Prediction of impaired performance in Trail Making Test in MCI patients with small vessel disease using DTI data

Stefano Ciulli, *Student Member, IEEE*, Luca Citi, *Member, IEEE*, Emilia Salvadori,
Raffaella Valenti, Anna Poggesi, Domenico Inzitari, Mario Mascalchi,
Nicola Toschi, Leonardo Pantoni and Stefano Diciotti*, *Member, IEEE*

Abstract—Mild cognitive impairment (MCI) is a common condition in patients with diffuse hyperintensities of cerebral white matter (WM) in T2-weighted magnetic resonance images and cerebral small vessel disease (SVD). In MCI due to SVD, the most prominent feature of cognitive impairment lies in degradation of executive functions, i.e. of processes which supervise the organization and execution of complex behavior. The Trail Making Test (TMT) is a widely employed test sensitive to cognitive processing speed and executive functioning. MCI due to SVD has been hypothesized to be the effect of WM damage and diffusion tensor imaging (DTI) is a well-established technique for *in-vivo* characterization of WM. We propose a machine learning scheme tailored to 1) predicting the impairment in executive functions in patients with MCI and SVD, and 2) examining the brain substrates of this impairment. We employed data from 40 MCI patients with SVD and created feature vectors by averaging mean diffusivity (MD) and fractional anisotropy maps within 50 WM regions of interest. We trained support vector machines (SVMs) with polynomial as well as radial basis function kernels using different DTI-derived features while simultaneously optimizing parameters in leave-one-out nested cross-validation. The best performance was obtained using MD features only and linear kernel SVMs, which were able to distinguish an impaired performance with high sensitivity (72.7%-89.5%), specificity (71.4%-83.3%) and accuracy (77.5%-80.0%). While brain substrates of executive functions are still debated, feature ranking confirm that MD in several WM regions, not limited to the frontal lobes, are truly predictive of executive functions.

Index Terms—MCI, machine learning, DTI, executive functions, neuroimaging

I. INTRODUCTION

S. Ciulli and M. Mascalchi are with the Department of Clinical and Experimental Biomedical Sciences, University of Florence, Florence, Italy.

S. Ciulli and L. Citi are with School of computer science and electronic engineering, University of Essex, Colchester, UK.

S. Ciulli and N. Toschi are with Medical Physics Section, Department of Biomedicine and Prevention, University of Rome "Tor Vergata", Rome, Italy.

E. Salvadori, R. Valenti, A. Poggesi, D. Inzitari and L. Pantoni are with the Department NEUROFARBA, University of Florence, Florence, Italy.

N. Toschi is also with Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging, Boston, MA, USA.

N. Toschi is also with Harvard Medical School, Boston, MA, USA.

*S. Diciotti, is with the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, Via Venezia 52, 47521, Cesena (FC), Italy (corresponding author email: stefano.diciotti@unibo.it).

THE term mild cognitive impairment (MCI) defines a transitional state between normal ageing and dementia and is thought to anticipate dementia. MCI represents an heterogeneous clinical entity that encompasses several possible etiologies. In a number of cases, MCI may be a prodromal phase of Alzheimers disease and carries the risk of conversion to AD-related dementia. However, pre-dementia stages are also acknowledged in cerebrovascular diseases, and cerebral small vessel disease (SVD) is widely recognized as one major cause of cognitive impairment and increases the risk of conversion to vascular dementia [1]. Within the elderly population, MCI is a common condition in patients with diffuse hyperintensities of cerebral white matter (WM) [visible in T2-weighted magnetic resonance (MR) images] and SVD [2]. In MCI due to SVD, cognitive impairment typically includes slowing of motor performance and information processing, impaired attention and impaired executive functions with partial preservation of memory [3], [4], [5], [6], [7]. The most prominent feature of vascular cognitive impairment (VCI) lies in degradation of executive functions, i.e. of processes which supervise the organization and execution of complex behavior, such as response inhibition (selective attention) and set shifting (divided attention) [8]. Still, the definition of exact criteria for the diagnosis of MCI as well as its operationalization are currently debated [9], and a comprehensive sampling of neuropsychological functions across several cognitive domains is necessary to better characterize cognitive impairment in MCI patients. In this context, one of the most widely employed second-level tests, sensitive to speed of cognitive processing and executive functioning and recommended in the current VCI harmonization standards, is the Trail Making Test (TMT) [10], [11]. The TMT is composed of two parts separately targeting phychomotor speed (part A) and divided attention (part B).

Intriguingly, the occurrence of MCI in the context of SVD has been hypothesized to be the effect of WM tract damage, which in turn can result in a *disconnection syndrome* [12], [13], [14] involving mainly transcallosal WM fibers [15]. Diffusion tensor imaging (DTI) is a well-established technique for in-vivo characterization of directional and overall diffusivity of water molecules in WM tracts which has been successfully employed to unveil microstructural changes and their connection to neurophysiological mechanisms underlying various neurodegenerative disorders (for a review, see Goveas

et al. [16]) as well as in elderly subjects with variable cognitive status [14], [17]. In this context, machine learning techniques have recently been identified as promising tools in neuroimaging data analysis (for a review, see Orrù *et al.* [18]). In particular, machine learning schemes can work on a single patient basis in predicting a group label (classification task) or a continuous value (regression task) as the output to a given set of input features. Importantly, compared to conventional group-based analysis, the predictive nature of machine learning strategies (i.e. their generalization properties) is more in line with the aim of clinical diagnosis and prognosis in the single patient.

In the present study, we propose a machine learning scheme tailored to 1) predicting, on a single-patient basis, the impairment in executive functions (as measured by the TMT-B) in patients with MCI and SVD, and 2) examining the brain substrates of this impairment. To this end, we employ DTI-derived features as well as the baseline neuropsychological data from 40 MCI patients with SVD recruited in the RehAtt trial which targets the rehabilitation of attention in patients with MCI and subcortical vascular changes.

II. MATERIALS AND METHODS

A. Subjects

RehAtt is a 3-year prospective, randomized, single-blinded clinical trial currently underway at the VAS-COG clinic of the Careggi University Hospital, Florence, Italy. The study was approved by local ethics committee, each participant gave a written informed consent, and all study procedures are carried out in accordance with the Helsinki Declaration. Study inclusion requires a diagnosis of MCI with an attentional deficit and SVD according to the following criteria: 1) MCI defined according to Winblad *et al.* criteria [19] and operationalized as detailed in [20]; 2) evidence of impairment across attention tests in a specifically developed neuropsychological battery [7] (at least one borderline score among attention/executive functions tests); 3) magnetic resonance imaging (MRI) evidence of subcortical vascular lesions, defined as moderate to severe age-related white matter hyperintensities in T2-weighted fluid attenuated inversion recovery (FLAIR) sequence according to a modified version of the Fazekas scale [2]. Exclusion criteria are: 1) inability or refusal to undergo brain MRI; 2) inability to give an informed consent; 3) age less than 18 years.

In this work we employed neuropsychological and MRI baseline data only. Fourty-four [15 women and 29 men, age 75.3±6.8 years (mean ± standard deviation) and years of education 8.3±4.3, patients with Fazekas grade 2 (moderate leuko-araiosis) = 30 and with grade 3 (severe leuko-araiosis) = 14] MCI patients with SVD were recruited according to the above criteria. At baseline, each patient underwent an extensive clinical, functional and neuropsychological assessment as well as and an MRI examination. Three subjects were excluded after neuropsychological assessment (see Section II-B) and one subject was excluded after MRI examination (see Section II-C). All further analyses were performed on the remaining 40 patients.

B. Neuropsychological assessment

All participants underwent testing through a comprehensive neuropsychological battery including several tests measuring executive functions [7], such as the TMT. During TMT-B the subject is required to draw a line between 8 sequentially numbered circles and 7 alphabetically lettered circles alternating between numbers and letters (i.e. connecting, in sequence, circles 1-A-2-B- etc.). Raw TMT-B scores correspond to the time required to complete the task - lower scores therefore indicate a better performance. TMT-B has a time limit: if the patient does not complete the task within 5 minutes, the examiner stops the administration and records a score of 300 seconds that corresponds to an abnormal performance. In this study, TMT-B was not administered if the patient did not complete the TMT-A (which measures psychomotor speed) in less than 300 seconds or did not remember the correct order of the letters of the alphabet. This condition occurred for 3 patients, who were thus excluded from further analysis.

Subsequently, TMT-B scores were transformed into equivalent scores (ES) [21] calculated through a quantile-based, nonparametric norming method. ES methodology allows to convert age and education adjusted scores into an ordinal 5-point scale (ranging from 0 to 4) [22]. The ES methodology aims to define the outer tolerance limit of the left tail of the adjusted scores, in order to evaluate [with a known risk of error ($<5\%$)] the cut-off which defines the bottom 5% of the population and is commonly interpreted as impaired performance (ES = 0). ES=1 indicates a borderline performance (an adjusted score between the outer and inner confidence limits for the 5th centile of the normal population), and the limits for ES = 2, 3 and 4 (which represent normal performance) are established by partitioning the distribution of adjusted scores between the 5th and the 50th quartiles into equal intervals. When the patient did not complete the task within 300 seconds, raw scores were not transformed in age and education adjusted scores, and an ES of 0 was assigned [7].
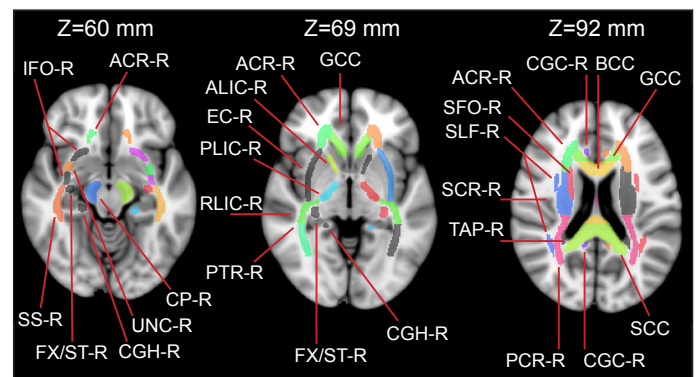

Fig. 1. Three axial views, at Z = 60 mm, Z = 69 mm and Z = 92 mm (MNI coordinates), of the WM atlas ICBM-DTI-81, showing a subset of the 50 ROIs (see Table II), overlaid on the MNI152 T1-weighted standard space. Bilateral ROIs are labeled on the right ("-R") side only.

C. MRI Image Acquisition

All subjects underwent MR examinations on a 1.5 T system (Magnetom Aera, Siemens Medical Solutions, Erlangen,

Germany) equipped with a 20-channel head and neck coil. The gradients had maximum strength of 45 mT/m and slew rate of 200 mT/m/ms. Axial diffusion weighted images were obtained with a single shot echo-planar imaging (EPI) sequence [repetition time (TR) = 10800 ms, echo time (TE) = 87 ms, acquisition matrix = 106×106, slice thickness = 2.4 mm, field of view (FOV) = 256 mm$\times$256 mm, 2.4-mm isotropic voxels, no inter-slice gap, number of excitations (NEX) = 3, GRAPPA acceleration factor = 2, number of slices=67]. Diffusion sensitizing gradients were applied along 30 non-collinear directions using b-value of 0 (b0 image) and 1000 s/mm^2. One patient was excluded after MR examination due to a technical problem during diffusion weighted acquisition.

D. Classification tasks

A dichotomic predictive classification between normal and impaired cognitive performance would have a significant impact in a clinical setting, and the characterization of brain substrates underlying impaired performance is of great interest. We therefore dichotomized the distribution of TMT-B performance (reformulated in terms of ES) in the 40 patients with MCI and SVD under study (see Table I). From a clinical

TABLE I
DISTRIBUTION OF TMT-B PERFORMANCE IN TERMS OF EQUIVALENT
SCORES (ES) OF THE 40 PATIENTS WITH MCI AND SVD.

TMT-B ES	Number of patients
0	19
1	3
2	3
3	5
4	10

point of view, in the dichotomization of cognitive performance, patients with borderline performance (ES=1) can be associated to either the impaired or non-impaired class. We therefore explored two different classification tasks: the "ES 0 vs. 1234" task [i.e. patients with impaired performance (ES=0) vs. patients with borderline or normal performance (ES$\geq$1)] and the "ES 01 vs. 234" task [patients with impaired or borderline performance (ES$\leq$1) vs. patients with normal performance (ES$\geq$2)].

In both classification tasks the sample size was well-balanced between classes ("ES 0 vs. 1234": 19 vs. 21 patients; "ES 01 vs 234": 22 vs. 18 patients) and age, gender and years of education were not statistically different between classes ["ES 0 vs. 1234": p=0.38 using Mann-Whitney U test (null hypothesis of equal medians) for age, p=0.38 using a χ^2 test (null hypothesis of equal proportions) for gender, p=0.06 using Mann-Whitney U test for years of education; "ES 01 vs. 234": p=0.08 using Mann-Whitney U test for age, p=0.97 at χ^2 test for gender, p=0.37 at Mann-Whitney U test for years of education].

E. Image processing

Diffusion weighted images were corrected for head motion and eddy current distortions using FDT (FMRIB's Diffusion

Toolbox; FMRIB, Oxford Center for Functional MRI of the Brain) part of FSL (FMRIB Software Library) version 5.0.2 [23] after which brain tissue was segmented using the brain extraction tool (BET), also part of FSL. Subsequently, the b-matrix was reoriented by applying the rotational part of the affine transformation employed in the eddy/motion-correction step [24]. A tensor model was then fitted to the raw data using a constrained nonlinear least squares procedure implemented in the software package CAMINO [25], and residual non-positive definite tensors (in regions where the nonlinear algorithm failed to converge) were removed by tensor interpolation in the log-euclidean domain [26] using custom-build code developed in Mathematica version 10 (Wolfram Research, Oxfordshire, UK). The following DTI-derived indexes were then computed from the estimated tensor field using the software package DTI-TK [27]: mean diffusivity (MD) and fractional anisotropy (FA).

In order to accurately transform diffusion tensor invariant maps into a standard space for subsequent processing, we employed a high-dimensional nonlinear image registration tool (SyN tool, available within the ANTs package [28] using symmetric diffeomorphic normalization transformation model with neighborhood cross correlation, regular sampling, gradient step size: 0.2, four multi-resolution levels, smoothing sigmas: 3, 2, 1, 0 voxels [fixed image space], shrink factors: 6, 4, 2, 1 voxels [fixed image space]). FA maps were nonlinearly co-registered to the FA template in MNI152 space included in FSL. These subject-specific transformations were subsequently applied to MD maps in order to also transform them into standard MNI152 space.

F. Feature extraction

Two input feature vectors were created by locally averaging MD and FA maps, respectively, in MNI152 within 50 regions of interest (ROIs) defined by the ICBM-DTI-81 atlas [29], [30] (available from http://www.loni.usc.edu/ICBM/Downloads/Downloads DTI-81.shtml). The version of the atlas we used (downloaded on November 1st 2014) did not suffer from the inaccuracies pointed out by Rohlfing [31]. The list of all WM ROIs is detailed in Table II.

In order to explore the differences in discrimination power between different DTI-derived features and their combinations, for each classification task, each feature vector was fed into the classifiers separately as well as pooled MD and FA data were merged before being fed into the classifiers.

G. Machine learning scheme

As base classifier we chose the particular type of kernel method machines known as support vector machines (SVMs), which have shown high classification abilities in neuroimaging applications to several neurological diseases (for a review, see Orrù et al. [18]). We studied the potential of MD and FA feature vectors in discriminating impaired TMT-B performance using a typical SVM classifier (C-SVM) with a) polynomial kernel of degree p, $K(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + 1)^p$, where $\mathbf{x}$ and $\mathbf{x}'$ are the feature vectors of two examples and b) radial basis

function (RBF) kernel, $K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \parallel (\mathbf{x} - \mathbf{x}') \parallel^2)$, where γ parameter is inversely proportional to the width of the RBF [32].

In order to enable proper comparison of performance between different classification strategies, all classification tasks were performed in a nested leave-one-out (LOO) cross-validation (CV) setting which included a loop for simultaneously parameter optimization. Parameter optimization for the polynomial kernel was performed by varying $C = 2^k$ according to $k = \{-10, -9, \ldots, 10\}$ and the polynomial degree p according to $p = \{1, 2, 3\}$. Parameter optimization for the RBF kernel was performed by varying $C = 2^k$ according to $k = \{-10, -9, \ldots, 10\}$ and the RBF γ parameter according to $\gamma = \{0.001, 0.01, 0.1\}$.

In this procedure, for each fold of the outer CV, the training set is used for an inner CV in order to evaluate the performance, in terms of accuracy, of the inner classifier while varying the parameters. Once the parameters with the best performance have been found in the inner CV, the model with those parameters' value is re-trained on the outer training set and tested on the example kept out from the outer CV. This procedure is repeated for each fold of the outer CV. Before each training of the SVM (both in the inner and in the outer CV), each feature is standardized (zero mean and unitary standard deviation) with reference to the training set only. In order to prevent an overestimation of the generalization capabilities of the classifier (sometimes called the "peeking" effect) [33], test-set data were not used in any way during the learning process. Performance was quantified in terms of sensitivity (the ability to correctly detect an impaired performance in patients who have this condition), specificity (the ability to correctly detect a non-impaired performance in patients who have this condition) and accuracy computed on the test set of the outer CV.

III. Results

The machine learning scheme was implemented by wrapping the Java classes included in the Weka Data Mining Software (http://www.cs.waikato.ac.nz/ml/weka, version 3.7.12) [34] in custom scripts developed in Matlab version 8.4 (Math-Works, Natick, Massachusetts, USA) and the Bash language under a Linux environment. The testing platform was a Dell PowerEdge T620 workstation equipped with two 8-core Intel Xeon E5-2640 v2, for a total of 32 CPU threads, and 64 GB RAM. In this work, we employed 10 cores using the Parallel Computing Matlab toolbox with a total computational time of about 1 minute and 50 seconds for each classification task.

Detailed results for all tests are reported in Table III. The best results in terms of sensitivity, specificity and accuracy were observed for SVM with polynomial kernel and MD features only. Specifically, in each fold, the optimization algorithm always selected the linear kernel (polynomial with degree $p = 1$). The optimized classification between patients with ES=0 (impaired performance) and all other ESs showed 89.5% sensitivity, 71.4% specificity and 80.0% accuracy. The optimized classification between patients with ES=0 or 1 (impaired or borderline performance) and all other ESs resulted in

TABLE II
LIST OF WM REGIONS OF INTEREST (ROI) ACCORDING TO THE WM ATLAS ICBM-DTI-81 PARCELLATION. FOR EACH ROI, ITS NUMBER, LABEL AND EXTENDED NAME ARE INDICATED. WM ROIS ARE GROUPED AS TRACTS IN THE BRAINSTEM, COMMISSURAL FIBERS, PROJECTION FIBERS AND ASSOCIATION FIBERS.

ROI number	ROI label	ROI name
Tracts in the brainstem		
1	MCP	Middle cerebellar peduncle
2	PCT	Pontine crossing tract (a part of MCP)
7/8	CST-R/L	Corticospinal tract R/L
9/10	ML-R/L	Medial lemniscus R/L
11/12	ICP-R/L	Inferior cerebellar peduncle R/L
13/14	SCP-R/L	Superior cerebellar peduncle R/L
Commissural fibers		
3	GCC	Genu of corpus callosum
4	BCC	Body of corpus callosum
5	SCC	Splenium of corpus callosum
49/50	TAP-R/L	Tapetum R/L
Projection fibers		
15/16	CP-R/L	Cerebral peduncle R/L
17/18	ALIC-R/L	Anterior limb of internal capsule R/L
19/20	PLIC-R/L	Posterior limb of internal capsule R/L
21/22	RLIC-R/L	Retrolenticular part of internal capsule R/L
23/24	ACR-R/L	Anterior corona radiata R/L
25/26	SCR-R/L	Superior corona radiata R/L
27/28	PCR-R/L	Posterior corona radiata R/L
29/30	PTR-R/L	Posterior thalamic radiation[1] R/L
Association fibers		
6	FX	Fornix (column and body of fornix)
31/32	SS-R/L	Sagittal stratum[2] R/L
33/34	EC-R/L	External capsule R/L
35/36	CGC-R/L	Cingulum (cingulate gyrus) R/L
37/38	CGH-R/L	Cingulum (hippocampus) R/L
39/40	FX/ST-R/L	Fornix (cres) / Stria terminalis[3] R/L
41/42	SLF-R/L	Superior longitudinal fasciculus R/L
43/44	SFO-R/L	Superior fronto-occipital fasciculus[4] R/L
45/46	IFO-R/L	Inferior fronto-occipital fasciculus R/L
47/48	UNC-R/L	Uncinate fasciculus R/L

L, left; R, right.
[1] include optic radiation.
[2] include inferior longitidinal fasciculus and inferior fronto-occipital fasciculus.
[3] can not be resolved with current resolution.
[4] could be a part of anterior internal capsule.

72.7% sensitivity, 83.3% specificity and 77.5% accuracy. The selection of a linear kernel enables a direct interpretation of the trained model: since the SVM weights of the trained model are given in the features space, this allows, if the features are standardized, to exploit the squared modulus of the SVM weights as feature importance. A ranking of all MD features according to their importance in determining the output class for the best SVM classifier was therefore created by averaging the squared modulus of the outer CV SVM weights over all outer CV folds (Figure 2) [35].

Figure 3 (a) shows an anatomical representation of the ROIs corresponding to the top ranks which contributed a total of 80% of the total squared modulus of all SVM weights.

IV. Discussion

In the present investigation, we have demonstrated that impaired performance in the TMT-B, which evaluates executive functions including visual recognition/scanning, and divided attention, can be predicted with high accuracy through an
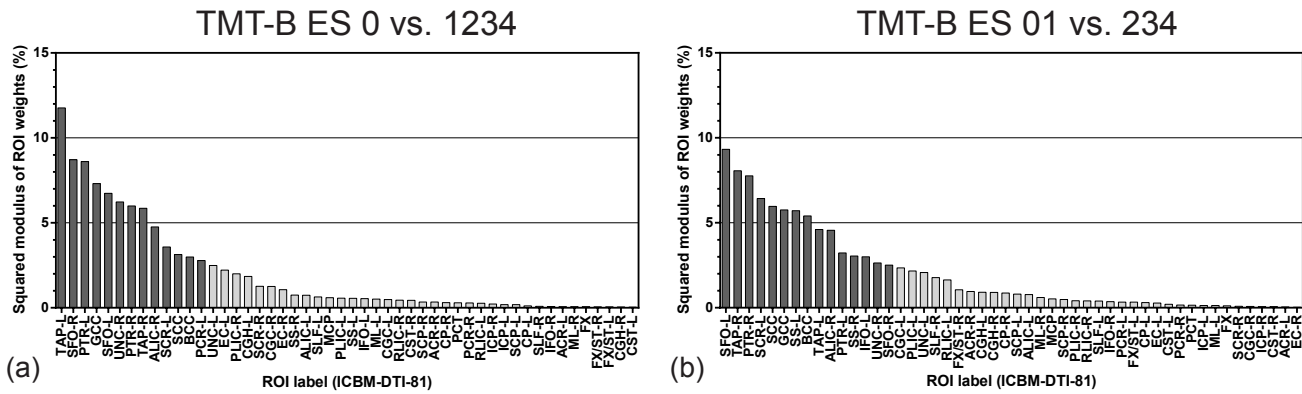
Fig. 2. MD feature ranking for the best SVM classifier (linear kernel) using the squared modulus of ROI weights (in decreasing order) for both classification tasks: patients with TMT-B ES 0 vs. 1234 (a) and patients with TMT-B ES 01 vs. 234 (b). The squared modulus of ROI weights are expressed in percentage with respect to the total squared modulus of all ROIs of the corresponding SVM [35]. Squared moduli are averaged over all trained SVMs of the outer CV; standard deviations were negligible and were not displayed. Bars of highly ranked ROIs contributing to reach the 80% of the total squared modulus are depicted in dark grey, while the remaining bars are in light grey.

TABLE III
CLASSIFICATION RESULTS (GROUP 1 VS. GROUP 2): SENSITIVITY AND SPECIFICITY ARE REFERRED TO THE IDENTIFICATION OF GROUP 1 SUBJECTS

Feature vectors	SVM kernel	Sensitivity (%)	Specificity (%)	Accuracy (%)
TMT-B ES 0 vs. 1234				
MD	polynomial	89.5	71.4	80.0
	RBF	78.9	71.4	75.0
FA	polynomial	29.3	47.6	37.5
	RBF	0.0	38.1	20.0
MD + FA	polynomial	42.1	52.4	47.5
	RBF	0.0	47.6	25.0
TMT-B ES 01 vs. 234				
MD	polynomial	72.7	83.3	77.5
	RBF	63.6	44.4	55.0
FA	polynomial	63.6	0.0	35.0
	RBF	100.0	0.0	55.0
MD + FA	polynomial	63.6	33.3	50.0
	RBF	77.3	0.0	42.5

SVM-based machine learning strategy in MCI patients with SVD using DTI data alone. One of the possible mechanisms underlying this impaired performance may be a disconnection between gray matter regions due to selective disruption of WM tracts, a condition which has been tentatively proposed as an explanation for deficits in higher-order functions [36]. Feature ranking for the best SVM classifiers showed that the ROIs with greatest prediction (i.e. generalization) performance (i.e. which, once ranked, contributed to a total of 80% of the total squared modulus of all SVM weights), include 1) commissural, 2) association and 3) projection fibers.

All commissural fibers [i.e. the corpus callosum, divided into genu (GCC), body (BCC), splenium (SCC) and bilateral tapetum (TAP-R/L)], ranked consistently high in both classification tasks. The involvement of the corpus callosum in higher-order cognitive functions such as attention and executive control has been corroborated by several studies (for a review, see Schulte *et al.* [37]). In particular, the genu has been associated with a frontal lobe contribution to executive

function and attention [38]. This involvement is in agreement with a previous large prospective study among 444 non demented elderly subjects with SVD, in which DTI-derived indexes in the corpus callosum showed significant associations with executive functions [39]. However, in our study this result was obtained in a cross-validation analysis, hence in a single-patient classification framework which substantially reduces the risk of overfitting.

Feature ranking also highlighted the importance of some long-range association fibers in discriminating impaired TMT-B performance. The uncinate fasciculus (UNC) connects the inferior frontal (IF) gyrus with the anterior temporal lobe. The IF gyrus plays an important part in inhibitory control [40], [41] and a more extensive role in executive functions [41]. Also, the bilateral superior fronto-occipital (SFO) fasciculus is functionally associated with executive functions by linking the dorsolateral prefrontal cortices with the parietal lobe [42]. Additionally, the right SFO, which links frontal and parietal lobes, is associated with visuospatial functions [42]. In this context, it should be noted that a recent paper which employed both tractography and WM microdissection put forward the hypothesis that the SFO, whose anatomy is well-documented in monkeys, may not exist in the same form in human brains, hence prompting a reconsideration of the anatomical localization of brain functions classically associated with the SFO [43].

The inferior fronto-occipital fasciculus (IFO), also included in the sagittal striatum (SS), also ranked highly amongst all features. The IFO links temporal and occipital lobes and is functionally associated with visual recognition [42].

Finally, the selected projection fibers include the left superior corona radiata (SCR-L) and the right anterior limb of the internal capsule (ALIC-R), corresponding to the anterior thalamic radiation. These fibers underlie connections to the frontal lobe and are commonly hyperintense in T2-weighted images [36]. The posterior thalamic radiation (PTR) and left posterior corona radiata (PCR-L) were also selected. The PTR has recently been associated with executive dysfunction in traumatic brain injury [44], and the authors demonstrated its

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2016.2537808, IEEE Journal of Biomedical and Health Informatics

6

contribution to executive functions along with frontal lobe structures, associative tracts, and corpus callosum.

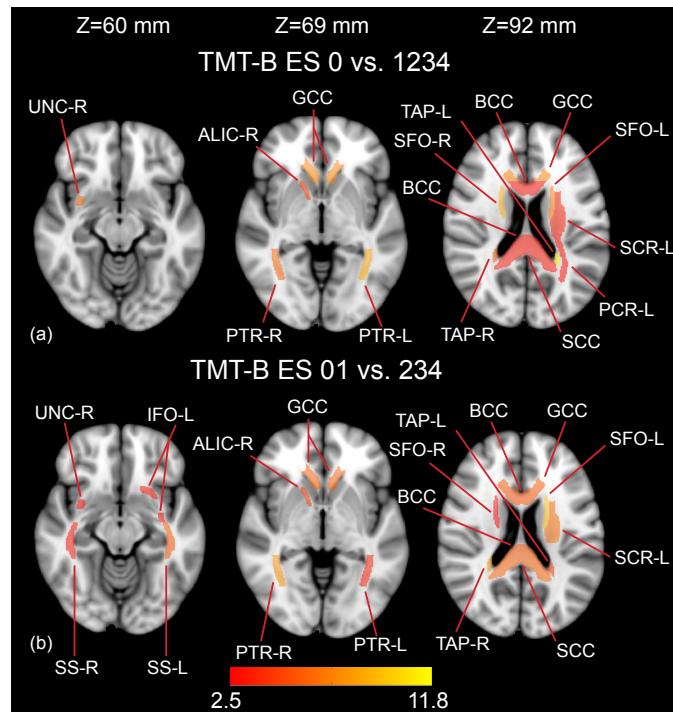

Fig. 3. The anatomical localization (MNI 152 standad space) of each ROI is visible on the underlying anatomical reference for both classification tasks. Each ROI is color-coded according to its importance, i.e. to the squared modulus of ROI weights, expressed in percentage with respect to the total squared modulus of all ROIs of the corresponding SVM [35]. For the sake of clarity, only ROIs that contribute to reach the 80% of the cumulative sum of squared modulus are represented (see Figure 2).

Overall, while brain substrates of executive functions are still debated, in line with our findings it has recently become evident that they are not supported by frontal lobe activity only [8] and, as pointed out by disconnection hypotheses, long range fibers may play an important role. In this context, additional objective neurobiological markers of executive functions would be desirable for monitoring, e.g., cognitive rehabilitation or disease progression.

An emerging body of literature is using predictive machine learning methods with neuroimaging data for classification and regression of disease states, cognitive characteristics or neuropsychological scores [18], [45]. Interestingly, these multivariate methods have also been used to unveil brain substrates underlying the conditions under study which could not be explored using univariate methods [45]. In our study, we chose to focus on designing possible biomarkers able to predict an impaired TMT-B performance, and therefore chose to work with dichotomized, standardized, age- and education-adjusted equivalent scores rather than predicting raw test results through a regression approach.

To the best of our knowledge, this is the first study that adopted machine learning techniques for a true prediction (i.e. performed on patients which remained unseen by the classifier during learning of the model) of impairment in a neuropsychological test in MCI patients with SVD. To this end, we employed mean MD and FA data within 50 WM ROIs

defined by the ICBM-DTI-81 atlas. Previous studies using DTI data in MCI patients with SVD and exploring executive functions employed only traditional correlation-based analysis on the entire data set, hence possibly incurring in overfitting [39], [46], [47], [48].

Differently from a previous (ICBM-DTI-81) atlas-based study in MCI patients with SVD [49], in our work, we used accurate high-dimensional nonlinear co-registrations of FA maps, as implemented in the ANTs package [50]. This ensured better inter-subject co-registration of single-subject maps, hence reducing the possibility of misalignment-related inaccuracies in feature generation.

In this paper, we have implemented a machine learning scheme with nested LOO in order to perform parametric optimization of the SVM classifiers without incurring in the peeking effect [33]. The number of folds of cross-validation is commonly chosen as a trade-off between bias and variance of cross-validation estimator. However, when the data set size is limited, the use of a LOO strategy may be the best choice to maximize the training set size in each fold at the cost of a higher computational time (an issue which we have tackled by using parallel computation) [51], [52], [32], [53]. Also, since the class distributions are well-balanced, we used accuracy as a performance metric in the inner CV loop.

Among MD, FA indexes and their combination, the SVMs trained using only MD data showed the best results. This is in accordance with current literature in which MD shows the stronger and more extensive (as compared to FA) correlations with executive functions (as measured by TMT-B) in elderly healthy subjects [36] and in various neurological diseases, including an inherited WM disease of vascular origin such as cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) [54] and traumatic brain injury [55]. Also, O'Sullivan *et al.* [14] found a significant correlation between MD (but not FA) in normal appearing WM and executive functions (measured through the Wisconsin card sorting test) in a group of patients with ischaemic leukoaraiosis; the same authors did not observe significant correlation between MD (or FA) and TMT.

Given that three patients had borderline performance in TMT-B, we explored two different classification tasks, associating the borderline group either to the impaired or to the non-impaired class. Both analyses yielded similarly high results in terms of accuracy (77.5% - 80.0%) and high sensitivity (72.7% - 89.5%) and specificity (71.4% - 83.3%). While the final ranking of features was not identical between the two analyses, the ROIs with greatest predictive power remained consistent. We defined a cutoff for our top ranks by thresholding the cumulative sum of the squared modulus of SVM weights at 80%. While this value could be further titrated to this specific application, we consider the choice of 80% a conservative strategy, and posit that the sensible neurobiological interpretation of the features selected through this approach as well as their consistency across the two classification tasks corroborates the robustness of this choice.

It is worth noting that parameter optimization in nested cross-validation is not commonly employed in MRI-based patient classification. The optimization procedure selected SVMs

with linear kernel, which are less prone to over-fitting (see [56] for details) in a limited size data set. Also, linear kernel has the major advantage of allowing a direct interpretation of the trained model by using the squared modulus of the SVM weights as feature relevance [35].

The main limitation of our study is the relatively small sample size, which also prevented a reliable employment of a wealth of other, complementary DTI-derived indexes without incurring in overfitting (which is likely to occurr when learning a high-dimensional problem from a small set of examples, i.e. "the curse of dimensionality" [57]). Furthermore, an automatic feature selection algorithm could be employed to automatically generate a reduced set of highly predictive features.

In conclusion, we have demonstrated the abilities of a machine learning scheme in predicting impaired performance in the TMT-B score in MCI patients with SVD using DTI data. Our results confirm that MD values in several WM fibers, not limited to the frontal lobes, are truly predictive of executive functions, and demonstrate the possibility of using machine learning strategies in elucidating brain substrates of cognitive dysfunction in studies with a large sample size.

References

[1] A. Zhou and J. Jia, "Different cognitive profiles between mild cognitive impairment due to cerebral small vessel disease and mild cognitive impairment of alzheimer's disease origin," *Journal of the International Neuropsychological Society : JINS*, vol. 15, pp. 898–905, 2009.

[2] L. Pantoni, A. M. Basile, G. Pracucci, K. Asplund, J. Bogousslavsky, H. Chabriat, T. Erkinjuntti, F. Fazekas, J. M. Ferro, M. Hennerici, J. O'Brien, P. Scheltens, M. C. Visser, L. O. Wahlund, G. Waldemar, A. Wallin, and D. Inzitari, "Impact of age-related cerebral white matter changes on the transition to disability – the LADIS study: rationale, design and methodology," *Neuroepidemiology*, vol. 24, no. 1-2, pp. 51–62, 2005.

[3] P. S. Sachdev, H. Brodaty, and J. C. Looi, "Vascular dementia: diagnosis, management and possible prevention," *Med J Aust*, vol. 170, no. 2, pp. 81–85, 1999.

[4] M. Lamar, C. C. Price, T. Giovannetti, R. Swenson, and D. J. Libon, "The dysexecutive syndrome associated with ischaemic vascular disease and related subcortical neuropathology: a Boston process approach," *Behav Neurol*, vol. 22, no. 1-2, pp. 53–62, 2010.

[5] L. Pantoni, "Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges," *Lancet Neurol*, vol. 9, no. 7, pp. 689–701, 2010.

[6] J. T. O'Brien, T. Erkinjuntti, B. Reisberg, G. Roman, T. Sawada, L. Pantoni, J. V. Bowler, C. Ballard, C. DeCarli, P. B. Gorelick, K. Rockwood, A. Burns, S. Gauthier, and S. T. DeKosky, "Vascular cognitive impairment," *Lancet Neurol*, vol. 2, no. 2, pp. 89–98, 2003.

[7] E. Salvadori, A. Poggesi, G. Pracucci, D. Inzitari, and L. Pantoni, "Development and psychometric properties of a neuropsychological battery for mild cognitive impairment with small vessel disease: The VMCI-Tuscany study," *J Alzheimers Dis*, vol. 43, no. 4, pp. 1313–1323, 2015.

[8] J. A. Alvarez and E. Emory, "Executive function and the frontal lobes: a meta-analytic review." *Neuropsychol Rev*, vol. 16, no. 1, pp. 17–42, 2006.

[9] M. W. Bondi and G. E. Smith, "Mild cognitive impairment: a concept and diagnostic entity in need of input from neuropsychology," *J Int Neuropsychol Soc*, vol. 20, no. 2, pp. 129–134, 2014.

[10] V. Hachinski, C. Iadecola, R. C. Petersen, M. M. Breteler, D. L. Nyenhuis, S. E. Black, W. J. Powers, C. DeCarli, J. G. Merino, R. N. Kalaria, H. V. Vinters, D. M. Holtzman, G. A. Rosenberg, M. Dichgans, J. R. Marler, and G. G. Leblanc, "National institute of neurological disorders and stroke-canadian stroke network vascular cognitive impairment harmonization standards," *Stroke*, vol. 37, no. 9, pp. 2220–2241, 2006.

[11] M. Nowrangi, C. Lyketsos, V. Rao, and C. Munro, "Systematic review of neuroimaging correlates of executive functioning: converging evidence from different clinical populations." *J Neuropsychiatry Clin Neurosci*, vol. 26, no. 2, pp. 114–125, 2014.

[12] M. Catani and D. ffytche, "The rises and falls of disconnection syndromes." *Brain*, vol. 128, no. 10, pp. 2224–2239, 2005.

[13] M. Lamar, M. Catani, C. C. Price, K. M. Heilman, and D. J. Libon, "The impact of region-specific leukoaraiosis on working memory deficits in dementia," *Neuropsychologia*, vol. 46, no. 10, pp. 2597–601, 2008.

[14] M. O'Sullivan, R. G. Morris, B. Huckstep, D. K. Jones, S. C. Williams, and H. S. Markus, "Diffusion tensor MRI correlates with executive dysfunction in patients with ischaemic leukoaraiosis," *J Neurol Neurosurg Psychiatry*, vol. 75, no. 3, pp. 441–447, 2004.

[15] Y. Otsuka, H. Yamauchi, N. Sawamoto, K. Iseki, H. Tomimoto, and H. Fukuyama, "Diffuse tract damage in the hemispheric deep white matter may correlate with global cognitive impairment and callosal atrophy in patients with extensive leukoaraiosis," *AJNR Am J Neuroradiol*, vol. 33, no. 4, pp. 726–732, 2012.

[16] J. Goveas, L. O'Dwyer, M. Mascalchi, M. Cosottini, S. Diciotti, S. De Santis, L. Passamonti, C. Tessa, N. Toschi, and M. Giannelli, "Diffusion-MRI in neurodegenerative disorders." *Magn Reson Imaging*, vol. 33, no. 7, pp. 853–876, 2015.

[17] M. Mascalchi, A. Ginestroni, A. Toschi, A. Poggesi, P. Cecchi, E. Salvadori, C. Tessa, M. Cosottini, N. De Stefano, G. Pracucci, L. Pantoni, D. Inzitari, S. Diciotti, and the VMCI Tuscany investigators, "The burden of microstructural damage modulates cortical activation in elderly subjects with MCI and leuko-araiosis. a DTI and fMRI study," *Hum Brain Mapp*, vol. 35, pp. 819–830, 2014.

[18] G. Orrù, W. Pettersson-Yeo, A. Marquand, G. Sartori, and A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review," *Neurosci Biobehav Rev*, vol. 36, pp. 1140–1152, 2012.

[19] B. Winblad, K. Palmer, M. Kivipelto, V. Jelic, L. Fratiglioni, L. Wahlund, A. Nordberg, L. Backman, M. Albert, O. Almkvist, H. Arai, H. Basun, K. Blennow, M. de Leon, C. DeCarli, T. Erkinjuntti, E. Giacobini, C. Graff, J. Hardy, C. Jack, A. Jorm, K. Ritchie, C. van Duijn, P. Visser, and R. Petersen, "Mild cognitive impairment–beyond controversies, towards a consensus: report of the international working group on mild cognitive impairment," *J Intern Med*, vol. 256, no. 3, pp. 240–246, 2004.

[20] E. Salvadori, A. Poggesi, R. Valenti, G. Pracucci, F. Pescini, M. Pasi, S. Nannucci, S. Marini, A. Del Bene, L. Ciolli, A. Ginestroni, S. Diciotti, G. Orlandi, I. Di Donato, N. De Stefano, M. Cosottini, A. Chiti, A. Federico, M. Dotti, U. Bonuccelli, D. Inzitari, and L. Pantoni, "Operationalizing mild cognitive impairment criteria in small vessel disease: the VMCI-Tuscany study." *Alzheimers Dement*, 2015, in press.

[21] E. Capitani and M. Laiacona, "Composite neuropsychological batteries and demographic correction: Standardization based on equivalent scores, with a review of published data. the italian group for the neuropsychological study of ageing," *J Clin Exp Neuropsychol*, vol. 19, pp. 795–809, 1997.

[22] A. R. Giovagnoli, M. Del Pesce, S. Mascheroni, M. Simoncelli, M. Laiacona, and E. Capitani, "Trail making test: normative values from 287 normal adult controls," *Ital J Neurol Sci*, vol. 17, no. 4, pp. 305–309, 1996.

[23] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews, "Advances in functional and structural mr image analysis and implementation as FSL," *Neuroimage*, vol. 23 Suppl 1, pp. S208–S219, 2004.

[24] A. Leemans and D. K. Jones, "The B-matrix must be rotated when correcting for subject motion in DTI data," *Magn Reson Med*, vol. 61, no. 6, pp. 1336–1349, 2009.

[25] P. A. Cook, Y. Bai, S. Nedjati-Gilani, K. K. Seunarine, M. G. Hall, G. J. Parker, and D. C. Alexander, "Camino: Open-source diffusion-mri reconstruction and processing," in *14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, 2006, p. 2759.

[26] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-euclidean metrics for fast and simple calculus on diffusion tensors," *Magn Reson Med*, vol. 56, no. 2, pp. 411–421, 2006.

[27] H. Zhang, B. B. Avants, P. A. Yushkevich, J. H. Woo, S. Wang, L. F. McCluskey, L. B. Elman, E. R. Melhem, and J. C. Gee, "High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis," *IEEE Trans Med Imaging*, vol. 26, no. 11, pp. 1585–1597, 2007.

[28] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ants similarity metric performance in brain image registration," *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2016.2537808, IEEE Journal of Biomedical and Health Informatics

8

[29] S. Mori, K. Oishi, H. Jiang, L. Jiang, X. Li, and K. Akhter, "Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template," *Neuroimage*, vol. 40, pp. 570–582, 2008.

[30] K. Oishi, K. Zilles, K. Amunts, A. Faria, H. Jiang, and X. Li, "Human brain white matter atlas: identification and assignment of common anatomical structures in superficial white matter," *Neuroimage*, vol. 43, pp. 447–457, 2008.

[31] T. Rohlfing, "Incorrect ICBM-DTI-81 atlas orientation and white matter labels," *Frontiers in neuroscience*, vol. 7, no. 4, 2013.

[32] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

[33] S. Diciotti, S. Ciulli, M. Mascalchi, M. Giannelli, and N. Toschi, "The "peeking" effect in supervised feature selection on diffusion tensor imaging data," *AJNR Am J Neuroradiol*, vol. 34, p. E107, 2013.

[34] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.

[35] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine learning*, vol. 46, pp. 389–422, 2002.

[36] M. O'Sullivan, D. K. Jones, P. E. Summers, R. G. Morris, S. C. Williams, and H. S. Markus, "Evidence for cortical "disconnection" as a mechanism of age-related cognitive decline," *Neurology*, vol. 57, no. 4, pp. 632–638, 2001.

[37] T. Schulte and E. Muller-Oehring, "Contribution of callosal connections to the interhemispheric integration of visuomotor and cognitive processes," *Neuropsychol Rev*, vol. 20, no. 2, pp. 174–190, 2010.

[38] H. Jokinen, C. Ryberg, H. Kalska, R. Ylikoski, E. Rostrup, M. Stegmann, G. Waldemar, S. Madureira, J. Ferro, E. van Straaten, P. Scheltens, F. Barkhof, F. Fazekas, R. Schmidt, G. Carlucci, L. Pantoni, D. Inzitari, T. Erkinjuntti, and LADIS group, "Corpus callosum atrophy is associated with mental slowing and executive deficits in subjects with age-related white matter hyperintensities: the ladis study." *J Neurol Neurosurg Psychiatry*, vol. 78, no. 5, pp. 491–496, 2007.

[39] A. Tuladhar, A. van Norden, K. de Laat, M. Zwiers, E. van Dijk, D. Norris, and F. de Leeuw, "White matter integrity in small vessel disease is related to cognition," *Neuroimage Clin*, vol. 7, pp. 518–524, 2015.

[40] A. Aron, T. Robbins, and R. Poldrack, "Inhibition and the right inferior frontal cortex," *Trends Cogn Sci*, vol. 8, no. 4, pp. 170–7, 2004.

[41] A. Hampshire, S. Chamberlain, M. Monti, J. Duncan, and A. Owen, "The role of the right inferior frontal gyrus: inhibition and attentional control," *Neuroimage*, vol. 50, no. 3, pp. 1313–1319, 2010.

[42] C. Filley, *The Behavioral Neurology of White Matter*. Oxford University Press, USA, 2001.

[43] A. Meola, A. Comert, F. Yeh, L. Stefaneanu, and J. Fernandez-Miranda, "The controversial existence of the human superior fronto-occipital fasciculus: Connectome-based tractographic study with microdissection validation," *Hum Brain Mapp*, In press.

[44] D. Little, M. Kraus, J. Joseph, E. Geary, T. Susmaras, X. Zhou, N. Pliskin, and P. Gorelick, "Thalamic integrity underlies executive dysfunction in traumatic brain injury." *Neurology*, vol. 74, no. 7, pp. 558–564, 2010.

[45] J. R. Cohen, R. F. Asarnow, F. W. Sabb, R. M. Bilder, S. Y. Bookheimer, B. J. Knowlton, and R. A. Poldrack, "Decoding continuous behavioral variables from neuroimaging data," *Frontiers in Neuroscience*, vol. 5, no. 75, 2011.

[46] M. O'Sullivan, P. E. Summers, D. K. Jones, J. M. Jarosz, S. C. Williams, and H. S. Markus, "Normal-appearing white matter in ischemic leukoaraiosis: a diffusion tensor mri study," *Neurology*, vol. 57, no. 12, pp. 2307–2310, 2001.

[47] Q. Xu, Y. Zhou, Y. S. Li, W. W. Cao, Y. Lin, Y. M. Pan, and S. D. Chen, "Diffusion tensor imaging changes correlate with cognition better than conventional MRI findings in patients with subcortical ischemic vascular disease," *Dement Geriatr Cogn Disord*, vol. 30, no. 4, pp. 317–326, 2010.

[48] R. D. Nave, S. Foresti, A. Pratesi, A. Ginestroni, M. Inzitari, E. Salvadori, M. Giannelli, S. Diciotti, D. Inzitari, and M. Mascalchi, "Whole-brain histogram and voxel-based analyses of diffusion tensor imaging in patients with leukoaraiosis. correlation with motor and cognitive impairment," *AJNR AM J Neuroradiol*, vol. 28, pp. 1313–1319, 2007.

[49] L. Lin, Y. Xue, Q. Duan, B. Sun, H. Lin, X. Chen, L. Luo, X. Wei, and Z. Zhang, "Microstructural white matter abnormalities and cognitive dysfunction in subcortical ischemic vascular disease: an atlas-based diffusion tensor analysis study." *J Mol Neurosci*, vol. 56, no. 2, pp. 363–370, 2015.

[50] A. Klein, J. Andersson, B. Ardekani, J. Ashburner, B. Avants, M. Chiang, G. Christensen, D. Collins, J. Gee, P. Hellier, J. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. Woods, J. Mann, and R. Parsey, "Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration," *Neuroimage*, vol. 46, no. 3, pp. 786–802, 2009.

[51] E. Alpaydin, *Introduction to Machine Learning*, 2nd ed. The MIT Press, 2010.

[52] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[53] E. F. I. H. Witten and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Third Edition. Morgan Kaufmann, Burlington, MA, USA, 2011.

[54] M. O'Sullivan, T. Barrick, R. Morris, C. Clark, and H. Markus, "Damage within a network of white matter regions underlies executive dysfunction in cadasil." *Neurology*, vol. 65, no. 10, pp. 1584–1590, 2005.

[55] K. Kinnunen, R. Greenwood, J. Powell, R. Leech, P. Hawkins, V. Bonnelle, M. Patel, S. Counsell, and D. Sharp, "White matter damage and cognitive impairment after traumatic brain injury," *Brain*, vol. 134, no. 2, pp. 449–463, 2011.

[56] K. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Neural Netw*, vol. 12, no. 2, pp. 181–201, 2001.

[57] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc., 1995.