**Predicting response to motor therapy in chronic stroke patients**

**using Machine Learning**

Running title: Predicting response to therapy in stroke

Ceren Tozlu1, Dylan Edwards2,3, Aaron Boes4, Douglas Labar5, K. Zoe Tsagaris3, Joshua Silverstein3, Heather Pepper Lane3, Mert R. Sabuncu6, Charles Liu7, Amy Kuceyeski1,8

1Department of Radiology and 8Brain and Mind Research Institute, Weill Cornell Medicine, New York, NY, USA

2Moss Rehabilitation Research Institute, Elkins Park, PA, USA; and Edith Cowan University, Joondalup, Australia

3Burke Neurological Institute, White Plains, NY, USA

4Iowa Neuroimaging and Noninvasive Brain Stimulation Laboratory, Departments of Pediatrics, Neurology & Psychiatry, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, Iowa City, IA 52242, USA

5Department of Neurology, Weill Cornell Medical College, New York, NY, USA

6School of Electrical and Computer Engineering, and Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA

7USC Neurorestoration Center, Los Angeles, CA; and Rancho Los Amigos National Rehabilitation Center, Downey, CA, USA

Corresponding Author: Amy Kuceyeski

Mailing Address: 526 Campus Road, Biotechnology building, Room 101D, Ithaca NY 14850

e-mail: amk2012@med.cornell.edu

**Supplementary material**

*Parameter setting*

The five methods require selection of hyperparameters that define the models. The best parameters for each method were chosen with grid-search analysis. The mean square error and classification error measurements were used respectively for the regression and classification analyses.

Elastic-net solves the following problem:

$$\min\_{β\_{0},β}[\frac{1}{2N} \sum\_{i=1}^{N}\left(y\_{i}-β\_{0}-x\_{i}^{T}β\right)^{2}+λ(\frac{\left(1-α\right)}{2}∥β∥\_{l\_{2}}^{2}+α∥β∥\_{l\_{1}})] $$

where N is the number of observations, $y\_{i }\in R $is the response vector, $x\_{i }\in R^{p}$is the predictor vector, and $β\_{0}$and $β$are the coefficients in the equation $E\left(X=x\right)=β\_{0}+x^{T}β$. $α$ takes the values between 0 and 1 as $α=0$ indicates the ridge penalty and $α=1$ indicates the Lasso penalty. The tuning parameter $λ $controls the strength of the penalty and is strictly higher than zero. A grid-searching approach was used where $α$ was in [0, 1] with steps of 0.1 and $λ$ was in [0, 3] with steps of 0.01.

SVM finds the hyperplane by minimizing

$$\frac{1}{2}∥w∥^{2}+C \sum\_{i=1}^{m}ξ\_{i}$$

 subject to $y\_{i}\left(w^{T}ϕ\left(x\_{i}\right)+b\right)\geq 1-ξ\_{i},$

 $ ξ\_{i}\geq 0$.

where $x\_{i}\in R $are the inputs, $y\_{i} $are the binary outputs, *C* is the misclassification error, and b is the bias. SVM with $ϵ$-regression aims to find a function *f(x)* as flat as possible with a deviation less than a given $ϵ$ for all the data. SVM estimate a linear regression $f(x)=<w,x>+b$

by minimizing $\frac{1}{2}∥w∥^{2}+\frac{C}{m}\sum\_{i=1}^{m}|y\_{i}-f(x\_{i})|$

subject to $\left\{\begin{array}{c}y\_{i}-<w,x\_{i}>-b \leq ϵ \\<w,x\_{i}>+b-y\_{i}\leq ϵ\end{array}\right.$

where C is the trade-off between the flatness of $f(x)$ and the amount up to which deviations larger than $ϵ$ are tolerated. A Gaussian Radial Basis kernel function $K (\vec{x\_{1}} ,\vec{x\_{2}}) =exp( -γ || \vec{x\_{1}} - \vec{x\_{2}} ||\^2) $was used for SVM where $γ $is a supplementary kernel parameter and $\vec{x\_{1}},\vec{x\_{2}}$ are observations. To optimize the kernel parameter $γ$ as well as the trade-off constant C, a grid-searching approach was used where C $\in $ (20, 21,22 …, 29) and $γ\in $(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1).

The maximum depth of the tree (*maxdepth*) and the minimum number of the observations in any terminal node (*minbucket*) were chosen as the hyperparameters for our CART analysis. The optimisation of *maxdepth* was searched in interval [2,10] with steps of 1 and *minbucket* in interval [5, 20] with steps of 5. CART choses the variable which minimizes the Gini Index (GI) in the classification analysis. GI at the node *t* is defined as

$$\sum\_{c=1}^{L}\hat{p\_{t^{c}}}(1-\hat{p\_{t^{c}}})$$

where $\hat{p\_{t^{c}}}$ is the proportion of the observation in class *c* at the node *t*.

For the ANN analysis, a single hidden layer was fitted and a grid-searching was used to optimize the number of nodes in interval [5, 10] with steps of 1 and a decay parameter $\in $ (0.0001, 0.001, 0.01, 0.1, 0.5).

In RF analysis, the number of variables used at each node (mrty) was set searched in interval [3, 10] with steps of 1, the number of trees optimized in the interval [10, 100] by increments of 10.

*Supplementary tables*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | EN | RF | ANN | SVM | CART |
| Pre- intervention FMA | 13.283 (1000) | 372.906 | 18.205 | 61.436 | 18401.267 |
| Difference in MT | 1.351 (1000) | 9.900 | 13.650 | 14.741 | 9092.622 |
| Absence or Presence of MEP | 0.265 (874) | 0.336 | 11.002 | 4.304 | 4894.644 |
| Gender | -0.087 (550) | 0.494 | 11.691 | 2.399 | 334.042 |
| Right Handed | 0.104 (428) | -0.014 | 10.904 | 2.270 | 317.954 |
| Time Since Stroke | 0.020 (81) | 0.714 | 12.532 | 4.713 | 1703.555 |
| Age | -0.004 (13) | 0.065 | 12.51 | 3.017 | 4283.188 |
| Stroke Hemisphere Side | 0.033 (13) | -0.127 | 9.505 | 1.760 | 485.003 |

***Supplementary Table 1.*** *Importance of the clinical variables only (demographics, clinical and neurophysiological measures) for all five machine learning methods in predicting post-intervention FMA.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | EN | RF | ANN | SVM | CART |
| Gender | -0.362 (1000) | -0.003 | 5.21 | 0.264 | 1.330 |
| Difference in MT | 0.098 (773) | 0.007 | 17.47 | 0.626 | 2.505 |
| Age | 0.023 (225) | 0.016 | 11.16 | 0.496 | 2.050 |
| Right Handed | 0.055 (207) | -0.001 | 15.74 | 0.821 | 0.618 |
| Time Since Stroke | -0.008 (125) | -0.001 | 10.00 | 0.392 | 0.542 |
| Pre- intervention FMA | 0.027 (117) | -0.008 | 15.08 | 0.790 | 6.305 |
| Stroke Hemisphere Side | -0.005 (78) | 0.000 | 21.47 | 0.657 | 0.542 |
| Absence or Presence of MEP  | 0.020 (53) | 0.001 | 3.84 | 0.189 | 0.616 |

***Supplementary Table 2.*** *Importance of the clinical variables only (demographics, clinical and neurophysiological measures) for all five machine learning methods in classifying the patients.*

|  |  |  |  |
| --- | --- | --- | --- |
| Method | Clinical Model | Clinical + regional disconnectivity Model | Clinical + pair-wisedisconnectivity Model |
| EN | 0.625 [0.447, 0.767] | 0.500 [0.500, 0.650] | 0.625 [0.560, 0.833] |
| RF | 0.433 [0.333, 0.533] | 0.500 [0.367, 0.625] | 0.500 [0.383, 0.683] |
| ANN | 0.500 [0.417, 0.750] | 0.688 [0.458, 0.792] | 0.667 [0.458, 0.700] |
| SVM | 0.583 [0.541, 0.708] | 0.458 [0.333, 0.567] | 0.417 [0.333, 0.500] |
| CART | 0.580 [0.467, 0.708] | 0.521 [0.375, 0.583] | 0.625 [0.500, 0.729] |

***Supplementary Table 3.*** *The classification results (AUC) for the five machine learning methods and three different input datasets. Values are presented as Median [1st quartile, 3rd quartile].*