Abstract— Successful post-stroke prognosis and recovery strategies are heavily dependent on our understanding about how the damage to one specific region may impact to other remote regions, as well as the various functional networks involved in efficient cognitive function. In this study, 27 consecutive ischemic stroke patients were recruited. Stroke patients underwent two complete neuropsychological assessments between the first 72 hours after stroke arrival and three months later. They were further evaluated with a MRI protocol at 3 months. Patients were splitted into two groups according to their level of cognitive recovery. A data mining technique was then applied to the probabilistic tractography data in order to determine whether the structural connectivity features can efficiently classify good from poor recovery. We found that the connectivity probability between the left Superior Parietal Gyrus and the left Angular Gyrus can describe the cognitive classification (good versus poor recovery) after stroke. Both regions are involved in higher cognitive functioning and their dysfunction has been related to mild cognitive impairment and dementia. Our findings suggest that cognitive prognosis, in stroke patients, mainly depends on the connection of these two regions. An accurate model for the early prediction of stroke recovery as the one presented herein is fundamental to develop early personalized rehabilitation strategies.

Keywords- Stroke, cognitive recovery, structural connectome, Naïve Bayesian Tree

I. INTRODUCTION

Acute ischemic stroke is the second most common cause of death worldwide and a major cause of disability in older adults [1]. Until now, the mechanisms underlying the post-stroke functional recovery, observed in some patients, remain unknown. Among the factors reported in the literature are the plasticity mechanisms within the surviving brain tissue [2]. Following a focal stroke, there are multiple
ways in which the structure and function of the brain may change. The region immediately surrounding a stroke undergoes potentially reversible structural changes and anterograde or retrograde degeneration of axons intersecting or connecting with a lesion site may occur [3]. Despite this knowledge, the responsible mechanisms for the potential for a good functional recovery remain elusive [4]. In stroke, clinical and cognitive deficits are not always easily explainable by the lesion localization itself [5], lesions of some areas tend to have more severe effects than others and individual outcomes seem to be more due to residual anatomy than to lesion localization [6].

Regarding stroke patients, two major unsolved questions arise. The first one is how the damage of a specific region can affect remote regions, while the second one considers which networks are more closely related with the efficient cognitive functions and how their damage have implications in prognosis and recovery. At this point, the ability to assess multiple networks at once as well as their interaction may be especially valuable, since stroke patients often show deficits in a wide range of cognitive functions, being executive function, processing speed and memory the most reported [7]. These functions are known to be subserved by distributed and integrated neural networks [8]. In a study about simulated targeted attack [9], it was found that dynamic lesion effects were particularly large and widespread when lesions included nodes or edges with high centrality. For example, in one study [3], it was reported that changes in brain structure occurred not only in remote regions following a focal damage in the affected hemisphere but in homologous locations in the contralesional hemisphere [9], suggesting that there was a disrupted information processing [10] across widely distributed regions.

In this scenario, it is difficult to predict which patient will recover and reintegrate into society. The main objective of this study was to examine if the structural connectivity in a sample of ischemic stroke patients can be used as a benchmark to predict cognitive recovery.

II. MATERIALS & METHODS

A. Participants

We performed a longitudinal study of 27 consecutive subjects that had suffered a first stroke. These subjects were recruited at the Neuroscience Service of the University Hospital Germans Trias i Pujol (Badalona) from September 2010 to May 2012. Inclusion and exclusion criteria for both the patients and the healthy controls have been previously defined, together with information about previous cognitive impairment [11]. Premorbid Intelligence was estimated using the vocabulary subtest of Wechsler Adults Intelligence Scale (WAIS-III-R) [12] at three months post-stroke. Barthel Index (BI) [13], the modified Rankin scale (mRS) [14] and the NIHSS [15] were collected by a trained neurologist. Eighteen healthy volunteers from the Barcelona-Asymptomatic Intracranial Atherosclerosis (AsIA) study [16] matched by age, sex, education, and handedness were recruited and second time reassessed (cognition and MRI) as the control group. The study was approved by the ethics committee of the University of Barcelona. All enrolled subjects gave their written consent to participate in the study, which was conducted according to the provisions of the Helsinki declaration.

B. Neuropsychological assessment and grouping criteria

Patients received neuropsychological examination at two different times: between 72 hours after arrival and at 3 months. The neuropsychological assessment included: the Digit Span Forward Test, the subtest of attention extracted from de Montreal Cognitive Test (MOCA), the Line Cancellation Test, the Digit Span Backwards (WAIS-III-R), the part B of the Trail Making Test, the verbal fluency test, the Semantic fluency test, the short version (15-items) of the Boston Naming Test, the Luria’s sequences test, the Rhythms subtest extracted from the MOCA test, the interference and inhibitory control subtest extracted from the Frontal Assessment Battery, the part A of the Trail Making Test and the Grooved Pegboard Test (For a detailed description of the tests used see [17]). Neuropsychological examinations also included the MMSE [18], as a global cognitive test and the Geriatric Depression Scale (GDS) [19].

Stroke patients were separated into two groups, the stroke group with poor recovery (SP) and the stroke group with good recovery (SG), using the following procedure: first, a paired t-test was run to determine those tests where the patients had undergone significant improvement at three months post-stroke; second, we decided that a subject had a better recovery if he/she had improved his/her scoring by 1.5 standard deviations in at least three of the tests that had shown significant improvement.

C. Lesion analysis

An experienced stroke neurologist and an experienced neuroradiologist, both blind to the neuropsychological data, evaluated the patients’ stroke location at baseline from CT or MRI. The side of the lesion was determined (left versus right), as well as lesion site (cortical involvement of the lesion versus exclusive subcortical involvement) and affected vascular territory. Lesion volume was calculated at three months post-stroke after manual tracing of the lesion on the T2-weighted images using Mricro software [20] on each slice showing the infarct, followed by multiplying lesion area by slice thickness in all slices showing the lesion. This method has been shown to have a high inter-rater reliability and is described in detail elsewhere [21].

D. Image Acquisition and Preprocessing

All images were acquired at 3 months after stroke. The MRI protocol included a set of magnetization prepared rapid gradient echo (MP-RAGE) T1-weighted images and two DWI runs in 30 noncollinear diffusion directions. Data preprocessing was performed using FSL’s tools.
(http://www.fmrib.ox.ac.uk/fsl/) and consisted in eddy current, motion artifact correction, calculation of the diffusion tensor and estimation of the probabilistic distribution of fiber orientations from each voxel.

**E. Probabilistic Tractography**

The 90 supratentorial regions defined by the Automated Anatomical Labeling (AAL) template [22] were established as network nodes. The parcellation process for each subject was conducted in the diffusion native space [23]. We performed probabilistic fiber tractography between each pair of the 90 AAL grey matter seed regions in every subject to estimate the connectivity between these node regions. We first estimated the local probability distribution of fiber direction at each voxel using the Bayesian framework proposed by [24]. We used a computation model capable of dealing with non-dominant fiber populations [25]. Probabilistic tractography was applied by sampling 5000 streamlines per voxel and performing 2000 steps per sample, while the anisotropy index was used to constrain tracking. In this method, the connectivity probability is interpreted as reliability that white matter fiber tracts exist between 2 given regions [24], [25]. We did not exclude the stroke lesion from this analysis since we found that in diffusion space they behaved as cerebrospinal fluid, therefore, potential streamlines would quickly spread in the lesion owing to the high uncertainty.

**F. Features and Feature Selection**

In order to detect if a stroke patient will have a good or poor cognitive recovery, we employed data mining techniques. For this procedure features based on structural connectivity were employed, resulting in 447 features in total. Since the data reduction in neuroscience is crucial from a computational perspective [26], from the initial connectome symmetric matrix (total cells: (90x90-90)/2=4005), we kept only 380 since the rest 3625 were equal to 0 for all the participants, so there is not any variance that can predict the cognitive recovery of the patients. Our classifier (see next section) was applied in the whole dataset of 447 features only in the features selected by a feature selection algorithm. For the feature selection the Best First search method was used, which usually starts with an empty set of features and searches forward (meaning adding and evaluating features), and was performed using the CfsSubsetEval function of WEKA [27] which evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them [28]. For the feature selection assessment the 10-fold cross validation method was used, where the initial dataset was randomly separated into 10 equal size sub-datasets, 9 of them were used as a training set and 1 as the validation set. The cross-validation procedure was then repeated 10 times with each one of the sub-datasets used exactly once as the validation set. Then for every validation time, our features were marked with 0s and 1s if a feature could describe/predict the validation set

**G. Naïve Bayesian Tree (NBTree)**

A Naïve Bayesian Classifier (NBC) is a simple classifier based on Bayes' theorem with enhanced independence assumptions[29]. In other words, a NBC assumes that the occurrence of a particular event is uncorrelated with the occurrence of any other event. Abstractly, the probability model for the NBC is a conditional model:

$$p(C | F_1, \ldots, F_n) = \frac{1}{Z} \prod_{i=1}^{n} p(F_i | C)$$

where $Z$ is a scaling factor dependent on feature variables $\{F_i\}_{i=1}^{n}$ and $C$ is the dependent class variable on $\{F_i\}_{i=1}^{n}$. The NBC uses the aforementioned model and combines it with a decision rule. The most common rule is to choose the most probable hypothesis, which is already known as the maximum a posteriori decision rule. The NBC is then described by the following function:

$$cl(f_1, \ldots, f_n) = \operatorname{argmax}_{c} p(C = c) \prod_{i=1}^{n} p(F_i = f_i | C = c)$$

The NBTree algorithm proposed by[30]is similar to the classical recursive partitioning schemes, except that the generated leaf nodes are Naïve-Bayes categorizers instead of nodes predicting a single class (for more details on this algorithm see [30]. For the training and evaluation of the NBTree we have used the 10-fold cross validation methodology. For the purposes of this study the NBTree implemented in WEKA toolbox [27] with the default parameter set was used.

**III. Results**

Only one feature, which is the connectivity probability between the left Superior Parietal Gyrus and the left Angular Gyrus, had 100% of success, meaning that for each one of the 10 validation times it was marked as a proper feature to describe the cognitive recovery after stroke. In order to verify the validity of the aforementioned feature selection procedure, the classification was performed two times, one with the whole feature-set and one with only the single aforementioned feature (Table 1).
We focused the rest of the analysis to the single feature results. As we can see from Table 1, NBTree managed to correctly classify 23/27 instances, while it missed only four instances. These four instances were patients with good cognitive recovery that were identified as poor, while NBTree managed to correctly classify all the patients with poor cognitive recovery. In order to compute specificity and sensitivity we defined the poor cognitive recovery as the positive condition and the good one as a negative, since the main challenge of our work is to detect which patients will have poor cognitive recovery. Taking into account the confusion matrix (Table 2), sensitivity is equal to 1, since all the SP cases were correctly classified, while the specificity is 0.7895, since we had 4 false negatives.

<table>
<thead>
<tr>
<th>Classified as</th>
<th>SG</th>
<th>SP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>SP</td>
<td>0</td>
<td>15</td>
</tr>
</tbody>
</table>

Fig.1 illustrates the classification tree for replicability purposes, as well as for further validation with new datasets of the presented classifier.

![Figure1. The NBTree for classifying the SG and SP classes.](image)

IV. DISCUSSION

The results extracted by WEKA were very promising, especially for the classification performed with the single feature. More precisely, we found that the classification accuracy attained acceptable levels, both for the feature-set (62.96% accuracy) and single feature (85.18%) cases.

This paper addresses the potential classification of patients with good cognitive recovery versus patients with poor cognitive recovery after ischemic stroke at 3 months. We found only one feature, the connectivity probability between the left Superior Parietal Gyrus (SPG) and the left Angular Gyrus (AG), which resulted in a 100% success rate, meaning that for each one of 10 validation runs, this connectivity feature was marked as a proper feature to describe the cognitive recovery after stroke. Both, the SPG and the AG are located in the parietal lobe and are connected to the medial temporal lobe (MTL), as well as to the dorsolateral prefrontal cortex (DLPFC) by means of the connections between the arcuate fasciculus and the Superior Longitudinal Fasciculus [29]. If we look at functional studies (www.neurosynth.org), we find that both regions belong to the same node in the DMN. In our study, the connectivity between these two structures predicts cognitive outcome in stroke patients. In fact, cytoarchitectonic maps obtained from post-mortem brains suggest that the human inferior parietal lobe has a more finely grained parcellation than previously suggested by the classical Brodmann map [30].

We postulate that the connectivity deficit found between these two structures could extend beyond and affect other regions highly connected in the DMN.

The SPG plays a pivotal role in many cognitive, perceptive and motor-related processes and it is structurally connected with dorsal-posterior temporal regions [31] and with the dorsolateral prefrontal cortex (DLPFC) by means of the Superior Longitudinal Fasciculus (SLF) [29]. The frontal, temporal and parietal lobes contain the majority of the tertiary association cortex, which are key substrates for higher cognition including executive function, language, memory and attention [31].

Abnormal connectivity of the DLPFC as well as functional disconnection involving this region have been observed in mild cognitive impairment (MCI) and Alzheimer’s disease AD, although the detailed connection patterns of the DLPFC is still unclear [32]. The SPG, an area highly connected with the DLPFC and the MTL, could be a key area to discriminate between stroke patients with good or poor recovery.

The AG, situated at the junction of the temporal, parietal, and occipital lobes, may be considered a heteromodal region. Resting-state functional magnetic resonance imaging (fMRI) and positron emission tomography studies have consistently identified the AG as a key parietal node of the DMN [33],[34],[35].Task-related deactivations have been widely reported in the AG [36], [37]. DMN disruption has been commonly observed by resting-state functional studies (fMRI-rs) in patients with mild cognitive impairment [38], [39], vascular cognitive impairment with subcortical lesions [40], patients with carotid stenosis [41], [42], and patients...
with stroke [43], [44]. The fact that both the SPG and AG have been related to MCI and dementia further reinforces our conclusion that reduced connectivity between those two regions is a key feature capable of classifying SG and SP participants.

These results are in agreement with previous research carried out by our group regarding cognitive recovery in stroke patients at 3 months after stroke. In one study, using graph theory analysis with fMRI-rs [44], we found that patients with ischemic stroke showed a disrupted DMN activity pattern in which the DLPFC, the MTL as well as the parietal lobe were disconnected. This disconnection was a key point not only in the disruption but in the cognitive execution of those patients. In a second study, in which we studied differences between healthy controls and stroke patients regarding their level of cognitive recovery, using Independent Component Analysis (ICA) with fMRI-rs [11], we found that the AG, between other regions, was an anatomical dysfunctional region involved, as an heteromodal region, in two different resting state network (basal ganglia and frontal network) as well as in cognitive function in stroke patients. Finally, in a third study we investigated white matter integrity status in stroke patients regarding their cognitive recovery using Tract Base Spatial Statistics [45]. We found that the left uncinate fasciculus and the left Superior Longitudinal Fasciculus were two regions impaired in stroke patients with poor cognitive recovery when compared with healthy controls.

Prediction of cognitive recovery in stroke is a research area of increasing interest due to the growing evidence that recovery not only happens in the first year as a result of spontaneous recovery [2] but it can occur in the long term as well [46]. Finding an accurate model for making an early prediction of limited stroke recovery can help us to better characterize those patients and to develop advanced and personalized rehabilitation techniques directed to better improve the different deficits in a moment when brain plasticity is still high enough to allow better remodeling of the remaining intact cortical and subcortical structures. It is time to go beyond; we do not have enough with merely understanding how the remaining structures can support the recovery of a function, we need to take advantage of our current knowledge and find strategies to improve stroke recovery in a clinical setting.

Although the results herein are robust and hopeful, we believe that the accuracy could be increased even more by deploying massive statistical models, like partial least square correlation [47], where the common information of both neuroimaging and neuropsychological data would be merged, revealing subtle relationships not detected by current classification approaches.

REFERENCES


