the home versus community classification  
322 problem, we used balanced accuracy to measure model performance, meaning the  
323 scores represent how the model accurately identified individuals on average across the  
324 home and community classes. Overall, RF performed the worst, achieving an average  
325 balanced accuracy score of 68.1% (SD 6.5%, range 50.6 – 87.3%) with selected  
326 features and 67.9% (SD 5.3%, range 58.6% - 82.7%) when using all features. SVM  
327 followed, achieving an average balanced accuracy score of 77.3% (SD 5.4%, range  
328 62.7% - 90.6%) with selected features and 73.1% (SD 5.7%, range 60.8% - 84.1%)  
329 when using all features. Finally, LR achieved the best overall balanced accuracy with an  
330 average score of 78.6% (SD 4.8%, range 68.7% - 90.1%) with selected features and  
331 75.6% (SD 5.5%, range 63.3% - 87.6%) when using all features. The similarities in  
332 model performance accuracies when comparing a model with all features to the

333 simplified model following regularization across all algorithms demonstrate that the  
334 regularization phase was effective in reducing the number of features while maintaining  
335 model performance accuracy.

336

337 **Fig 3. Model performance for the home vs. community threshold (2500 steps/day;**  
338 **A- upper figure) and aerobic threshold (5500 steps/day; B- bottom figure).** Model  
339 performance for each algorithm is displayed with all features included (AF) and with  
340 feature selection (FS) that occurred as a result of the regularization step. Circles  
341 represent individual accuracy results for model performance during the 100 different  
342 train-tests splits of the data. Diamonds represent outliers. A higher accuracy score  
343 reflects better model performance.

344

## 345 **Results for aerobic threshold**

346 After the regularization process, both linear SVM and LR produced the exact  
347 same results, with 16 features retained and carried forwards into the second stage of  
348 the analysis (see Y axis of Fig. 4).

349 The drop column procedure was then run using these 16 variables over 100  
350 random train-test splits of the data for all three algorithms. From these results, the only  
351 feature found to be important to all three algorithms was speed modulation, suggesting  
352 that the ability to change walking speed is critically important for predicting the aerobic

353 step threshold in stroke. The full results of the drop column procedure for the aerobic  
354 threshold are displayed in Figure 4.

355

356 **Fig 4. Drop column feature importance for aerobic threshold (5500 steps/day).**

357 Red markers show mean feature importance with 95% bootstrapped confidence  
358 interval. Speed modulation was the only feature found to be important across all three  
359 algorithms. *Abbreviations: ABC- Activities Specific Balance Confidence Scale, ADI\_N-*  
360 *Area Deprivation Index (national percentile), BMI- body mass index, CCI- Charlson*  
361 *Comorbidity Index (age-adjusted), PHQ-9 (Patient Health Questionnaire-9),*  
362 *Readiness\_Stage- Readiness to change stage score, 6MWT- 6-Minute Walk Test,*  
363 *TSIS- time since initial stroke, LR- Logistic regression, SVM- Support vector machine,*  
364 *RF- Random forest.*

365

366 Figure 3B displays the model performances for the aerobic threshold. All models  
367 demonstrated a moderate level of test-set accuracy. Overall, SVM performed the worst,  
368 achieving an average standard accuracy score of 71.1% (SD 4.6%, range 58.0% -  
369 81.5%) with selected features and 71.1% (SD 4.4%, range 60.5% - 82.7%) when using  
370 all features. LR and RF performed marginally better than SVM, with LR achieving  
371 standard accuracy scores of 73.9% (SD 4.5%, range 59.3% - 82.7%) with feature  
372 selection and 73.1% (SD 4.2%, range 59.3% - 81.5%) when using all features. RF  
373 achieved standard accuracy scores of 73.3% (SD 4.6%, range 6.8% - 81.5%) with  
374 feature selection and 73.5% (SD 4.1%, range 61.7% - 81.5%) when using all features.

375 Again, the regularization phase was effective in reducing the number of features while  
376 maintaining model performance accuracy.

## 377 **Discussion**

378 We determined that 6MWT was the only variable found to be an important  
379 predictor across all three algorithms in distinguishing home versus community  
380 ambulators. We also found that speed modulation was the only variable deemed  
381 important across all algorithms in distinguishing between stroke survivors who meet  
382 physical activity guidelines versus those who do not. Considering that 6MWT and speed  
383 modulation are measures of a stroke survivor's physical capacity and were the only  
384 variables found to be important across all three mathematically unique algorithms led us  
385 to conclude that measures of physical capacity are *primary characteristics* that  
386 distinguish between groups of stroke survivors using these binary step thresholds.

387 Our finding that 6MWT was a primary characteristic for distinguishing between  
388 home versus community ambulators is in agreement with previous studies  
389 demonstrating that the 6MWT discriminates between functional walking categories in  
390 individuals with stroke [3, 26]. This finding is also aligned with several studies reporting  
391 that measures of physical capacity are, in general, important predictors of functional  
392 walking categories [3, 14, 26, 27] and physical activity [20] in individuals with stroke. As  
393 past work has generally utilized a single analytical approach, our results extend this  
394 work by demonstrating that the 6MWT prevails as an important predictor across multiple  
395 settings and further suggests that walking endurance is important for predicting whether  
396 a stroke survivor will achieve community mobility status using a 2500-step threshold.

397 These results are clinically important and suggest that clinicians should target the  
398 individual's walking endurance to achieve the goal of community ambulation. However,  
399 as can be observed in Figure 2, additional features were deemed important across  
400 some algorithms, but not all. The Area Deprivation Index, Patient Health Questionnaire-  
401 9, readiness to change relapse score, self-selected walking speed, assistive device use,  
402 and years of education were found to be important for both SVM and LR, but not RF.  
403 Taken together, these findings suggest that addressing walking endurance is likely  
404 *necessary but not sufficient* for achieving a 2500-step threshold and that *ancillary*  
405 *features* (defined as predictors that were important to one or two algorithms, but not all  
406 three), including depressive symptoms and readiness to change activity behavior, may  
407 need to be addressed to fully achieve this step threshold.

408 For the aerobic threshold, speed modulation most consistently improved the  
409 prediction when using standard accuracy. Previous work has demonstrated that speed  
410 modulation is related to fall status [56] and daily walking activity [57] in older adults.  
411 Here we showed that speed modulation is also an important predictor of whether a  
412 stroke survivor will achieve a daily step threshold reflecting physical activity guidelines.

413 Since there is no uniform consensus on what magnitude of class imbalance  
414 warrants the use of standard versus balanced accuracy, we also assessed whether  
415 using balanced accuracy for the aerobic threshold would change our result. In this  
416 analysis, we found that both 6MWT and speed modulation were primary characteristics  
417 for achieving the aerobic step threshold (see S1 Fig). This similarity in result regardless  
418 of the use of standard accuracy (in which the 6MWT was close to being considered a  
419 primary characteristic) or balanced accuracy increases our confidence in our

420 methodology and suggests that measures of physical capacity are critically important for  
421 achieving the aerobic step threshold.

422         However, our results demonstrate that targeting physical capacity is likely  
423 *necessary but not sufficient* for achieving a 5500-step threshold. This is supported by  
424 the fact that some algorithms, but not all, found that ancillary features were important in  
425 predicting the outcome. For example, both LR and SVM found that readiness to change  
426 activity behavior and the number of medications taken, were also important for  
427 predicting the aerobic step threshold, suggesting that readiness to change activity  
428 behavior and physical health status, may also need to be addressed to fully achieve this  
429 threshold.

430         We examined predictors of both 2500 and 5500-step thresholds as past work  
431 suggests that these thresholds likely provide different information. The 2500-step  
432 threshold is intended to distinguish between stroke survivors who walk primarily within  
433 their home setting from those who walk within their community [3]. As community  
434 ambulation typically involves greater walking distances [58], sufficient walking  
435 endurance is likely necessary to traverse community distances, lending credence to our  
436 finding that 6MWT is a primary characteristic of the home versus community threshold.  
437 In contrast, the 5500-step threshold is intended to distinguish between those who meet  
438 physical activity guidelines through walking and those who do not [10]. Physical activity  
439 guidelines are expressed in terms of exercise intensity and increasing one's walking  
440 speed is one approach to increasing exercise intensity [9]. Therefore, it logically follows  
441 that the ability to modulate walking speed is a primary characteristic of those who meet  
442 intensity-based physical activity guidelines through walking. However, when using

443 balanced accuracy for the aerobic threshold, we found the 6MWT was also a primary  
444 characteristic and increasing one's exercise duration is another approach to meeting  
445 physical activity guidelines. We therefore conclude that measures of physical capacity  
446 are *primary characteristics* necessary to achieve these step thresholds and that  
447 associated *ancillary characteristics* may need to be addressed to fully achieve these  
448 step thresholds.

449         Importantly, these results should not be interpreted as the 6MWT and speed  
450 modulation represent the *best* single predictors of home versus community or aerobic  
451 step thresholds (see S2 Fig for predictive accuracy of 6MWT and speed modulation  
452 alone in predicting thresholds). This kind of inference would require determining that  
453 they hold the greatest predictive power over the other variables, which was not the  
454 purpose of this study. What our results do indicate, however, is that 6MWT and speed  
455 modulation are primary characteristics that distinguish stroke survivors who do and do  
456 not meet these step thresholds. Similarly, it is worth noting that neither 6MWT nor  
457 speed modulation had the greatest importance across all three algorithms in either  
458 classification setting (see Figs 2 and 4). Thus, our results extend upon literature  
459 examining predictors of stepping activity in stroke by demonstrating that there are  
460 features that are *primary characteristics that should be addressed* as well as *ancillary*  
461 *features that may need to be addressed* based on the unique circumstances of the  
462 individual.

463         With respect to model performance, RF was outperformed by SVM and LR in the  
464 home versus community threshold, with LR marginally achieving the best results over  
465 SVM. In the case of the aerobic threshold, RF and LR outperformed SVM with average

466 accuracy scores within less than 1% of each other both with and without selected  
467 features. It is important to note that we did minimal hyperparameter tuning in this  
468 assessment, only optimizing the parameter for class weights in the case of the home  
469 versus community threshold due to the class imbalance. After tuning just this single  
470 hyperparameter, LR and SVM saw improvements in average performance of 7.3-13.5%  
471 after tuning, where RF's performance did not improve with balanced class weights. It is  
472 possible that with more exhaustive hyperparameter tuning, model performance in the  
473 case of both thresholds could be improved.

474         Importantly, model performance accuracies were similar or better for a model  
475 with only features retained after regularization compared to a model with all features for  
476 all three algorithms and both thresholds. This validated the regularization phase, as we  
477 were able to reduce the number of variables to a relevant subset without compromising  
478 model performance. More directly, we were able to predict the home versus community  
479 and aerobic threshold classifications as well with 7 and 16 features, respectively, as we  
480 were when using all 25 features.

## 481 **Limitations**

482         First, the class imbalance in the case of the home versus community threshold  
483 may have limited the performance of our algorithms. Consequently, model performance  
484 metrics across samples could suffer from high variability. In addition, with only 58 home  
485 ambulators, those in our sample may not be representative of the stroke population, as  
486 individuals were excluded if they walked at a self-selected gait speed of <0.3 m/s.  
487 However, we aimed to address this limitation by using balanced accuracy as our metric



488 of model performance for the home versus community threshold to avoid potential  
489 scenarios resulting in inflated measures of accuracy caused by the class imbalance  
490 (e.g., a constant model classifying all points as community ambulators). Second,  
491 although we used step thresholds endorsed by previous studies, we recognize that  
492 there are likely individuals who may not fit these exact criteria. For example, stroke  
493 survivors who achieve the step threshold to be considered community ambulators may  
494 take these steps primarily within their home environment.

## 495 **Conclusions/Implications**

496 A stroke survivor's physical capacity to walk is a *primary characteristic* that can  
497 be used to determine whether they will achieve step thresholds corresponding to home  
498 versus community ambulation and physical activity guidelines. However, measures of  
499 physical capacity were not necessarily the single best predictors of achieving these  
500 thresholds. Thus, addressing physical capacity is *necessary but not sufficient* for  
501 achieving these thresholds and *ancillary factors*, such as readiness to change activity  
502 behavior and physical health status, among others, may need to be addressed based  
503 on the individual's unique clinical presentation. Future work on larger sample sizes that  
504 contain greater representation of home ambulators and other potentially relevant  
505 variables, such as fatigue and quality of life, is necessary.

506

## 507 **Acknowledgements**

508 None.

## 509 **References**

510

511 1. Winstein CJ, Stein J, Arena R, Bates B, Cherney LR, Cramer SC, et al.  
512 Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare  
513 Professionals From the American Heart Association/American Stroke Association.  
514 Stroke. 2016;47(6):e98-e169. Epub 2016/05/06. doi: 10.1161/str.0000000000000098.  
515 PubMed PMID: 27145936.

516 2. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, et  
517 al. Heart Disease and Stroke Statistics-2018 Update: A Report From the American  
518 Heart Association. Circulation. 2018;137(12):e67-e492. Epub 2018/02/02. doi:  
519 10.1161/cir.0000000000000558. PubMed PMID: 29386200.

520 3. Fulk GD, He Y, Boyne P, Dunning K. Predicting Home and Community Walking  
521 Activity Poststroke. Stroke. 2017;48(2):406-11. Epub 2017/01/07. doi:  
522 10.1161/strokeaha.116.015309. PubMed PMID: 28057807.

523 4. Michael KM, Allen JK, Macko RF. Reduced ambulatory activity after stroke: the  
524 role of balance, gait, and cardiovascular fitness. Archives of physical medicine and  
525 rehabilitation. 2005;86(8):1552-6. Epub 2005/08/09. doi: 10.1016/j.apmr.2004.12.026.  
526 PubMed PMID: 16084807.

527 5. Barclay R, Ripat J, Mayo N. Factors describing community ambulation after  
528 stroke: a mixed-methods study. Clin Rehabil. 2015;29(5):509-21. Epub 2014/08/31. doi:  
529 10.1177/0269215514546769. PubMed PMID: 25172087.

530 6. Ezekiel L, Collett J, Mayo NE, Pang L, Field L, Dawes H. Factors Associated  
531 With Participation in Life Situations for Adults With Stroke: A Systematic Review.

- 532 Archives of physical medicine and rehabilitation. 2019;100(5):945-55. Epub 2018/07/08.  
533 doi: 10.1016/j.apmr.2018.06.017. PubMed PMID: 29981316.
- 534 7. Durcan S, Flavin E, Horgan F. Factors associated with community ambulation in  
535 chronic stroke. Disability and rehabilitation. 2016;38(3):245-9. Epub 2015/04/10. doi:  
536 10.3109/09638288.2015.1035460. PubMed PMID: 25856203.
- 537 8. Fini NA, Holland AE, Keating J, Simek J, Bernhardt J. How Physically Active Are  
538 People Following Stroke? Systematic Review and Quantitative Synthesis. Physical  
539 therapy. 2017;97(7):707-17. Epub 2017/04/27. doi: 10.1093/ptj/pzx038. PubMed PMID:  
540 28444348.
- 541 9. Billinger SA, Arena R, Bernhardt J, Eng JJ, Franklin BA, Johnson CM, et al.  
542 Physical activity and exercise recommendations for stroke survivors: a statement for  
543 healthcare professionals from the American Heart Association/American Stroke  
544 Association. Stroke. 2014;45(8):2532-53. Epub 2014/05/23. doi:  
545 10.1161/str.0000000000000022. PubMed PMID: 24846875.
- 546 10. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, De Bourdeaudhuij I, et  
547 al. How many steps/day are enough? For older adults and special populations. Int J  
548 Behav Nutr Phys Act. 2011;8:80. Epub 2011/07/30. doi: 10.1186/1479-5868-8-80.  
549 PubMed PMID: 21798044; PubMed Central PMCID: PMC3169444.
- 550 11. Bohannon RW, Horton MG, Wikholm JB. Importance of four variables of walking  
551 to patients with stroke. Int J Rehabil Res. 1991;14(3):246-50. Epub 1991/01/01. doi:  
552 10.1097/00004356-199109000-00010. PubMed PMID: 1938039.
- 553 12. Lang CE, Macdonald JR, Reisman DS, Boyd L, Jacobson Kimberley T,  
554 Schindler-Ivens SM, et al. Observation of amounts of movement practice provided

- 555 during stroke rehabilitation. Archives of physical medicine and rehabilitation.  
556 2009;90(10):1692-8. Epub 2009/10/06. doi: 10.1016/j.apmr.2009.04.005. PubMed  
557 PMID: 19801058; PubMed Central PMCID: PMC3008558.
- 558 13. Mulder M, Nijland RH, van de Port IG, van Wegen EE, Kwakkel G. Prospectively  
559 Classifying Community Walkers After Stroke: Who Are They? Archives of physical  
560 medicine and rehabilitation. 2019;100(11):2113-8. Epub 2019/06/04. doi:  
561 10.1016/j.apmr.2019.04.017. PubMed PMID: 31153852.
- 562 14. Perry J, Garrett M, Gronley JK, Mulroy SJ. Classification of walking handicap in  
563 the stroke population. Stroke. 1995;26(6):982-9. Epub 1995/06/01. doi:  
564 10.1161/01.str.26.6.982. PubMed PMID: 7762050.
- 565 15. Handlery R, Fulk G, Pellegrini C, Stewart JC, Monroe C, Fritz S. Stepping After  
566 Stroke: Walking Characteristics in People With Chronic Stroke Differ on the Basis of  
567 Walking Speed, Walking Endurance, and Daily Steps. Physical therapy.  
568 2020;100(5):807-17. Epub 2020/01/30. doi: 10.1093/ptj/pzaa020. PubMed PMID:  
569 31995194.
- 570 16. Saint-Maurice PF, Troiano RP, Bassett DR, Jr., Graubard BI, Carlson SA,  
571 Shiroma EJ, et al. Association of Daily Step Count and Step Intensity With Mortality  
572 Among US Adults. Jama. 2020;323(12):1151-60. Epub 2020/03/25. doi:  
573 10.1001/jama.2020.1382. PubMed PMID: 32207799; PubMed Central PMCID:  
574 PMCPMC7093766.
- 575 17. Loprinzi PD, Addoh O. Accelerometer-Determined Physical Activity and All-  
576 Cause Mortality in a National Prospective Cohort Study of Adults Post-Acute Stroke. Am

- 577 J Health Promot. 2018;32(1):24-7. Epub 2017/07/19. doi: 10.1177/0890117117720061.  
578 PubMed PMID: 28718295.
- 579 18. Willey JZ, Moon YP, Sacco RL, Greenlee H, Diaz KM, Wright CB, et al. Physical  
580 inactivity is a strong risk factor for stroke in the oldest old: Findings from a multi-ethnic  
581 population (the Northern Manhattan Study). Int J Stroke. 2017;12(2):197-200. Epub  
582 2017/01/18. doi: 10.1177/1747493016676614. PubMed PMID: 28093966; PubMed  
583 Central PMCID: PMCPMC5490071.
- 584 19. Bell EJ, Lutsey PL, Windham BG, Folsom AR. Physical activity and  
585 cardiovascular disease in African Americans in Atherosclerosis Risk in Communities.  
586 Medicine and science in sports and exercise. 2013;45(5):901-7. Epub 2012/12/19. doi:  
587 10.1249/MSS.0b013e31827d87ec. PubMed PMID: 23247714; PubMed Central PMCID:  
588 PMCPMC3622814.
- 589 20. Thilarajah S, Mentiplay BF, Bower KJ, Tan D, Pua YH, Williams G, et al. Factors  
590 Associated With Post-Stroke Physical Activity: A Systematic Review and Meta-Analysis.  
591 Archives of physical medicine and rehabilitation. 2018;99(9):1876-89. Epub 2017/10/24.  
592 doi: 10.1016/j.apmr.2017.09.117. PubMed PMID: 29056502.
- 593 21. Aguiar LT, Nadeau S, Martins JC, Teixeira-Salmela LF, Britto RR, Faria C.  
594 Efficacy of interventions aimed at improving physical activity in individuals with stroke: a  
595 systematic review. Disability and rehabilitation. 2018:1-16. Epub 2018/11/20. doi:  
596 10.1080/09638288.2018.1511755. PubMed PMID: 30451539.
- 597 22. Narayanan A, Desai F, Stewart T, Duncan S, Mackay L. Application of Raw  
598 Accelerometer Data and Machine-Learning Techniques to Characterize Human

- 599 Movement Behavior: A Systematic Scoping Review. *J Phys Act Health*. 2020:1-24.  
600 Epub 2020/02/09. doi: 10.1123/jpah.2019-0088. PubMed PMID: 32035416.
- 601 23. Stinear CM, Byblow WD, Ackerley SJ, Smith MC, Borges VM, Barber PA.  
602 PREP2: A biomarker-based algorithm for predicting upper limb function after stroke.  
603 *Ann Clin Transl Neurol*. 2017;4(11):811-20. Epub 2017/11/22. doi: 10.1002/acn3.488.  
604 PubMed PMID: 29159193; PubMed Central PMCID: PMC5682112.
- 605 24. Tozlu C, Edwards D, Boes A, Labar D, Tsagaris KZ, Silverstein J, et al. Machine  
606 Learning Methods Predict Individual Upper-Limb Motor Impairment Following Therapy  
607 in Chronic Stroke. *Neurorehabil Neural Repair*. 2020:1545968320909796. Epub  
608 2020/03/21. doi: 10.1177/1545968320909796. PubMed PMID: 32193984.
- 609 25. Lin CH, Hsu KC, Johnson KR, Fann YC, Tsai CH, Sun Y, et al. Evaluation of  
610 machine learning methods to stroke outcome prediction using a nationwide disease  
611 registry. *Comput Methods Programs Biomed*. 2020;190:105381. Epub 2020/02/12. doi:  
612 10.1016/j.cmpb.2020.105381. PubMed PMID: 32044620.
- 613 26. Fulk GD, Reynolds C, Mondal S, Deutsch JE. Predicting home and community  
614 walking activity in people with stroke. *Archives of physical medicine and rehabilitation*.  
615 2010;91(10):1582-6. Epub 2010/09/30. doi: 10.1016/j.apmr.2010.07.005. PubMed  
616 PMID: 20875518.
- 617 27. Bowden MG, Balasubramanian CK, Behrman AL, Kautz SA. Validation of a  
618 speed-based classification system using quantitative measures of walking performance  
619 poststroke. *Neurorehabil Neural Repair*. 2008;22(6):672-5. Epub 2008/10/31. doi:  
620 10.1177/1545968308318837. PubMed PMID: 18971382; PubMed Central PMCID:  
621 PMCPMC2587153.

- 622 28. Danks KA, Pohlig RT, Roos M, Wright TR, Reisman DS. Relationship Between  
623 Walking Capacity, Biopsychosocial Factors, Self-efficacy, and Walking Activity in  
624 Persons Poststroke. *J Neurol Phys Ther.* 2016;40(4):232-8. Epub 2016/08/23. doi:  
625 10.1097/NPT.000000000000143. PubMed PMID: 27548750; PubMed Central PMCID:  
626 PMCPMC5025374.
- 627 29. French MA, Moore MF, Pohlig R, Reisman D. Self-efficacy Mediates the  
628 Relationship between Balance/Walking Performance, Activity, and Participation after  
629 Stroke. *Topics in stroke rehabilitation.* 2016;23(2):77-83. Epub 2015/12/15. doi:  
630 10.1080/10749357.2015.1110306. PubMed PMID: 26653764; PubMed Central PMCID:  
631 PMCPMC4833556.
- 632 30. Holleran CL, Bland MD, Reisman DS, Ellis TD, Earhart GM, Lang CE. Day-to-  
633 Day Variability of Walking Performance Measures in Individuals Poststroke and  
634 Individuals With Parkinson Disease. *J Neurol Phys Ther.* 2020;44(4):241-7. Epub  
635 2020/08/10. doi: 10.1097/npt.0000000000000327. PubMed PMID: 32769671; PubMed  
636 Central PMCID: PMCPMC7486249.
- 637 31. Miller A, Pohlig RT, Reisman DS. Social and physical environmental factors in  
638 daily stepping activity in those with chronic stroke. *Topics in stroke rehabilitation.*  
639 2020:1-9. Epub 2020/08/11. doi: 10.1080/10749357.2020.1803571. PubMed PMID:  
640 32772823.
- 641 32. Robinson CA, Matsuda PN, Ciol MA, Shumway-Cook A. Participation in  
642 community walking following stroke: the influence of self-perceived environmental  
643 barriers. *Physical therapy.* 2013;93(5):620-7. Epub 2013/01/19. doi:  
644 10.2522/ptj.20110217. PubMed PMID: 23329558.

- 645 33. Shumway-Cook A, Patla AE, Stewart A, Ferrucci L, Ciol MA, Guralnik JM.  
646 Environmental demands associated with community mobility in older adults with and  
647 without mobility disabilities. *Physical therapy*. 2002;82(7):670-81. Epub 2002/06/29.  
648 PubMed PMID: 12088464.
- 649 34. Wright H, Wright T, Pohlig RT, Kasner SE, Raser-Schramm J, Reisman D.  
650 Protocol for promoting recovery optimization of walking activity in stroke (PROWALKS):  
651 a randomized controlled trial. *BMC Neurol*. 2018;18(1):39. Epub 2018/04/14. doi:  
652 10.1186/s12883-018-1044-1. PubMed PMID: 29649992; PubMed Central PMCID:  
653 PMCPMC5898044.
- 654 35. Eng JJ, Dawson AS, Chu KS. Submaximal exercise in persons with stroke: test-  
655 retest reliability and concurrent validity with maximal oxygen consumption. *Archives of*  
656 *physical medicine and rehabilitation*. 2004;85(1):113-8. Epub 2004/02/19. doi:  
657 10.1016/s0003-9993(03)00436-2. PubMed PMID: 14970978; PubMed Central PMCID:  
658 PMCPMC3167868.
- 659 36. Flansbjerg UB, Holmback AM, Downham D, Patten C, Lexell J. Reliability of gait  
660 performance tests in men and women with hemiparesis after stroke. *J Rehabil Med*.  
661 2005;37(2):75-82. Epub 2005/03/25. doi: 10.1080/16501970410017215. PubMed PMID:  
662 15788341.
- 663 37. Pendlebury ST, Mariz J, Bull L, Mehta Z, Rothwell PM. MoCA, ACE-R, and  
664 MMSE versus the National Institute of Neurological Disorders and Stroke-Canadian  
665 Stroke Network Vascular Cognitive Impairment Harmonization Standards  
666 Neuropsychological Battery after TIA and stroke. *Stroke*. 2012;43(2):464-9. Epub



667 2011/12/14. doi: 10.1161/strokeaha.111.633586. PubMed PMID: 22156700; PubMed

668 Central PMCID: PMCPMC5390857.

669 38. Tessier A, Finch L, Daskalopoulou SS, Mayo NE. Validation of the Charlson

670 Comorbidity Index for predicting functional outcome of stroke. Archives of physical

671 medicine and rehabilitation. 2008;89(7):1276-83. Epub 2008/07/01. doi:

672 10.1016/j.apmr.2007.11.049. PubMed PMID: 18586129.

673 39. de Man-van Ginkel JM, Gooskens F, Schepers VP, Schuurmans MJ, Lindeman

674 E, Hafsteinsdottir TB. Screening for poststroke depression using the patient health

675 questionnaire. Nurs Res. 2012;61(5):333-41. Epub 2012/06/20. doi:

676 10.1097/NNR.0b013e31825d9e9e. PubMed PMID: 22710475.

677 40. Salbach NM, Mayo NE, Hanley JA, Richards CL, Wood-Dauphinee S.

678 Psychometric evaluation of the original and Canadian French version of the activities-

679 specific balance confidence scale among people with stroke. Archives of physical

680 medicine and rehabilitation. 2006;87(12):1597-604. Epub 2006/12/05. doi:

681 10.1016/j.apmr.2006.08.336. PubMed PMID: 17141639.

682 41. Kind AJ, Jencks S, Brock J, Yu M, Bartels C, Ehlenbach W, et al. Neighborhood

683 socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study.

684 Ann Intern Med. 2014;161(11):765-74. Epub 2014/12/02. doi: 10.7326/m13-2946.

685 PubMed PMID: 25437404; PubMed Central PMCID: PMCPMC4251560.

686 42. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics

687 Accessible - The Neighborhood Atlas. The New England journal of medicine.

688 2018;378(26):2456-8. Epub 2018/06/28. doi: 10.1056/NEJMp1802313. PubMed PMID:

689 29949490; PubMed Central PMCID: PMCPMC6051533.

- 690 43. Miller A, Wright T, Wright H, Thompson E, Pohlig RT, Reisman DS. Readiness to  
691 Change is Related to Real-World Walking and Depressive Symptoms in Chronic Stroke.  
692 J Neurol Phys Ther. 2021;45(1):28-35. Epub 2020/12/15. doi:  
693 10.1097/npt.0000000000000345. PubMed PMID: 33315834; PubMed Central PMCID:  
694 PMCPMC7739270.
- 695 44. Rosenkranz RR, Duncan MJ, Caperchione CM, Kolt GS, Vandelanotte C,  
696 Maeder AJ, et al. Validity of the Stages of Change in Steps instrument (SoC-Step) for  
697 achieving the physical activity goal of 10,000 steps per day. BMC Public Health.  
698 2015;15:1197. Epub 2015/12/02. doi: 10.1186/s12889-015-2539-y. PubMed PMID:  
699 26620188; PubMed Central PMCID: PMCPMC4666193.
- 700 45. Fulk GD, Combs SA, Danks KA, Nirider CD, Raja B, Reisman DS. Accuracy of 2  
701 activity monitors in detecting steps in people with stroke and traumatic brain injury.  
702 Physical therapy. 2014;94(2):222-9. Epub 2013/09/21. doi: 10.2522/ptj.20120525.  
703 PubMed PMID: 24052577.
- 704 46. Hui J, Heyden R, Bao T, Accettone N, McBay C, Richardson J, et al. Validity of  
705 the Fitbit One for Measuring Activity in Community-Dwelling Stroke Survivors.  
706 Physiother Can. 2018;70(1):81-9. Epub 2018/02/13. doi: 10.3138/ptc.2016-40.ep.  
707 PubMed PMID: 29434422; PubMed Central PMCID: PMCPMC5802948.
- 708 47. Klassen TD, Semrau JA, Dukelow SP, Bayley MT, Hill MD, Eng JJ. Consumer-  
709 Based Physical Activity Monitor as a Practical Way to Measure Walking Intensity During  
710 Inpatient Stroke Rehabilitation. Stroke. 2017;48(9):2614-7. Epub 2017/08/09. doi:  
711 10.1161/strokeaha.117.018175. PubMed PMID: 28784922.

- 712 48. Klassen TD, Simpson LA, Lim SB, Louie DR, Parappilly B, Sakakibara BM, et al.  
713 "Stepping Up" Activity Poststroke: Ankle-Positioned Accelerometer Can Accurately  
714 Record Steps During Slow Walking. *Physical therapy*. 2016;96(3):355-60. Epub  
715 2015/08/08. doi: 10.2522/ptj.20140611. PubMed PMID: 26251478; PubMed Central  
716 PMCID: PMC4774387.
- 717 49. Tudor-Locke C, Burkett L, Reis JP, Ainsworth BE, Macera CA, Wilson DK. How  
718 many days of pedometer monitoring predict weekly physical activity in adults? *Prev*  
719 *Med*. 2005;40(3):293-8. Epub 2004/11/10. doi: 10.1016/j.ypmed.2004.06.003. PubMed  
720 PMID: 15533542.
- 721 50. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research  
722 electronic data capture (REDCap)--a metadata-driven methodology and workflow  
723 process for providing translational research informatics support. *J Biomed Inform*.  
724 2009;42(2):377-81. Epub 2008/10/22. doi: 10.1016/j.jbi.2008.08.010. PubMed PMID:  
725 18929686; PubMed Central PMCID: PMC2700030.
- 726 51. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al.  
727 Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*.  
728 2011;12:2825-30.
- 729 52. Brick TR, Koffer RE, Gerstorff D, Ram N. Feature Selection Methods for Optimal  
730 Design of Studies for Developmental Inquiry. *J Gerontol B Psychol Sci Soc Sci*.  
731 2017;73(1):113-23. Epub 2017/02/07. doi: 10.1093/geronb/gbx008. PubMed PMID:  
732 28164232; PubMed Central PMCID: PMC6075467.
- 733 53. Olson RS, Cava W, Mustahsan Z, Varik A, Moore JH. Data-driven advice for  
734 applying machine learning to bioinformatics problems. *Pac Symp Biocomput*.

735 2018;23:192-203. Epub 2017/12/09. PubMed PMID: 29218881; PubMed Central  
736 PMCID: PMCPMC5890912.

737 54. Chen C, Liaw A, Breiman L. Using Random Forest to Learn Imbalanced Data.  
738 [Tech Report]. In press 2004.

739 55. Krawczyk B. Learning from imbalanced data: open challenges and future  
740 directions. *Progress in Artificial Intelligence*. 2016;5(4):221-32. doi: 10.1007/s13748-  
741 016-0094-0.

742 56. Middleton A, Fulk GD, Herter TM, Beets MW, Donley J, Fritz SL. Self-Selected  
743 and Maximal Walking Speeds Provide Greater Insight Into Fall Status Than Walking  
744 Speed Reserve Among Community-Dwelling Older Adults. *American journal of physical  
745 medicine & rehabilitation*. 2016;95(7):475-82. Epub 2016/03/24. doi:  
746 10.1097/phm.0000000000000488. PubMed PMID: 27003205; PubMed Central PMCID:  
747 PMCPMC4912425.

748 57. Middleton A, Fulk GD, Beets MW, Herter TM, Fritz SL. Self-Selected Walking  
749 Speed is Predictive of Daily Ambulatory Activity in Older Adults. *Journal of aging and  
750 physical activity*. 2016;24(2):214-22. Epub 2015/09/16. doi: 10.1123/japa.2015-0104.  
751 PubMed PMID: 26371593; PubMed Central PMCID: PMCPMC4792803.

752 58. Andrews AW, Chinworth SA, Bourassa M, Garvin M, Benton D, Tanner S.  
753 Update on distance and velocity requirements for community ambulation. *J Geriatr Phys  
754 Ther*. 2010;33(3):128-34. Epub 2010/12/16. PubMed PMID: 21155508.

755

756

## 757 **Supporting information**

### 758 **S1 Fig. Drop column feature importance for aerobic threshold (5500 steps/day)**

759 **using balanced accuracy.** Red markers show mean feature importance with 95%  
760 bootstrapped confidence interval. 6MWT and speed modulation were the only features  
761 found to be important across all three algorithms. *Abbreviations: ABC- Activities Specific*  
762 *Balance Confidence Scale, ADI\_N- Area Deprivation Index (national percentile), BMI-*  
763 *body mass index, CCI- Charlson Comorbidity Index (age-adjusted), PHQ-9- Patient*  
764 *Health Questionnaire-9, Readiness\_Stage- Readiness to change stage score, 6MWT-*  
765 *6-Minute Walk Test, TSIS- time since initial stroke, LR- Logistic regression, SVM-*  
766 *Support vector machine, RF- Random forest.*

767

### 768 **S2 Fig. Precision-recall curve for predicting home versus community ambulation**

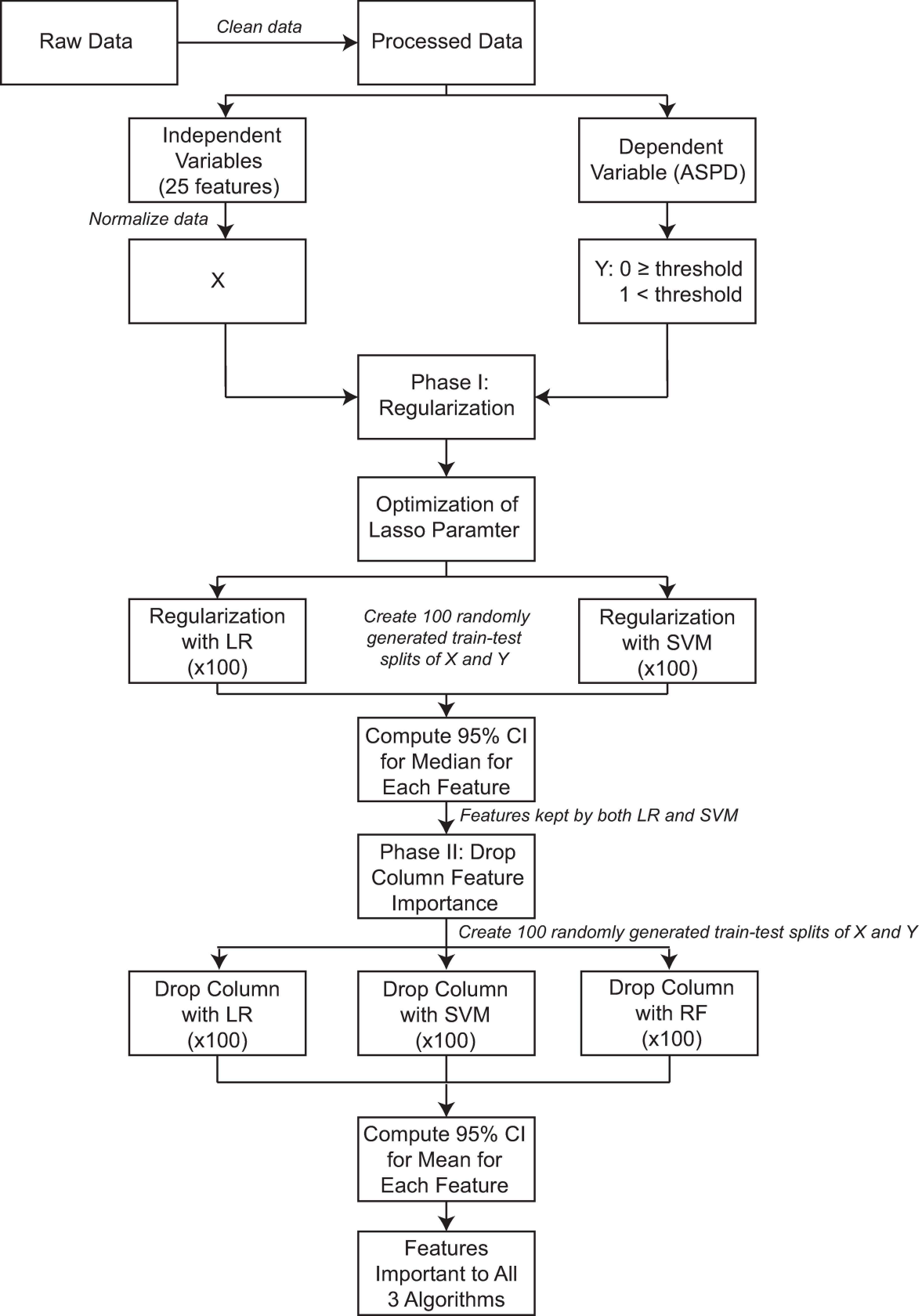
769 **using the 6-minute walk test (A) and aerobic threshold using speed modulation**

770 **(B).** Precision is defined as:  $\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False}$   
771  $\text{Positives})$ . Recall is defined as:  $\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False}$   
772  $\text{Negatives})$ . The solid line plots the precision-recall curve, and the dash line reflects a  
773 no-skill classifier (a model that cannot discriminate between classes). *Abbreviations:*  
774 *AUC- Area Under Curve.*

775

776 **S1 Supplementary Material. Supplement\_10-6-2021.**

777

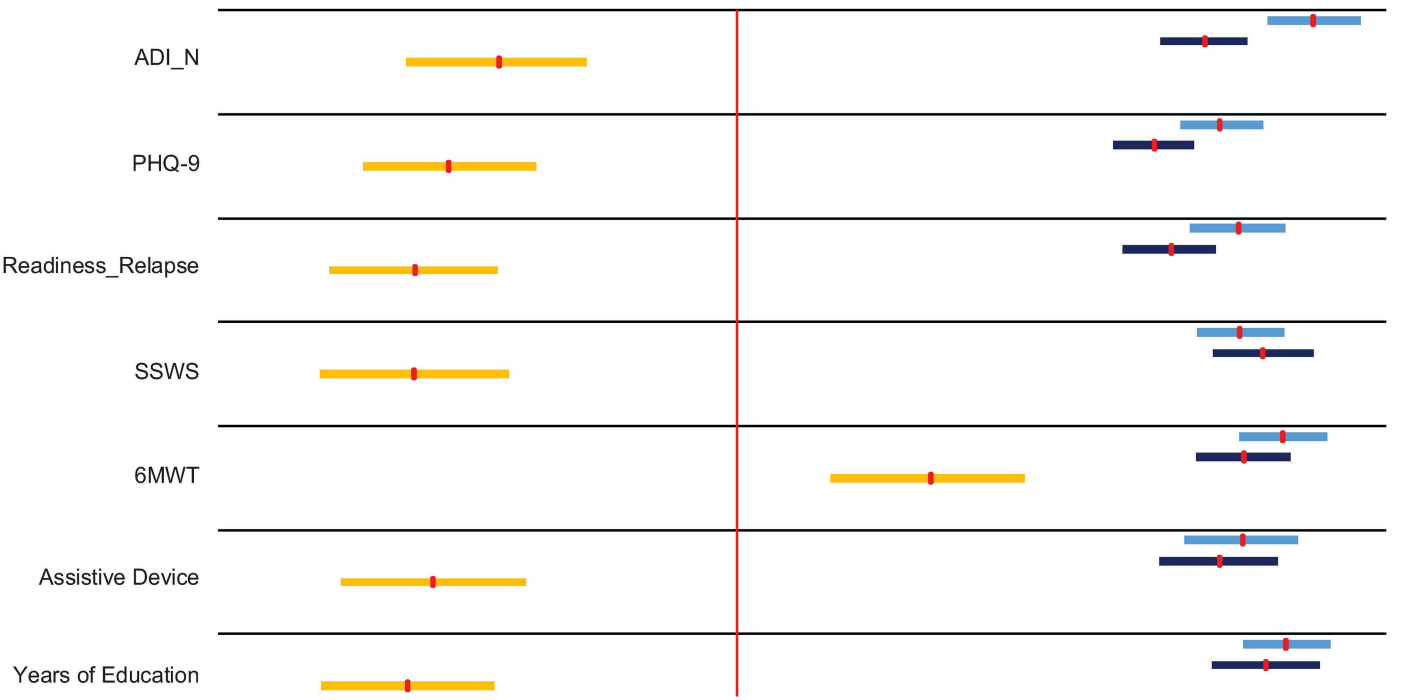


*Not important to the model*

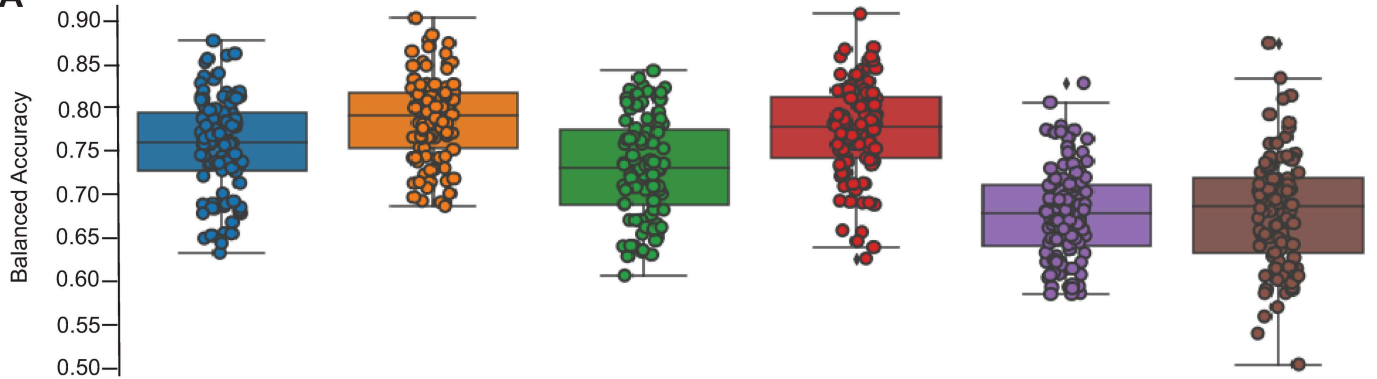
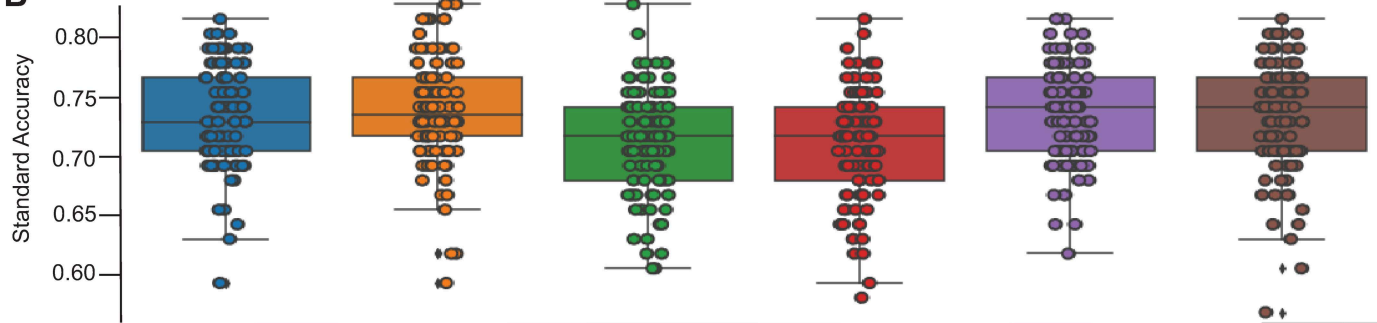
*Important to the model*



-0.08    -0.06    -0.04    -0.02    0    0.02    0.04    0.06    0.08    0.1



■ LR ■ SVM ■ RF

**A****B**

Logistic  
Regression  
(AF)

Logistic  
Regression  
(FS)

Support Vector  
Machine  
(AF)

Support Vector  
Machine  
(FS)

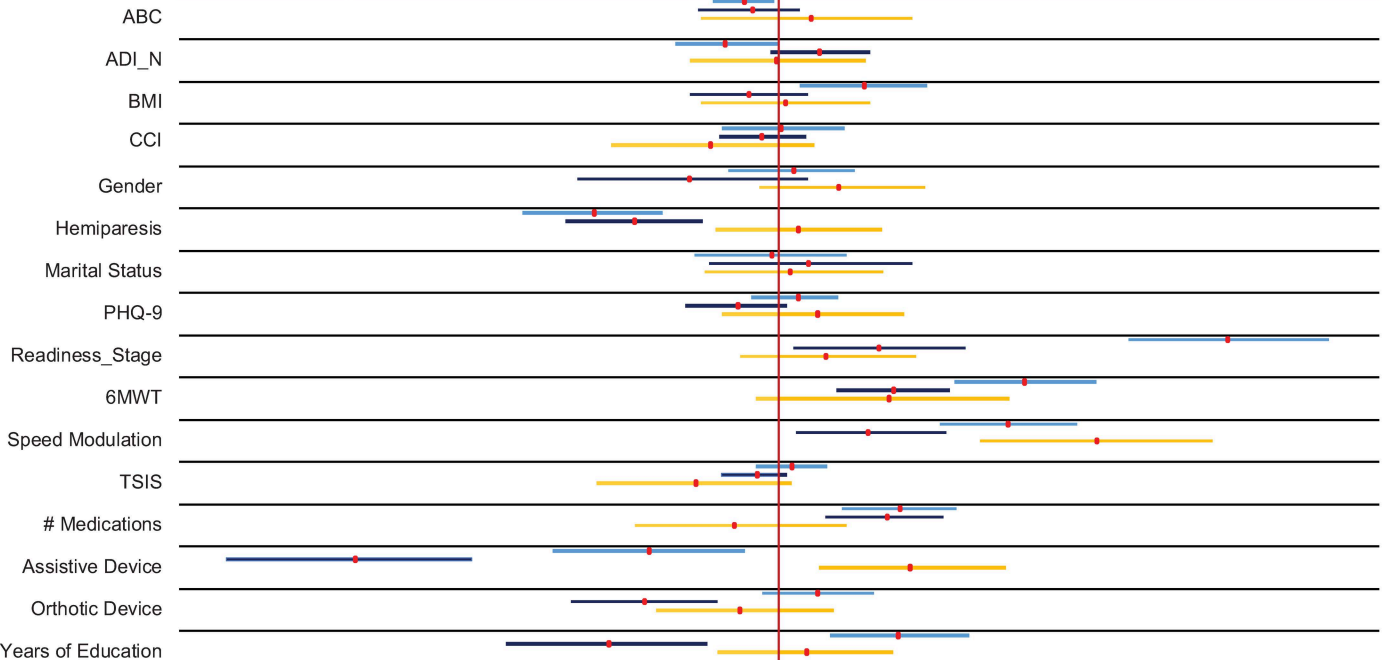
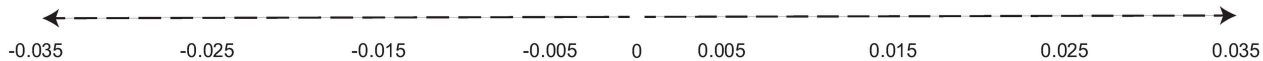
Random  
Forest  
(AF)

Random  
Forest  
(FS)



Not important to the model

Important to the model



■ LR ■ SVM ■ RF

1 A machine learning approach to identifying important features for achieving step  
2 thresholds in individuals with chronic stroke: Supplementary material

### 3 **Machine learning algorithms**

4 This section of the supplement is intended to provide a brief overview of the  
5 machine learning algorithms used in this study for readers unfamiliar with them. The  
6 three machine learning (ML) algorithms used in this analysis were Logistic Regression  
7 (LR), Support Vector Machine (SVM) with a radial basis function (RBF) kernel, and  
8 Random Forest (RF) classifier. These were chosen for this analysis because they are  
9 all commonly used supervised ML algorithms that are mathematically distinct from one  
10 another. A supervised classification algorithm is one that is trained by being provided  
11 with both data (X) and labels (Y), similar to a flash card. The way that supervised  
12 classification algorithms generally work is that they use the training data and labels to  
13 construct a function,  $f(x)$ , to make predictions about which class other points in the  
14 same population belong to. Here, each point  $x = (x^1, x^2, \dots, x^m)$  where  $x^k$  is the value of  
15  $x$  that corresponds with the  $k^{\text{th}}$  feature. This is done by solving an optimization problem  
16 to find the best parameters for  $f(x)$  based on what quantity is being optimized. The form  
17 of the predictor  $f(x)$  and the optimization process varies across ML algorithms.

18 In the case of LR with binomial class labels, as we have in our analysis, the  
19 general idea is that, based on our sample data, X, and labels, Y, we want to model the  
20 probability that  $Y=1$ , given X takes some value  $x$ , i.e.

$$21 \quad p(x) = P[Y = 1|X = x]$$

22 Then we also have that the probability that  $Y = 0$  given  $X = x$  is  $1 - p(x)$ . Now, to model  
23 this probability,  $p(x)$ , LR uses the logistic function which takes values in  $(-\infty, \infty)$  and  
24 maps them to the interval  $[0, 1]$  (the values of the probabilities). During this process, we  
25 want to find some set of parameters,  $\beta$  (the vector of beta weights), that fit the data in  
26 some optimal way where the logistic function to model this probability is:

$$27 \quad p(x, \beta) = \frac{e^{x^T \beta}}{1 + e^{x^T \beta}}$$

28 Given our training data has  $n$  observations, the goal of LR is find the values of  $\beta$  that  
29 maximize the log-likelihood function:

$$30 \quad LL(\beta) = \ln \left( \prod_{i=1}^n p(x_i, \beta)^{y_i} (1 - p(x_i, \beta))^{1-y_i} \right)$$
$$31 \quad = \sum_{i=1}^n y_i x_i^T \beta - \ln(1 + e^{x_i^T \beta})$$

32 One important thing to note is that of the three models used in our analysis, LR is the  
33 most closely related to a linear model in the sense that the term  $x^T \beta$  is a linear  
34 combination of the values of the vector  $x$  with the beta weights. This means that LR is  
35 only able to detect linear relationships between the features, where non-linear ones  
36 might exist.

37 For SVM, rather than computing the probability that a given observation belongs  
38 to some class, the goal is to establish boundaries with an allotted margin of error in the  
39 space the data comes from. These boundaries, which are defined by linear  
40 hyperplanes, separate the space the data comes from into regions. Then, a point would

41 be classified depending on which of these regions the point lies in. In general, if our  
42 data set consists of  $m$  variables, then our data points live in an  $m$ -dimensional space. A  
43 hyperplane in an  $m$ -dimensional space would be defined by:

$$44 \quad \beta_0 + \beta_1 x^1 + \dots + \beta_m x^m = 0$$

45 Which, in vector form, would be:

$$46 \quad \beta_0 + \beta^T x = 0$$

47 Where  $x$  is a single data point,  $\beta_0$  is the bias term, and  $\beta$  is the vector of beta weights.  
48 Notice that, similar to LR, the linear hyperplane in the  $m$ -dimensional space is only able  
49 to use linear relationships between the variables, but it is very possible that the data is  
50 not linearly separable. This is where one of the benefits of SVM, the use of a kernel,  
51 comes into play. A kernel, such as RBF, allows for additional “new” variables to be  
52 constructed using non-linear combinations of the existing variables. This allows the data  
53 to be projected into higher,  $m+d$  dimensional space, where  $d$  is the number of “new”  
54 variables, in which the data may now be separable by a linear hyperplane in this new,  
55 higher dimensional space. The way this would change the formulation of the boundary  
56 would be:

$$57 \quad \beta_0 + \beta^T \phi(x) = 0$$

58 where  $\phi$  is an unknown function of  $x$  that is related to the transformation into higher  
59 dimensional space. The term  $\phi(x)$  is what would be the non-linear part of the definition  
60 of the hyperplane in  $m$ -dimensional space.

61 The RBF kernel, which is defined by:

62

$$K(x_i, x_j) = e^{-\gamma(x_i - x_j)^2}$$

63 allows the data to be projected into an infinite dimensional space, but rather than ever  
64 actually making the physical transformation, the kernel functions allow for these higher  
65 dimensional relationships between the points to be computed because the kernel  
66 function acts as a dot product in that space. In this way, the RBF kernel can be thought  
67 of as a measure of how “close” two observations  $x_i$  and  $x_j$  are in infinite dimensional  
68 space. What makes SVM so robust is not only that it can use a kernel but also that,  
69 when establishing this boundary, there are allowances for misclassification of training  
70 data to handle noise. The optimization problem then becomes maximizing the distance  
71 between the points and the hyperplane while trying to minimize misclassification error:

72

$$\max_{\beta_0, \beta, M, \varepsilon_1, \dots, \varepsilon_n} M$$

73 Subject to the constraints:

74

$$\|\beta\| = 1$$

75

$$(2y_i - 1)(\beta_0 + \beta^T \phi(x_i)) \geq M(1 - \varepsilon_i)$$

76

$$\sum_{i=1}^n \varepsilon_i \leq C$$

77 Where  $M$  is the size of the margin,  $(2y_i - 1)(\beta_0 + \beta^T \phi(x_i))$  is the distance from the  
78 observation  $x_i$  to the hyperplane in the higher dimensional space, and  $C$  and  $\varepsilon_i \geq 0$  are  
79 the total error tolerance and the “slack” allowed for each observation,  $x_i$ , respectively.  
80 This optimization problem can be solved using a Lagrange multiplier, which eliminates  
81 the dependency on the unknown function  $\phi(x)$ . The end result is a potentially non-linear

82 boundary that is as far away as possible from most points in each class, thus creating a  
83 robust classifier in the physical feature space.

84 Finally, RF is what is referred to as an “ensemble” because it is much more a  
85 collection of individual models, rather than a single model on its own. Simply put, a  
86 random forest is exactly what it sounds like: it is a collection (or “forest”) of decision  
87 trees which are each constructed using random subsets of the training data. Once a  
88 new point is passed into the RF model, each tree in the forest will classify that point and  
89 then RF will choose the class with the most “votes” in the forest. Each decision tree is  
90 made up of a collection of nodes that act like a path where each node is defined by a  
91 single feature with some threshold. The first node is the root node and connected to that  
92 are branches and then leaves. If we think of each of these nodes as possible stops on  
93 the tree’s path to a decision about a point ,  $x = (x^1, x^2, \dots, x^m)$ , at each stop, the value of  
94 one of  $x$ ’s entries will determine the next node it will stop at until it reaches a leaf node  
95 where the path ends and a decision is made based off of where it ended up. When  
96 growing each tree, a bootstrap sample of the training data is used and at each new  
97 node, a random subset of the variables is chosen, from which the best single predictor  
98 will be used to define the threshold that splits that node. This optimization of the  
99 individual decision tree is more of a greedy algorithm than an actual “optimization”  
100 problem in the classical sense. Consider an example where we are deciding how to  
101 best split a node with a random subset of  $k \leq m$  features. For each feature  $j$  in  $\{1, \dots, k\}$   
102 choosing a threshold,  $t$ , will create two regions, one for the points where the value of the  
103 feature  $j$  is less than or equal to  $t$  and one where the value is greater than  $t$  :

104 
$$R_1 = \{\text{points } x \text{ with } x_j \leq t\} \text{ and } R_2 = \{\text{points } x \text{ with } x_j > t\}$$

105 Then the feature  $j$  and the threshold  $t$  are chosen to be those that best classify the  
106 training data, i.e., those that minimize the classification error across the regions by  
107 satisfying:

$$108 \quad \min_{j,t} \left[ \min_{x_i \text{ in } R_1} \sum error_{(j,s)}(x_i) + \sum_{x_i \text{ in } R_2} error_{(j,s)}(x_i) \right]$$

109 Where  $error_{R(j,s)}(x)$  is the quantification of the error of classifying point  $x_i$  in region R  
110 using the variable  $j$  with threshold  $s$ . For classification, a popular metric to use here is  
111 the Gini Index. The end result is a robust model that is known to perform well in a wider  
112 variety of situations, but that is not very interpretable.

## 113 Lasso regularization

114 In this analysis, Lasso (Least Absolute Shrinkage and Selection Operator)  
115 regularization was used to perform dimensionality reduction in the first phase of the  
116 feature importance process. This was done to reduce noise and redundancy among the  
117 full set of variables; thus, the goal was to reduce the number of variables while still  
118 retaining the same amount of information. As seen in the results section of this paper,  
119 this goal was achieved.

120 When fitting a linear model to data, as mentioned above, there is an optimization  
121 problem being solved. In general, for a linear model of the form:

$$122 \quad f(x, \beta) = \beta_0 + \beta_1 x^1 + \dots + \beta_m x^m$$

123 You want to find the values of the coefficients,  $\beta_j$ , such that you are minimizing the loss  
124 function,  $L(\beta)$ . What Lasso regularization does is place a penalty on the size of the  $\beta$

125 coefficients where the optimization problem changes to finding the coefficients

126  $\beta_1, \dots, \beta_m$  such that we want to solve:

127 
$$\min_{\beta} L(\beta) + \lambda \sum_{j=1}^m |\beta_j|$$

128 Where  $\lambda > 0$  [1]. Note that in our analysis, the values of the regularization parameters

129 can be interpreted as  $\frac{1}{\lambda}$ .

130 The reason Lasso regularization was chosen for this analysis is that the penalty it  
131 places on the size of the  $\beta$  coefficients shrinks them in a way that forces some of the  
132 coefficients to 0 [1]. In doing this, the remaining features with non-zero coefficients can  
133 be interpreted as the subset of variables that were “chosen” by lasso regularization.

## 134 **Balanced accuracy for imbalanced data**

135 We used two metrics of model performance throughout the analysis: standard  
136 accuracy in the case of the aerobic threshold (5500 steps/day), which is the total  
137 proportion of correctly classified points in the test set:

138 
$$\frac{\text{true positives} + \text{true negatives}}{\text{true positives} + \text{true negatives} + \text{false positives} + \text{false negatives}}$$

139 and balanced accuracy in the case of the home vs. community threshold (2500  
140 steps/day), which is the arithmetic mean of the recall (or sensitivity) and specificity:

141 
$$\frac{\text{true positives}}{\text{true positive} + \text{false negatives}} + \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$



142 These metrics were chosen to ensure that the model performance metric used was  
143 representative of how well the model was performing on the target class.

144 The class distribution of each threshold was 58 (21.64%) home ambulators and  
145 210 (78.36%) community ambulators in the case of the home vs. community threshold  
146 and 185 (69.03%) below the minimum aerobic activity threshold and 83 (30.97%) above  
147 for the aerobic threshold. While class distributions are not perfectly balanced in each  
148 case, the class imbalance in the case of the home vs. community threshold is starker at  
149 1:4 and, more importantly, in favor of the negative class, unlike the aerobic threshold  
150 [2]. In the case of any significant class imbalance where the target class (the positive  
151 class) is the minority, the metric of standard accuracy could misrepresent how well the  
152 model is performing on the target class because standard accuracy may still be high,  
153 even if the model is not performing well on the target class [2]. In cases like this, we can  
154 use a metric like balanced accuracy, which allows the performance on the minority class  
155 to hold equal weight to that of the majority class and is also robust to misclassification  
156 noise [3].

## 157 **Results for the aerobic threshold (5500 steps/day)**

### 158 **using balanced accuracy**

159 When using the metric of balanced accuracy for the aerobic threshold, the  
160 optimized regularization parameters were 2.9 for LR and 0.99 for linear SVM. Given that  
161 when we use standard accuracy, the optimal regularization parameters were very  
162 similar at 2.9 and 0.9, it follows that the results of the lasso regularization stage when  
163 using balanced accuracy were identical to the 16 features reported in the primary

164 analysis: 6MWT, speed modulation, CCI age-adjusted score, PHQ-9, readiness to  
165 change stage score, ABC, usual orthotic and assistive device, marital status, years of  
166 education, gender, BMI, side of hemiparesis, time since initial stroke, number of  
167 medications, and ADI\_N.

168 In the second stage, the drop column procedure was run using these 16 features.  
169 For LR, 10 features were found to be important, 6 of which were the important features  
170 found when using standard accuracy. In order of importance, these features were:  
171 readiness to change stage score, speed modulation, 6MWT, usual assistive device,  
172 number of medications, usual orthotic device, years of education, BMI, CCI age-  
173 adjusted score, and time since initial stroke. For SVM, the same four features were  
174 found to be important as when using standard accuracy. In order of importance, these  
175 features were: readiness to change stage score, number of medications, speed  
176 modulation, and 6MWT. Like SVM, RF also had four features found to be important,  
177 which, in order of importance were: speed modulation, usual assistive device, 6MWT,  
178 and PHQ-9. From these results, speed modulation was still found to be a primary  
179 characteristic for the aerobic threshold, in addition to 6MWT. BMI, CCI, PHQ-9,  
180 readiness to change stage score, time since initial stroke, number of medications taken,  
181 assistive and orthotic device use, and years of education were found to be ancillary  
182 characteristics. Supplemental Figure 1 displays the results for the drop column  
183 procedure for the aerobic threshold using balanced accuracy.

184

185 **S1 Fig. Drop column feature importance for aerobic threshold (5500 steps/day)**  
186 **using balanced accuracy.** Red markers show mean feature importance with 95%

187 bootstrapped confidence interval. 6MWT and speed modulation were the only features  
188 found to be important across all three algorithms. *Abbreviations: ABC- Activities Specific*  
189 *Balance Confidence Scale, ADI\_N- Area Deprivation Index (national percentile), BMI-*  
190 *body mass index, CCI- Charlson Comorbidity Index (age-adjusted), PHQ-9- Patient*  
191 *Health Questionnaire-9, Readiness\_Stage- Readiness to change stage score, 6MWT-*  
192 *6-Minute Walk Test, TSIS- time since initial stroke, LR- Logistic regression, SVM-*  
193 *Support vector machine, RF- Random forest.*

194

195       Importantly, these results are very similar to the results found when using  
196 standard accuracy. Note that when standard accuracy was used, 6MWT was found to  
197 be important for both SVM and LR but the 95% confidence interval for RF was only  
198 barely negative, resulting in 6MWT not being found to be a primary characteristic. It is  
199 encouraging that these results were found in this analysis because it is reflective of the  
200 fact that, even with mild class imbalance, standard accuracy was an appropriate choice  
201 of metric for this threshold.

## 202 **6-minute walk test as a single predictor of home vs.** 203 **community thresholds (2500 steps/day)**

204       A precision-recall curve was used to assess the predictive ability of the 6MWT  
205 alone in predicting the home vs. community threshold (2500 steps/day; Supplemental  
206 Figure 2A). Similar to an ROC analysis, a precision-recall curve can be used to assess  
207 the skill of a prediction model, where:

208 
$$\textit{Precision} = \frac{\textit{True Positives}}{(\textit{True Positives} + \textit{False Positives})}$$

209 and

211 
$$\textit{Recall} = \frac{\textit{True Positives}}{(\textit{True Positives} + \textit{False Negatives})}$$

210

212 **S2 Fig. Precision-recall curve for predicting home vs. community ambulation**

213 **using the 6-minute walk test (A) and aerobic threshold using the speed**

214 **modulation (B).** The solid line plots the precision-recall curve, and the dashed line

215 reflects a no-skill classifier (i.e., a model that cannot discriminate between classes).

216 *Abbreviations: AUC- Area Under Curve.*

217

218 Recall is also referred to as sensitivity. As depicted in Supplemental Figure 2, a  
219 precision-recall curves plots the precision on the Y axis and recall on the X axis, where  
220 a point at (1,1) in the upper right corner of the figure would reflect a model with perfect  
221 precision and perfect recall. These curves can often be used instead of ROC curves in  
222 cases of class imbalance, as they are more sensitive to measures of recall. Similar to  
223 an ROC analysis, the area under the curve (AUC) can be computed to provide a  
224 measure of how well a model is performing where the AUC can have a maximum value  
225 of 1. A no-skill classifier (i.e., a model that randomly “guesses” the class) is depicted as  
226 a dashed line and changes based on the distribution of positive and negative cases:

227 
$$\textit{Precision (of a no-skill classifier)} = \frac{\textit{Positive Cases}}{(\textit{Positive Cases} + \textit{Negative Cases})}$$

228 Previous work suggests that a precision-recall plot is more informative than an ROC plot  
229 when evaluating a binary classifier in the presence of imbalanced data [4, 5], as was the  
230 case with the home vs. community threshold. For comparison purposes, we present the  
231 precision-recall curves for the 6-Minute Walk Test and speed modulation for predicting  
232 home vs. community and aerobic thresholds, respectively (S2 Fig).

233 As shown in Supplemental Figure 2A, the AUC for the 6MWT in predicting home  
234 vs. community ambulation is 0.642 which exceeds that of a no-skill classifier (precision  
235 of no-skill classifier =  $58/(210 + 58) = 0.2164$ ).

## 236 **Speed modulation as a single predictor of aerobic** 237 **threshold (5500 steps/day)**

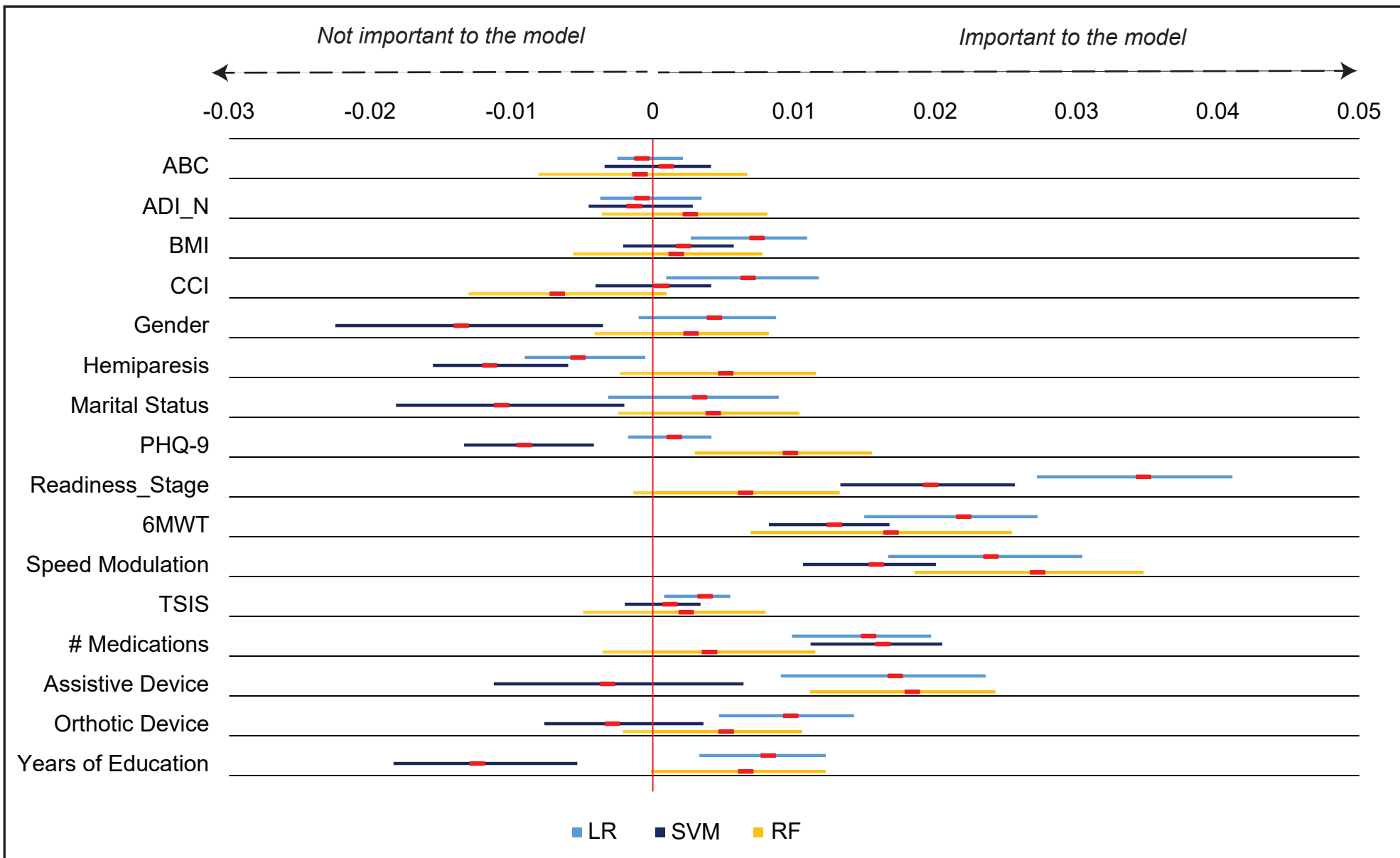
238 Supplemental Figure 2B displays the precision-recall curve for speed modulation  
239 in predicting the aerobic threshold of 5500 steps/day. The dashed line reflects a no-skill  
240 classifier and is calculated as: precision =  $185/(83 + 185) = 0.6903$ . The AUC of 0.849  
241 exceeds that of the no-skill classifier. When comparing the two precision-recall curves  
242 (and specifically the AUC values), it can be observed that speed modulation is a  
243 stronger predictor of the aerobic threshold than the 6MWT is in predicting the home vs.  
244 community threshold.

245 We refer the reader to the reference list below for further reading on these topics.

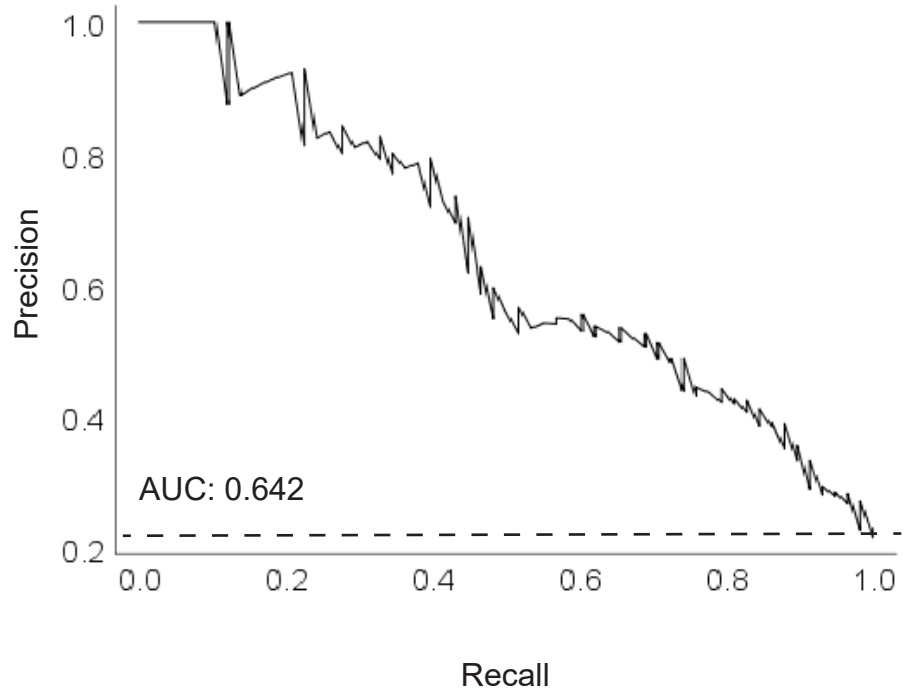
## 246 **References**

- 247 1. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning: Springer  
248 series in statistics New York; 2001.
- 249 2. Krawczyk B. Learning from imbalanced data: open challenges and future  
250 directions. Progress in Artificial Intelligence. 2016;5(4):221-32. doi: 10.1007/s13748-  
251 016-0094-0.
- 252 3. Ferri C, Hernández-Orallo J, Modroiu R. An experimental comparison of  
253 performance measures for classification. Pattern Recognition Letters. 2009;30(1):27-38.  
254 doi: 10.1016/j.patrec.2008.08.010.
- 255 4. He H, Garcia EA. Learning from Imbalanced Data. IEEE Transactions on  
256 Knowledge and Data Engineering. 2009;21(9):1263-84. doi: 10.1109/TKDE.2008.239.
- 257 5. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the  
258 ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PloS one.  
259 2015;10(3):e0118432. doi: 10.1371/journal.pone.0118432.

260



**A)**



**B)**

