Decoding post-stroke motor function from structural brain imaging

Jane M Rondina [a], Maurizio Filippone [b], Mark Girolami [c], Nick S Ward [a]

[a] *Sobell Department of Motor Neuroscience, Institute of Neurology, University College London, UK*

[b] *Department of Data Science, EURECOM, France*

[c] *Department of Statistics, University of Warwick, UK*

[*] *Corresponding author at:*
Sobell Department of Motor Neuroscience, University College London, Institute of Neurology
33 Queen Square, London WC1N 3BG, UK. Tel.: *+44 (0) 20 3448 8773*
E-mail address: j.rondina@ucl.ac.uk (J.M.Rondina)

Abstract

Clinical research based on neuroimaging data has benefited from machine learning methods, which have the ability to provide individualized predictions and to account for the interaction among units of information in the brain. Application of machine learning in structural imaging to investigate diseases that involve brain injury presents an additional challenge, especially in conditions like stroke, due to the high variability across patients regarding characteristics of the lesions. Extracting data from anatomical images in a way that translates brain damage information into features to be used as input to learning algorithms is still an open question. One of the most common approaches to capture regional information from brain injury is to obtain the lesion load per region (i.e. the proportion of voxels in anatomical structures that are considered to be damaged). However, no systematic evaluation has yet been performed to compare this approach with using patterns of voxels (i.e. considering each voxel as a single feature). In this paper we compared both approaches applying Gaussian Process Regression to decode motor scores in 50 chronic stroke patients based solely on data derived from structural MRI. For both approaches we compared different ways to delimit anatomical areas: regions of interest from an anatomical atlas, the corticospinal tract, a mask obtained from fMRI analysis with a motor task in healthy controls and regions selected unsing lesion-symptom mapping. Our analysis showed that extracting features through patterns of voxels that represent lesion probability produced better results than quantifying the lesion load per region. In particular, from the different ways to delimit anatomical areas compared, the best performance was obtained with a combination of a range of cortical and subcortical motor areas as well as the corticospinal tract. These results will inform the appropriate methodology for predicting long term motor outcomes from early post-stroke structural brain imaging.

Keywords: Stroke, Motor impairment, Lesion patterns, Machine learning, Gaussian processes, Multiple kernel learning, Features extraction, Patterns of lesion probability, Lesion load.

1. Introduction

The ability to predict long term outcome after stroke is urgently required in order to facilitate a stratified approach to clinical decision making (Ward, 2015). It has long been known that information encoded in brain lesions (e.g. extent and location) can explain variability in post-stroke outcomes (Bayona et al., 2005; Geva et al., 2011; Särkämö et al., 2009; Schiemanck et al., 2005; Yang et al., 2008; Zhu et al., 2010), but no approaches have been routinely incorporated into clinical practice.

Machine learning (ML) techniques are potentially useful for clinical applications, aiming to provide sensitive and specific diagnostic and prognostic indicators for individuals, as opposed to analysing statistical group differences (Wang et al., 2010). In neuroimaging, clinical applications of ML methods have initially focused mainly on binary classification of disease states (Davatzikos et al., 2005; Teipel et al., 2007; Fu et al., 2008; Klöppel et al., 2008; Vemuri et al., 2010). More recently, decoding of outcomes represented by continuous scales has also become increasingly common in several neurological and psychiatric conditions through predictive multivariate regression methods (Cohen et al., 2011). In our study, the outcome is a motor score that the patient attained at a chronic stage.

The extraction of features from brain images in a way that is meaningfully related to the clinical condition being studied is a fundamental step in a predictive analysis framework. In the context of stroke, feature extraction from structural neuroimaging is additionally challenging due to the high variability in anatomical location and extension of brain injury. Although lesion characteristics can potentially contribute towards making accurate predictions of the likely level of impairment and recovery, there is currently no consensus on how to quantify these characteristics.

Progress has been made in predicting language outcomes using features derived from stroke lesions (Payabvash et al. 2010; Hope et al. 2013, 2015) but predicting motor outcomes is lagging behind. One of the most common approaches to quantify characteristics from lesions is summarizing the proportion of voxels in each region of interest (ROI) that are considered to be part of a lesion. This information is commonly referred to as *lesion load* and it is obtained using anatomical masks to

define the ROIs and a method to segment the lesions, either manually (Kim et al., 2014) or automatically (Hope et al. 2013, 2015). Recently, voxel-based lesion symptom mapping (VLSM, (Bates et al., 2003)) has also been proposed as a way to extract features from stroke lesions to be used as input to machine learning models. Voxel-based lesion symptom values are obtained for each voxel through a statistical test on the continuous scores representing the symptom between two groups (which are defined according to having or not lesion in that particular voxel). The voxelwise maps resulting from this method are used as a way to define a mask to restrict voxels (Munsch et al., 2015) or to build symptom or condition specific ROIs (Forkert et al., 2015).

In this paper we have directly compared a range of approaches for assessing the relationship between structural brain damage and long term outcome in the same cohort of chronic stroke patients. Using structural MRI images from 50 patients we derived lesion probability images (i.e., images where each voxel is assigned a value between 0 and 1 representing the likelihood of being part of injured tissue). We wanted to investigate which features have the highest power to decode the individual level of motor impairment. There are two key questions: firstly, what type of data should be extracted from the images? Secondly, which are the key brain regions from which data should be extracted? To investigate the first question, we used two strategies to extract information from images: *i*) *patterns of voxels*, where each feature corresponds to a single voxel representing lesion probability, and *ii)* anatomical summarization, where each feature corresponds to the lesion load in an ROI. To investigate the second question, we employed a number of different approaches to define anatomical regions: *i*) regions of interest (ROIs) from an anatomical atlas; *ii*) a mask delimiting the corticospinal tract (CST); *iii*) combination of all ROIs and the CST; *iv*) a subset of the ROIs expected to be related to motor function; *v*) combination of the subset of ROIs and the CST; *vi*) active voxels from fMRI acquired with a motor task in healthy controls; *vii*) a mask restricting voxels to lesions; *viii*) voxels selected through lesion-symptom mapping. Additionally, we did a secondary analysis, applying multiple kernel learning techniques using kernels extracted from brain regions to investigate the possibility of learning the relevance of each anatomical pattern.

2. Material and Methods

2.1. Study population

Fifty patients that had their first stroke at least three months before the collection of the data (mean 29.06, std 31.05 months) participated in the study. The patients had mean age 54.22 years (std 12.62), Seventeen patients were female and 18 patients had right hand affected. Complete demographic and clinical characteristics of each patient can be found in the supplementary material (table S1). The extent and location of the lesions for each patient (figure S1) is also presented. A control group was composed by 23 age-matched healthy subjects who reported no history of neurological or psychiatric illness, vascular disease or hypertension. All subjects provided full written consent in accordance with the Declaration of Helsinki. The study was approved by the Joint Ethics Committee of the UCL Institute of Neurology, the National Hospital for Neurology and Neurosurgery and UCL Hospitals NHS Foundation Trust.

2.1.1 Motor scores

Measures of motor impairment in the contralesional upper limb were obtained through four different assessment scales: Action Research Arm Test (ARAT) (Lyle, 1981), grip strength (GS) (Sunderland et al., 1989), Motricity Index (MI) (Bohannon 1999) and Nine-Hole Peg Test (NHPT) (Mathiowetz et al., 1985).

As the different motor scores are correlated but also complementary, a single representative measure was calculated using principal component analysis (PCA). Considering Y as a matrix of 50 examples and 4 labels (corresponding to the number of patients and motor scores, respectively), the PCA was obtained using the following steps:

1. Calculate the mean of each score across patients and subtract it from Y (*zero mean Y*);
2. Obtain the covariance matrix from *zero mean Y (cov zero mean Y);*
3. Find the eigenvalues of *cov zero mean Y.*

A vector $\mathbf{y} = [y_1,\ldots,y_m]$ where m is the number of subjects represents the first principal component (FPC) of the four scores, which accounts for the greatest possible variance across them. Additionally, it contributes to reduce the problem potentially caused by floor and ceiling effects in the individual scores.

2.2. Images acquisition and pre-processing

T1-weighted high resolution magnetic resonance images were acquired using a 3T Allegra system (Siemens AG, Erlangen, Germany) with the following protocol: number of slices = 176, slice thickness = 1 mm, matrix size = 224 × 256, in-plane resolution = 1 mm × 1 mm.

The origin of each image was set at the anterior commissure. Images from patients that had injury predominantly in the left hemisphere were flipped in relation to the mid-sagittal plane so that all scans presented lesion in the right hemisphere. Images from all subjects were segmented into grey matter, white matter, cerebrospinal fluid and then normalised using the New Segment routine in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/).

Lesion probability images were obtained from the high-resolution T1-weighted volumetric MRI scans using an automatic method for detection of outlier voxels (Seghier et al., 2008). This method is based on the assumption that lesions are characterized as atypical voxels regarding expected brain tissues (grey matter, white matter and cerebrospinal fluid). The characterization of tissues uses the unified segmentation-normalization approach (Ashburner and Friston, 2005) modified to include an extra tissue to account for the perturbation introduced by lesions. Grey and white matter segmented tissues from patients are compared with the corresponding tissues from healthy control subjects in a voxel by voxel way. As a result, each voxel is represented by a value between 0 and 1 that quantifies the likelihood of it being part of injured tissue.

2.3. Segmentation of lesions

Figure 1 presents the steps that were performed to obtain binary images of the lesions. Images representing lesion probability were derived from T1 anatomical images according with the procedure described in the previous section. In order to segment the lesions we applied a threshold selecting the voxels with probability of being part of injured tissue greater than 0.3, producing binary images. Finally we selected only contiguous clusters with 100 or more voxels. Please see (Seghier et al., 2008) for a detailed explanation regarding the rationale behind both parameters (threshold value and cluster size) and comparison with manually traced lesions. We also performed additional tests to check the adequacy of these parameters to segment lesions in our images (supplementary material, item 2).

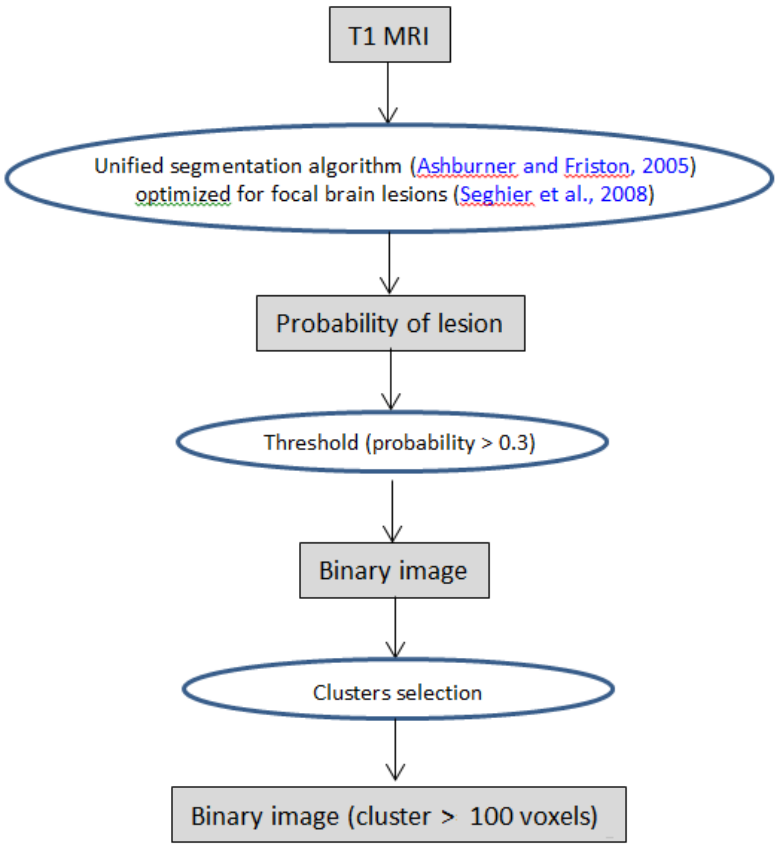
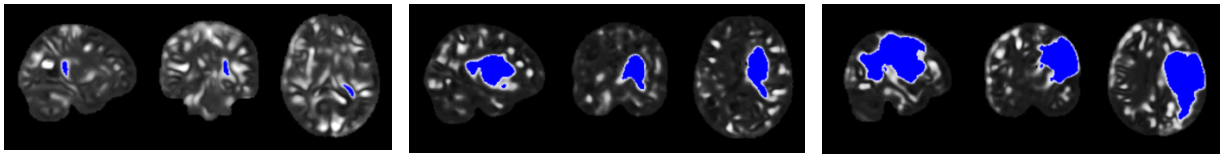


Figure 1. Steps for segmentation of lesions

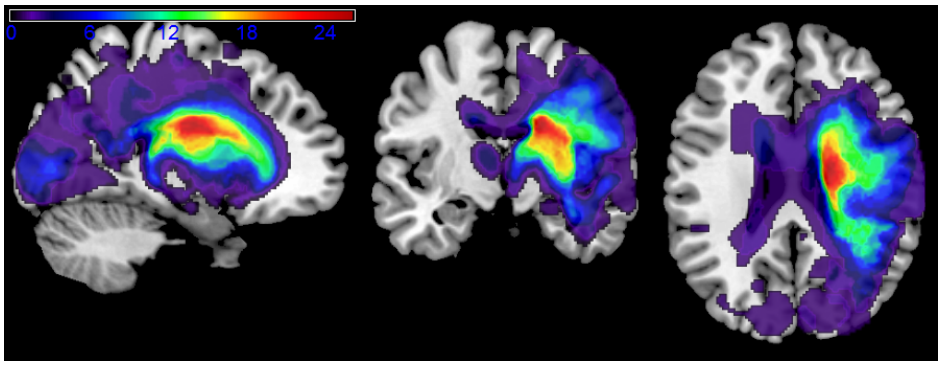
Figure 2 shows examples of lesions segmented according to the described approach. The binary images corresponding to lesions (visualized in blue) were overlaid on the lesion probability images (visualized in grayscale).



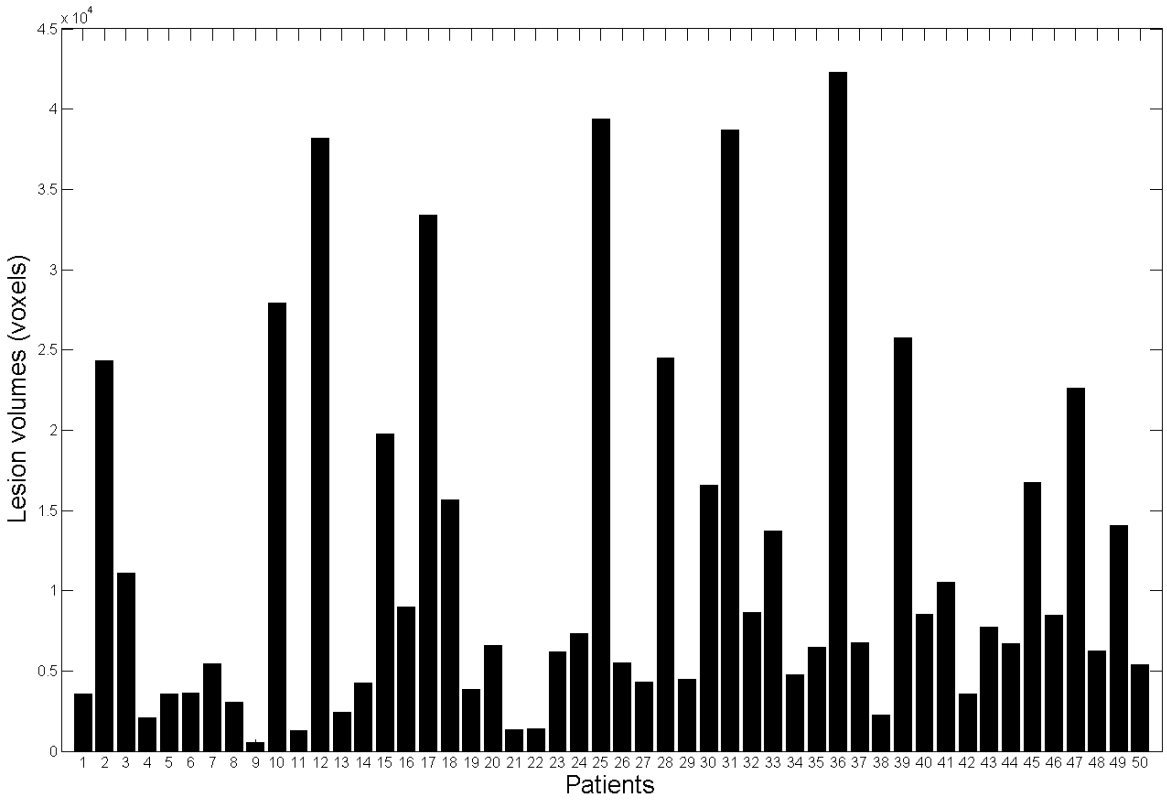
(a) (b) (c)

Figure 2. Examples of segmented lesions: (a) small; (b) medium; (c) large

Figure 3a presents a map illustrating the overlap of segmented lesions across all patients. The colour map represents the incidence of lesions in each voxel, ranging from purple (lesion in 1 subject only) to red (lesion in 26 out of the 50 subjects). Figure 3b presents a plot showing the volume of the segmented lesion for each patient according to the procedure described above. This plot illustrates the variability of the sample regarding the extent of the lesions across the patients. In the supplementary material, we provide a figure that shows the distribution of the lesions in each patient across brain regions (figure S1).



(a)



(b)

Figure 3 Lesion volumes per patient

2.4. Features extraction

In the following sections we describe the strategies that we used to extract features. In the section 2.4.1 we describe the approaches implemented to extract information from images of lesion probability (through patterns of voxels or lesion load). In the section 2.4.2 we describe the approaches used to limit ROIs in the brain.

2.4.1. Anatomical patterns and summarizations

Patterns of voxels

Through this approach, each voxel inside a binary mask corresponds to a particular feature that represents lesion probability. It is important to notice that independent of the strategy to delimit a mask, each voxel inside the delimited region corresponds to a particular feature in a pattern.

Anatomical summarization (lesion load)

In this approach we used labelled masks to delimit ROIs and applied a summarization technique to transform the information encoded in each ROI into a single value. The summarization was based on the *lesion load*, which is the proportion of voxels in each region that were considered to be part of a lesion according to the segmentation algorithm described in the section 2.3. Each summarization unit corresponds to a single feature.

The Figure 4 illustrates the difference between the approaches to extract features. The area circulated in red illustrates a mask corresponding to an ROI. The grid in the panel (a) indicates that the mask encompasses a set of voxels corresponding to a probabilistic value. In the panel (b) the white region represents a lesion. The intersection between the lesion and the ROI divided by the size of the ROI corresponds to the lesion load, which is a single feature (as opposed to several features in the former approach).

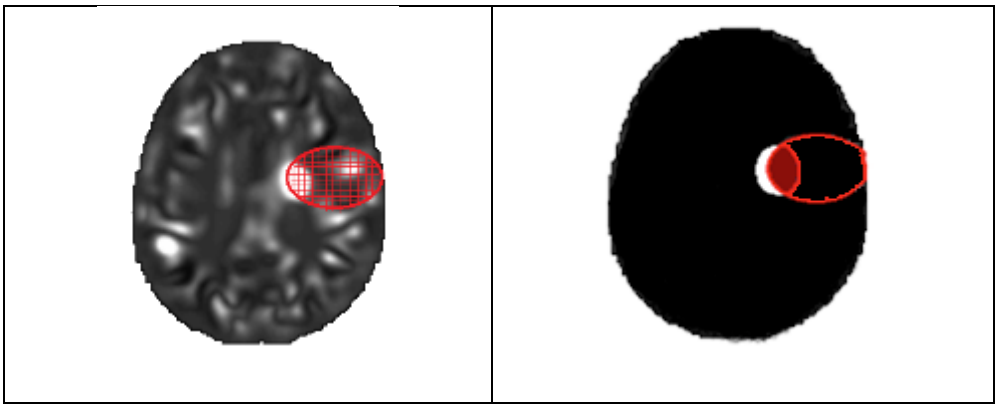


Figure 4. Extraction of features through patterns of voxels representing lesion probability (a) and anatomical summarization (lesion load) (b)

2.4.2. Approaches to delimit regions of interest

Atlas-based ROIs

Regions of interest were obtained automatically from the AAL atlas (Tzourio-Mazoyer et al., 2002), which comprises 116 disjoint regions corresponding to cortical and subcortical anatomical structures. Figure 5a illustrates the atlas overlapped on a structural brain template.

Motor ROIs

We hypothesized that some brain regions might be more relevant than others for motor function. Thus we selected a subset of ROIs from the AAL atlas that correspond to regions expected to be related to motor and sensorial function according to literature (Alexander, Crutcher, and DeLong 1990; Kim et al. 1993; Cao et al. 1998; Hauk, Johnsrude, and Pulvermüller 2004). The regions (illustrated in the Figure 5b) are the following: Postcentral gyrus, Precentral gyrus, Supplementary motor area, Superior frontal gyrus, Middle frontal gyrus, Inferior and Superior parietal regions, Thalamus, Caudate, Putamen and Pallidum.

Corticospinal tract (CST)

Recent studies have proposed that the CST plays a relevant role in recovery of motor function post-stroke (Byblow et al., 2015; Stinear, 2010; Zarahn et al., 2011), as the reorganization of motor function ultimately requires access from primary motor cortex areas to muscles.

A CST mask was obtained by probabilistic tractography from nine age-matched healthy volunteers in a previous study (see details in Ward, Swayne, and Newton 2008; Schulz et al. 2012). The mask is illustrated in Figure 5c.

Functional ROIs

With this mask we aimed to evaluate the effect of atypical tissue in both the lesional and contralesional hemispheres in areas involved in normal motor function. The voxels were selected through the application of General Linear Model (GLM) to fMRI data from healthy controls performing a motor task (hand grip using the dominant hand). Figure 5d illustrates the functional mask obtained from the contrast grasp versus rest (with family-wise error control $p<0.005$, cluster size = 100 voxels). The mask was mirrored so that both hemispheres were considered, as in all other approaches. It is important to observe that the mask was obtained from healthy controls, so concerns are not applicable to the effect of lesion side and dominant hand in activation.

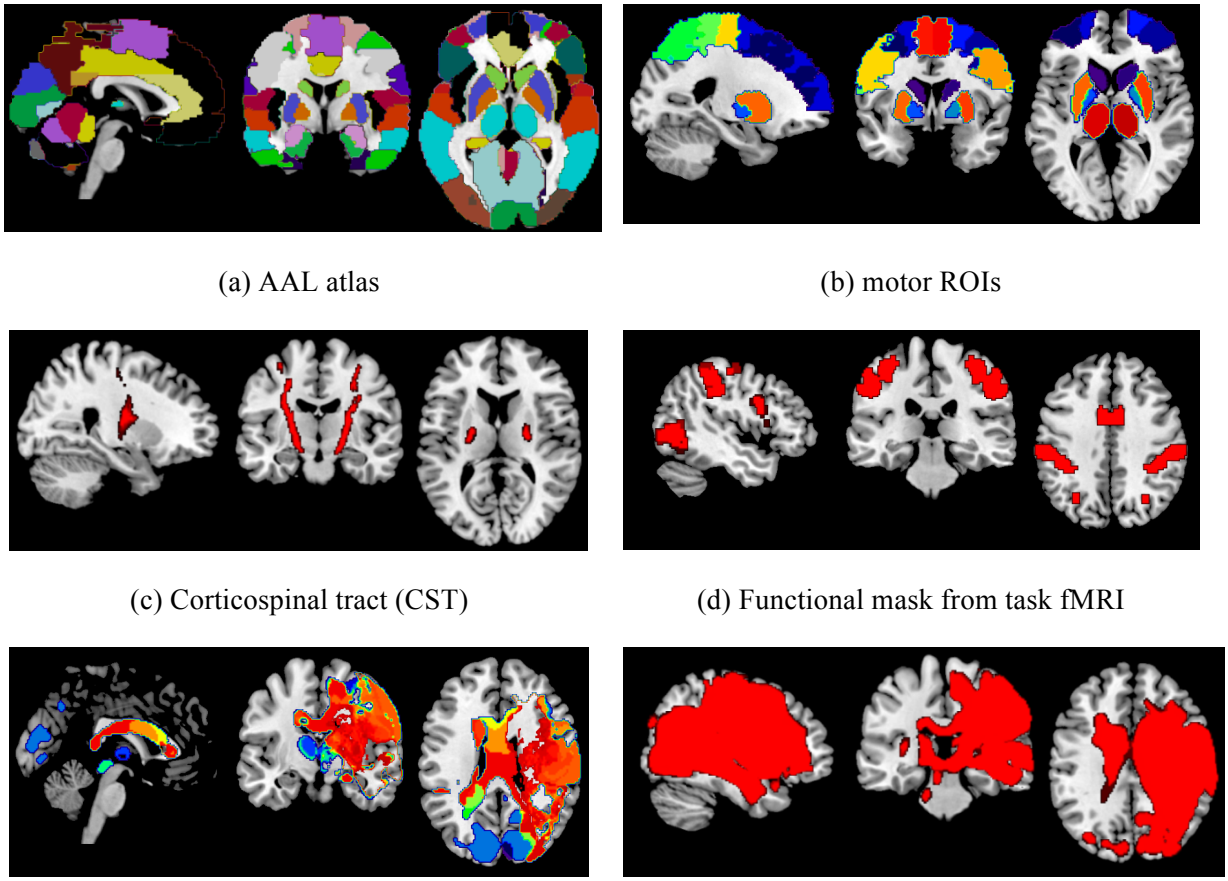
Lesion-symptom mapping ROIs

Another way of obtaining ROIs that represent the relation between structure and function is through voxel-based lesion-symptom map (VLSM) (Bates et al., 2003; Forkert et al., 2015). A mask can be obtained by applying *t*-test in each voxel to compare the average of the motor score between the subjects that have lesion and those who do not have lesion in that particular voxel.

We implemented a variation on the approach proposed by (Forkert et al., 2015). The authors built problem-specific ROIs assigning labels to voxels according to the difference in the median modified Rankin Scale (mRS) of each group. The mRS is a commonly used scale for measuring the degree of disability or dependence in the daily activities of people who have suffered neurological disability (Bonita and Beaglehole, 1988). It varies from 0 (meaning perfect health without symptoms) to 6 (meaning death). They produced four problem-specific ROIs: voxels with median mRS difference $d>2$ corresponding to the first ROI, voxels with $1<d< 2$ corresponding to the second ROI, voxels with $0< d <1$ corresponding to the third ROI and all the remaining voxels within the brain corresponding to the fourth ROI. As our scale (the first principal component of the motor scores) is continuous with values ranging from -2.0237 to 1.5815, we converted them to a discreet range, rounding each fractional value to the closest integer number (-2, -1, 0, 1 and 2). For each voxel we obtained the median of the rounded score of each subject that had lesion in that particular voxel. This procedure resulted in 5 ROIs illustrated in the Figure 5e.

Lesions-bounded ROI

For completeness, we also defined a mask considering only voxels that were part of a lesion in at least one patient (Figure 5f).



(a) AAL atlas (b) motor ROIs

(c) Corticospinal tract (CST) (d) Functional mask from task fMRI

(e) Lesion-symptom mapping (f) Segmented lesions in all patients

Figure 5. Delimitation of regions of interest

2.5. Prediction analysis

Data for each analysis can be represented by a matrix *X,* where each column corresponds to a particular feature and each row contains the set of features corresponding to a particular example (in our case, data derived from structural image of a single patient).

The data matrices were standardized in the following way: each feature vector (row) was normalized by the Euclidean distance. Then each feature was scaled to zero mean and unit variance. A linear function (dot product) was used to build a covariance matrix *C,* with dimensions corresponding to the number of examples (50x50). Covariance is a measure of how much two variables change together, and the covariance function (or kernel) describes the spatial covariance of a random variable.

$$C = X X^T \tag{1}$$

We applied Gaussian Process Regression (GPR) (Filippone et al., 2012; Marquand et al., 2010; Rasmussen, 2006) using the covariance matrices to predict the representative score for motor impairment (first principal component from ARAT, Grip, Motricity index and NHPT). The score was also standardized according to the same procedure applied to the features.

A Gaussian process (GP) is formally defined as a set of random variables such that any subset of them is jointly Gaussian. These Gaussian distributions need the definition of a mean and a covariance, and this is typically done by associating each of the GP random variables with elements of some input domain, say R^d, and by defining mean and covariance as functions of the input domain. For instance, a covariance function that makes random variables highly correlated when the corresponding inputs are close by, makes the GP random variables smoothly change over the input domain. Different covariance functions determine different behaviours of these random variables over the input domain, suggesting that GPs can be interpreted as a distribution over functions.

Gaussian process regression uses GPs as priors over functions, and attempts to learn a posterior distribution over these functions after data are observed using Bayesian inference techniques. In GP regression, the assumption is that data are Gaussian distributed around an unobserved function *f(x)* that is modelled using GPs, namely $y = f(x) + \varepsilon$, with $\varepsilon \sim N(0,\sigma n2)$. GP regression can be seen as an extension of standard linear regression, where the unobserved functions are assumed to be linear combinations of the features $f(x) = x^T w$. In GP regression, obtaining the posterior distribution over the unobserved function *f(x,)* and learning any parameter of the covariance function (e.g., the smoothness

from data), requires standard algebraic operations involving the covariance matrix, that is the matrix of all pairwise covariances of the random variables at the inputs where observations are available.

In order to check whether the results obtained with GPR would be similarly true for other predictive multivariate regression method, we repeated all analyses using Support Vector Regression (SVR) (Drucker et al., 1997; Smola et al., 2004) with the same linear kernel.

Learning from multiple sources

As a secondary analysis, we wanted to investigate the relevance of each ROI based on patterns of voxels representing lesion probability, following a similar procedure as in (Filippone et al., 2012).

Given the covariance matrices C_i associated to M sources of information, it is possible to build the covariance of the Gaussian process in the following way:

$$K = \sum_i^M C_i \, w_i \tag{2}$$

Considering that the covariance matrices were obtained by scaling all the sources in the same way it is possible to interpret the weights as the relevance of the associated sources.

The weights w_i were optimized by the standard multivariate Gaussian log-likelihood:

$$L = -\frac{1}{2}\log|K| - \frac{1}{2} y^T K^{-1} y + const \tag{3}$$

Multiple kernel learning (MKL) can be used in different ways (Gönen and Alpaydın, 2011). Different kernels (covariance matrices) can correspond to different notions of similarity (e.g. linear, polynomial, Gaussian, exponential) or to different representations of data (sources). In neuroimaging, different data sources may comprise different imaging modalities (e.g., T1-MRI or DWI-MRI), different ways of extracting data from a same modality (e.g., volumetric or cortical thickness in structural MRI; ADC in DWI), or different subsets of features. In this study we applied the latter approach using anatomical criteria to define the subsets of features from patterns of voxels. Thus the

number of sources *M* was 118, corresponding to 116 regions from the AAL atlas and the CST divided into left and right portions (to be consistent with the structures from the atlas, which are unilateral).

2.6. Results evaluation

To assess how the results of the analysis generalize to an independent data set we used cross-validation. One round of cross-validation involves partitioning the data sample into disjoint subsets of examples, performing the analysis on one subset (the training set), and validating the analysis on the other subset (the validation or testing set). To reduce the variability, multiple rounds of cross-validation are performed using different partitions, and the results are averaged over the rounds.

In this study, we implemented a 10-fold cross-validation, which involves separating 10% of the examples from the original sample for test while the remaining examples are used for training. This splitting is repeated so that each example in the sample is used once for validation. After repeating this process (leaving out all examples), the final accuracy is quantified as the average of accuracies obtained across all 10 folds. It is important to notice that all additional operations that involve data from different examples (as scaling labels, calculating the principal component of the motor scores and obtaining problem-specific ROIs) were performed within the cross-validation folds without using any information from the test data, thus ensuring the validity of the reported performance scores.

The measure of accuracy in each analysis was obtained through the correlation between real and predicted labels. Correlation values were obtained in each fold and the final measure (R) corresponds to the average across the folds. Additionally, we present the root mean squared error (RMSE) according to the following equation, where N is the number of examples per fold, f_i is the predicted label and y_i is the real label:

$$MSE = \sqrt{\frac{1}{N}\sum(f_i - y_i)^2}$$

To evaluate whether the difference in RMSE across folds resulting from different models is statistically significant we used Wilcoxon rank sum test.

3. Results

3.1. Learning from single sources: patterns of voxels representing lesion probability

In this section, we present results of GPR analyses using a single source (only one covariance matrix) for which each feature corresponds to a single voxel representing the probability of being atypical tissue. Results of all strategies to delimit ROIs as anatomical patterns are presented in Table 1. For each ROI we show the corresponding number of features (NF) used as input to the model (that corresponds to the number of voxels in the mask), the correlation between real and predicted labels (R) and the root mean squared error (RMSE). The measures of correlation and the RMSE resulted from the average of the folds in the cross-validation framework. We also present the result of the same analysis performed without any mask to delimit voxels (model M1, whole brain approach).

Limiting the analysis to voxels belonging to all ROIs from AAL atlas produced the same result as using all voxels in the whole brain (models M1 and M1.1: R = 0.72; RMSE = 0.73). A similar result (R = 0.73; RMSE = 0.72) was obtained when adding the CST to the AAL ROIs (model M1.3). Limiting the analysis to voxels in the CST (model M1.2) led to R = 0.65 (RMSE = 0.75). The accuracy increases to R = 0.80 (RMSE = 0.70) when limiting the analysis to the areas selected from the AAL atlas that are believed to be the more specifically related to motor function according to literature (model M1.4). The best prediction was obtained joining motor ROIs and CST (model M1.5: R=0.83, RMSE = 0.68). The remaining masks - areas resulting from motor activation in fMRI (model M1.6), lesion-symptom mapping (model M1.7) and lesion-bounded (model M1.8) led to accuracies lower than the whole brain (respectively R = 0.67; RMSE = 0.78, R = 0.66; RMSE = 0.76, R = 0.65; RMSE = 0.76).

Table 1 Predicting motor impairment based on different masks to limit subsets of voxels. NF = number of features; R = correlation between real and predicted motor scores; RMSE = Root Mean Squared Error.

Model	Features	NF	R	RMSE
M1	Whole brain	630786	0.72	0.73
M1.1	Voxels limited by AAL atlas	451318	0.72	0.73
M1.2	Voxels limited by the Corticospinal tract (CST)	4421	0.65	0.75
M1.3	Voxels limited by AAL atlas + CST	457384	0.73	0.72
M1.4	Voxels limited by motor ROIs	120793	0.80	0.70
M1.5	Voxels limited by motor ROIs and CST	125214	0.83	0.68
M1.6	Voxels limited by mask from task fMRI in healthy controls	35545	0.67	0.78
M1.7	Voxels limited by lesion-symptom mapping	** 9991.1*	0.66	0.76
M1.8	Voxels limited by lesion in at least 1 patient	158907	0.68	0.75

** Average across cross-validation folds*

3.2. Learning from single sources: lesion load per regions

In this section we present results of GPR analyses for which each feature corresponds to the summarization of all voxels belonging to a particular region. The strategies to delimit regions are the same used in the approach described in the previous section. However, in this approach, labelled ROIs are used so that each feature is the lesion load in a particular region (i.e., the proportion of voxels in an ROI that are considered to be part of a lesion). Thus, the number of features corresponds to the number of regions. Table 2 presents the results of each strategy to define the ROIs.

Prediction based on the lesion load in the whole brain (i.e. the proportion of damage in the brain, model M2) resulted in R = 0.30 (RMSE = 0.94). When considering the lesion load in each of the 116 ROIs from the AAL atlas (model M2.1), the accuracy decreases to R = 0.20 (RMSE = 8.04). When adding the CST to the AAL ROIs (model M2.3), the accuracy increases to R = 0.25 (RMSE = 6.47).

The best accuracy in this approach was obtained limiting the analysis to the CST alone (model M2.2: R = 0.51; RMSE = 0.84).

As opposed to the previous approach, prediction based on the lesion load in the motor ROIs did not improve the accuracy in relation to the whole brain (model 2.4: R = 0.21; RMSE = 1.10). However, it increases slightly when including the lesion load in the CST (model M2.5, R = 0.26; RMSE = 1.09). Predictions based on lesion load in the ROIs defined by lesion-symptom mapping and in the ROI based on lesions in at least one patient (models M2.7 and M2.8) produced results similar to the lesion load in the whole brain.

Table 2 Predicting motor impairment based on different labelled ROIs to extract features by summarization of regions (lesion load). NF = number of features; R = correlation between real and predicted motor scores; RMSE = Mean Squared Error

Model	Features	NF	R	RMSE
M2	Lesion load in the whole brain	1	0.30	0.94
M2.1	Lesion load in ROIs from AAL atlas	116	0.20	8.04
M2.2	Lesion load in Corticospinal tract	1	0.51	0.84
M2.3	Lesion load in ROIs from AAL atlas + CST	117	0.25	6.47
M2.4	Lesion load in motor ROIs	22	0.21	1.10
M2.5	Lesion load in motor ROIs + CST	23	0.26	1.09
M2.6	Lesion load in Functional mask from task fMRI	1	0.17	0.95
M2.7	Lesion load in ROIs defined by Lesion-symptom mapping (median PCA)	5	0.31	0.92
M2.8	Lesion load in ROI from lesion in at least 1 patient	1	0.30	0.94

The extraction of features through patterns of voxels produced better results than extracting lesion load for all strategies used to delimit regions. According to the Wilcoxon rank sum test based on the root mean squared error (RMSE) across folds, the difference between categories was statistically significant when comparing the models involving all ROIs and CST (M1.1 with M2.1 and M1.3 with M2.3, $p < 0.001$) and when comparing the models restricting ROIs to motor regions (M1.4 with M2.4,

p = 0.023) and motor regions + CST (M1.5 with M2.5, p = 0.018). Comparisons of analyses using different ROIs within the same approach were not statistically significant.

Figure 6 shows a scatter plot of the real and predicted scores resulting from the model that produced the best accuracy (M1.5 - pattern of voxels limited by motor ROIs and CST).

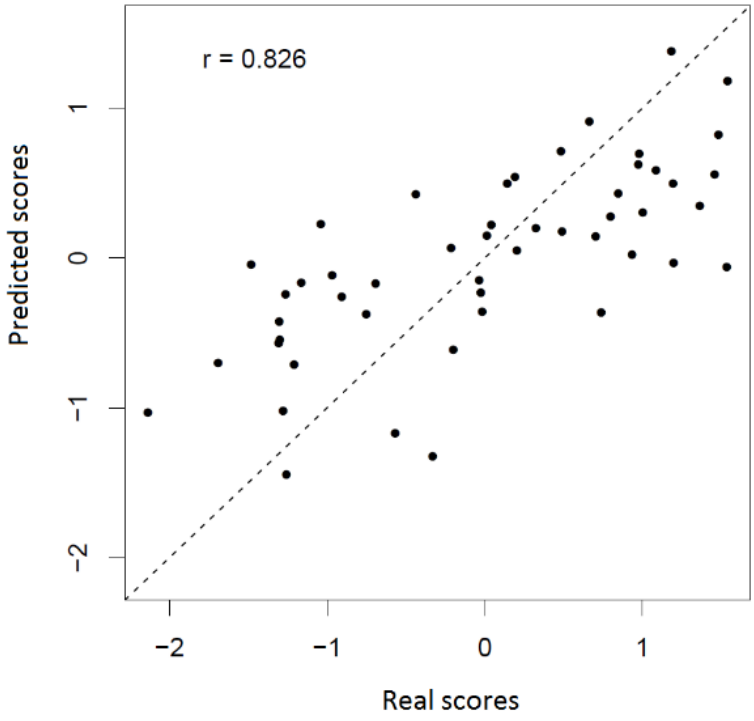


Figure 6 Prediction of first principal component of motor impairment scores based on voxels limited by motor ROIs and CST

We performed additional analyses using SVM with all the approaches to extract features. The results were well in line with the prediction derived from the GPR analyses, as the differences across strategies to define ROIs within each approach to extract features and the differences between the approaches remain similar to the differences described above. Please see the tables S4 and S5 in the section 3 of the supplementary material for detailed results.

3.3. Learning from multiple sources

The correlation between actual and predicted scores using MKL was 0.66 (RMSE = 0.79). The figure S3 in the supplementary material displays a plot of each ROI ranked in descending order of weight. The ROI corresponding to the highest weight (0.0267) was the right caudate nucleus, followed by the left inferior frontal gyrus, pars opercularis (weight 0.0264) and the right corticospinal tract (weight 0.0262). A table with the complete list of all 118 regions sorted by descending order of weights is also presented in the supplementary material (item 4, table S6). Note that this analysis is based on the maximisation of the log-marginal likelihood of the GPR approach, thus it does not give a full account of the uncertainty in this assessment. In a follow-up study we intend to carry out a deeper analysis of the uncertainty associated with these weights; the aim of this study was to gain some insights into the different ways to extract features and define regions of interest.

4. Discussion

We have applied machine learning techniques to ask how accurately upper limb motor impairment can be decoded from data derived from structural brain images in chronic stroke patients. We have compared a number of different approaches to extract data in the same subjects. There are two key findings: (i) Firstly, approaches that extract features through patterns of voxels produced better results than those extracting lesion load per region; (ii) secondly, using only data from the corticospinal tract was not sufficient to produce the most accurate results..

For all approaches applied to define ROIs, the extraction of features corresponding to patterns of voxels produced higher accuracy in comparison to summarizing the information encoded in the voxels through lesion load per region. The differences between models using patterns of voxels and lesion load were statistically different when using all ROIs from AAL atlas, motor ROIs, combining AAL ROIs and CST and combining motor ROIs and CST. The best result was obtained using patterns of voxels from motor regions and CST together, suggesting that having some knowledge about the

likelihood of damage to cortical regions (especially those thought to be motor related) may be important in accounting for variability in motor impairment.

Most studies investigating the anatomical correlates of motor outcomes after stroke have concentrated on quantifying damage in the corticospinal tract (Burke Quinlan et al., 2015; Byblow et al., 2015; Lindenberg et al., 2010; Schulz et al., 2012; Zhu et al., 2010). These studies have shown that damage in CST correlates well with motor outcomes. Our results are in alignment with these previous findings. Whilst the information contained in the CST lesion load is less powerful than when represented as a pattern of voxels representing lesion probability, it is noteworthy that the lesion load in the CST presented higher predictive power than any other ROI compared. On the other hand, many of the findings in previous studies have concentrated on patients with subcortical infarcts. The patients in our study are diverse regarding both extension (Figure 3) and location of lesions (figure S1, supplementary material). Recent results suggest that predicting the motor behavioural consequences of stroke damage in more heterogeneous groups of patients is likely to require information about cortical as well as corticospinal tract damage (Park et al, 2016), and this is once again borne out by our findings (as results from models M1.4 and M1.5 suggest).

Another key point is that although damage to these motor regions (especially CST) can account for initial impairment, the anatomical correlates of subsequent recovery patterns are not yet known. It may be that survival of non-motor regions that are important for sensory processing, sustained attention, memory and learning have a key impact on recovery processes. The medium weights for predicting outcomes given to a variety of cortical regions in our MKL approach would be in keeping with this idea. However, the accuracy obtained through this method was lower than most of the approaches for learning from a single source using patterns of voxels. Thus further investigations would need to be carried out to investigate the potential of MKL to learn non-trivial combinations of features. Additional insights could be obtained with further investigations to select the most relevant data sources using some criterion to produce sparse results (i.e., assigning weight zero to some kernels during the optimization process).

In summary, this study provides a thorough comparison of approaches to extract features based on patterns of voxels and lesion load to decode post-stroke impairment. It is also the first attempt to use multiple kernel learning to investigate the relevance of regions in predicting a score based on structural images from injured brain. Although our investigation is based on a cross-sectional analysis to ask whether current motor impairment can be explained by lesion characteristics, it provides insights towards predicting long term outcome from imaging data acquired in the early stages after stroke. With respect to structural images we expect that data extracted from lesion characteristics will not change substantially over time. Nevertheless, this assumption needs to be tested in a prospective study.

It is important to notice that although the way through which the features are extracted is very important, there are no clear cut answers on the best way to segment lesions. In this paper we opted for defining damaged voxels through lesion probability as a measure of atypicality regarding the expected tissues in healthy controls (Seghier et al., 2008), but there are other ways to define lesions. An interesting extension of this work would be to replicate the findings with other automatic segmentation approach and ultimately with manual tracing of lesions.

From the methodological perspective, there are some advances that we left for future work, given that the main focus of this work is to demonstrate the predictive capabilities of different ways to extract features and regions of interest. In particular, we intend to carry out a fully probabilistic analysis of the proposed GPR approach in order to obtain calibrated uncertainty levels on the weights associated with different data sources. This advance may provide more specific information regarding which brain regions (besides the areas traditionally associated with motor function) can also play an important role in recovery of movement. Another aspect that we aim to investigate is the predictability of the individual motor scores addressing the fact that the motor scales have ceiling and flooring effects. From the biological perspective, we intend to investigate whether the same strategies to extract features which resulted in the best accuracies to explain motor impairment can also be successful to predict future outcomes.

Acknowledgment

This project is funded by the Efficacy and Mechanism Evaluation (EME) Program, an MRC and NIHR partnership.

References

Alexander, G.E., Crutcher, M.D., DeLong, M.R., 1990. Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. Prog. Brain Res. 85, 119–46. doi:10.1016/S0079-6123(08)62678-3

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851. doi:10.1016/j.neuroimage.2005.02.018

Bates, E., Wilson, S.M., Saygin, A.P., Dick, F., Sereno, M.I., Knight, R.T., Dronkers, N.F., 2003. Voxel-based lesion-symptom mapping. Nat. Neurosci. 6, 448–450. doi:10.1038/nn1050

Bayona, N. a, Bitensky, J., Foley, N., Teasell, R., 2005. Intrinsic factors influencing post stroke brain reorganization. Top. Stroke Rehabil. 12, 27–36. doi:10.1310/3BXL-18W0-FPJ4-F1GY

Bonita, R., Beaglehole, R., 1988. Recovery of motor function after stroke. Stroke 19, 1497–1500. doi:10.1161/01.STR.19.12.1497

Bohannon, QR., 1999. Motricity Index Scores are Valid Indicators of Paretic Upper Extremity Strength Following Stroke. J. Phys. Ther. Sci. 11, 59–61. doi:10.1589/jpts.11.59

Burke Quinlan, E., Dodakian, L., See, J., McKenzie, A., Le, V., Wojnowicz, M., Shahbaba, B., Cramer, S.C., 2015. Neural function, injury, and stroke subtype predict treatment gains after stroke. Ann. Neurol. 77, 132–145. doi:10.1002/ana.24309

Byblow, W.D., Stinear, C.M., Barber, P.A., Petoe, M.A., Ackerley, S.J., 2015. Proportional recovery after stroke depends on corticomotor integrity. Ann. Neurol. doi:10.1002/ana.24472

Cao, Y., D'Olhaberriague, L., Vikingstad, E.M., Levine, S.R., Welch, K.M., 1998. Pilot study of functional MRI to assess cerebral activation of motor function after poststroke hemiparesis. Stroke. 29, 112–122. doi:10.1161/01.STR.29.1.112

Cohen, J.R., Asarnow, R.F., Sabb, F.W., Bilder, R.M., Bookheimer, S.Y., Knowlton, B.J., Poldrack, R.A., 2011. Decoding continuous variables from neuroimaging data: basic and clinical applications. Front. Neurosci. 5, 75. doi:10.3389/fnins.2011.00075

Davatzikos, C., Shen, D., Gur, R.C., Wu, X., Liu, D., Fan, Y., Hughett, P., Turetsky, B.I., Gur, R.E., 2005. Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. Arch. Gen. Psychiatry 62, 1218–1227. doi:10.1001/archpsyc.62.11.1218

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V., 1997. Support vector regression machines. Adv. Neural Inf. Process. Syst. 9, 155–161. doi:10.1.1.10.4845

Filippone, M., Marquand, a. F., Blain, C.R. V, Williams, S.C.R., Mourão-Miranda, J., Girolami, M., 2012. Probabilistic prediction of neurological disorders with a statistical assessment of neuroimaging data modalities. Ann. Appl. Stat. 6, 1883–1905. doi:10.1214/12-AOAS562

Forkert, N.D., Verleger, T., Cheng, B., Thomalla, G., Hilgetag, C.C., Fiehler, J., 2015. Multiclass Support Vector Machine-Based Lesion Mapping Predicts Functional Outcome in Ischemic Stroke Patients. PLoS One 10, e0129569. doi:10.1371/journal.pone.0129569

Fu, C.H.Y., Mourao-Miranda, J., Costafreda, S.G., Khanna, A., Marquand, A.F., Williams, S.C.R., Brammer, M.J., 2008. Pattern Classification of Sad Facial Processing: Toward the Development of Neurobiological Markers in Depression. Biol. Psychiatry 63, 656–662. doi:10.1016/j.biopsych.2007.08.020

Geva, S., Jones, P.S., Crinion, J.T., Price, C.J., Baron, J.C., Warburton, E. a., 2011. The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. Brain 134, 3071–3082. doi:10.1093/brain/awr232

Gönen, M., Alpaydın, E., 2011. Multiple kernel learning algorithms. J. Mach. Learn. Res. 12, 2211–2268.

Hauk, O., Johnsrude, I., Pulvermüller, F., 2004. Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. Neuron 41, 301–307. doi:10.1016/S0896-6273(03)00838-9

Hope, T.M.H., Parker Jones, I., Grogan, A., Crinion, J., Rae, J., Ruffle, L., Leff, a. P., Seghier, M.L., Price, C.J., Green, D.W., 2015. Comparing language outcomes in monolingual and bilingual stroke patients. Brain 1 – 14. doi:10.1093/brain/awv020

Hope, T.M.H., Seghier, M.L., Leff, A.P., Price, C.J., 2013. Predicting outcome and recovery after stroke with lesions extracted from MRI images. NeuroImage Clin. 2, 424–433.

doi:10.1016/j.nicl.2013.03.005

Kim, B.J., Kim, Y., Kim, N., Kwon, S.U., Kim, S.J., Kim, J.S., Kang, D., Disclosure, F., 2014. PLOS ONE Lesion Location-Based Prediction of Visual Field Recovery after Cerebral Infarction 1–15. doi:10.1371/journal.pone.0143882

Kim, S.G., Ashe, J., Hendrich, K., Ellermann, J.M., Merkle, H., Uğurbil, K., Georgopoulos, A.P., 1993. Functional magnetic resonance imaging of motor cortex: Hemispheric asymmetry and handedness. Science (80-.). 261, 615–7. doi:http://dx.doi.org/10.1126/science.8342027

Klöppel, S., Stonnington, C.M., Barnes, J., Chen, F., Chu, C., Good, C.D., Mader, I., Mitchell, L.A., Patel, A.C., Roberts, C.C., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S.J., 2008. Accuracy of dementia diagnosis - A direct comparison between radiologists and a computerized method. Brain 131, 2969–2974. doi:10.1093/brain/awn239

Lindenberg, R., Renga, V., Zhu, L.L., Betzler, F., Alsop, D., Schlaug, G., 2010. Structural integrity of corticospinal motor fibers predicts motor impairment in chronic stroke. Neurology 74, 280–287. doi:10.1212/WNL.0b013e3181ccc6d9

Lyle, R.C., 1981. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. Int. J. Rehabil. Res. 4, 483–492.

Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., Mourão-Miranda, J., 2010. Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. Neuroimage 49, 2178–89. doi:10.1016/j.neuroimage.2009.10.072

Mathiowetz, V., Weber, K., Kashman, N., Volland, G., 1985. Adult norms for the Nine Hole Peg Test of finger dexterity. Occup. Ther. J. Res. 5, 24–38. doi:10.1177/153944928500500102

Munsch, F., Sagnier, S., Asselineau, J., Bigourdan, A., Guttmann, C.R., Debruxelles, S., Poli, M., Renou, P., Perez, P., Dousset, V., Sibon, I., 2015. Stroke Location Is an Independent Predictor of Cognitive Outcome. doi:10.1161/STROKEAHA.115.011242

Park CH, Kou, N, Ward NS. The contribution of lesion location to upper limb deficit after stroke. Journal of Neurology, Neurosurgery and Psychiatry 2016 (Epub ahead of print).

Payabvash, S., Kamalian, S., Fung, S., Wang, Y., Passanese, J., Kamalian, S., Souza, L.C.S., Kemmling, a., Harris, G.J., Halpern, E.F., González, R.G., Furie, K.L., Lev, M.H., 2010. Predicting language improvement in acute stroke patients presenting with aphasia: A multivariate logistic model using location-weighted atlas-based analysis of admission CT

perfusion scans. Am. J. Neuroradiol. 31, 1661–1668. doi:10.3174/ajnr.A2125

Rasmussen, C., 2006. Gaussian processes for machine learning. Int. J. Neural Syst. 14, 69–106. doi:10.1142/S0129065704001899

Särkämö, T., Tervaniemi, M., Soinila, S., Autti, T., Silvennoinen, H.M., Laine, M., Hietanen, M., 2009. Cognitive deficits associated with acquired amusia after stroke: A neuropsychological follow-up study. Neuropsychologia 47, 2642–2651. doi:10.1016/j.neuropsychologia.2009.05.015

Schiemanck, S.K., Post, M.W.M., Kwakkel, G., Witkamp, T.D., Kappelle, L.J., Prevo, A.J.H., 2005. Ischemic lesion volume correlates with long-term functional outcome and quality of life of middle cerebral artery stroke survivors. Restor. Neurol. Neurosci. 23, 257–263.

Schulz, R., Park, C.-H., Boudrias, M.-H., Gerloff, C., Hummel, F.C., Ward, N.S., 2012. Assessing the integrity of corticospinal pathways from primary and secondary cortical motor areas after stroke. Stroke. 43, 2248–51. doi:10.1161/STROKEAHA.112.662619

Seghier, M.L., Ramlackhansingh, A., Crinion, J., Leff, A.P., Price, C.J., 2008. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. Neuroimage 41, 1253–1266. doi:10.1016/j.neuroimage.2008.03.028

Smola, A.J., Sch, B., Schölkopf, B., 2004. A Tutorial on Support Vector Regression. Stat. Comput. 14, 199–222. doi:10.1023/B:STCO.0000035301.49549.88

Stinear, C., 2010. Prediction of recovery of motor function after stroke. Lancet Neurol. doi:10.1016/S1474-4422(10)70247-7

Sunderland, A., Tinson, D., Bradley, L., Hewer, R.L., 1989. Arm function after stroke. An evaluation of grip strength as a measure of recovery and a prognostic indicator. J. Neurol. Neurosurg. Psychiatry 52, 1267–72. doi:10.1136/jnnp.52.11.1267

Teipel, S.J., Born, C., Ewers, M., Bokde, A.L., Reiser, M.F., Moller, H.J., Hampel, H., 2007. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. Neuroimage 38, 13–24. doi:10.1016/j.neuroimage.2007.07.008

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289. doi:10.1006/nimg.2001.0978

Vemuri, P., Wiste, H.J., Weigand, S.D., Knopman, D.S., Trojanowski, J.Q., Shaw, L.M., Bernstein, M.A., Aisen, P.S., Weiner, M., Petersen, R.C., Jack, C.R., 2010. Serial MRI and CSF biomarkers in normal aging, MCI, and AD. Neurology 75, 143–151. doi:10.1212/WNL.0b013e3181e7ca82

Wang, Y., Fan, Y., Bhatt, P., Davatzikos, C., 2010. High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables. Neuroimage 50, 1519–1535. doi:10.1016/j.neuroimage.2009.12.092

Ward, N.S., 2015. Does neuroimaging help to deliver better recovery of movement after stroke? Curr. Opin. Neurol. 44, 323–329. doi:10.1097/WCO.0000000000000223

Ward, N.S., Swayne, O.B.C., Newton, J.M., 2008. Age-dependent changes in the neural correlates of force modulation: An fMRI study. Neurobiol. Aging 29, 1434–1446. doi:10.1016/j.neurobiolaging.2007.04.017

Yang, Z.-H., Zhao, X.-Q., Wang, C.-X., Chen, H.-Y., Zhang, Y.-M., 2008. Neuroanatomic correlation of the post-stroke aphasias studied with imaging. Neurol. Res. 30, 356–360. doi:10.1179/174313208X300332

Zarahn, E., Alon, L., Ryan, S.L., Lazar, R.M., Vry, M.S., Weiller, C., Marshall, R.S., Krakauer, J.W., 2011. Prediction of motor recovery using initial impairment and fMRI 48 h poststroke. Cereb. Cortex 21, 2712–2721. doi:10.1093/cercor/bhr047

Zhu, L.L., Lindenberg, R., Alexander, M.P., Schlaug, G., 2010. Lesion load of the corticospinal tract predicts motor impairment in chronic stroke. Stroke 41, 910–915. doi:10.1161/STROKEAHA.109.577023

Supplementary Material

1. Characteristics of the sample

Table S1 presents characteristics of each patient enrolled in the study. The quantitative summary of each characteristic is presented in the last row. Gender is summarized as the total number of females (F) and the affected hand is summarized as the number of patients with right hand impairment (R). The other characteristics are summarized with average (standard deviation) of the sample.

The location of the lesions in each patient is presented in the Table S1 (that also shows the proportion of voxels considered to be part of the lesion in each ROI).

Table S1. Demographic and clinical characteristics of the patients

Patient ID	Gender	Age (yrs)	Affected hand	Time since stroke (months)	Lesion volume (voxels)	ARAT	Grip	MI	NHPT
1	M	60	L	6	3589	26	24.9	48	5.1
2	M	60	L	41	24326	39	20.1	65	0
3	M	48	L	116	11068	54	87.6	88	70.8
4	M	46	R	8	2070	57	106.2	100	104.5
5	F	62	L	3	3568	52	111.7	93	60.6
6	M	69	R	9	3599	57	80.5	100	69.7
7	M	44	L	8	5437	36	78.6	81	5.1
8	M	33	R	63	3048	57	71.7	100	68.9
9	M	51	R	4	569	57	64.2	100	89.6
10	M	48	L	7	27919	31	35.3	42	0
11	M	58	L	7	1270	57	93.3	100	74.5
12	M	59	L	3	38204	52	64.4	42	14.9
13	M	22	R	3	2448	57	93.1	100	94.5
14	F	69	L	30	4225	56	84	85	19.8
15	M	42	L	26	19753	44	80.6	85	0
16	M	70	L	3	9011	57	81.3	100	81.3
17	F	50	L	18	33415	28	41.3	66	8.2
18	F	40	L	21	15679	52	58	93	25.5
19	M	80	L	20	3858	26	56.3	74	4
20	M	75	L	6	6600	57	96.6	100	73.7
21	F	55	L	5	1320	55	64	93	97
22	F	53	L	31	1391	50	40	91	50
23	F	59	L	165	6210	29	18.2	68	0
24	F	77	R	26	7332	38	57.2	77	9
25	M	58	L	11	39409	57	88.2	100	87
26	F	56	R	16	5486	36	44	77	9
27	M	62	L	38	4325	36	31	72	8
28	M	59	L	79	24482	21	50.3	73	0
29	M	63	R	8	4459	57	91	100	77.9
30	M	58	L	13	16587	37	66	64	9

31	M	66	R	26	38714	35	81	65	39
32	M	53	R	12	8653	48	52	91	53
33	F	43	R	20	13736	41	71	91	31
34	M	36	R	20	4745	54	81.9	93	31
35	F	45	L	29	6450	45	51	85	35
36	F	52	R	83	42304	27	31	73	0
37	M	51	R	60	6778	45	104	92	31
38	M	61	L	13	2226	45	51.1	65	19.7
39	F	42	L	22	25769	45	65	91	21
40	F	56	R	32	8526	39	41	77	9
41	M	57	R	17	10528	57	84	84	71
42	M	52	L	24	3579	42	35	88	9
43	M	50	R	13	7744	57	68	100	60
44	F	69	L	38	6690	57	94	93	50
45	F	40	L	25	16720	55	60	93	30
46	M	46	L	26	8461	46	52.3	84	38
47	M	56	R	13	22623	37	107	91	40
48	M	66	L	76	6233	21	46.6	72	0
49	F	18	L	5	14037	44	27.7	93	5.9
50	M	66	L	5	5409	57	63.4	92.5	98.2
Summary	**17 F**	**54.22 (12.62)**	**18 R**	**27.06 (31.05)**	**11811.64 (11365.58)**	**45.30 (11.28)**	**64.93 (24.58)**	**83.81 (15.30)**	**37.78 (33.13)**

The figure S1 presents a matrix showing the proportion of each region affected by lesions in each patient. Each column (from 1 to 50) relates to a particular patient and each row corresponds to an ROI from the AAL atlas. Only ROIs that were affected in at least one patient are presented. The colour scale (from 0 to 1) represents the lesion load (i.e., the number of voxels in the ROI considered to be part of the lesion divided by the ROI size).

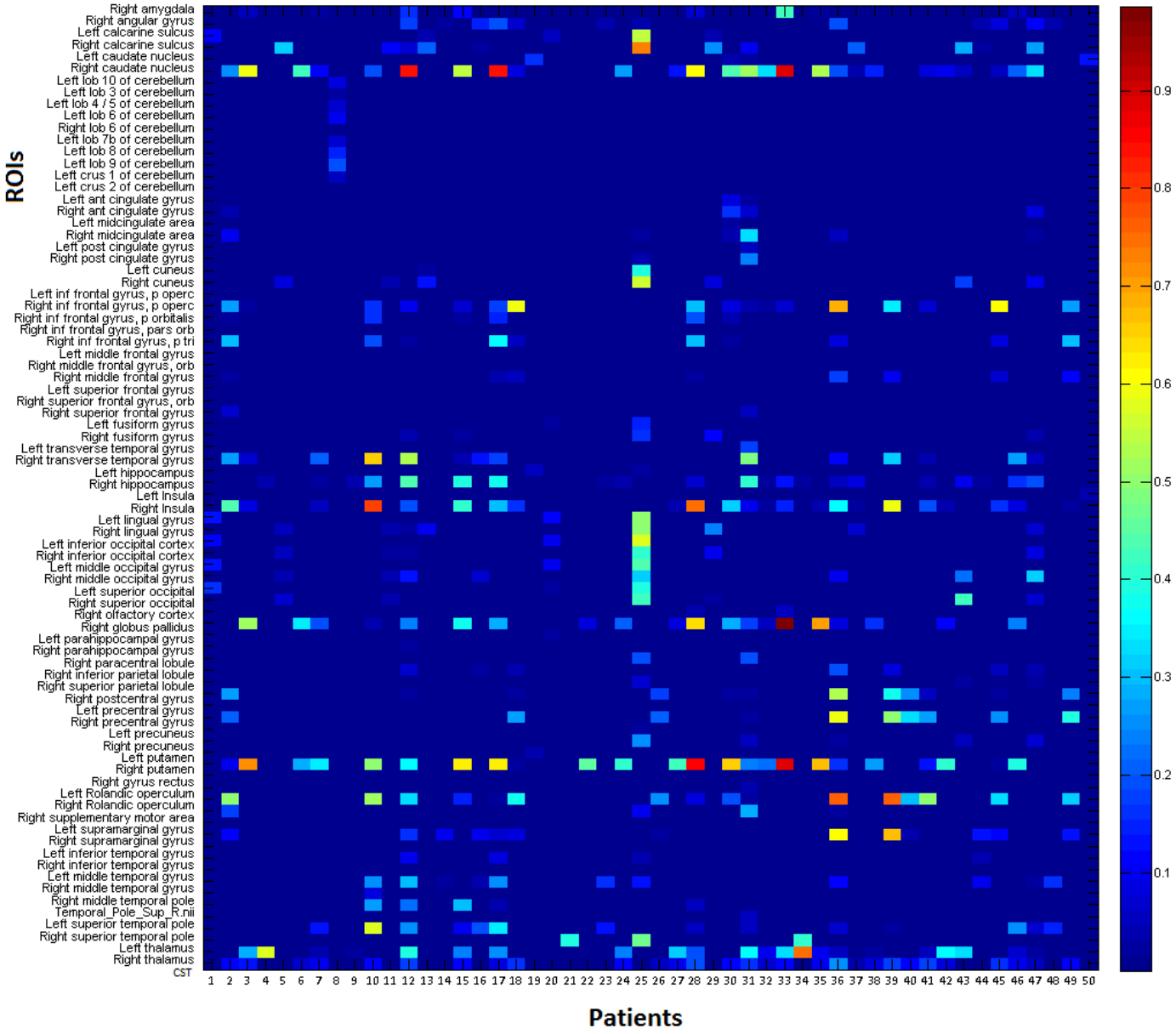


Figure S1. Lesion load per ROI for each patient

2. Lesion segmentation

We evaluated whether the filter to eliminate small clusters of contiguous voxels in the binary images resulting from segmentation was necessary in our sample. Given that stroke lesions typically involve neighbouring voxels, a considerable amount of single, false-positive voxels would not be expected. However, we checked the segmentation of each patient and noticed that the filter was indeed necessary in our images, as suggested by (Seghier et al., 2008). The figure S2 illustrates an example, where false positive voxels (represented in red) were above the threshold, but they formed small clusters with less than 100 voxels, thus being excluded. Additionally, in order to check whether this filter is not eliminating small lesions, we checked the segmentation of all patients. The smallest lesion is a contiguous volume with 569 voxels.

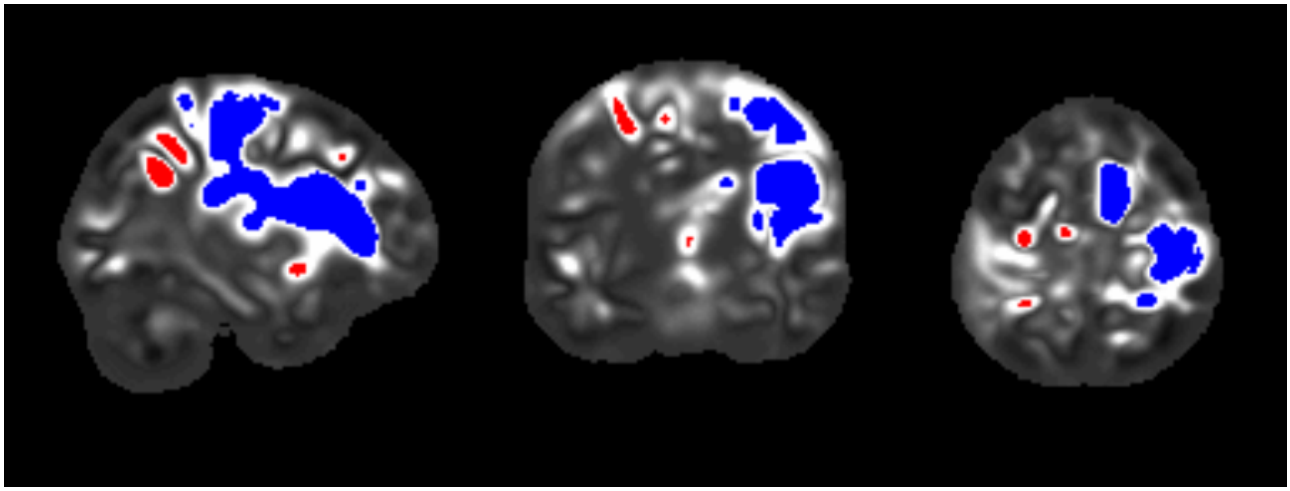


Figure S2. Example of segmentation without filtering small clusters (represented in red)

Regarding the threshold level, we did segmentations with alternative values slightly below and above 0.30 (the level suggested by (Seghier et al., 2008)) and repeated the regression analysis, whose results are presented in Table S2 (threshold 0.25) and Table S3Table S3 (threshold 0.35). These results did not show systematic improvement regarding results from analysis using the suggested threshold (presented in the manuscript, table 3).

Table S2. Prediction of motor scores based on lesion load segmented with threshold 0.25

Model	Features	NF	R	MSE
M2	Lesion load in the whole brain	1	0.28	0.95
M2.1	Lesion load in ROIs from AAL atlas	116	0.31	2.27
M2.2	Lesion load in Corticospinal tract	1	0.50	0.84
M2.3	Lesion load in ROIs from AAL atlas + CST	117	0.23	1.87
M2.4	Lesion load in motor ROIs	22	0.32	3.48
M2.5	Lesion load in motor ROIs + CST	23	0.31	3.60
M2.6	Lesion load in Functional mask from task fMRI	1	0.25	0.92
M2.7	Lesion load in ROIs defined by Lesion-symptom mapping	5	0.30	0.94
M2.8	Lesion load in ROI from lesion in at least 1 patient	1	0.29	0.94

Table S3. Prediction of motor scores based on lesion load segmented with threshold 0.35

Model	Features	NF	R	MSE
M2	Lesion load in the whole brain	1	0.30	0.93
M2.1	Lesion load in ROIs from AAL atlas	116	0.10	120.95
M2.2	Lesion load in Corticospinal tract	1	0.49	0.84
M2.3	Lesion load in ROIs from AAL atlas + CST	117	0.29	182.10
M2.4	Lesion load in motor ROIs	22	0.11	1.14
M2.5	Lesion load in motor ROIs + CST	23	0.16	1.13
M2.6	Lesion load in Functional mask from task fMRI	1	0.25	0.91
M2.7	Lesion load in ROIs defined by Lesion-symptom mapping	5	0.30	0.92
M2.8	Lesion load in ROI from lesion in at least 1 patient	1	0.30	0.93

3. Support vector regression

The results of the application of SVR with the same kernel are presented in the tables Table S4 and Table S5.

Table S4. SVR prediction of motor scores based on patterns of voxels representing lesion probability

Model	Features	NF	R	MSE
M1	Whole brain	630786	0.67	0.61
M1.1	Voxels limited by AAL atlas	451318	0.69	0.59
M1.2	Voxels limited by the Corticospinal tract (CST)	4421	0.56	0.67
M1.3	Voxels limited by AAL atlas + CST	457384	0.70	0.58
M1.4	Voxels limited by motor ROIs	120793	0.74	0.57
M1.5	Voxels limited by motor ROIs and CST	125214	0.77	0.54
M1.6	Voxels limited by mask from task fMRI in healthy controls	35545	0.67	0.72
M1.7	Voxels limited by lesion-symptom mapping	** 9991.1*	0.57	0.62
M1.8	Voxels limited by lesion in at least 1 patient	158907	0.66	0.57

Table S5. SVR prediction of motor scores based on lesion load

Model	Features	NF	R	MSE
M2	Lesion load in the whole brain	1	0.29	1.02
M2.1	Lesion load in ROIs from AAL atlas	116	0.47	0.91
M2.2	Lesion load in Corticospinal tract	1	0.50	0.86
M2.3	Lesion load in ROIs from AAL atlas + CST	117	0.43	0.92
M2.4	Lesion load in motor ROIs	22	0.36	1.23
M2.5	Lesion load in motor ROIs + CST	23	0.37	1.29
M2.6	Lesion load in Functional mask from task fMRI	1	0.17	0.95
M2.7	Lesion load in ROIs defined by Lesion-symptom mapping	5	0.40	0.86
M2.8	Lesion load in ROI from lesion in at least 1 patient	1	0.29	1.02

4. Multiple kernel learning applied to ROIs

Table S6 presents a list of all ROIs (116 areas from the AAL atlas and the CST divided into left and right portions). The ROIs are ranked in descending order according to the weight assigned to them by the multiple kernel learning analysis. The figure S3 gives an intuition of the distribution of the weights across the ROIs through a bar graph.

Table S6. Rois sorted by weight in multiple kernel learning analysis

Ranking	ROI name	Weight
1	Right caudate nucleus	0.0267
2	Left inferior frontal gyrus, pars opercularis	0.0264
3	Right corticospinal tract	0.0262
4	Left middle temporal pole	0.0259
5	Right lobule 7b of cerebellar hemisphere	0.0256
6	Right superior temporal pole	0.0247
7	Right lobule 9 of cerebellar hemisphere	0.0239
8	Left Cuneus	0.0236
9	Left caudate nucleus	0.0233
10	Lobule VII of vermis	0.0231
11	Right lobule 10 of cerebellar hemisphere	0.0230
12	Left calcarine sulcus	0.0227
13	Lobule VI of vermis	0.0226
14	Lobule VIII of vermis	0.0223
15	Right cuneus	0.0223
16	Left globus pallidus	0.0222
17	Right crus II of cerebellar hemisphere	0.0218
18	Lobule IV and V of vermis	0.0216
19	Left middle frontal gyrus, orbital part	0.0214
20	Right Amygdala	0.0209
21	Left medial frontal gyrus	0.0208
22	Right Hippocampus	0.0207
23	Left lobule 7b of cerebellar hemisphere	0.0201
24	Lobule X of vermis	0.0200
25	Left inferior temporal gyrus	0.0191
26	Right rectus gyrus	0.0189

27	Left lobule 4_5 of cerebellar hemisphere	0.0188
28	Left transverse temporal gyrus	0.0186
29	Right globus pallidus	0.0184
30	Left superior temporal gyrus	0.0183
31	Right middle occipital gyrus	0.0183
32	Lobule I and II of vermis	0.0182
33	Left superior parietal lobule	0.0180
34	Left supplementary motor area	0.0180
35	Right paracentral lobule	0.0180
36	Left putamen	0.0179
37	Left corticospinal tract	0.0178
38	Left angular gyrus	0.0178
39	Right superior temporal gyrus	0.0177
40	Right inferior frontal gyrus, pars opercularis	0.0172
41	Left olfactory cortex	0.0171
42	Right lobule 8 of cerebellar hemisphere	0.0167
43	Left fusiform gyrus	0.0166
44	Left lingual gyrus	0.0165
45	Left superior occipital	0.0165
46	Right superior occipital	0.0165
47	Right ParaHippocampal gyrus	0.0164
48	Right inferior frontal gyrus, pars triangularis	0.0164
49	Left Rolandic operculum	0.0164
50	Right Rolandic operculum	0.0162
51	Right precentral gyrus	0.0161
52	Left Thalamus	0.0160
53	Right supramarginal gyrus	0.0159
54	Left supramarginal gyrus	0.0159
55	Right inferior temporal gyrus	0.0158
56	Left lobule 10 of cerebellar hemisphere	0.0156
57	Left precuneus	0.0154
58	Left gyrus rectus	0.0154
59	Right inferior parietal lobule	0.0152
60	Right Fusiform gyrus	0.0152
61	Right posterior cingulate gyrus	0.0151
62	Right postcentral gyrus	0.0148

63	Left lobule 3 of cerebellar hemisphere	0.0147
64	Right middle frontal gyrus, orbital part	0.0146
65	Left lobule 9 of cerebellar hemisphere	0.0146
66	Right lobule 4_5 of cerebellar hemisphere	0.0145
67	Left crus I of cerebellar hemisphere	0.0144
68	Right crus I of cerebellar hemisphere	0.0143
69	Right Thalamus	0.0143
70	Right supplementary motor area	0.0141
71	Left precentral gyrus	0.0140
72	Left superior frontal gyrus, orbital part	0.0140
73	Right inferior occipital cortex	0.0139
74	Right inferior frontal gyrus, pars orbitalis	0.0139
75	Right calcarine sulcus	0.0136
76	Right middle temporal pole	0.0135
77	Right insula	0.0133
78	Right Putamen	0.0133
79	Right transverse temporal gyrus	0.0132
80	Left medial frontal gyrus	0.0132
81	Left lobule 8 of cerebellar hemisphere	0.0130
82	Left ParaHippocampal gyrus	0.0129
83	Right lingual gyrus	0.0129
84	Left superior temporal pole	0.0128
85	Left Hippocampus	0.0127
86	Right lobule 6 of cerebellar hemisphere	0.0127
87	Left Amygdala	0.0126
88	Right superior frontal gyrus	0.0124
89	Lobule IX of vermis	0.0124
90	Right middle temporal gyrus	0.0124
91	Left inferior occipital cortex	0.0120
92	Left superior frontal gyrus	0.0120
93	Right superior parietal lobule	0.0118
94	Left lobule 6 of cerebellar hemisphere	0.0117
95	Right olfactory cortex	0.0116
96	Left middle occipital gyrus	0.0116
97	Left midcingulate area	0.0116
98	Left middle temporal gyrus	0.0113

99	Right lobule 3 of cerebellar hemisphere	0.0113
100	Right angular gyrus	0.0112
101	Left middle frontal gyrus, orbital part	0.0111
102	Left postcentral gyrus	0.0110
103	Lobule III of vermis	0.0108
104	Left crus II of cerebellar hemisphere	0.0107
105	Right middle frontal gyrus, orbital part	0.0104
106	Left paracentral lobule	0.0103
107	Right precuneus	0.0103
108	Right anterior cingulate gyrus	0.0099
109	Right medial frontal gyrus	0.0099
110	Right midcingulate area	0.0098
111	Right superior frontal gyrus, orbital part	0.0097
112	Right middle frontal gyrus	0.0097
113	Left inferior frontal gyrus, pars orbitalis	0.0096
114	Left insula	0.0095
115	Left anterior cingulate gyrus	0.0094
116	Left inferior parietal lobule	0.0087
117	Left posterior cingulate gyrus	0.0083
118	Left inferior frontal gyrus, pars triangularis	0.0081

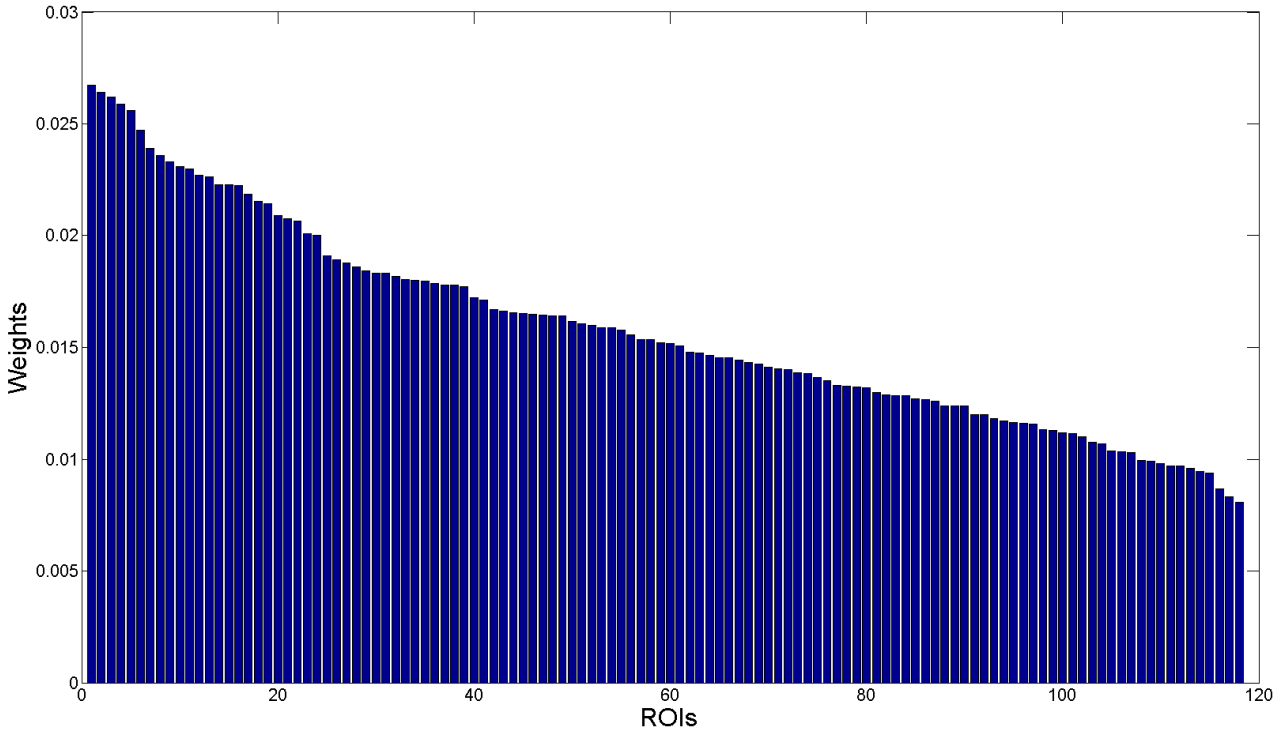


Figure S3 Weight of each pattern ROI to predict motor impairment using multiple kernel learning

5. Number of features and performance

In order to investigate how the number of features used relates to the performance of the regression algorithm independent of the anatomical information, we selected a subset of ROIs from the AAL atlas in a random way. ROIs were randomly selected until the total number of voxels reached the order of the model M1.4 (table 1, 120793 voxels). Using the pattern of voxels delimited by this mask and applying GPR to predict the motor score, the correlation between real and predicted labels was R = 0.61; MSE = 0.77. The accuracy was lower than using the pattern of voxels delimited by both the whole brain (model M1: R = 0.72; MSE = 0.73) and the motor ROIs (model M1.4: R=0.80; MSE = 0.70). This result suggests that the influence of the number of features in the performance is not independent of the anatomical information.