Published in final edited form as:

Neurorehabil Neural Repair. 2020 May ; 34(5): 428–439. doi:10.1177/1545968320909796.

Machine Learning methods predict individual upper limb motor impairment following therapy in chronic stroke

Ceren Tozlu, (Ph.D.)¹, Dylan Edwards, (Ph.D.)^{2,3}, Aaron Boes, (MD, Ph.D.)⁴, Douglas Labar, (MD)⁵, K. Zoe Tsagaris, (M.S.)³, Joshua Silverstein, (B.S.)³, Heather Pepper Lane, (M.A.)³, Mert R. Sabuncu, (Ph.D.)⁶, Charles Liu, (MD, Ph.D.)⁷, Amy Kuceyeski, (Ph.D.)^{1,8}

¹Department of Radiology, 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

⁸Brain and Mind Research Institute, Weill Cornell Medicine, New York, NY, USA

Abstract

Background: Accurate prediction of clinical impairment in upper extremity motor function following therapy in chronic stroke patients is a difficult task for clinicians, but is key in prescribing appropriate therapeutic strategies. Machine learning is a highly promising avenue with which to improve prediction accuracy in clinical practice.

Objectives: The objective was to evaluate the performance of five machine learning methods in predicting post-intervention upper extremity motor impairment in chronic stroke patients, using demographic, clinical, neurophysiological, and imaging input variables.

Methods: One hundred and two patients (female: 31%, age 61±11 years) were included. The upper extremity Fugl-Meyer Assessment (UE-FMA) was used to assess motor impairment of the

Corresponding Author: Amy Kuceyeski, Ph.D., Mailing Address: 526 Campus Road, Biotechnology building, Room 101D, Ithaca NY 14850, amk2012@med.cornell.edu - Phone: 330 340 5847.

Competing interests:

The authors declare that they have no competing interest.

Data availability statement:

The deidentified data that support the findings of this study are confidential. The codes that were used in this study are available. Please see <https://github.com/cerent/Stroke-UE-FMA>

upper limb before and after intervention. Elastic-Net (EN), Support Vector Machines (SVM), Artificial Neural Networks (ANN), Classification and Regression Trees (CART), and Random Forest (RF) were used to predict post-intervention UE-FMA. The performances of methods were compared using cross-validated r-squared (R^2).

Results: Elastic-Net performed significantly better than other methods in predicting post-intervention UE-FMA using demographic and baseline clinical data (median $R_{EN}^2 = 0.91$, $R_{RF}^2 = 0.88$, $R_{ANN}^2 = 0.83$, $R_{SVM}^2 = 0.79$, $R_{CART}^2 = 0.70$; $p < 0.05$). Pre-intervention UE-FMA and the difference in motor threshold (MT) between the affected and unaffected hemispheres were the strongest predictors. The difference in MT had greater importance than the absence or presence of a motor-evoked potential (MEP) in the affected hemisphere.

Conclusion: Machine learning methods may enable clinicians to accurately predict a chronic stroke patient's post-intervention UE-FMA. Inter-hemispheric difference in MT is an important predictor of chronic stroke patients' response to therapy, therefore could be included in prospective studies.

Keywords

chronic stroke; predictive models; Fugl-Meyer Assessment; machine learning; white matter disconnectivity

Introduction

Stroke is one of the most common diseases causing functional impairment worldwide and the most common cause of the morbidity in many developed countries.^{1,2} Stroke survivors frequently have residual motor impairments which impact negatively their quality of life.³ One of the difficult challenges facing clinicians is to determine whether or not a patient will benefit from a certain type of treatment, particularly in the case of chronic stroke where gains may be more incremental than in the acute to the sub-acute stage. Initial clinical scores, demographic information, and imaging data have been shown to be important predictors of outcome in both acute and chronic stroke patients, yet at the individual level, these remain unsatisfactory.^{1,4-7} When assessed in the acute to subacute post-stroke period, the severity of motor impairment and neurophysiological inputs, including the presence of motor evoked potentials (MEP) in response to transcranial magnetic stimulation (TMS) in the affected hemisphere⁸⁻¹⁰ and the motor threshold (MT)¹¹ were found to be important predictors of clinical improvement.

Many recent studies have shown the importance of MRI-based metrics, including lesion size/location and connectome disruption, i.e. the brain's functional and structural connections, in post-stroke clinical impairment and recovery.¹²⁻¹⁸ Here, we investigate the additive value of MRI-based measurements of structural connectome disruption that have been shown to be predictive of baseline impairment and recovery in our previous work.^{17,18} In particular, those studies used the Network Modification (NeMo) Tool¹⁹, which quantifies each gray matter region's white matter disconnectivity to the rest of the brain as well as the amount of disconnectivity between pairs of regions. The regional measurements, called Change in Connectivity (ChaCo) scores, represent the percent of disrupted white matter

fibers connecting to a given region. The pairwise disconnectivity measures represent the change in the number of fibers connecting any given pair of regions after removing those fibers that pass through the area of lesion.

Machine learning methods are highly promising for quantitative predictions of post-stroke recovery as well as assessment of demographic or imaging variables that are important in these predictions.^{20,21} Support vector machine (SVM) and artificial neural networks (ANN) have previously shown accurate results in predicting clinical scores in chronic stroke patients.^{22,23} SVM and tree-based methods (Classification and Regression Tree [CART] and bagging forest) have also been used to classify patients into groups that did and did not have improvement in motor function and to identify significant predictors of these classes.^{10,24} A recent study reported that a classic regression method using Elastic-Net (EN) regularization performed just as well as SVM, ANN and Random Forest (RF) in predicting the clinical and radiological outcomes after endovascular therapy in acute ischemic stroke patients.²⁵

The primary aim of the present study was to assess the performance of common machine learning methods (SVM, ANN, CART, RF, EN) in predicting chronic stroke patients' upper extremity motor function after six weeks of intervention based on demographic, clinical, neurophysiological and imaging metrics. An advantage of EN compared to other methods is that variable selection and model fitting are performed simultaneously, which can be useful in high-dimensional datasets.²⁶ The upper extremity Fugl-Meyer Assessment (UE-FMA), a comprehensive and consistent clinical scale, was used to assess upper limb motor impairment before and after intervention.^{27,28} The secondary aim was to assess the importance of variables in predicting post-intervention UE-FMA. We hypothesized that pre-intervention UE-FMA and the neurophysiological variables are the most important predictors in predicting the post-intervention UE-FMA. The third aim of our study was to use the previously discussed machine learning methods to classify patients into two classes, i.e. responders versus non-responders, as opposed to the regression methods predicting continuous outcomes (post-intervention UE-FMA) in the first aim. The classification of the patients into binary classes based on significant change in upper extremity motor function, while a harder task overall because of the distribution of the change in UE-FMA, may be more easily utilized by clinicians in the clinical setting.

Materials and Methods

Subjects

The data for this paper were extracted and further analyzed from a subset of a broader multisite intervention trial with the principal results reported elsewhere.²⁹ Institutional Review Board approval and individual consent were obtained at each site. The authors were able to obtain access to data collected from two out of the twelve institutions that participated in the original study, resulting in a subset of 102 subjects (45 from the Burke Neurological Institute and 57 from Rancho Los Amigos rehabilitation center) out of the original study's total of 199 subjects. We had two subjects with missing data; one had no identifiable lesion on MRI and one subject was lost the follow-up. Adult (>18years) subjects with residual hemiparesis from a first-time ischemic or hemorrhagic stroke within 3–12 months prior were enrolled in the study. Some patients sometimes had impairments in

multiple domains. Patients scoring >1 for limb ataxia, >2 for best language, or a score of 2 for extinction and inattention on the National Institutes of Health stroke scale were excluded, see the original study for details on exclusion and inclusion criteria.²⁹ Handedness refers to pre-stroke.

Intervention and TMS protocols

The detailed information of the rTMS and physical therapy can be found in Harvey et al., 2018, the paper that originally presented these data.²⁹ In brief, participants completed 18 therapeutic sessions over a 6-week period (typically 3 × per week). Each session consisted of 1) 20 minutes of warm-up (stretching/strengthening) focusing on shoulder and elbow mobilization, 2) 10-minute rest, 3) repetitive TMS (active or sham, depending on the individuals' group assignment) 4) 5–10 minute rest and 5) a 60-minute task-oriented rehabilitation therapy session that focused on arm and hand practice graded according to participants' current Chedoke arm stage. We delivered TMS using a navigated brain therapy device (NBT, Nexstim Corporation, Finland). Active TMS was delivered using 1 Hz frequency to the non-injured motor cortex for > 15 minutes per assigned protocol. Sham TMS aimed to provide no stimulation at the target site but instead delivered weak stimulation over a wide doughnut-shape region around the target area, thus minimize stimulation to hand-arm primary motor cortex. One third of patients received sham stimulation (2:1 randomization, N = 34) and two thirds received real stimulation (N = 68); a previous study of the superset of 199 patients found there were no differences in the change in UE-FMA between the sham and real stimulation groups.²⁹ We verified the same finding by performing a t-test of the change in UE-FMA between our data's sham and real stimulation groups and found there was no significant difference ($p = 0.46$). Since determining the impact of sham versus real stimulation was not a goal of our work and there were no differences in improvement in UE-FMA between the groups, we did not use this variable as a predictor. TMS-based neurophysiological information (difference in MT in the unaffected versus affected hemispheres and absence or presence of a MEP in the affected hemisphere) was collected for each patient. MT was determined for the extensor digitorum communis using the Nexstim software algorithm.³⁰ Muscles were at rest during MT determination. The TMS operator monitored ongoing EMG and excluded any MEPs in which EMG activation was $25\mu\text{V}$ within 100ms of stimulus onset. MT was determined by a minimum of 16 trials, using 50 μV minimum amplitude criteria.³¹ If MT could not be established (i.e. no MEP at the upper limit of stimulator output) in the affected hemisphere, patients were assigned a value of 100 (% Maximum Stimulator Output).³²

Image acquisition and processing

Structural MRIs were acquired on Siemens 3T scanners prior to commencement of the intervention (T1 MPRAGE, 1-mm iso-voxel at both sites). Lesion masks were hand-traced on patients' native T1 scans. T1 images and the associated lesion masks were transformed to MNI space using both linear (FLIRT) and non-linear transformation techniques (FNIRT) available in FSL (<http://www.fmrib.ox.ac.uk/fsl/index.html>). Lesion borders were confirmed by a neurologist (A.B.) who was blinded to the behavioral data, both in native space and again after transformation to MNI space. Left-hemisphere lesions were flipped on the x-axis so that all lesions across both cohorts appeared in the right hemisphere in order to increase

the signal to noise ratio of the imaging measures. The lesion masks were then processed through the NeMo Tool software, which estimates the amount of regional disconnectivity and disconnectivity between pairs of regions based on a database of 73 healthy control brains' white matter connectivity maps (40 men, 33 women, 30.2±6.7 years). Regional disconnectivity measurements (ChaCo scores) were quantified as the percent of white matter streamlines passing through a lesion divided by the total number connecting to that region, while pair-wise disconnectivity measurements are given as z-scores quantifying the number of white matter fibers between pairs of regions that pass through a lesion.

Statistical Analysis

For the regression analyses, five machine learning methods (EN²⁶, ANN³³, SVM³⁴, CART³⁵, and RF³⁶ – see Supplementary Materials for details) were implemented to predict post-intervention UE-FMA. The input dataset included three demographic variables (age, sex, handedness before stroke), three clinical variables (time since stroke, left vs. right hemisphere stroke, pre-intervention UE-FMA), two TMS-based neurophysiological measures (difference in MT in the unaffected versus affected hemispheres and absence or presence of a MEP in the affected hemisphere), and structural disconnectivity measurements for 86 regions. While the dataset including the demographic, clinical, and TMS-based information only has a relatively low number of input features (7 total), the other two datasets incorporating imaging biomarkers have a high number of features (86 + 7 = 93 total). Therefore, we chose to implement ML methods on all of the datasets in order to be able to compare directly the results of both low and high dimensional input datasets.

For the classification analyses, the five machine learning methods that were used for the regression analysis, and a basic and popular classification method Logistic Regression (LR) were used to classify the patients into those that improved by a clinically meaningful amount versus those that did not, using the clinical, neurophysiological, and imaging metrics. The same variables were used for both regression and classification analysis. The motivation for performing classification in addition to regression is that it may be more useful clinically to provide a dichotomous outcome (responder versus non-responder) to aid in treatment decisions. The minimal clinically important difference (MCID) in the FMA has been reported previously to be 5.5³⁷; thus we used this number to separate our patients into the two groups with post-intervention FMA minus pre-FMA < 5.5 (non-responder) or ≥ 5.5 (responder). Pre-intervention UE-FMA, time since stroke, age, and the difference in MT were standardized to facilitate the interpretation of EN method's coefficients.

Three datasets for each regression/classification analysis were constructed using different sets of input variables. First, a “clinical” dataset was used that included only demographics, clinical, and TMS-based neurophysiological information. Next, two “combined” datasets were used that contained the variables in the clinical dataset plus 1) ChaCo (regional disconnectivity) scores or 2) pair-wise disconnection scores from the NeMo Tool. Machine learning methods' performance for the regression task was assessed using the root of mean squared error (RMSE) and R-squared, defined as $R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$ where $\hat{y}_i$ is the prediction of y_i and $\bar{y}$ is the mean. Method performance for the classification was mainly

assessed using the area under Receiver Operating Characteristic curve (AUC). Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were also calculated to facilitate comparison to previous findings in the literature.

Each method was trained with two loops (outer and inner) of k-fold cross validation ($k = 10$) to optimize the hyperparameters (see Supplementary Material for details and the values of the best hyperparameters selected in the inner loop) and test the performance of the methods. The outer loop provided a training set (9/10 folds) for machine learning methods's hyperparameters optimization and test set (1/10 folds) for method's performance assessment using the best hyperparameters. Thus, the training dataset included approximately 90 patients while the testing dataset included about 10 patients. The folds that were created in the inner and outer loop were stratified, thus each fold contained the same proportion of the patients who had a MCID in the FMA and those who did not with the original dataset. The inner loop (repeated over 10 different partitions of the training dataset only) performed grid-search to find the set of method hyperparameters that minimized the average hold-out RMSE for regression or maximized AUC for classification (see the supplementary material for the intervals of the hyperparameters). A method was fitted using those optimal hyperparameters on the entire training dataset and assessed on the hold-out test set from the outer loop. The outer loop was repeated for 100 different random partitions of the data (see Supplementary Figure 1 for a visualization of the cross-validation scheme). The mean test R^2 and RMSE in the regression analysis, and the mean test AUC in the classification analysis (over all 10 folds $\times$ 100 iterations = 1000 test sets) were calculated to assess the performance of the methods for each dataset. The performance metrics were compared across methods and input datasets with the Kruskal-Wallis and Wilcoxon rank sum test, and considered significantly different when $p < 0.05$. False discovery rate (FDR) correction was used to address the issue of multiple comparisons.

When the data is imbalanced in a classification task, i.e. one class has more subjects than the other, machine learning methods may tend to favor the class with the majority observation. Therefore, the methods may fail to make an accurate prediction for the minority class that has less information. Due to the class imbalance in our data (42% that had clinically significant improvement vs 58% that did not), the over-sampling approach Synthetic Majority Over-sampling Technique (SMOTE)³⁸ was used to obtain a balanced training dataset during the cross-validation and to improve the prediction accuracy for the minority class. SMOTE compensates for imbalanced classes by creating synthetic examples using nearest neighbor information instead of creating copies from the minority class and has been shown to be among the most robust and accurate methods with which to control for imbalanced data.³⁹

We considered the importance of the variables in EN to be the magnitude of the regression coefficient of the method averaged over all 1000 results (100 iterations of the outer loop $\times$ 10 test datasets for each iteration). The importance of the variables for CART was considered to be the sum of squared error for all the splits in which a variable is used.³⁵ In RF, each variable is randomly permuted and the difference in the new MSE and the original MSE is considered that variable's importance.⁴⁰ The importance of the variables in ANN is calculated by taking the absolute value of the input-hidden layer connection weight and

dividing that by the sum of the absolute value of the input-hidden layer connection weight of all input neurons.⁴¹ The difference in output for the maximum and minimum values of a given variable across the subjects, while other variables are held constant at the mean across the subjects, gives the importance of the variables for SVM.⁴² All statistical analyses and graphs were performed using R (<https://www.r-project.org>), version 3.4.4.

Results

Patient Characteristics

Table 1 shows patient demographics, clinical information, lesion volume and neurophysiological data. Supplementary Figure 2 shows the lesion heatmap over the individuals. Pre-intervention UE-FMA was significantly lower than post-intervention UE-FMA (p-value < 0.01). The majority of the patients were right handed (90 of 102 patients). The number of the patients with left hemisphere stroke was similar to the number of the patients with right hemisphere stroke (53 vs 49). Forty-three patients had clinically significant increases in UE-FMA (responders) while 59 did not (non-responders).

Prediction of Post-intervention UE-FMA

Figure 1 shows a violin plot (probability density function where white point = median, thick black bar = Interquartile Range [IQR]) of each method/datasets' RMSE over the 1000 results [from 100 iterations of the outer loop $\times$ 10 test datasets for each iteration], while Table 2 lists the median and IQR for each method/datasets' R^2 and RMSE. Table 2 and Figure 1 illustrate that the EN method had the highest explained variance and lowest RMSE for all three input datasets. The addition of regional disconnectivity to the clinical variables resulted in less explained variance for SVM and RF, and more explained variance for ANN and CART. SVM and CART showed significantly higher R^2 values with the clinical and pair-wise disconnection dataset compared to the clinical dataset only while RF and ANN had greater explained variance. The addition of both disconnectivity measurements to the clinical dataset resulted in the same amount of explained variance for EN. The RMSE results largely agree with the R^2 results as the EN method trained on the clinical data had the lowest RMSE. Most of RMSE results were significantly different across the different sets of input variables. Only the results of SVM and CART were not significantly different for the clinical + regional disconnectivity dataset (p-value=0.932). The RMSE of clinical + regional disconnectivity and clinical + pair-wise disconnectivity datasets were significantly greater than the clinical dataset for all methods.

CART showed the lowest R^2 results and EN the highest R^2 results for the dataset based on clinical input variables. Figure 2 shows the observed post-intervention UE-FMA versus the average ($\pm$ standard deviation) of each subject's predictions over all 1000 models created during the cross-validation procedure. We see that, compared to the CART predictions, EN predictions are closer to the line of identity and the standard deviations are smaller, meaning there was less uncertainty in the prediction. The uncertainty for the CART method predictions for individuals with observed UE-FMA under 30 is particularly large.

Figure 3 shows bar plots indicating variable importance for each of the five methods that were trained on the clinical dataset. Unsurprisingly, pre-intervention UE-FMA had the highest importance for all of the methods, followed by the difference in MT as the next most important predictor. The absence or presence of MEP in the affected hemisphere and time since stroke were also found to be important predictors. Pre-intervention UE-FMA and difference in MT were chosen 1000 times over 1000 results (100 iterations of the outer loop $\times$ 10 test datasets for each iteration) performed with EN. The absence or presence of MEP in the affected hemisphere, sex and right handedness were selected more than 400 times over 1000 results by EN method (see Supplementary Table I).

Because of the importance of the pre-intervention UE-FMA and the difference in MT measures in the analyses predicting post-intervention UE-FMA, we wanted to fully quantify the value of adding the TMS measures to a dataset including only pre-intervention UE-FMA. Therefore, in a post-hoc analysis, we specifically investigated the role of the neurophysiological data in improving the post-intervention UE-FMA predictions. We compared linear regression methods based on 1) pre-intervention UE-FMA only, 2) pre-intervention UE-FMA and difference in MT and 3) pre-intervention UE-FMA and the absence or presence of the MEP at the affected hemisphere. The performance of the linear regression method for these three datasets was compared using the weighted Akaike Information Criterion (wAIC).⁴³ wAIC gives the probability that each of the possible models is the best model for the given data. The model based on pre-intervention UE-FMA and difference in MT was the best model, with a probability higher than for the model based only on pre-intervention UE-FMA (wAIC=0.95 vs 0.05). The wAIC values show that the model that contains difference in MT is 19 ($=0.95/0.05$) times more likely to be a better model than the model based only on pre-intervention UE-FMA. However, the wAICs were close when the absence or presence of MEP was added to the model based only on pre-intervention UE-FMA (wAIC=0.57 vs 0.43).

Figure 4 and Table 3 show that RF and ANN gave satisfactory results, having at least one AUC greater than 0.6 over the different sets of input variables. EN, RF, and LR trained on the clinical data performed similarly with an AUC of 0.55 and 0.56, while SVM and ANN performed better than the other methods ($AUC > 0.58$). However, RF and ANN performed better than other methods ($AUC=0.63$ vs $AUC < 0.57$) when trained on the dataset including the clinical + pair-wise dysconnectivity dataset. EN showed greater AUC compared to LR when trained on higher-dimensionality clinical + pair-wise disconnectivity data ($AUC_{EN}=0.57$ vs $AUC_{LR}=0.50$, $p < 0.05$), while they performed similarly on the other input datasets ($AUC_{EN}=0.56$ vs $AUC_{LR}=0.55$ for the clinical dataset and $AUC_{EN}=0.50$ vs $AUC_{LR}=0.50$ for the clinical dataset, $p > 0.05$ for both). RF and ANN performed significantly better with the clinical + pair-wise disconnectivity data compared to the clinical dataset (Wilcoxon rank sum test, $p\text{-value} < 0.05$ for RF and $p\text{-value} = 0.014$ for ANN), while SVM and LR performed the best on the clinical dataset. EN, RF, ANN performed well in classifying the patients who did not have a significant increase in UE-FMA with a sensitivity greater than 60% with the datasets including pair-wise disconnectivity information, while SVM showed the best specificity (above 80% with the input datasets including disconnectivity measurements) in classifying the patients with a clinically significant increase in UE-FMA (see Supplementary Table III). The importance of the variables was

also analyzed for the classification task. RF, ANN, and CART showed pre-intervention UE-FMA as the best predictor in classifying the patients who did have a significant increase in UE-FMA and those who had not (see Supplementary Table II). The neurophysiological inputs were also found as important variables by some of the methods. The difference in MT was shown as the best predictor by SVM and LR, and as the second strongest predictor by CART, while the absence or presence of MEP was given as the second most important predictor by ANN in classifying the patients who did have a significant increase in UE-FMA and those who had not. We want to emphasize that the classification task is not merely predicting a binarized version of the regression task's outcome variable. The regression task was to predict post-intervention UE-FMA (for example: the patient #1 was predicted to have UE-FMA 40 after the intervention) while the classification task was to predict the classes defined using the change between the pre- and post- intervention UE-FMA (for example: the patient #2 was predicted to be in the class that indicated a significant change in UE-FMA between pre and post-intervention). The latter is a much more difficult task, as evidenced by the moderate AUC in the classification task when compared to the relatively high explained variance in the regression task.

Discussion

Here, we applied five machine learning methods to predict the response to intervention in a large cohort of chronic stroke patients. We predicted both a continuous measure of post-intervention UE-FMA as well as whether or not a patient would achieve a clinically meaningful improvement in UE-FMA. This study is one of the largest to date that uses various machine learning methods on demographic, clinical, neurophysiological and imaging data to predict post-intervention UE-FMA in chronic stroke. Our main findings were that 1) EN performed best when predicting the continuous-valued post-intervention UE-FMA, 2) ANN and RF performed the best when predicting the binary-value of clinically meaningful change (or not) in UE-FMA and 3) the difference in MT between affected and unaffected hemispheres had more importance than other metrics, including age, time since stroke and imaging-based measures of structural disconnection in the predictive models. We conjecture that the EN was generally better in the regression task than others since it optimizes for sparsity; there were only a few variables in our dataset that were strong predictors of post-intervention UE-FMA. Important predictors in our dataset, including pre-intervention UE-FMA and difference in MT, largely agree with previous findings.^{32,44–47} Pre-intervention UE-FMA highly correlated with post-intervention UE-FMA. However, using other (clinical, demographic, neurophysiological, and imaging) variables together with pre-intervention UE-FMA in the same dataset provided the relative importance of the variables. In particular, the post-hoc analysis showed that adding difference in MT to a dataset based only on pre-intervention UE-FMA significantly improves the method's performance. Therefore, our results suggest the difference in MT could be included in prospective studies for chronic stroke post-intervention UE-FMA prediction. This may also have relevance in acute studies that typically use absence or presence of MEP in the affected hemisphere for outcome prediction, although this remains to be systemically tested. All methods showed higher RMSE results when the regional or pair-wise structural disconnectivity measurements were added into the clinical dataset, while RF and ANN

showed better classification performance when pair-wise structural disconnectivity measurement were used compared to the clinical dataset.

There has been increasing interest in using machine learning methods for predicting impairment and recovery from stroke.^{10,17,18,48–51} One such study developed a model that had promising accuracy of acute post-stroke outcome prediction after the intra-arterial therapy with ANN and SVM using demographic and comorbidity information.⁶ Rehme et al.⁵² used SVM to classify patients with respect to acute post-stroke motor impairment based on the resting-state functional MRI; they reported a classification rate of over 80%. In addition to SVM and ANN methods, decision tree based classification algorithm was used to predict upper-extremity outcome at 3 months and 2 years post-stroke using baseline demographic, neurophysiological, and MRI features.^{10,53–55} The important predictors identified in the CART analysis were used to create a decision tree based Predicting Recovery Potential (PREP) and the Predicting Recovery Potential 2 (PREP 2) algorithms that can classify patients into categories of upper limb recovery.^{10,53–55} These studies largely found that imaging metrics did not significantly improve accuracy of PREP 2 algorithm (75%) and the accuracy was the same when MRI biomarkers were available, which agrees with our current findings. The PREP and PREP2 algorithms gave better classification accuracy than our CART analysis. However, our EN method trained on clinical + regional dysconnectivity data had higher sensitivity than the PREP algorithm. The PPV of the PREP 2 algorithm for patients in the good recovery category was similar to the PPV of our CART analysis (58% vs 60%). The same publications reported that the absence or presence of a MEP, acute motor functionality and age were most important in predicting recovery, which are slightly different than the most important variables found in our study. This could be due to many differences between our studies, the most likely of which is that our predictions were for changes in chronic stroke due to intervention and not in predicting acute recovery. The former is a much more difficult task, as the changes in the chronic stage of stroke are most often not as large and thus have a lower signal-to-noise ratio. Attempts have been made to predict response of chronic stroke subjects to treatment or interventions, although most are in moderate sample sizes or are correlation based and do not utilize machine learning techniques.^{56,57} One study, in particular, found that measures of functional and structural connectivity were important in predicting motor gains from therapy in chronic stroke; however their overall variance explained by the predictors was moderate ($R^2 = 0.44$) and their sample size relatively small ($N = 29$).⁵⁸

The present study on chronic, intervention-related response prediction, supports findings in studies of post-intervention UE-FMA prediction from the early phase, that baseline impairment and ipsilesional corticospinal excitability are key predictive factors. An important addition to the body of knowledge for the present study, as well as examining a distinctly different phase of recovery post stroke than is most often reported (chronic), is one of method. Namely, that the present best practice of ipsilesional MEP presence, while an important predictor here, was inferior to a bilateral recoding of resting motor threshold. This subtle but important distinction would indicate that at least in the chronic phase, bilateral assessment of MEPs could be conducted for response prediction. A consideration for use of the motor evoked potential as an outcome predictor, is the influence of amplitude and probability criteria for MEP *presence* determination, which requires further exploration.⁵⁹

The post-intervention UE-FMA of the patients was tracked at 1, 3, and 6 months of the therapy. The UE-FMA at 1, 3, and 6 months of the intervention was significantly correlated with the post-intervention UE-FMA just after the intervention (Pearson's coefficient correlation >0.90 for all). Also, we have observed that 80% of the patients who showed MCID in UE-FMA immediately after the six weeks of intervention showed persistent recovery with MCID ≥ 5.5 at 6 months after the intervention.

One limitation of this study is that we had only the patient's structural T1 scans. It may be more informative to have access to stroke individual's diffusion or functional MRI, which have been shown to be important for extracting biomarkers that can predict impairment and recovery after stroke.^{7,15,60} One recent study showed increased baseline functional connectivity in certain regions in stroke patients who recovered better than those that did not.⁶⁰ Other studies have shown biomarkers of structural white matter integrity, particularly in the motor tracts, are predictive of recovery.^{7,61,62} Future studies will focus on collecting multi-modal imaging data at baseline and post-intervention to make better predictions of response to treatment as well as detect recovery-relevant changes that may shed light on neurological mechanisms of motor improvements. Beside the imaging variables, dexterity, attention, visuo-spatial neglect, sensory deficits motivation/depression might be important predictors of motor impairment and will be added in future studies. Another limitation of this study was the sample size; while relatively large for studies of this nature, there are machine learning techniques (such as deep learning) that may improve accuracy of predictions but are best implemented on larger sets of data. Finally, while the NeMo Tool has been applied in previous studies involving older subjects^{16,17,63}, one of its limitations is that the healthy controls it is based on are quite younger than the stroke population. However, it should be noted that the ChaCo scores it produces are a percent of disrupted connections, a measure which is likely to a large extent preserved in healthy aging subjects.

In summary, the machine learning methods employed here, gave highly promising results in predicting post-intervention UE-FMA using demographic, clinical and neurophysiological data. The same methods could effectively classify patients into responders versus non-responders based on change in UE-FMA, yet, had lower accuracy. Thorough validation of these types of machine learning methods using larger, multi-modal sets of data is needed. If successful, the final model will be easily accessible by clinicians and provide them a valuable tool (application or software) that can improve the accuracy of prognoses and response to treatment after rTMS and physical therapy as indicated in our study, which in turn can assist in developing personalized therapeutic plans at the chronic stage.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

C.T. carried out the statistical analyses, drafted and wrote the article.

D.E. performed study oversight, data interpretation, manuscript development

A.B. checked the data, commented and reviewed the article.

K. Z. T., CL and H. P. L. organized, checked, and managed the data collection, commented and reviewed the article.

J.S. performed neurophysiology data acquisition, processing and interpretation

M.S. supervised the analysis and reviewed the article.

A.K. organized the study, commented and reviewed the article.

Funding:

This work was supported by the NIH R21 NS104634-01 (A.K.), NIH R01 NS102646-01A1 (A.K.), NIH R01 grants (R01LM012719 and R01AG053949) (M.S.), the NSF NeuroNex grant 1707312 (M.S.), and NSF CAREER grant (1748377) (M.S.). D.J.E. serves on the advisory board for Nexstim Ltd.

Abbreviations:

ANN	Artificial Neural Networks
AUC	Area Under Receiver Operating Characteristic Curve
CART	Classification and Regression Trees
ChaCo	Change in Connectivity
EN	Elastic-Net
GI	Gini Index
LR	Logistic Regression
MCID	Minimal Clinically Important Difference
MEP	Motor Evoked Potentials
MSE	Mean Squared Error
MT	Motor Threshold
NPV	Negative Predictive Value
NeMo Tool	Network Modification Tool
PPV	Positive Predictive Value
PREP	Predicting Recovery Potential
RMSE	Root of Mean Squared Error
RF	Random Forest
SVM	Support Vector Machine
TEP	Transcranial magnetic stimulation evoked potential
TMS	Transcranial magnetic stimulation
UE-FMA	Upper extremity Fugl-Meyer Assessment

wAIC

Weighted Akaike Information Criterion

References:

1. Coupar F, Pollock A, Rowe P, Weir C, Langhorne P. Predictors of upper limb recovery after stroke: A systematic review and meta-analysis. *Clin Rehabil.* 2012;26(4):291–313. doi:10.1177/0269215511420305 [PubMed: 22023891]
2. Karahan AY, Kucuksen S, Yilmaz H, Salli A, Gungor T, Sahin M. Effects of Rehabilitation Services on Anxiety, Depression, Care-Giving Burden and Perceived Social Support of Stroke Caregivers. *Acta Medica (Hradec Kral Czech Republic).* 2014;57(2):68–72. doi:10.14712/18059694.2014.42
3. Dobkin BH. Rehabilitation after stroke. *N Engl J Med.* 1990;352(16):1677–1684.
4. Kim B, Winstein C. Can Neurological Biomarkers of Brain Impairment Be Used to Predict Poststroke Motor Recovery? A Systematic Review. *Neurorehabil Neural Repair.* 2017;31(1):3–24. doi:10.1177/1545968316662708 [PubMed: 27503908]
5. Burke E, Cramer SC. Biomarkers and predictors of restorative therapy effects after stroke. *Curr Neurol Neurosci Rep.* 2013;13(2):1–14. doi:10.1007/s11910-012-0329-9.Biomarkers
6. Asadi H, Dowling R, Yan B, Mitchell P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PLoS One.* 2014;9(2):14–19. doi:10.1371/journal.pone.0088225
7. Stinear CM, Barber PA, Smale PR, Coxon JP, Fleming MK, Byblow WD. Functional potential in chronic stroke patients depends on corticospinal tract integrity. *Brain.* 2007;130(1):170–180. doi:10.1093/brain/awl333 [PubMed: 17148468]
8. Escudero JV, Sancho J, Bautista D, Escudero M, Lopez-Trigo J. Prognostic Value of Motor Evoked Potential Obtained by Transcranial Magnetic Brain Stimulation in Motor Function Recovery in Patients With Acute Ischemic Stroke. *Stroke.* 1998;29(9):1854–1859. doi:10.1161/01.STR.29.9.1854 [PubMed: 9731608]
9. Hendricks HT, Pasman JW, van Limbeek J, Zwarts MJ. Motor Evoked Potentials in Predicting Recovery from Upper Extremity Paralysis after Acute Stroke. *Cerebrovasc Dis.* 2003;16(3):265–271. doi:10.1159/000071126 [PubMed: 12865615]
10. Stinear CM, Byblow WD, Ackerley SJ, Smith M-C, Borges VM, Barber PA. PREP2: A biomarker-based algorithm for predicting upper limb function after stroke. *Ann Clin Transl Neurol.* 2017;4(11):811–820. doi:10.1002/acn3.488 [PubMed: 29159193]
11. Kemlin C, Moulton E, Lamy JC, Rosso C. Resting motor threshold is a biomarker for motor stroke recovery. *Ann Phys Rehabil Med.* 2018;61:e26. doi:10.1016/j.rehab.2018.05.057
12. Grefkes C, Fink GR. Reorganization of cerebral networks after stroke: New insights from neuroimaging with connectivity approaches. *Brain.* 2011;134(5):1264–1276. doi:10.1093/brain/awr033 [PubMed: 21414995]
13. Westlake KP, Nagarajan SS. Functional Connectivity in Relation to Motor Performance and Recovery After Stroke. *Front Syst Neurosci.* 2011;5(March):1–12. doi:10.3389/fnsys.2011.00008 [PubMed: 21347218]
14. Carter AR, Astafiev SV, Lang CE, et al. Resting Inter-hemispheric fMRI Connectivity Predicts Performance after Stroke. *Ann Neurol.* 2010;67(3):365–375. doi:10.1002/ana.21905.Resting [PubMed: 20373348]
15. Siegel JS, Ramsey LE, Snyder AZ, et al. Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. *Proc Natl Acad Sci.* 2016;113(30):E4367–E4376. doi:10.1073/pnas.1521083113 [PubMed: 27402738]
16. Kuceyeski A, Kamel H, Navi BB, Raj A, Iadecola C. Predicting future brain tissue loss from white matter connectivity disruption in ischemic stroke. *Stroke.* 2014;45(3):717–722. [PubMed: 24523041]
17. Kuceyeski A, Navi BB, Kamel H, et al. Exploring the brain's structural connectome: a quantitative stroke lesion-dysfunction mapping study. *Hum Brain Mapp.* 2015;36(6):2147–2160. doi:10.1158/1541-7786.MCR-15-0224.Loss [PubMed: 25655204]
18. Kuceyeski A, Navi BB, Kamel H, et al. Structural connectome disruption at baseline predicts 6-months post-stroke outcome. *Hum Brain Mapp.* 2016;37(7). doi:10.1002/hbm.23198

19. Kuceyeski A, Maruta J, Relkin N, Raj A. The Network Modification (NeMo) Tool: Elucidating the Effect of White Matter Integrity Changes on Cortical and Subcortical Structural Connectivity. *Brain Connect.* 2013;3(5):451–463. doi:10.1089/brain.2013.0147 [PubMed: 23855491]
20. Wang Y, Fan Y, Bhatt P, Davatzikos C. High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables. *Neuroimage.* 2010;50(4):1519–1535. doi:10.1016/j.neuroimage.2009.12.092 [PubMed: 20056158]
21. Cohen JR, Asarnow RF, Sabb FW, et al. Decoding continuous variables from neuroimaging data: Basic and clinical applications. *Front Neurosci.* 2011;5(JUN):1–12. doi:10.3389/fnins.2011.00075 [PubMed: 21390287]
22. Wang J, Yu L, Wang J, Guo L, Gu X, Fang Q. Automated Fugl-Meyer Assessment using SVR model. 2014 IEEE Int Symp Bioelectron Bioinformatics, IEEE ISBB 2014 2014:0–3. doi:10.1109/ISBB.2014.6820907
23. Kim WS, Cho S, Baek D, Bang H, Paik NJ. Upper extremity functional evaluation by Fugl-Meyer assessment scoring using depth-sensing camera in hemiplegic stroke patients. *PLoS One.* 2016;11(7):1–13. doi:10.1371/journal.pone.0158640
24. Rondina JM, Filippone M, Girolami M, Ward NS. Decoding post-stroke motor function from structural brain imaging. *NeuroImage Clin.* 2016;12:372–380. doi:10.1016/j.nicl.2016.07.014 [PubMed: 27595065]
25. van Os HJA, Ramos LA, Hilbert A, et al. Predicting Outcome of Endovascular Treatment for Acute Ischemic Stroke: Potential Value of Machine Learning Algorithms. *Front Neurol.* 2018;9:784. doi:10.3389/fneur.2018.00784 [PubMed: 30319525]
26. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B (Statistical Methodol.)* 2005;67(2):301–320. doi:10.1111/j.1467-9868.2005.00503.x
27. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties. *Am Soc Neurorehabilitation.* 2002;16(3):232–240.
28. Duncan PW, Propst M, Nelson SG. Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. *Phys Ther.* 1983. doi:10.1093/ptj/63.10.1606
29. Harvey RL, Edwards D, Dunning K, et al. Randomized Sham-Controlled Trial of Navigated Repetitive Transcranial Magnetic Stimulation for Motor Recovery in Stroke. *Stroke.* 2018;49(9):2138–2146. doi:10.1161/STROKEAHA.117.020607 [PubMed: 30354990]
30. Hannula H, Ilmoniemi RJ. Basic Principles of Navigated TMS In: Navigated Transcranial Magnetic Stimulation in Neurosurgery. Cham: Springer International Publishing; 2017:3–29. doi:10.1007/978-3-319-54918-7_1
31. Rossini PM, Burke D, Chen R, et al. Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application: An updated report from an I.F.C.N. Committee. *Clin Neurophysiol.* 2015;126(6):1071–1107. doi:10.1016/j.clinph.2015.02.001 [PubMed: 25797650]
32. Manganotti P, Acler M, Masiero S, Del Felice A. TMS-evoked N100 responses as a prognostic factor in acute stroke. *Funct Neurol.* 2015;30(2):125–130. doi:10.11138/FNeur/2015.30.2.125 [PubMed: 26415785]
33. Jain AK, Jianchang Mao, Mohiuddin KM. Artificial neural networks: a tutorial. *Computer (Long Beach Calif).* 1996;29(3):31–44. doi:10.1109/2.485891
34. Hsu Chih-Wei, Chang Chih-Chung and C-J L. A Practical Guide to Support Vector Classification. *BJU Int.* 2008;101(1):1396–1400. doi:10.1177/02632760022050997 [PubMed: 18190633]
35. Breiman L, Friedman J, J.Stone C, Olshen RA. Classification Algorithms and Regression Trees.; 1984 <https://rafalab.github.io/pages/649/section-11.pdf>.
36. Breiman L. Random Forests. *Mach Learn.* 2001;45:5–32.
37. Page SJ, Fulk GD, Boyne P. Clinically Important Differences for the Upper-Extremity Fugl-Meyer Scale in People With Minimal to Moderate Impairment Due to Chronic Stroke. *Am Phys Ther Assoc.* 2012;92(6):791–798. doi:10.2522/ptj.20110009
38. Chawla NV., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res.* 2002;16:321–357. doi:10.1613/jair.953

39. Santos MS, Soares JP, Abreu PH, Araujo H, Santos J. Cross-Validation for Imbalanced Datasets: Avoiding Overoptimistic and Overfitting Approaches [Research Frontier]. *IEEE Comput Intell Mag.* 2018;13(4):59–76. doi:10.1109/MCI.2018.2866730
40. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw.* 2008;28(5):1–26. doi:10.1053/j.sodo.2009.03.002 [PubMed: 27774042]
41. Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol Model* 160. 2003;160:249–264.
42. Cortez P, Embrechts MJ. Using sensitivity analysis and visualization techniques to open black box data mining models. *Inf Sci (Ny).* 2013;225:1–17. doi:10.1016/J.INS.2012.10.039
43. Burnham KP, Anderson DR. Multimodel inference: Understanding AIC and BIC in model selection. *Sociol Methods Res.* 2004;33(2):261–304. doi:10.1177/0049124104268644
44. Hallett M. Transcranial magnetic stimulation and the human brain. *Nature.* 2000;406(6792):147–150. doi:10.1038/35018000 [PubMed: 10910346]
45. Talelli P, Greenwood RJ, Rothwell JC. Arm function after stroke: Neurophysiological correlates and recovery mechanisms assessed by transcranial magnetic stimulation. *Clin Neurophysiol.* 2006;117(8):1641–1659. doi:10.1016/j.clinph.2006.01.016 [PubMed: 16595189]
46. Thickbroom GW, Byrnes ML, Archer SA, Mastaglia FL. Motor outcome after subcortical stroke correlates with the degree of cortical reorganization. *Clin Neurophysiol.* 2004;115(9):2144–2150. doi:10.1016/j.clinph.2004.04.001 [PubMed: 15294217]
47. Jo JY, Lee A, Kim MS, et al. Prediction of Motor Recovery Using Quantitative Parameters of Motor Evoked Potential in Patients With Stroke. *Ann Rehabil Med.* 2016;40(5):806–815. doi:10.5535/arm.2016.40.5.806 [PubMed: 27847710]
48. Hope TMH, Seghier ML, Leff AP, Price CJ. Predicting outcome and recovery after stroke with lesions extracted from MRI images. *NeuroImage Clin.* 2013;2(null):424–433. doi:10.1016/j.nicl.2013.03.005 [PubMed: 24179796]
49. Varoquaux G, Baronnet F, Kleinschmidt A, Fillard P, Thirion B. Detection of Brain Functional-Connectivity Difference in Post-stroke Patients Using Group-Level Covariance Modeling. In: Springer, Berlin, Heidelberg; 2010:200–208. doi:10.1007/978-3-642-15705-9_25
50. Prabhakaran S, Zarahn E, Riley C, et al. Inter-individual Variability in the Capacity for Motor Recovery After Ischemic Stroke. *Neurorehabil Neural Repair.* 2008;22(1):64–71. doi:10.1177/1545968307305302 [PubMed: 17687024]
51. Stinear CM. Prediction of motor recovery after stroke: advances in biomarkers. *Lancet Neurol.* 2017;16(10):826–836. doi:10.1016/S1474-4422(17)30283-1 [PubMed: 28920888]
52. Rehme AK, Volz LJ, Feis DL, et al. Identifying neuroimaging markers of motor disability in acute stroke by machine learning techniques. *Cereb Cortex.* 2015;25(9):3046–3056. doi:10.1093/cercor/bhu100 [PubMed: 24836690]
53. Stinear CM, Byblow WD, Ackerley SJ, Barber PA, Smith MC. Predicting Recovery Potential for Individual Stroke Patients Increases Rehabilitation Efficiency. *Stroke.* 2017;48(4):1011–1019. doi:10.1161/STROKEAHA.116.015790 [PubMed: 28280137]
54. Stinear CM, Barber PA, Petoe M, Anwar S, Byblow WD. The PREP algorithm predicts potential for upper limb recovery after stroke. *Brain.* 2012;135(8):2527–2535. doi:10.1093/brain/aww146 [PubMed: 22689909]
55. Smith M- C, Ackerley SJ, Barber PA, Byblow WD, Stinear CM. PREP2 Algorithm Predictions Are Correct at 2 Years Poststroke for Most Patients. *Neurorehabil Neural Repair.* 7 2019:154596831986048. doi:10.1177/1545968319860481
56. Gauthier LV., Taub E, Mark VW, Barghi A, Uswatte G. Atrophy of Spared Gray Matter Tissue Predicts Poorer Motor Recovery and Rehabilitation Response in Chronic Stroke. *Stroke.* 2012;43(2):453–457. doi:10.1161/STROKEAHA.111.633255 [PubMed: 22096036]
57. Ramos-Murguialday A, Broetz D, Rea M, et al. Brain-machine interface in chronic stroke rehabilitation: A controlled study. *Ann Neurol.* 2013;74(1):100–108. doi:10.1002/ana.23879 [PubMed: 23494615]
58. Abdelnour F, Voss HU, Raj A. Network diffusion accurately models the relationship between structural and functional brain connectivity networks. *Neuroimage.* 2014;90:335–347. doi:10.1016/j.neuroimage.2013.12.039 [PubMed: 24384152]

59. Edwards DJ, Cortes M, Rykman-Peltz A, et al. Clinical improvement with intensive robot-assisted arm training in chronic stroke is unchanged by supplementary tDCS. *Restor Neurol Neurosci*. 2019;37(2):167–180. doi:10.3233/RNN-180869 [PubMed: 30932903]
60. Puig J, Blasco G, Alberich-Bayarri A, et al. Resting-State Functional Connectivity Magnetic Resonance Imaging and Outcome After Acute Stroke. *Stroke*. 2018;49(10):2353–2360. doi:10.1161/STROKEAHA.118.021319 [PubMed: 30355087]
61. Lindenberg R, Zhu LL, Rüber T, Schlaug G. Predicting functional motor potential in chronic stroke patients using diffusion tensor imaging. *Hum Brain Mapp*. 2012;33(5):1040–1051. doi:10.1002/hbm.21266 [PubMed: 21538700]
62. Rüber T, Schlaug G, Lindenberg R. Compensatory role of the cortico-rubro-spinal tract in motor recovery after stroke. *Neurology*. 2012;79(6):515–522. doi:10.1212/WNL.0b013e31826356e8 [PubMed: 22843266]
63. Kuceyeski A, Navi BB, Kamel H, et al. Structural connectome disruption at baseline predicts 6-months post-stroke outcome. *Hum Brain Mapp*. 2016;37(7):2587–2601. doi:10.1002/hbm.23198 [PubMed: 27016287]

Figure 1. Root of mean squared error (RMSE) results of regression analysis in predicting post-intervention UE-FMA.

Violin plots show the median, 1st and 3rd quartile, minimum and maximum value of the RMSE distribution calculated by five machine learning methods. (Elastic-Net (EN) = green, Random Forest (RF) = yellow, Artificial Neural Network (ANN) = purple, Support Vector Machines (SVM) = blue and Classification And Regression Trees (CART) = pink). The panels represent the RMSE over three sets of input variables (A) clinical, (B) clinical + regional disconnectivity and (C) clinical + pair-wise disconnectivity. A significant difference between the performances of two methods was shown with a star above the violin plots.

Figure 2. The observed versus predicted post-intervention UE-FMA.

Classification and Regression Tree (A) and Elastic-Net (B) methods were trained on clinical input variables, and the observed and predicted post-intervention UE-FMA are presented. Points represent the average prediction while the bars represent the standard deviation over 100 iterations in the outer loop.

Figure 3. The importance of variables extracted from the methods using the clinical dataset. Importance of the clinical variables only (demographics, clinical and neurophysiological measures) for all five machine learning methods. (Elastic-Net (EN) = green, Random Forest (RF) = yellow, Artificial Neural Network (ANN) = purple, Support Vector Machines (SVM) = blue and Classification And Regression Trees (CART) = pink). For visualization purposes, the weights of the variables' importance are rescaled to be relative to pre-intervention UE-FMA.

Figure 4. AUC results of classification analysis (minimal clinically important difference < 5.5 versus 5.5).

Violin plots show the median, 1st and 3rd quartile, minimum and maximum value of the AUC distribution calculated by five machine learning methods (Elastic-Net (EN) = green, Random Forest (RF) = yellow, Artificial Neural Network (ANN) = purple, Support Vector Machines (SVM) = blue and Classification And Regression Trees (CART) = pink, Logistic Regression (LR) = gray) using (A) clinical, (B) clinical + regional disconnectivity and (C) clinical + pair-wise disconnectivity datasets. A significant difference between the performances of two methods was shown with a star above the violin plots.

Table 1.

Patient demographics, clinical and neurophysiological characteristics (N = 102).

Variable	Value
Age	60.5 [54.0, 66.0]
Female	32
Right Handed	90
Number of patients with left hemisphere stroke	53
Time since stroke (in days)	293.0 [197.2, 347.2]
Lesion volume (in cm ³)	2.9 [1.2, 14.7]
Pre-intervention UE-FMA	33.0 [24.0, 45.0]
Post-intervention UE-FMA	41.5 [27.8, 52.0]
Difference in MT	-12.0 [-34.5, -0.3]
Number of patients with no MEP in the affected hemisphere	21
Number of patients who showed clinically significant increase in UE-FMA	43

Values are presented as mean (standard deviation) for the continuous variables except for lesion volume which is presented as median and inter-quartile range, and as the number of patients with that characteristic for the binary variables.

Table 2.**The R^2 and RMSE results of the methods in predicting post-intervention UE-FMA.**

The R^2 and RMSE results of the clinical, clinical + regional disconnectivity and clinical + pair-wise disconnectivity datasets in predicting post-intervention UE-FMA.

	Method	Clinical dataset	Clinical + regional disconnectivity dataset	Clinical + pair-wise disconnectivity dataset
R^2	EN	0.910 [0.866, 0.940]	0.901 [0.882, 0.929]	0.902 [0.890, 0.925]
	RF	0.882 [0.841, 0.917]	0.449 [0.295, 0.510]	0.064 [-0.148, 0.199]
	ANN	0.836 [0.703, 0.889]	0.372 [-0.104, 0.650]	0.546 [0.230, 0.720]
	SVM	0.797 [0.729, 0.855]	0.624 [0.448, 0.734]	0.846 [0.744, 0.894]
	CART	0.703 [0.625, 0.743]	0.790 [0.652, 0.795]	0.823 [0.578, 0.841]
RMSE	EN	0.362 [0.222, 1.015]	1.357 [0.772, 1.493]	1.359 [1.011, 1.632]
	RF	1.014 [0.641, 1.357]	4.183 [2.193, 5.806]	2.633 [0.891, 4.444]
	ANN	1.065 [0.476, 1.803]	2.982 [1.291, 5.146]	2.101 [1.018, 3.658]
	SVM	1.165 [0.512, 1.799]	1.633 [0.540, 3.869]	1.542 [1.175, 2.009]
	CART	1.373 [1.242, 2.346]	1.807 [0.829, 3.286]	2.003 [0.684, 3.008]

Values are presented as Median [1st quartile, 3rd quartile].

Table 3.

The classification results (AUC) for the five machine learning methods and Logistic Regression using three different input datasets.

Method	Clinical dataset	Clinical + regional disconnectivity dataset	Clinical + pair-wise disconnectivity dataset
EN	0.56 [0.42, 0.67]	0.50 [0.50, 0.61]	0.57 [0.43, 0.71]
RF	0.55 [0.43, 0.67]	0.63 [0.50, 0.75]	0.63 [0.50, 0.75]
ANN	0.60 [0.46, 0.71]	0.58 [0.43, 0.71]	0.63 [0.50, 0.73]
SVM	0.58 [0.47, 0.71]	0.50 [0.38, 0.63]	0.47 [0.38, 0.60]
CART	0.53 [0.42, 0.65]	0.52 [0.40, 0.67]	0.55 [0.42, 0.65]
LR	0.55 [0.42, 0.67]	0.50 [0.42, 0.63]	0.50 [0.38, 0.60]

Values are presented as Median [1st quartile, 3rd quartile].