Article:
Fasano, F., Mitolo, M., Gardini, S. et al. (3 more authors) (2018) Combining Structural Magnetic Resonance Imaging and Visuospatial Tests to Classify Mild Cognitive Impairment. Current Alzheimer Research, 15 (3). pp. 237-246. ISSN 1567-2050

https://doi.org/10.2174/1567205014666171030112339

Send Orders for Reprints to reprints@benthamscience.ae

Combining structural magnetic resonance imaging and Visuospatial tests to classify Mild Cognitive Impairment.

Fabrizio Fasano[1,2,*], Micaela Mitolo[3], Simona Gardini[1,4], Annalena Venneri[3,4], Paolo Caffarra[1,5], Francesca Pazzaglia[6]

[1]Neuroscience Department, Parma University, Parma, Italy; [2]IRCCS SDN, Napoli, Italy; [3]IRCCS Fondazione Ospedale San Camillo, Venice, Italy; [4]Department of Neuroscience, University of Sheffield, Sheffield, UK; [5]Centre for Cognitive Disorders, AUSL, Parma, Italy; [6]Department of General Psychology, University of Padua, Padua, Italy.

Abstract: Recently, efforts have been made to combine complementary perspectives in the assessment of Alzheimer type dementia. Of particular interest is the definition of the fingerprints of an early stage of the disease known as Mild Cognitive Impairment or prodromal Alzheimer's Disease. Machine learning approaches have been shown to be extremely suitable for the implementation of such a combination. In the present pilot study we combined the machine learning approach with structural magnetic resonance imaging and cognitive test assessments to classify a small cohort of 11 healthy participants and 11 patients experiencing Mild Cognitive Impairment. Cognitive assessment included a battery of standardized tests and a battery of experimental visuospatial memory tests. Correct classification was achieved in 100% of the participants, suggesting that the combination of neuroimaging with more complex cognitive tests is suitable for early detection of Alzheimer Disease. In particular, the results highlighted the importance of the experimental visuospatial memory test battery in the efficiency of classification, suggesting that the high-level brain computational framework underpinning the participant's performance in these ecological tests may represent a "natural filter" in the exploration of cognitive patterns of information able to identify early signs of the disease.

Keywords: Visuospatial Memory, Spatial Abilities, Support Vector Machine, Magnetic Resonance Imaging, Mild Cognitive Impairment, Classification.

1. INTRODUCTION

1.1. Alzheimer's disease and Mild Cognitive Impairment

Recent pharmaceutical trials have demonstrated that slowing or reversing pathology in Alzheimer's disease (AD) is likely to be possible only in the earliest stages of disease, perhaps even before significant symptoms develop [1]. The development and use of diagnostic tools for the early detection of AD, therefore, might be helpful for clinicians. An ideal diagnostic tool must be sensitive to the earliest disease changes, not too expensive and should also be able to differentiate the very early stage of AD from normal ageing. A long-standing literature has addressed the question of what deficits can be taken as early predictors of AD.

AD is the most frequent cause of dementia. Cognitive impairments commonly begin with difficulties in remembering recent events as well as semantic memory impairments [2]. Although visuospatial deficits have also been described in Mild Cognitive Impairment (MCI), that may represent the prodromal phase of AD, this multi-faceted domain of cognition is not extensively assessed and fully investigated in clinical practice [3].

The term MCI was proposed by Petersen and colleagues [4] to define a nosological entity referring to elderly persons with mild cognitive deficits without dementia. MCI encompasses patients with and without memory impairment. Of those with memory loss, some have memory impairment as their only deficit (single domain amnestic MCI) whereas others have impairments of memory plus changes in other cognitive domains (multiple amnestic MCI). Of those without any memory loss, some patients have deficits only in one domain (single non-amnestic MCI), such as visuospatial abilities, executive functions, praxis or language, whereas others have deficits in several domains, excluding memory (multiple non-amnestic MCI) [5].

Some studies have investigated the influence of normal and pathological ageing on visuospatial performance showing an age-related decline in normal ageing, which is more severe and earlier in pathological conditions such as AD or Dementia with Lewy Bodies [6-8]. These patients frequently have difficulties with spatial orientation in everyday life and may fail to find their way in unfamiliar environments when facing entirely new spatial settings [9].

1.2. Visuospatial abilities

Spatial ability is defined as the skill in representing, transforming, generating, and recalling symbolic, nonlinguistic information [10]. It is not considered a unitary function, and, in fact, the plural – spatial abilities – is frequently used. Spatial abilities include a wide range of processes and components, such as spatial perception, visualization and mental rotation [10,11], and these skills are essential in spatial navigation, even if the two functions can be considered as partially distinct [12,13]. Spatial navigation is the ability to find and follow a route from one place to another [14] and relies onto two different, egocentric and allocentric, strategies, respectively related to distinct spatial representations. The egocentric strategy is based on route spatial representation, a body-centered representation, where

the position of landmarks is encoded with respect to the position of the subject; the allocentric strategy is associated with a survey, world-cantered representation, where locations and distances among landmarks are encoded and maintained independently of the position of the subject. Different cortical areas serve the execution of spatial tasks and spatial navigation [15]. In particular it seems that different memory systems are involved in small and large scale environments [16,17], in object locations and navigational memory [18], map learning and navigation [19], and in navigation using allocentric or egocentric strategies [20]. A recent meta-analysis [21] showed an age related decrease in the performance on standardized test of spatial perception, spatial visualization and mental rotation. Spatial navigation is subjected to a decrement with ageing, particularly when allocentric strategies are required [22], and in unfamiliar contexts [23]. The decrement is more severe in pathological ageing, and topographical disorientation is described as one of the distinctive signs of AD [9]. Moreover, several authors consider spatial navigation and spatial ability deficits as first signs of incipient AD in MCI patients [2,20].

The assessment of spatial and navigation abilities is also a valid tool for discriminating between healthy ageing and MCI individuals. Mitolo et al (2013) [3] administered a battery of spatial tasks to a group of MCI patients and a group of healthy controls and they found that the discriminative power of the spatial tasks in identifying MCI was high and superior in comparison to the other neuropsychological tests frequently used in clinical settings, including the Stroop test, Wisconsin card sorting test, and category fluency test. Moreover, the same authors found distinct patterns of significant correlations between grey matter volume and performance on spatial tests in the MCI group and in the healthy control [3].

Overall, the picture described so far highlights the importance of adding spatial tests to neuropsychological batteries, and of identifying reliable tools for assessing spatial and navigation ability in clinical settings. In this context Mitolo et al. (2015) [24] have devised a battery of spatial tests and questionnaires designed to assess spatial ability in normal and pathological ageing. The battery is suitable for use in clinical settings and comprised tests for the assessment of three different abilities within the spatial domain: route learning, map learning and objects location. Further, it includes self-report scales, designed to obtain measures of navigation abilities, as they appear in every-day life, with questionnaires evaluating sense of direction, spatial anxiety, self-efficacy, and attitude toward spatial tasks. So far, several studies have shown that subjective measures of spatial efficiency are reliable indicators of actual performance in navigation and orienting task [25-28]. The battery, administered to a sample of 90 healthy aged participants, appeared to have good internal reliability, and, as expected, to tap different abilities within the spatial domain. Furthermore, significant correlations were found between the spatial tests and the questionnaires, suggesting a relationship between the abilities measured by the spatial tests and spatial performance in every-day life (Mitolo et al. 2015) [24].

A long-standing literature has addressed the question of what deficits can be taken as early predictors of AD. So far, the greatest attention has been paid to verbally-mediated memory disorders, specifically episodic and semantic memory that are traditionally considered the earliest and most severe deficits [29]. Visuospatial deficits, even in the early stages of AD, have long been recognized but have not been extensively studied [30].

AD causes circumscribed atrophy in distinct neural networks, and they impact visuospatial abilties in different ways. These effects depend on topographic patterns of brain pathology, and it is, therefore, important to use anatomically-specific methods for diagnosis and for monitoring disease progression [31].

1.3. Searching for Biomarkers: magnetic resonance imaging

At present, several types of biomarkers have been found to be sufficiently sensitive to discriminate the MCI status from normal ageing, including brain atrophy measured using magnetic resonance imaging (MRI) [32], hypometabolism measured by FDG-PET imaging [33], and quantification of specific proteins measured in the cerebrospinal fluid [34].

MRI is considered a promising technique for obtaining new feasible biomarkers. The most popular method to analyse structural MRI data is voxel based morphometry (VBM) [35], in which statistical significance in between-group analyses is assessed voxel by voxel on the maps derived from the comparison of high resolution structural T1-weighted (T1w) images. A big effort in the direction of deriving relevant structural diagnostic features based on state of the art structural MRI acquisitions is represented by the Alzheimer's Disease Neuroimaging Initiative (ADNI, http://adni.loni.usc.edu). The ADNI data repository includes high resolution T1w images collected from 400 subjects experiencing MCI. Based on these data, numerous studies have identified regional grey matter atrophy patterns associated with MCI and AD, by computing morphological indices as modulated grey matter (modGM) [36] and cortical thickness [37]. Nevertheless the translational shift from group studies to single subject approach remains still unaccomplished.

1.4. The machine learning approach

Recently, in the search for operative biomarkers, the analysis of MRI data has benefitted from machine learning based techniques [38-41]. If the first approach is mainly focused on finding group differences, and to explain them, the second one is much more focused on single subject assessment (classification). For a comprehensive review about the single subject approach, benefits and pitfalls, see [42]. Differently from univariate methodologies like VBM, machine learning approaches are based on multivariate analysis of data. The rational of a machine learning approach is to search for patterns of similarities/dissimilarities contained in the entire data set (e.g. the whole modGM map or the entire cognitive test scores set) of subjects classified as belonging to same/different groups (e.g. HC subjects and patients suffering from MCI). Once the pattern is "learned" by the "machine", the result is generalised to predict which group a new unidentified subject belongs to. In 2014-15 the computer-aided diagnosis of dementia challenge (CADDementia Challenge; http://caddementia.grand-challenge.org) took place, in which the performance of 29 newly developed algorithms was compared. The comparison was implemented

on a structural MRI dataset composed of T1w volumes acquired in more than 300 subjects. The goal was to classify subjects in three possible groups, healthy volunteers, patients with MCI and patients with AD. The best results, allowing 63% correct classification [43], were still far from the identification of a golden standard algorithm, but the challenge clearly evidenced that the best performance was obtained by combining different types of features.

1.5. Support vector machine and multi kernel learning.

In this study we combine the visuospatial ecological tasks described in Mitolo et al [3,24], standard cognitive tests and MRI derived modGM to test the accuracy of prediction of MCI diagnosis by using a machine learning approach based on support vector machine (SVM)[44]. SVM has been shown to be a feasible method to analyze large feature small sample data sets [45]. In fact a standard modGM map contains a number of voxels (features) of the order of 100.000 while a classical MRI clinical study only includes few tens of subjects. To find the solution to the learning problem the SVM approach associates a mathematical entity called Kernel to the whole data set (features set) (See Supplementary Materials for more details on SVM).

Although in principle machine learning methods allow better accuracy in classification than univariate methods, differently from univariate methods they miss in the identification of those regions/voxels responsible for the classification performance. In fact their principal goal is not to analyse statistical differences between single regions/voxels but to recognise the existence of some patterns, not necessarily exploring within them. Among the methods proposed to overcome this limitation a new one makes use of the so-called Multi Kernel Learning (MKL) approach [46], an extension of the single Kernel approach (see Supplementary Materials). The MKL approach intrinsically offers the opportunity to refine the results by questioning the relevance/irrelevance of different feature subsets. The whole features set is divided in different subsets, and a different Kernel is associated to each subset. This can be useful to include, in the same optimisation problem, features of different nature (for example, as in the case of the present study, neuroimaging data and cognitive test scores). The method can be also useful to select the most important subsets by assigning relative weights to the Kernels associated with each subset. The algorithm can be forced to assign a zero weight to irrelevant Kernels (i.e. to retain only a small portion of feature subsets).

The last potentiality was used in this study in order to reduce the subset of features among the whole set we analysed and to try to determine the driving feature sets in the prediction performance.

In the present study the PRoNTo software [47-49] was used to implement the SVM MKL approach. PRoNTo is a user friendly tool recently developed at University College of London which allows MKL analysis in addition to standard classification optimization procedures. The MKL implementation in PRoNTo is grounded on the SimpleMKL algorithm developed by Alain Rakotomamonjy and co-workers at CNRS/Universite´ de Technologie de Compie`gne (France) [50].

1.6. Aims

Given the evidence reported in previous studies [3, 24], in the present study we aimed to combine the data derived from the neuropsychological tests battery with those derived from structural MRI, to verify their discriminant power in identifying MCI patients. We expected that the combination of behavioural and MRI data would have more discriminant power than individual measures alone, and that the more complex experimental visuospatial tests would be better classifiers than standard cognitive measures

*Address correspondence to this author at the Department of Neuroscience of Parma University, Via Volturno 39/E, 43100, Parma (PR), Italy; Tel/Fax: +39-0521-903963; E-mail: fabrizio.fasano@nemo.unipr.it

2. MATERIALS AND METHOD

2.1. Participants

Eleven patients with MCI (73 ± 6 years, 5 females, 6 males) and eleven healthy control (HC) participants (68 ± 5 years, 7 females, 4 males) were included in this study (see Table 1). All participants had a complete neuropsychological assessment using standardised tests [3]. The MCI patients were diagnosed according to Petersen's criteria [51], which include: subjective or proxy cognitive complaint; objective impairment in memory and/or in other cognitive domains; relatively intact functional ability. The HC group scored in the normal range on all tasks.

In detail, the MCI patients were all of the amnestic type: 5 patients were amnestic single-domain and 6 patients were amnestic multidomain. HC participants were in good general physical and cognitive health and had a Mini Mental State Evaluation (MMSE) [52] score higher than 27. Participants with neurological or mood disorders were excluded.

All participants were also assessed with a battery of experimental visuospatial memory tasks, self-rating spatial questionnaires, and high resolution 3D MRI T1w brain scanning.

The study was approved by the local ethics committee, and all participants provided written informed consent before study initiation.

2.2. Cognitive assessment

All participants underwent a battery of standardised neuropsychological tests and a battery of new experimental tests, as described in two previous studies [3,24]. The standardised neuropsychological battery included the following tests: Mini Mental State Examination (MMSE); IQ (intelligence quotient) tests: Raven's Coloured Progressive Matrices, TIB (Test di intelligenza breve), Vocabulary test (WAIS sub-test); verbal memory tests: Prose Memory Test, Rey Auditory Verbal Learning Test, Verbal semantic encoding and recognition, Digit Span forward; language tests: Boston Naming Test, Verbal Associative Fluency Test, Category Words Fluency Test; executive function tests: Stroop test, Wisconsin Card Sorting Test (WCST), Tower of

London, Dual task, Multiple feature target cancellation, Digit Span backward; visuospatial and visual perception tests: Corsi Block Tapping test, Rey Osterrieth Complex Figure, Visuospatial supra span, Visual Object and Space Perception battery (VOSP), Mental Rotation test.

The experimental battery included object location recognition, map learning and route learning, self-administered questionnaires evaluating sense of direction, attitude toward environment orientation tasks, spatial orientation anxiety and self-efficacy toward environment orientation tasks. A total of 38 scores were recorded for each participant from the standardised test battery, and a total of 14 scores from the experimental one, 10 objective test scores and 4 self-evaluative test scores (see Table 1 for the experimental test battery scores).

2.2.1. Object location task

This task assesses object recognition, recall and location skills. It is divided into two sub-tests that involve recognising, recalling and locating some objects in a picture. In the recognition sub-test, participants are shown six objects (elephant, lamp, slipper, guitar, bottle and hat) and asked to memorise them. Then, for each object, participants are asked to recognise the target among three options. The total number of items correctly recognised is recorded. The second sub-test, which is assumed to require spatial memory for locations, involves memorising a picture (42 cm X 30 cm) of a room containing twelve objects (table, cat, chessboard, guitar, etc.) for 1 min, then recalling all the objects and locating them in a picture of an empty room immediately afterwards, by writing the name of the object in its location. The resulting score corresponds to the number of objects recalled and correctly located.

2.2.2. Map learning task

This task was developed to assess the respondent's ability to memorise a map. It involves remembering the names and locations of eight landmarks on a map (21 cm X 30 cm), i.e., pharmacy, school, cinema, hospital, bakery, park, bar and dairy. Immediately after being exposed to the map for 5 min, participants have to write the names of the landmarks in the right position on a blank map. The learning phase (and subsequent recall and localisation phase) is repeated, and we calculated the number of landmarks recalled and located in the right position after the first and second learning trials.

2.2.3. Route learning task

This task assesses memory for routes. Similar to the procedure used in a previous work [17], this task involves memorising routes within a matrix of 25 squares (5 X 5) located on the floor; each square is 15 X 15 cm, and the distance between the squares is 30 cm. The task is divided into three sub-tests, and, for each one, participants have to remember increasingly longer routes. In the first sub-test (route learning from action), the participant first learns the routes by stepping on the sequence of squares with the examiner and is asked to repeat each route immediately afterwards. In the second sub-test (route learning from vision), the participant is asked to watch while the examiner covers a route and to repeat it immediately afterwards. In the third sub-test (route learning from a map), participants learn each route on a map and then reproduce it on the matrix. Each sub-test begins with a route of just two segments, and then the routes become gradually longer (including three segments, four segments, and so on). For our study, two sequences were used for each length, and the test came to an end when a participant was unable to reproduce both sequences of the same length. The longest route that a participant reproduced correctly in at least one of the two trials was taken as the score for each sub-test.

Table 1. Demographic data, mean (and standard deviation) spatial scores in HC and MCI groups. The Spatial Questionnaires and Visuospatial tests scores are indicated respectively as "CTS14-self evaluative" and "CTS14-objective" in the article.

	HC		MCI	
	Mean (SD)		Mean (SD)	
Demographic data				
Age (years)	68	(5)	73	(6)
Education (years)	9	(5)	8	(3)
MMSE	30	(1)	26	(3)
Spatial questionnaires				
Spatial Anxiety Scale	14	(4)	16	(5)
Spatial Attitude Scale	22	(4)	18	(4)
Self-Efficacy Scale	18	(5)	13	(3)
Sense of Direction Scale	55	(8)	43	(12)
Visuospatial tests				
Objects recognition	5	(1)	3	(1)
Objects recall	10	(1)	6	(3)

Objects location	11	(1)	7	(3)
Map recall 1	7	(1)	3	(2)
Map location 1	6	(1)	2	(1)
Map recall 2	8	(1)	4	(2)
Map location 2	8	(1)	3	(3)
Learning from action	5	(1)	4	(1)
Learning from vision	6	(1)	4	(1)
Learning from map	6	(1)	4	(1)

2.3 MRI

2.3.1. Acquisition

All participants underwent the same MRI protocol on a 3 Tesla General Electric MR750 scanner, equipped with 8-channel phased array receiver head coil. The protocol included an Inversion Recovery Prepared Fast Spoiled Gradient Recall 3D high resolution T1w sequence (0.9 × 0.9 × 0.9 mm^3, TR/TE = 9.7/4 ms).

2.3.2. Modulated grey matter maps

For each subject the original (subject native space) T1w images were skull stripped and segmented to obtain the grey matter (GM) maps in native space (nsGM). The skull stripped T1w images were diffeomorphically registered to a template previously created in our laboratory from 10 healthy volunteers (5 females and 5 males) and 10 age and gender matched patients with MCI [53]. The resulting template space T1w images were in turn segmented to obtain the template space GM maps, and the last ones were averaged to obtain a study population GM template. The diffeomorphic registration between the nsGM and the study population GM template were computed and applied to the (weakly smoothed, 1mm full width half maximum) nsGM images, obtaining for each participant its GM to template grey matter probability map. For each participant the Jacobian of the diffeomorphic transformation was multiplied for the GM to template probability map and logarithmic scaled obtaining the modGM maps. All modGM maps were under-sampled from the original 0.9 x 0.9 x 0.9 mm^3 to 2 x 2 x 2 mm^3.

2.3.3. Regions of interest

On the population template 81 regions of interest were extracted by applying FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) [54,55] automated parcellation procedure and a study Atlas was created by labelling the region of interests with integers values ranging 1-81 (see Figure 1 and Supplementary Materials for label names and voxel counts for each region). A study mask including all the regions was also created. The study Atlas and the study mask were under-sampled to 2 x 2 x 2 mm^3 to match the modGM maps.

2.4. Age effect correction

Age was found to be significantly different between the two groups (68±5y for the MCI group and 73±6y for the HC group, p=0.019). Differently from univariate approaches, multivariate analysis does not allow covariates integration in the classification procedures. A way to overcome the problem of nuisance variables (e.g. age) in multivariate procedures is to perform a prior general linear model (GLM) analysis to compute the nuisance variable effect, and then to correct the original features by applying the final classification analysis to the GLM corrected features [56]. Following this procedure, we corrected for age effect the modGM maps and the cognitive tests scores. More details are provided in Supplementary Materials.

The results obtained with the corrected data were very similar to the ones obtained without correction (see Supplementary Materials), suggesting that the small group difference in age had no significant influence on the results. In the following sections we report the results obtained by using as source data for our analysis (features) the original (uncorrected) data. Nevertheless, in Table 2 the prediction accuracy of GLM age corrected data for the whole feature set is also reported.

2.5. Multivariate analysis

2.5.1. The features

The 38 test scores (CTS38) of the standard battery, the 14 test scores (CTS14) of the visuospatial battery and the 77715 modGM values in the voxels included in the regions selected for the study Atlas were used as input features in a multivariate analysis to assess the classification power of the combination of cognitive tests and structural MRI.

2.5.2. Multi-kernel learning classification procedure

An L2-regularized L1-loss SVM MKL approach [50] was applied, as implemented in the PRoNTo software [47-49]. In solving the MKL optimization problem, one linear kernel was assigned to each brain region in the Atlas (81 kernels), one linear kernel was assigned to the standard test battery, one linear kernel was assigned to the self-evaluation spatial tests and one linear kernel was assigned to the objective visuospatial tests, for a total of 84 kernels (see Supplementary Materials). A leave-one-out cross validation was used to determine the accuracy of classification, and a nested 5-fold cross validation was used to choose the optimal soft-margin parameter in SVM (by exponential spanning [0.1, 1, 10]). We implemented a Matlab (The MathWorks Inc., Natick, MA, 2000) home made script to integrate the cognitive tests results with the neuroimaging (modGM) ones in the PRoNTo framework. Due to the high variability of features among brain regions and the reduced number of features in the cognitive tests, kernels were normalized. The significance

level of classification accuracy was assessed by permutation tests (1000 repetitions), with significance threshold set at p=0.01.

2.5.3 Best Kernels selection

As discussed in the introduction, the PRoNTo software MKL approach intrinsically offers the opportunity to refine the results by questioning the relevance/irrelevance of each feature set, i.e. by optimizing the Kernel weights and forcing (L1-loss) the assignation of zero weights to irrelevant Kernels. We applied this analysis to extract the relevant features sets (i.e. brain regions and/or cognitive test sets) from the whole features set.

3. RESULTS AND DISCUSSIONS:

3.1. Evaluation of prediction accuracy

Preliminarily, nested 5-fold cross validation revealed that the best soft-margin parameter was 1. Classification accuracy of the whole features set (including modGM in the 81 brain regions and the whole battery of cognitive tests CTS14 and CTS38) was found to be 100% (sensitivity = specificity = 1). For all the tests the prediction accuracy estimation, as determined by the permutation test, was found to be significant at p=0.01 level (Table 2, first row). The classification scores obtained with the exclusion of the CTS38 was still found to be 100%. On the contrary the exclusion of CTS14 or of modGM scores worsened the accuracy. See Table 2 for more details. The results imply that the CTS14 scores were determinant to improve the classification accuracy.

3.2. Feature sets contribution results

The MKL ranking procedure clearly identified a group of 10 scores of the objective visuospatial memory tests (CTS14 objective) as the driving feature set for classification, with the associated Kernel contributing for the 63,62% to the model prediction. The other regions found to have non zero weight associated Kernel were: left pars opercularis (12,27%), right entorhinal cortex (8,73%), left pericalcarine (8,14%), Brain Stem (3,97%), left Nucleus Pallidum (1,32%), right Putamen (0,57%). right Amygdala (0,57%), right Putamen (0,56), right Caudate Nucleus (0,29%), right Thalamus (0,15%), right caudal middle frontal cortex (0,13%), right Accumbens area (0,09%), left pars orbitalis (0,07%), CTS14 self-evaluative (0,04%), left transverse temporal cortex (0,04%). See Figure 1. The contribution, the exponential rankings and feature set size (number of features) of each region/cognitive test set are reported in the Supplementary Materials.

As already suggested by the classification accuracies results obtained by combining different feature sets (see previous paragraph and Table 2), the visuospatial tests appeared to clearly drive the performance of the classifier. See Discussion for a more detailed analysis of the implications of the results and for an interpretation which highlights the technical approach benefits and pitfalls, and the necessary interpretative caution given the small size cohort of participants.

Table 2. For different combinations of feature sub-sets, the weight (%) of the feature sub-set in the prediction, the accuracy of prediction, true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) as well as sensitivity and specificity are reported. The estimation of balanced significance of the prediction score as assessed by the permutation test is also reported (BA p-value). Where for a given combination of feature sub-sets one sub-set is absent, a dash sign (-) is reported in the table. Bold is used to highlight the weight of the most contributive feature set. The GLM age corrected data results for the whole feature set are also reported.

modGM	CTS38	CTS14 objective	CTS14 self eval.	TP	TN	FP	FN	sensitivity	specificity	BA p-value
				Original data						
36,34%	0%	**63,62%**	0,04%	11	11	0	0	1,00	1,00	0.001
100,00%	-	-	-	9	11	2	0	1,00	0,85	0.006
-	0,59%	**99,35%**	0,06%	11	9	0	2	0,85	1,00	0,001
71,48%	28,52%	-	-	9	10	2	1	0,90	0,83	0.001
36,67%	-	**63,30%**	0,03%	11	11	0	0	1,00	1,00	0.001
				GLM age corrected data						
33,69%	1,03%	**65,27%**	0,01%	10	10	1	1	0.91	0.91	0.004

3.3. Discussion

In the Introduction an important concept related to the use of machine learning approaches in the "individual level assessment challenge" has been extensively highlighted: when applied to multimodal investigations, a multivariate analysis causes a paradigm shift from "using multiple modalities to interpret individual specific differences between two classes" (e.g. different regional grey matter density in patients and HC participants) to "using multiple modalities to identify the presence of patterns able to differentiate the two classes". It is important to restate here that, due to this reason, the new paradigm may determine a loss of interpretability of results, although it may allow, in principle, gains in sensitivity and specificity.

This pilot study confirmed these expectations. The results, while obtained in a small cohort of participants, suggest that performance scores in objective visuospatial memory tests and the regional values of modGM (an index assessing GM integrity) may contain patterns of information which, added to the structural information provided by MRI, allow a better classification between the early stages of cognitive decline and healthy ageing than standard cognitive tests.

Looking at Table 2 we can see that the single modalities contributed differently in sensitivity and specificity when taken individually. When used together their combination resulted in increased classification performance. Thus the present results support the use of the combination of proper cognitive explorations and GM integrity features to detect the early stages of abnormal cognitive decline. In particular the present investigation suggested the specific advantage that may be offered to the classification procedure by adding a "complex objective measure" as the one represented by the visuospatial abilities to the feature set.

In the current study estimation of Kernel weights in a MKL approach was suggested as a way to overcome the intrinsic lack of interpretability of any encoding method results [42,57,58], and to try to rank the contribution of the single scores set used in the study (brain regions assessed by MRI and cognitive tests). A clear indication of the driving role of visuospatial memory tests is shown by such analysis. In fact their contribution to the classification power is always higher than that of MRI and standard cognitive tests feature sets. This is clearly depicted by the results in Table 2 and Figure 1, especially looking at the third row of Table 2, showing the classification accuracy of the sub-sets combination including only CTS38, CTS14 objective and CTS14 self-evaluative (i.e. excluding MRI data). The results are in accordance with our expectations and with the findings reported by Mitolo et al. (2013). It is also important to recognise that self-evaluative visuospatial test scores (possibly biased by patients' poor awareness of their impaired abilities) were not as effective as objective measures in driving the process of classification.

For MRI, the analysis of Kernel weights reveals a picture of brain regions that is only partially in agreement with expectations based on previous knowledge from studies of MCI converters vs healthy subjects [59-61]. This is most likely due to the high variability of results through the different cross validation folds (see following discussion and Supplementary Material on SVM for more details), to the small cohort of participants and to the overshadowing of some Kernels (regions) by others Kernels (regions) containing the same patterns of information [49]. This latter is a well-known issue in machine learning neuroscientific applications.

A more in-depth analysis on these issues, as allowed by the Kernels method and described with more technical details in Supplementary Materials, supports the idea that the exclusion of those areas from the set of the best performing classifiers may be related to high inter-folds variability (probably due in part to the small data sample) and/or to the correlation of the information that they contain with the information contained in other areas (among the best performing ones, e.g. right Amygdala, right Caudate Nucleus, right Putamen and right Thalamus, right enthorinal cortex). Most likely the "CTS14 objective" sub-set of features itself might also share the same pattern of information of some excluded areas, denoting a correlation between visuospatial test performance and atrophy in some brain regions [3]. This last hypothesis is also supported by the results obtained on the true brain regions only (i.e. by excluding CTS14 from analysis). In fact in that case most of those areas that were found to have zero weights when GM areas were analysed in combination with the visuospatial tests gain nonzero weights (see Figure 1 and Supplementary Materials).

One obvious question which needs addressing is why visuospatial objective test scores have such high classificatory power in this sample. In addition to their possible high sensitivity to early subtle visuospatial breakdown due to the insidious brain region specific effects of AD, it is possible that these tests might have higher sensitivity to subtle cognitive decline because of their potential load on other cognitive abilities. These tests, due to their ecological value and their complexity, may require the involvement of concurrent abilities (e.g. executive functions, memory, perception, semantic processing) as well as spatial orientation. We suggest that this multi-componential involvement would make them more sensitive to the breakdown of function of multiple brain networks even in a heterogeneous sample of different MCI phenotypes. In this case, these tests would not operate, therefore, as a specific biomarker of AD, but rather as a marker of a diseased brain. On the other hand, it is also equally possible that their high sensitivity may reflect specific breakdowns in fronto-parietal and default mode networks caused by the insidious regional encroachment of AD pathology over a long period of time. Finally, in relation to the machine learning procedure adopted in this investigation, we might argue that visuospatial tests may have represented a good natural "feature selection filter". In fact, the very first important feature selection occurs in the feature choice phase. The extensive neuropsychological clinical experience rounding the tests definition may be seen as a complex ecological observation-based learning procedure, able to select those tests best suited to evaluate the high-level brain computational framework underpinning the participant disease.

In the present article we reported the results obtained with age uncorrected data. While in principle the correction procedure may help to exclude the age related differences from being included together with the pathology related ones, the GLM based correction procedure is quite invasive and may introduce errors, especially for those features in which there is not a strong correlation between data and age. For that reason we decided to report the results based on original data

in the main body of this article and interested reader are referred to the supplementary material for the analyses of corrected data. Nevertheless we are aware of the limitation inherent in the use of unbalanced data. In fact, for those variables in which the unbalancing is not peculiar to the true nature of the phenomenon under investigation (i.e. if we are not interested to include them in the prediction/classification procedure), the best approach would be to have a balanced dataset. Anyway, being the age corrected and age uncorrected results very similar we can be quite confident that age didn't play an important role in determining the good performance of the machine learning algorithm classification power, at least in this case.

It has to be remarked here, that the two groups of participants could in principle be discriminated by standard group analysis procedures, as for example by the MMSE score (see Table 1). Nevertheless, the aim of the machine learning procedure adopted in the study was to shift from the group level to the single-subject level analysis. As shown in Table 2, the addition of standard tests (CTS38, including MMSE) did not contribute to the classification performance as well as the addition of the visuospatial tests (CTS14). Even more explicative of the potential limitations represented by single scores like the MMSE in a single-subject oriented approach are the false positive and false negative rates illustrated in Table 2. In fact, when the only modGM features were used to classify the participants, two false positive were found, but when the CTS38 were added to help classification procedure, in addition to the two false positive a false negative added as well. The MCI subject presented an MMSE score of 30. In this case the feature represented by the MMSE score degraded in some way the classification accuracy. This doesn't mean that the MMSE would not be effective in "helping" participant classification, especially when they achieve an MMSE score clearly under or above a normality defined threshold, but this means that the MMSE is not sufficient to classify them. This is a well known issue in early AD classification, not limited to the MMSE score. Unsatisfactory false negative and/or false positive rates are common to all the state of the art biomarkers.

Given the small dataset that does not allow wide generalisation of results, this should mainly be considered a pilot study, the results of which suggest the potential represented by the combination of neuroimaging and visuospatial test scores for obtaining accurate classification of disease status. In particular the very high percentage of success in the classification has to be taken with a degree of caution given the number of participants under investigation. The significance of predictivity, assessed by the permutation test was found to be always higher than 99% ($p<0.01$), making us quite confident that what we obtained was a reliable result. Nevertheless, due to the small number of participants, we cannot be sufficiently confident that the study populations are representative of the actual MCI patients and healthy people populations, that is they could embody sub-populations in which the predictors worked very well by chance [62]. Due to the small number of participants included in the training and testing set of each cross validation fold (respectively 21 and 1 in the leave-one-out cross validation), some sort of "quantisation" of the detectable classification accuracies there exists (e.g. for 22/22 we have 100% accuracy, for 21/22 we would have 95% accuracy, and for 20/22 we would have 91% accuracy). It means that correct/incorrect classifications of few subjects may cause a dramatic percentage change of the estimated predictivity. We should also account for that when in the classification procedure we compare the results obtained by considering only one modality (i.e. modGM or cognitive tests alone, see Table 2) with the results obtained by including both modalities (modGM plus cognitive tests).

CONCLUSION

The present pilot study applied a combined approach of neuroimaging, standard cognitive test and visuospatial cognitive test assessment to a small cohort of patient experiencing MCI and a HC participant group. To investigate the classification power of the combined assessments we used an SVM approach based on an MKL algorithm recently integrated in the PRoNTo software. The classification procedure was found to be generalisable with a very high precision (100%), highlighting the superior performance of the feature mixture of a variety of features in identifying patterns that differentiate healthy participants from patients with MCI. The MKL approach may represent an optimal strategy to combine features which are very different in nature as neuroimaging indices and cognitive test scores. Although more work should be done to optimise Kernel selection (to avoid redundant region being erroneously excluded), the presented procedure appears also appropriate to identify those feature sets (brain regions and cognitive test batteries) best defining the patterns of information which can differentiate healthy brain structural substrates and functional performance from pathological ones.

If our findings are confirmed in a larger population of participants, visuospatial memory tests and cerebral atrophy indices might be assumed to be reliable detectors of the pathological decline associated with the preclinical stage of AD, and would represent valid biomarkers.

Ultimately, the present study, by combining a SVM MKL analysis with cognitive tests and neuroimaging investigation modalities, suggests a general feasible approach in the search for a good candidate biomarker for the early detection of AD.

LIST OF ABBREVIATIONS

AD: Alzheimer's Disease; ADNI: Alzheimer's Disease Neuroimaging Initiative; CTS: Cognitive Test Score; ER: Expected Ranking; GLM: general linear model; GM: Grey Matter; MCI: Mild Cognitive Impairment; MKL: Multi Kernel Learning; MMSE: Mini Mental Evaluation Score; modGM: Modulated Grey Matter; MRI: Magnetic Resonance Imaging; nsGM: native space Grey Matter; SVM: Support Vector Machine; TE: echo time; TR: repetition time; T1w: T1-weighted; VBM: Voxel Based Morphometry; 3D: 3-dimensional;

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

REFERENCES

1. Knight MJ, McCann B, Kauppinen RA, Coulthard EJ. Magnetic Resonance Imaging to Detect Ealry Molecular and Cellular Changes in Alzheimer's Disease. Front Aging Neurosci 8:139

2. Iachini I, Iavarone A, Senese VP, Ruotolo F, Ruggiero G. Visuospatial memory in healthy elderly, AD and MCI: a review. Curr Aging Sci 2:43–59 (2009).

3. Mitolo M, Gardini S, Fasano F, Crisi G, Pelosi A, Pazzaglia F, et al. Visuospatial memory and neuroimaging correlates in mild cognitive impairment. J Alzheimers Dis 35:75–90 (2013).

4. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 6:303–8 (1999).

5. Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 56: 1133–42 (2001).

6. Tiraboschi P, Salmon DP, Hansen LA, Hofstetter RC, Thal LJ, Corey-Bloom J. What best differentiates Lewy body from Alzheimer's disease in early-stage dementia? Brain 129:729–35 (2006).

7. Mitolo M, Salmon DP, Gardini S, Galasko D, Grossi E, Caffarra P. The new Qualitative Scoring MMSE Pentagon Test (QSPT) as a valid screening tool between autopsy-confirmed dementia with Lewy bodies and Alzheimer's disease. J Alzheimers Dis 39:823–32 (2014).

8. Cagnin A, Bussè C, Jelcic N, Gnoato F, Mitolo M, Caffarra P. High specificity of MMSE pentagon scoring for diagnosis of prodromal dementia with Lewy bodies. Parkinsonism Relat Disord 21:303-5 (2015).

9. Pai M-C, Jacobs WJ. Topographical disorientation in community-residing patients with Alzheimer's disease. Int J Geriatr Psychiatry 19:250–5 (2004).

10. Linn MC, Petersen AC. Emergence and characterization of sex differences in spatial ability: a meta-analysis. Child Dev 56:1479–98 (1985).

11. Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. Psychol Bull 117:250-70 (1995).

12. Allen GL, Kirasic KC, Dobson SH, Long RG, Beck S. Predicting environmental learning from spatial abilities: An indirect route. Intelligence 22:327-55 (1996).

13. Hegarty M, Montello DR, Richardson AE, Ishikawa T, Lovelace K. Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. Intelligence 34:151–76 (2006).

14. Gallistel CR Eds. The Organization of Learning. Cambridge: MIT Press (1990).

15. Waller D, Nadel L Eds. Handbook of Spatial Cognition. Washington: Amer Psychological Association (2013).

16. Piccardi L, Berthoz A, Baulac M, Denos M, Dupont S, Samson S, et al. Different spatial memory systems are involved in small- and large-scale environments: evidence from patients with temporal lobe epilepsy. Exp Brain Res 206(2):171–7 (2010).

17. Piccardi L, Iaria G, Ricci M, Bianchini F, Zompanti L, Guariglia C. Walking in the Corsi test: which type of memory do you need? Neurosci Lett 432(2):127–31 (2008).

18. Lambrey S, Samson S, Dupont S, Baulac M, Berthoz A. Reference frames and cognitive strategies during navigation: is the left hippocampal formation involved in the sequential aspects of route memory? Int Congr Ser 1250:261–74 (2003).

19. Yamamoto N, Degirolamo GJ. Differential effects of aging on spatial learning through exploratory navigation and map reading. Front Aging Neurosci 12: 4:14 (2012).

20. Gazova I, Vlcek K, Laczó J, Nedelska Z, Hyncicova E, Mokrisova I, et al. Spatial navigation-a unique window into physiological and pathological aging. Front Aging Neurosci 21:4-16 (2012).

21. Techentin C, Voyer D, Voyer SD. Spatial Abilities and Aging: A Meta-Analysis. Exp Aging Res 40(4):395–425 (2014).

22. Wiener JM, Kmecova H, de Condappa O. Route repetition and route retracing: effects of cognitive aging. Front Aging Neurosci 21:4-7 (2012).

23. Devlin AS Eds. Mind and maze: Spatial cognition and environmental behavior. Westport: Praeger Publishers (2001).

24. Mitolo M, Gardini S, Caffarra P, Ronconi L, Venneri A, Pazzaglia F. Relationship between spatial ability,

visuospatial working memory and self-assessed spatial orientation ability: a study in older adults. Cogn Process 16:165–76 (2015).

25. Kozlowski LT, Bryant KJ. Sense of direction, spatial orientation, and cognitive maps. J Exp Psychol Human 3:590–8 (1977).

26. Hegarty M. Development of a self-report measure of environmental spatial ability. Intelligence 30:425–47 (2002).

27. Pazzaglia F, Taylor HA. Perspective, Instruction, and Cognitive Style in Spatial Representation of a Virtual Environment. Spat Cogn Comput 7:349–64 (2008).

28. Labate E, Pazzaglia F, Hegarty M. What working memory subcomponents are needed in the acquisition of survey knowledge? Evidence from direction estimation and shortcut tasks. J Environ Psychology 37:73–9 (2014).

29. Fox NC, Warrington EK, Seiffer AL, Agnew SK, Rossor MN. Presymptomatic cognitive deficits in individuals at risk of familial Alzheimer's disease. A longitudinal prospective study. Brain 121:1631-9 (1998)

30. Mendez MF, Mendez MA, Martin R, Smyth KA, Whitehouse PJ. Complex visual disturbances in Alzheimer's disease. Neurology 40:439-43 (1990)

31. Possin KL. Visual Spatial Cognition in Neurodegenerative Disease. Neurocase 16:466-87 (2010)

32. McEvoy LK, Fennema-Notestine C, Roddey JC, Hagler DJ, Holland D, Karow DS, et al. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology 251:195–205 (2009).

33. De Santi S, de Leon MJ, Rusinek H, Convit A, Tarshish CY, Roche A, et al. Hippocampal formation glucose metabolism and volume losses in MCI and AD. Neurobiol Aging 22:529–39 (2001).

34. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol 65:403–13 (2009).

35. Ashburner J, Friston KJ. Voxel-based morphometry—the methods. Neuroimage 11: 805-21 (2000).

36. Mechelli A, Price CJ, Friston KJ. Voxel-based morphometry of the human brain: methods and applications. Current Medical Imaging Rev 1:105–13 (2005).

37. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. PNAS 97:11050–5 (2000).

38. Escudero J, Zajicek JP, Ifeachor E, Alzheimer's Disease Neuroimaging Initiative. Machine Learning classification of MRI features of Alzheimer's disease and mild cognitive impairment subjects to reduce the sample size in clinical trials. Conf Proc IEEE Eng Med Biol Soc 2011:7957–60 (2011).

39. Sabuncu MR, Konukoglu E, Alzheimer's Disease Neuroimaging Initiative. Clinical prediction from structural brain MRI scans: a large-scale empirical study. Neuroinform 13:31–46 (2015).

40. Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I, et al. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. Front Neurosci 9:307 (2015). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4555016/

41. Beltrachini L, De Marco M, Taylor ZA, Lotjonen J, Frangi AF, Venneri A. Integration of Cognitive Tests and Resting State fMRI for the Individual Identification of Mild Cognitive Impairment. Curr Alzheimer Res 12:592–603 (2015).

42. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single Subject prediction of brain disorders in neuroimaging. Promises and pitfalls. Neuroimage 2016 (in press).

43. Bron EE, Smits M, van der Flier WM, Vrenken H, Barkhof F, Scheltens P, et al. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. Neuroimage 111:562–79 (2015).

44. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory; 1992 July 27-29; Pittsburgh, USA.

45. Zhang L, Lin X. Some considerations of classification for high dimension low-sample size data. Stat Methods Med Res 22:537–50 (2013).

46. Lanckriet GRG, De Bie T, Cristianini N, Jordan MI, Noble WS. A statistical framework for genomic data fusion. Bioinformatics 20:2626–35 (2004).

47. Schrouff J, Rosa MJ, Rondina JM, Marquand AF, Chu C, Ashburner J, et al. PRoNTo: pattern recognition for neuroimaging toolbox. Neuroinform 11:319–37 (2013).

48. Schrouff J, Monteiro J, Joao Rosa M, Portugal L, Phillips C, Mourao-Miranda J. Can we interpret linear kernel machine learning models using anatomically labelled regions? Organization for Human Brain Mapping; 2014 June 8-12; Hamburg, Germany.

49. Schrouff J, Cremers J, Garraux G, Baldassarre L, Mourao-Miranda J, Phillips C. Localizing and Comparing Weight Maps Generated from Linear Kernel Machine Learning Models. International Workshop on Pattern Recognition in Neuroimaging; 2013 June 22-24; Philadelphia, USA.

50. Rakotomamonjy A, Bach F, Canu S. SimpleMKL. J Mach Learn Res 9:2491-521 (2008).

51. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer"s disease Report of the NINCDS ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer"s Disease. Neurology 34:939–9 (1984).

52. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." J Psychiatr Res 12:189–98 (1975).

53. Fasano F, Ganazzoli C, Gardini S, Sambataro F, Concari L, Caffarra P. SyN based multimodal investigation on a small cohort of patients affected with Amnesic Mild Cognitive Impairment. Proceeding of International Society of Magnetic Resonance in Medicine; 2010 May 9-13; Montreal, Canada.

54. Fischl B, Salat DH, van der Kouwe AJW, Makris N, Ségonne F, Quinn BT, et al. Sequence-independent segmentation of magnetic resonance images. NeuroImage 2:S69–84 (2004).

55. Fischl B, Salat D, Busa E, Albert M, Dieterich M. Whole Brain Segmentation Automated Labeling of Neuroanatomical Structures in the Human Brain. Neuron 33:341-55 (2002).

56. Dukart J, Schroeter ML, Mueller K, The Alzheimer's Disease Neuroimaging Initiative. Age Correction in Dementia – Matching to a Healthy Brain. Valdes-Sosa PA, Valdes-Sosa PA, editors. PLoS ONE 6:e22193–9 (2011).

57. Naselaris T, Kay KN, Nishimoto S, Gallant JL. Encoding and decoding in fMRI. Neuroimage 56:400–10 (2011).

58. Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J-D, Blankertz B, et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage 87:96–110 (2014).

59. Pennanen C, Kivipelto M, Tuomainen S, Hartikainen P, Hänninen T, Laakso MP, et al. Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. Neurobiol Aging 25:303–10 (2004).

60. Bozzali M, Filippi M, Magnani G, Cercignani M, Franceschi M, Schiatti E, et al. The contribution of voxel-based morphometry in staging patients with mild cognitive impairment. Neurology 67:453–60 (2006);

61. Cardenas VA, Tosun D, Chao LL, Fletcher PT, Joshi S, Weiner MW, et al. Voxel-Wise Co-analysis of Macro- and Microstructural Brain Alteration in Mild Cognitive Impairment and Alzheimer's Disease Using Anatomical and Diffusion MRI. J Neuroimaging 24:435–43 (2013).

62. Sheskin DJ Eds. Handbook of Parametric and Nonparametric Statistical Procedures. Boca Raton: CRC Press (2003).

Received: March 20, 2014 Revised: April 16, 2014 Accepted: April 20, 2014

SUPPLEMENTARY MATERIALS

SVM

SVMs are a set of frequently used methods for classification and regression in supervised machine learning applications. Supervised learning means that the learning is grounded on prior knowledge. For example, in our study we collected neuroimaging data (the value of the modGM map in 77715 voxels of grey matter) and cognitive test data (the values of 52 test scores, see Methods) for a cohort of 22 participants. For each participant we knew the group they belonged to, thus we worked within a "supervised" learning framework. Our goal was to understand if what we learned from the group could be extended

(generalised) to other individuals. The generalisation is at the basis of the cross validation procedures. Those procedures learn on a training subset of data (in our case on 21 participants) and try to extend the result to a testing subset of data (in our case on 1 participant). The operation is repeated for all possible combinations of training set and testing sets (in our case 22). Each permutation is called fold and each fold supplies a percentage of successes/failures (in our case 100% or 0%). The average across all success percentage of all fold permutations represents the prediction accuracy.

SVM has been shown to be a feasible method to analyse large feature small sample data sets [40]. In fact a standard modGM map contains a number of voxels (features) of the order of 100.000 while a classical MRI clinical study only includes few tens of subjects. These kinds of dataset are prone to overfitting problems, that is, by using a too complex model to solve a problem, "too many parameters to explain few sample data", the risk is to fit noise, determining a good but low generalisable solution. SVM theory lies on the use of so-called Kernels, mathematical entities that are able to increase the dimensionality of the feature space used to solve the problem by preserving a low complexity for the model.

Mathematically, the SVM learning problem consists of determining α_i^* and b^* parameters in the following equation:

$$f(x) = \sum_{i=1}^{l} \alpha_i^* K(x, x_i) + b^* \qquad [A1]$$

where f(x) is a function to be learned on the basis of previous knowledge of l previously collected examples: $\{x_i, y_i\}$ with i=1,..,l (in our study l=77767, i.e. 52 cognitive test scores plus 77715 modGM map values, see Methods). K represents the chosen kernel function. For example, in case of linear Kernel (the one used in the present study) $K(x, x') = \sum_{i=1}^{l} x_i * x'_i + c$. Each y_i can get one of two values, i.e. +1 or -1 (corresponding for example to the MCI or HC group).

In MKL formulation of the learning problem the $K(x, x')$ of Eq. A1 is replaced by $\sum_{m=1}^{M} d_m K(x, x')$, being d_m (m=1,…,M) the weights to be determined by optimisation procedure (in our case M=84, i.e. 81 brain regions plus 3 cognitive scores sub-sets and d_m are the weights assigned to the 84 associated Kernels and used to assess the contribution to the prediction accuracy of each sub-set.

Age correction procedure.

The 22 modGM maps were corrected voxel by voxel. Firstly, we evaluated the effect of age on the healthy participants by GLM:

y_HC(i)=b0(i) + b1(i) x ageHC [A2]

where yHC(i) is the vector composed by the 11 uncorrected modGM values in the i-th voxel, ageHC is the vector of the ages of the eleven healthy control participants and b0(i) and b1(i) are the two GLM parameters to be estimated for the i-th voxel. Secondly we applied voxel by voxel the estimated correction to all the participants scores (including patients):

corr_yHC(i)=yHC(i) - b1(i) x ageHC [A3a]

corr_yMCI(i)=yMCI(i) - b1(i) x ageMCI [A3b]

where corr_yHC(i) and corr_yMCI(i) are the corrected 11 elements vectors for the i-th voxel respectively for the control participants and for the patients. The same procedure was applied to the cognitive test scores.

In-depth analysis of Results (weights and ER)

Variability of results through the different cross validation folds, small cohort of participants and overshadowing of some Kernels (regions) by others Kernels (regions) containing the same patterns of information (see Ref. 44) are important issues that can be explored more in-depth with the MKL approach implemented by the PRoNTo software. The latter issue specifically is a well-known one in machine learning neuroscientific applications, i.e. it is the reason why this approach cannot easily provide an answer to the following general problem: "Given that the information/activity contained in a brain region A is shown to predict a clinical/behavioural state B, we want to select the sub-region of A mostly responsible for that clinical/behavioural state". The Kernels (regions) found to have non-zero weights can be assumed to be part of the informative pattern, but it cannot be excluded that some 0 weighted Kernels (regions) might also be contributing. To help the identification of those Kernels (regions) that have a stable performance among folds (individuals), the PRoNTo software calculates the "Expected Ranking" index (ER) that is a measure of how much the Kernel (region) rating in each fold is close to its averaged value across fold (see Ref. 44). If we look at the ER scores of the list of regions having non-zero weights (see Supplementary Materials) we find that the "CTS14 objective" ER is 0.95. Furthermore, the ranking was found to be 1 for most of the folds (data not reported). This confirms the high reliability of the "CTS14 objective" as the sub-set of features driving the classification performance. A similar consistency is found for the Brain Stem region, as well as for other brain areas among the most relevant ones (as ranked by weights). "CTS14 self" on the contrary, shows a very high ER, thus confirming its lower stability and relevance in classification, as already suggested by its ranking and its low weight. While among the first ranked regions we find quite good ER scores, there are more discrepancies in

the lower ranked regions between rank/weight and ER, including Hippocampus, left Caudate Nucleus, Putamen, and Thalamus, as well as left Amygdala.

MRI only feature set results

In the following table we report the results obtained when the MRI data only are analysed. We report for each brain region the number of features (N), i.e. number of voxels included in the associated Kernel, the normalised contribution of feature subset in classification performance, i.e. its weight (%w), and the Expected Ranking (ER). Data are sorted in decreasing order from higher %w to lower %w.

	Original Data		
CTS and TBR Labels	N	%w	ER
Brain-Stem	**25,73**	**3622**	**1,55**
ctx-rh-entorhinal	**23,74**	**248**	**1,91**
Left-Pallidum	**21,59**	**338**	**2,41**
ctx-lh-pericalcarine	**7,50**	**286**	**7,68**
ctx-lh-inferiortemporal	**5,37**	**1709**	**11,23**
ctx-rh-transversetemporal	**4,94**	**147**	**9,23**
Right-Caudate	**4,35**	**690**	**6,73**
ctx-rh-paracentral	**1,61**	**685**	**37,95**
Left-Thalamus-Proper	**1,06**	**1149**	**8,86**
ctx-rh-parstriangularis	**0,98**	**536**	**42,59**
ctx-rh-frontalpole	**0,72**	**83**	**59,14**
ctx-lh-parstriangularis	**0,55**	**374**	**27,59**
ctx-lh-paracentral	**0,49**	**522**	**29,23**
Right-Thalamus-Proper	**0,37**	**1142**	**13,73**
Right-Amygdala	**0,26**	**221**	**18,00**
ctx-lh-parsorbitalis	**0,15**	**335**	**33,41**
ctx-lh-parsopercularis	**0,15**	**751**	**31,45**
Left-Hippocampus	**0,10**	**592**	**11,82**
ctx-lh-transversetemporal	**0,08**	**205**	**44,09**
Left-Amygdala	**0,07**	**194**	**12,64**
Right-Accumbens-area	**0,06**	**103**	**19,00**
ctx-rh-temporalpole	**0,06**	**266**	**73,14**
ctx-rh-caudalanteriorcingulate	**0,04**	**338**	**47,77**
ctx-lh-middletemporal	**0,02**	**1663**	**30,09**
ctx-lh-lateralorbitofrontal	0,00	1167	27,41
Left-Caudate	0,00	665	10,36
Left-Putamen	0,00	965	11,32
Left-Accumbens-area	0,00	93	14,00
Right-Putamen	0,00	910	15,82
Right-Pallidum	0,00	306	16,77
Right-Hippocampus	0,00	593	17,73
ctx-lh-caudalanteriorcingulate	0,00	311	20,50
ctx-lh-caudalmiddlefrontal	0,00	1075	21,45
ctx-lh-cuneus	0,00	453	22,41

ctx-lh-entorhinal	0,00	292	23,36
ctx-lh-fusiform	0,00	1398	24,32
ctx-lh-inferiorparietal	0,00	1674	25,27
ctx-lh-isthmuscingulate	0,00	367	26,50
ctx-lh-lateraloccipital	0,00	1605	27,45
ctx-lh-lingual	0,00	889	29,32
ctx-lh-medialorbitofrontal	0,00	555	30,27
ctx-lh-parahippocampal	0,00	327	32,14
ctx-lh-postcentral	0,00	1389	36,45
ctx-lh-posteriorcingulate	0,00	522	37,41
ctx-lh-precentral	0,00	2158	38,36
ctx-lh-precuneus	0,00	1265	39,32
ctx-lh-rostralanteriorcingulate	0,00	400	40,27
ctx-lh-rostralmiddlefrontal	0,00	1870	41,23
ctx-lh-superiorfrontal	0,00	3076	42,18
ctx-lh-superiorparietal	0,00	1584	43,14
ctx-lh-superiortemporal	0,00	1865	44,09
ctx-lh-supramarginal	0,00	1475	45,05
ctx-lh-frontalpole	0,00	66	46,00
ctx-lh-temporalpole	0,00	239	46,95
ctx-lh-insula	0,00	1062	48,77
ctx-rh-caudalmiddlefrontal	0,00	990	50,64
ctx-rh-cuneus	0,00	493	51,59
ctx-rh-fusiform	0,00	1411	52,55
ctx-rh-inferiorparietal	0,00	2139	53,50
ctx-rh-inferiortemporal	0,00	1463	54,45
ctx-rh-isthmuscingulate	0,00	357	55,41
ctx-rh-lateraloccipital	0,00	1515	56,36
ctx-rh-lateralorbitofrontal	0,00	1060	57,32
ctx-rh-lingual	0,00	1014	58,27
ctx-rh-medialorbitofrontal	0,00	626	59,23
ctx-rh-middletemporal	0,00	1853	60,18
ctx-rh-parahippocampal	0,00	349	61,14
ctx-rh-parsopercularis	0,00	677	62,64
ctx-rh-parsorbitalis	0,00	336	63,59
ctx-rh-pericalcarine	0,00	333	65,14
ctx-rh-postcentral	0,00	1268	66,09
ctx-rh-posteriorcingulate	0,00	646	67,05
ctx-rh-precentral	0,00	2080	68,00
ctx-rh-precuneus	0,00	1416	68,95
ctx-rh-rostralanteriorcingulate	0,00	288	69,91
ctx-rh-rostralmiddlefrontal	0,00	2004	70,86
ctx-rh-superiorfrontal	0,00	2786	71,82
ctx-rh-superiorparietal	0,00	1566	72,77
ctx-rh-superiortemporal	0,00	1824	73,73
ctx-rh-supramarginal	0,00	1381	74,68

ctx-rh-insula	0,00	1025	77,32

Comparison between GLM age corrected and uncorrected results

In the following table, for both original data and age corrected data, we report, for each cognitive test score and true brain region, the number of features (N), i.e. number of test scores / number of voxels included in the associated Kernel, the normalised contribution of feature subset in classification performance, i.e. its weight (%w), and the Expected Ranking (ER). Data are sorted in decreasing order from higher %w to lower %w as assessed in the original data.

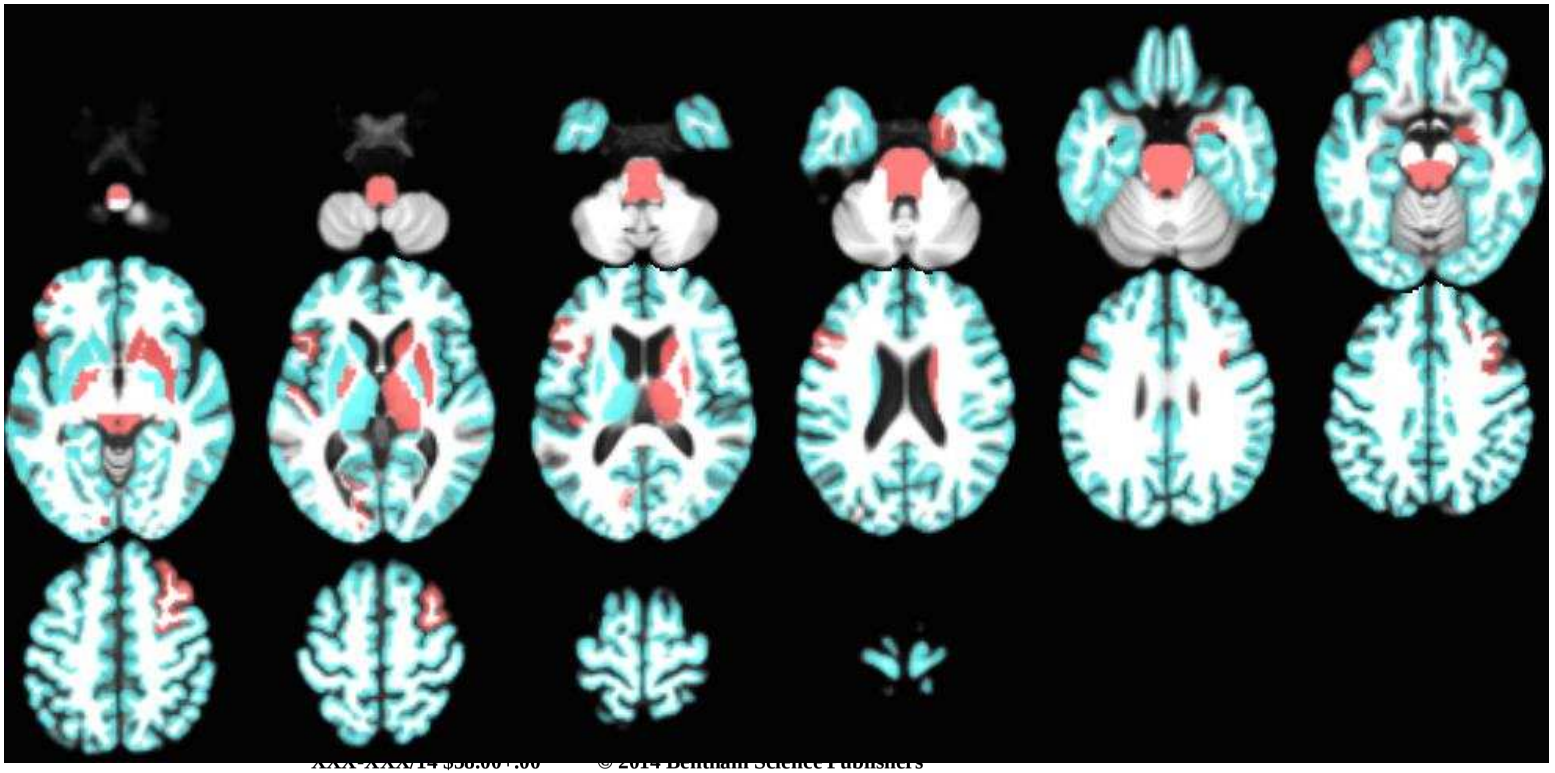
		Original Data		Age Corrected Data	
CTS and TBR Labels	N	%w	ER	%w	ER
CTS14 objective	**10**	**63,62**	**0,95**	**65,27**	**0,95**
ctx-lh-parsopercularis	**751**	**12,28**	**6,32**	**13,94**	**13,09**
ctx-rh-entorhinal	**248**	**8,74**	**10,14**	**4,51**	**26,36**
ctx-lh-pericalcarine	**286**	**8,14**	**7,86**	**3,81**	**16,86**
Brain-Stem	**3622**	**3,97**	**5,32**	**0,01**	**9,55**
Left-Pallidum	**338**	**1,32**	**7,41**	**0,69**	**8,32**
Right-Amygdala	**221**	**0,57**	**16,27**	**1,16**	**15,27**
Right-Putamen	910	0,56	10,27	0,00	15,23
Right-Caudate	**690**	**0,29**	**11,73**	**0,18**	**13,82**
Right-Thalamus-Proper	1142	0,15	12,32	0,00	13,36
ctx-rh-caudalmiddlefrontal	990	0,13	49,55	0,00	50,73
Right-Accumbens-area	**103**	**0,09**	**17,23**	**0,06**	**18,23**
ctx-lh-parsorbitalis	**335**	**0,07**	**33,23**	**0,33**	**34,09**
ctx-lh-transversetemporal	205	0,04	44,73	0,00	47,86
CTS14 self-evaluative	**4**	**0,04**	**75,68**	**0,12**	**79,68**
Left-Thalamus-Proper	1149	0,00	6,32	0,29	5,45
Left-Caudate	665	0,00	7,27	0,00	6,73
Left-Putamen	965	0,00	8,23	0,00	7,68
Left-Hippocampus	592	0,00	10,05	0,00	10,50
Left-Amygdala	194	0,00	11,00	0,00	11,45
Left-Accumbens-area	93	0,00	11,95	0,00	12,41
Right-Pallidum	306	0,00	15,05	0,00	16,18
Right-Hippocampus	593	0,00	16,00	0,00	17,14
ctx-lh-caudalanteriorcingulate	311	0,00	18,77	0,00	19,77
ctx-lh-caudalmiddlefrontal	1075	0,00	19,73	0,00	20,73
ctx-lh-cuneus	453	0,00	20,68	0,00	21,68
ctx-lh-entorhinal	292	0,00	21,64	0,00	22,64
ctx-lh-fusiform	1398	0,00	22,59	0,00	23,59
ctx-lh-inferiorparietal	1674	0,00	23,55	0,00	24,55
ctx-lh-inferiortemporal	1709	0,00	24,50	4,19	15,95
ctx-lh-isthmuscingulate	367	0,00	25,45	0,00	26,09
ctx-lh-lateraloccipital	1605	0,00	26,41	0,00	27,05
ctx-lh-lateralorbitofrontal	1167	0,00	27,36	0,00	28,00
ctx-lh-lingual	889	0,00	28,32	0,00	28,95
ctx-lh-medialorbitofrontal	555	0,00	29,27	0,00	29,91
ctx-lh-middletemporal	1663	0,00	30,23	0,00	30,86
ctx-lh-parahippocampal	327	0,00	31,18	0,00	31,82

ctx-lh-paracentral	522	0,00	32,14	0,00	32,77
ctx-lh-parstriangularis	374	0,00	34,18	0,00	35,05
ctx-lh-postcentral	1389	0,00	35,27	0,00	36,41
ctx-lh-posteriorcingulate	522	0,00	36,23	0,00	37,36
ctx-lh-precentral	2158	0,00	37,18	0,00	38,32
ctx-lh-precuneus	1265	0,00	38,14	0,00	39,27
ctx-lh-rostralanteriorcingulate	400	0,00	39,09	0,00	40,23
ctx-lh-rostralmiddlefrontal	1870	0,00	40,05	0,00	41,18
ctx-lh-superiorfrontal	3076	0,00	41,00	0,00	42,14
ctx-lh-superiorparietal	1584	0,00	41,95	0,00	43,09
ctx-lh-superiortemporal	1865	0,00	42,91	0,00	44,05
ctx-lh-supramarginal	1475	0,00	43,86	0,00	45,00
ctx-lh-frontalpole	66	0,00	44,82	0,00	45,95
ctx-lh-temporalpole	239	0,00	45,77	0,00	46,91
ctx-lh-insula	1062	0,00	47,64	0,00	48,82
ctx-rh-caudalanteriorcingulate	338	0,00	48,59	0,00	49,77
ctx-rh-cuneus	493	0,00	50,50	0,00	51,68
ctx-rh-fusiform	1411	0,00	51,59	0,00	53,09
ctx-rh-inferiorparietal	2139	0,00	52,55	0,00	54,05
ctx-rh-inferiortemporal	1463	0,00	53,50	0,00	55,00
ctx-rh-isthmuscingulate	357	0,00	54,45	0,00	55,95
ctx-rh-lateraloccipital	1515	0,00	55,41	0,00	56,91
ctx-rh-lateralorbitofrontal	1060	0,00	56,36	0,00	57,86
ctx-rh-lingual	1014	0,00	57,32	0,00	58,82
ctx-rh-medialorbitofrontal	626	0,00	58,27	0,00	59,77
ctx-rh-middletemporal	1853	0,00	59,23	0,00	60,73
ctx-rh-parahippocampal	349	0,00	60,18	0,61	45,55
ctx-rh-paracentral	685	0,00	61,14	0,00	62,36
ctx-rh-parsopercularis	677	0,00	62,09	2,27	35,55
ctx-rh-parsorbitalis	336	0,00	63,05	0,00	63,82
ctx-rh-parstriangularis	536	0,00	64,00	1,48	44,86
ctx-rh-pericalcarine	333	0,00	64,95	0,00	65,41
ctx-rh-postcentral	1268	0,00	65,91	0,00	66,36
ctx-rh-posteriorcingulate	646	0,00	66,86	0,00	67,32
ctx-rh-precentral	2080	0,00	67,82	0,00	68,27
ctx-rh-precuneus	1416	0,00	68,77	0,00	69,23
ctx-rh-rostralanteriorcingulate	288	0,00	69,73	0,00	70,18
ctx-rh-rostralmiddlefrontal	2004	0,00	70,68	0,00	71,14
ctx-rh-superiorfrontal	2786	0,00	71,64	0,00	72,09
ctx-rh-superiorparietal	1566	0,00	72,59	0,00	73,05
ctx-rh-superiortemporal	1824	0,00	73,55	0,00	74,00
ctx-rh-supramarginal	1381	0,00	74,50	0,00	74,95
ctx-rh-frontalpole	83	0,00	75,45	0,00	75,91
ctx-rh-temporalpole	266	0,00	76,41	0,00	76,86

ctx-rh-transversetemporal	147	0,00	77,36	0,05	74,27
ctx-rh-insula	1025	0,00	78,32	0,00	78,73
CTS38	38	0,00	80,18	1,03	45,09

		Thalamus(Proper.	Caudate.	Putamen.	Pallidum.	Hippocampus.	Amygdala.	Accumbens(area.	ctx(caudalanteriorcingulate.	ctx(caudalmiddlefrontal.	ctx(cuneus.	ctx(entorhinal.	ctx(fusiform.	ctx(inferiorparietal.	ctx(inferiortemporal.	ctx(isthmuscingulate.	ctx(lateraloccipital.	ctx(lateralorbitofrontal.	ctx(lingual.	ctx(medialorbitofrontal.	ctx(middletemporal.	ctx(parahippocampal.	ctx(paracentral.	ctx(parsopercularis.	ctx(parsorbitalis.	ctx(parstriangularis.	ctx(pericalcarine.	ctx(postcentral.	ctx(posteriorcingulate.	ctx(precentral.	ctx(precuneus.	ctx(rostralanteriorcingulate.	ctx(rostralmiddlefrontal.	ctx(superiorfrontal.	ctx(superiorparietal.	ctx(superiortemporal.	ctx(supramarginal.	ctx(frontalpole.	ctx(temporalpole.	ctx(transversetemporal.	ctx(insula.
LEFT	%w	0,00	0,00	0,00	1,32	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	12,28	0,07	0,00	8,14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,04	0,00
	ER	6,32	7,27	8,23	7,41	10,05	11,00	11,95	18,77	19,73	20,68	21,64	22,59	23,55	24,50	25,45	26,41	27,36	28,32	29,27	30,23	31,18	32,14	6,32	33,23	34,18	7,86	35,27	36,23	37,18	38,14	39,09	40,05	41,00	41,95	42,91	43,86	44,82	45,77	44,73	47,64
RIGHT	%w	0,15	0,29	0,56	0,00	0,00	0,57	0,09	0,00	0,13	0,00	8,74	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	ER	12,32	11,73	10,27	15,05	16,00	16,27	17,23	48,59	49,55	50,50	10,14	51,59	52,55	53,50	54,45	55,41	56,36	57,32	58,27	59,23	60,18	61,14	62,09	63,05	64,00	64,95	65,91	66,86	67,82	68,77	69,73	70,68	71,64	72,59	73,55	74,50	75,45	76,41	77,36	78,32
	N	1149	665	965	338	592	194	93	311	1075	453	292	1398	1674	1709	367	1605	1167	889	555	1663	327	522	751	335	374	286	1389	522	2158	1265	400	1870	3076	1584	1865	1475	66	239	205	1062

	BrainStem.	CT514.self.	CT514.objective.	CT538
%w	3,97	0,04	63,62	0,00
ER	5,32	75,68	0,95	80,18
N	3622	4	10	38



XXX-XXX/14 $58.00+.00

Figure 1. For each features subset (81 FreeSurfer extracted cerebral structures and three sets of CTS) the normalised weight (%w) assigned by the MKL optimization procedure to the associated Kernel is reported, as well as the Expected Ranking (ER) and the Number of features (N) included in the subset. The feature sets for which the weight was different from zero are highlighted in red. Fifteen axial brain slices of the study population template are also shown, with superimposed cyan overlay on structural scans representing the cerebral structures included in the analysis but excluded by the optimisation procedure (zero associated Kernel weight), and superimposed red overlay representing the structures with non-zero associated Kernel weight.