



Individual prediction and classification of cognitive impairment in patients with white matter lesions based on gray matter volume

Jinfang Wang^{1,2#}, Cui Zhao^{3,4#}, Jing Wei^{3,4}, Chunlin Li^{3,4}, Xu Zhang^{3,4}, Ying Liang^{3,4}, Yumei Zhang¹

¹Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, China National Clinical Research Center for Neurological Diseases, Center of Stroke, Beijing Institute for Brain Disorders, Beijing Key Laboratory of Translational Medicine for Cerebrovascular Disease, Beijing, China; ²Department of Neurology, General Hospital of The Yang Tze River Shipping, Wuhan Brain Hospital, Wuhan, China; ³School of Biomedical Engineering, Capital Medical University, Beijing, China; ⁴Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing, China

Contributions: (I) Conception and design: Y Liang, Y Zhang; (II) Administrative support: Y Liang, Y Zhang; (III) Provision of study materials or patients: J Wang; (IV) Collection and assembly of data: J Wang; (V) Data analysis and interpretation: C Zhao, C Li, J Wei, X Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Yumei Zhang. Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, China National Clinical Research Center for Neurological Diseases, Center of Stroke, Beijing Institute for Brain Disorders, Beijing Key Laboratory of Translational Medicine for Cerebrovascular Disease, Beijing, 100053, China. Email: zhangyumei95@aliyun.com; Ying Liang. School of Biomedical Engineering, Capital Medical University, Beijing 100069, China. Email: yingliang@ccmu.edu.cn.

Background: Vascular risk factors like white matter lesions (WMLs) are increasingly recognized as risk factors for vascular dementia (VaD) and can predict Alzheimer's disease (AD) at least a decade before the clinical stage of the disease. This study aimed to predict cognitive decline and use machine learning techniques to classify older individuals (aged 50 years or older) with WMLs as having vascular mild cognitive impairment (VaMCI), VaD, or in good cognitive health (CH).

Methods: A total of 79 individuals with WMLs were selected for this study and categorized into the following 3 groups: CH (n=25), VaMCI (n=33), and VaD (n=21). Data from the entire cohort was then divided into a training dataset (n=56) and testing dataset (n=23). The data were extracted from gray matter (GM) segmentations using voxel-based morphometry (VBM). A relevance vector regression (RVR) approach was used to test the relationship between the structural brain images and clinical scores. To predict the individual-level subtypes, we applied 2 different machine learning-based classifiers: support vector machine (SVM) and Gaussian process classification (GPC). All predictive models were trained on the training dataset and then validated on the testing dataset of age-matched participants.

Results: Multi-domain cognitive performance could be predicted based on the pattern of GM atrophy in older people with WMLs using a RVR approach. The classification of VaD versus CH (cross-validation accuracy =93.94%, test set accuracy =76.92%) and VaMCI versus CH (cross-validation accuracy =95.24%, test set accuracy =87.50%) could be successfully achieved using both SVM and GPC. However, SVM (cross-validation accuracy =67.57%, test set accuracy =70.59%) performed better than GPC in the classification of VaD versus VaMCI.

Conclusions: Based on the patterns of gray matter and RVR-based model could achieve prediction of cognitive test scores, and SVM and GPC could classify the severity of cognitive impairment in older people with WMLs.

Keywords: Mild cognitive impairment; dementia; aging; machine learning; white matter lesions

Submitted Jul 08, 2021. Accepted for publication Nov 26, 2021.

doi: 10.21037/atm-21-3571

View this article at: <https://dx.doi.org/10.21037/atm-21-3571>

Introduction

White matter lesions (WMLs), also referred to as age-related white matter hyperintensities on T2 weighted images, are prevalent in the elderly, especially in individuals with cardiovascular risk factors. Mounting evidence indicates that WMLs contribute to cognitive dysfunction in multiple domains, especially executive function, processing speed, and memory (1-3). Evidence has also shown that WMLs contribute to a spectrum of vascular mild cognitive impairments (VaMCI) (4) and may even predict Alzheimer's disease (AD) at least a decade before the clinical stage of the disease, independently of AD pathology biomarkers (5). Although many studies have observed an association between the severity of WMLs and cognitive dysfunction, there have been no consistent conclusions.

Structural magnetic resonance imaging (MRI) studies have found that older adults with WMLs have a significantly reduced gray matter (GM) volume and cortical thickness (6-8). The cortical alterations caused by WMLs may lead to cognitive decline and future dementia (9,10). Our previous study found that WMLs caused changes in GM density (1). Therefore, it may be possible to predict the severity of cognitive impairment based on the properties of the GM in older people with WMLs. As the diagnosis of cognitive function is primarily based on neuropsychological assessment, it is easy for diagnostic errors to occur when there is a lack of cooperation due to a patient's educational level or serious cognitive impairment. Automatically determining the severity of cognitive impairment could facilitate the formulation of an appropriate clinical diagnosis and treatment plan.

In the last decade, machine-learning methods have provided the opportunity to perform quantitative predictions of individual clinical assessments and disease classifications (11) and have been proposed as an aid in the early diagnosis of dementia (12). One of the advantages of machine learning approaches is that they can analyze many variables simultaneously and observe inherent patterns in the data (13). In addition to this, machine learning algorithms are also sensitive to the subtle, spatially distributed differences in brain MRI which have great promise in deriving individualized neuroimaging features of brain anatomy and providing an ideal framework to investigate psychiatric disorders (14). Relevance vector regression (RVR), a multivariate machine learning technique (15), has been used to quantitatively predict variables of interest in several neuroimaging studies

(16-18). Support vector machine (SVM) and Gaussian process classification (GPC) models have achieved high accuracy in AD classification, even with relatively small training sample sizes (19-21). Detecting dementia at the prodromal stages is one advantage of these classifiers, as it can predict the conversion from mild cognitive impairment (MCI) to dementia.

Applying these machine learning methods to predict individual cognitive assessments of older individuals with WMLs could provide an early warning of cognitive impairment and assist in clinical diagnosis. The current study aimed to investigate the diagnostic accuracy of machine learning classifiers in distinguishing between VaMCI, vascular dementia (VaD), and cognitive health (CH) and in predicting cognitive decline in older individuals with WMLs, based on the pattern of GM density maps.

We present the following article in accordance with the STARD reporting checklist (available at <https://atm.amegroups.com/article/view/10.21037/atm-21-3571/rc>).

Methods

Ethics statement

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of the Beijing Tiantan Hospital, Capital Medical University, China (KYSB2016-023). Written informed consent was provided by all participants.

Participants

We initially enrolled 79 older individuals with WMLs from the Beijing Tiantan Hospital, Capital Medical University, from January 2011 to December 2016. The WMLs were diagnosed from T2-fluid attenuated inversion recovery (T2-FLAIR) images separately by 2 radiologists without knowledge of the participants' clinical profiles. The inclusion criteria for this study were as follows: (I) aged between 50 and 85 years, (II) showed white matter hyperintensities of presumed vascular origin on T2-FLAIR MR images with a Fazekas score of ≥ 1 (22), (III) presence of a contactable informant throughout the study. The exclusion criteria were as follows: (I) cardiac or renal failure, cancer, or other severe systemic diseases; (II) unrelated neurological diseases such as epilepsy, traumatic brain injury, or multiple sclerosis; (III) chronic cerebral infarction

Table 1 Summary of grouping criteria

Group	CDR	MMSE	MoCA
CH (n=25)	CDR =0	MMSE \geq 27 with \geq 6 years of education, or MMSE \geq 24 with <6 years of education, or MMSE \geq 21 with 0 years of education	MoCA \geq 26
VaMCI (n=33)	CDR =0.5	24 \leq MMSE <27 with \geq 6 years of education, or 20 \leq MMSE <24 with <6 years of education, or 17 \leq MMSE <21 with 0 years of education	22 \leq MoCA <26
VaD (n=21)	CDR \geq 1	MMSE <24 with \geq 6 years of education, MMSE <20 with <6 years of education, or MMSE <17 with 0 years of education	MoCA <22

CH, good cognitive health; VaMCI, vascular mild cognitive impairment; VaD, vascular dementia; CDR, clinical dementia rating; MMSE, Chinese version of the Mini-Mental State Examination; MoCA, Beijing version of the Montreal Cognitive Assessment.

or other lesions; (IV) leukoencephalopathy of non-vascular origin; (V) dementia of non-vascular origin (imaging with posterior cingulate and neocortical temporoparietal cortical losses, or medial temporal-lobe atrophy); (VI) psychiatric diseases or drug addiction; (VII) consciousness disruption or aphasia; (VIII) inability or refusal to undergo a brain MRI. We followed the definition of WMLs described by Wardlaw *et al.* (23).

Clinical cognitive assessment

All participants were instructed to complete the Chinese version of the Mini-Mental State Examination (MMSE) (24), the Beijing version of the Montreal Cognitive Assessment (MoCA) (25), and the Clinical Dementia Rating (CDR) under the supervision of a physician. The tests were completed in a quiet room and in strict order, according to standard protocols. The following education-specific reference cut-off values for MMSE scores were used: 27 for middle and high school, 24 for elementary school, and 21 for preliterate. The MoCA cut-off value for cognitive impairment was <26 (26). In addition, 1 point was added to the raw MoCA score for patients with fewer than 12 years of education (27). Based on the results of these cognitive tests, participants were divided into the following 3 groups: VaD, VaMCI, and CH. Detailed grouping criteria is summarized in *Table 1*.

Acquisition of brain MRI data

The MRI data were acquired using a Siemens Magnetom Verio 3T superconducting MRI system (Siemens Healthineers, Erlangen, Germany) in the Department of Radiology at Beijing Tiantan Hospital. A T2W-FLAIR sequence was used to detect white matter hyperintensities.

A standard T1-weighted 3D magnetization-prepared rapid gradient-echo sequence was applied using the following parameters: repetition time (TR) =2,300 ms, echo time (TE) =3.28 ms, inversion time (TI) =1,200 ms, field of view =204×240 mm², matrix size =256×256, flip angle (FA) =9°, slice thickness =1 mm, inter slice gap =0.5 mm, and number of slices =256.

Imaging preprocessing and data extraction

Identical imaging processing procedures were used for all participants. Imaging data were preprocessed using the Statistical Parametric Mapping 12 (SPM12) (Institute of Neurology, London, UK, <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>) and the CAT12 toolbox (<http://www.neuro.uni-jena.de/cat12-html/cat.html>) running on MATLAB version 2018a (MathWorks, Natick, MA, USA). The GM and white matter were segmented for each participant using a unified tissue-segmentation procedure after image-intensity nonuniformity correction. In this study, we used only the GM volume images. These segmented GM images were spatially normalized to 1 mm³ voxels in the Montreal Neurological Institute (MNI) standard space. The resulting GM images were smoothed using an 8 mm full-width at half-maximum (FWHM) Gaussian kernel (21,28). Each image underwent visual quality control after segmentation and transformation. In our previous study, we illustrated that older individuals with WMLs and VaD or VaMCI tended to have a smaller whole-brain GM volume compared to CH control participants (29). The GM atrophy was tested in the different groups of participants with WMLs enrolled in our study using an analysis of covariance (ANCOVA) test. Older individuals with VaD had significant cortex atrophy compared to VaMCI and CH participants (*Figure S1* in

Supplementary materials).

In the present study, voxel-based (VB) and region of interest-based (ROI-B) data were extracted from the GM imaging data (28,30,31). The VB data corresponded to all voxels on the GM images for each participant. The GM densities of each voxel were used as input data for the regression or classification models. We also computed the ROI-B data, incorporating pre-existing atlas-based anatomical information to test whether an anatomical knowledge of GM volume can benefit the predictive performance of a machine learning model. The GM volume of each participant was divided into distinct cortical regions according to the automated anatomical labelling atlas 3 (AAL3) (32). There are a total of 166 parcellations in AAL3, including a number of brain areas not defined in previous versions, such as the subdivision of the anterior cingulate cortex, the thalamus, and the other subcortical nuclei. The ROI-B data corresponded to the average GM density, computed in a set of ROIs obtained from the AAL3 atlas. Regions belonging to the cerebellum were excluded from this study as they were unlikely to be linked to the neuropathology of VaD and VaMCI. Therefore, for each participant, we obtained 140 pieces of data from the GM images, herein referred to as the AAL3 data.

Machine learning methods

Relevance vector regression

To investigate whether the GM images were predictive of cognitive performance in older individuals with WMLs, a relevance vector regression (RVR) algorithm was implemented in the Pattern Recognition for Neuroimaging Toolbox (PRoNT) software program (33). The RVR model is a sparse kernel method based on a probabilistic Bayesian framework with zero-mean Gaussian priors for the model weights governed by hyperparameters (15). It takes the computed data derived from GM images as input vectors, and the performance on a given neuropsychological test as the target. The posterior distributions of many of the model weights were sharply peaked at zero when estimated using the training data. Non-zero weights were taken as “relevance vectors”, which were then used as the model weight vectors to predict the target.

The RVR model predicted scores for given neuropsychological tests based on the individual GM structures of participants. The significance of the predictive performance was assessed using Pearson’s correlation coefficient (CORR), the coefficient of determination (R^2),

and the mean-squared error (MSE), which are shown below as functions [1], [2], [3]. A permutation test was performed to assess the stability of the proposed model. Each model was retrained 1,000 times and P values for the CORR, R^2 , and MSE were obtained from the performance statistics, which were considered significant if the P value <0.05 .

The CORR provides a measure of the linear dependence between the predicted scores and the clinically measured scores and reflects the predictive accuracy of the RVR model. A CORR of 1 would indicate that the predicted score is identical to the actual measured one. The CORR was determined using the following formula:

$$CORR = \frac{\sum_n (y_n - \mu_y)(f(x_n) - \mu_f)}{\sqrt{\left[\sum_n (y_n - \mu_y)^2 \sum_n (f(x_n) - \mu_f)^2\right]}} \quad [1]$$

The R^2 is also a measure for predicting the accuracy of the RVR model, and was calculated using the following formula:

$$R^2 = CORR^2 \quad [2]$$

The MSE is a standard measure for regression models, and in our study, reflected how well a multivariate pattern of GM predicted the scores of clinical cognitive tests in older individuals with WMLs. There are different scales to determine the value of the MSE for different clinical tests. In different models predicting the same test scores, the higher the MSE, the less accurate the predictions. The MSE was calculated using the following formula:

$$MSE = \frac{1}{N} \sum_n (y_n - f(x_n))^2 \quad [3]$$

In these formulae, y_n and $f(x_n)$ denote the targets and predictions corresponding to the input predictors, and x_n , μ_y , and μ_f are the sample means of the targets and predictions, respectively. N is the total number of samples.

Classification

We used 2 different binary classifiers available in PRoNT (linear SVM and GPC) to evaluate whether it is possible to classify different groups of participants with WMLs and different levels of cognitive impairment. We used these 2 algorithms to determine whether GM patterns are predictive of the rate of cognitive decline in older individuals with WMLs. The linear SVM classifier is 1 of the most popular methods for neuroimaging classification problems and can achieve a good separation of the different

classes by constructing a hyperplane in a high or infinite dimensional space (34). The decision function of the linear SVM was defined as follows:

$$f(x) = \text{sign} \left[\sum_{i=1}^N \alpha_i y_i K(x_i, x) - b \right] \quad [4]$$

In this formula, N is the number of training examples, α_i is the contribution of training example x_i to the final classification, y_i is the class label of x_i , and b is a bias term. α_i and b were determined by the following equation:

$$\begin{aligned} \max(\alpha_i) - \frac{1}{2} \sum_{i,j} y_i y_j x_i \cdot x_j \alpha_i \alpha_j + \sum_{i=1}^N \alpha_i \quad \text{subject to} \\ \sum_{i=1}^N \alpha_i y_i = 0 \quad \text{and} \quad 0 \leq \alpha_i \leq C \end{aligned} \quad [5]$$

C is the regularization parameter which controls the distance between the hyperplane and the support vector. Here, we set $C=1$.

The GPC is a probabilistic pattern recognition model based on Laplace approximation that can achieve probabilistic classification and an equivalent performance to SVM (35). Making GPC predictions is a 2-step process. First, the distribution of the latent variable at training points is computed, and then its expectation to produce a probabilistic prediction is computed. Class probabilities of GPC are derived from integrating over the entire distribution for the latent function at the test data point. In this study, classification accuracy (ACC) was calculated to evaluate the performance of the classifying model. The area under curve (AUC) of the receiver operating characteristic (ROC) curve and confusion matrix were also reported in this study to assess the performance of the different classifiers. Each classifier was retrained 1,000 times, and P values were computed for prediction ACC. A P value <0.05 was considered statistically significant.

Computation of weight map for prediction models

The model weights represented the contribution of each dataset (voxel- or region-based) for the predictive model. In this study, we computed the model weights and plotted them as brain images to display the decision functions of the predictive models for regression and classification. The decision function was defined as follows:

$$f(x_i) = (w \cdot x_i) + b \quad [6]$$

Here, $(w \cdot x_i)$ represents the dot product between the

weight vector $w \in R^1$ corresponding to a data vector x_i , and b is a bias term. Thus, for the model using VB data, each voxel corresponded to a weight vector, and the weights of the brain voxels were averaged in each region of the AAL3 atlas. Only regions with a positive contribution to the predictive model were shown, and these regions were ranked in ascending order based on their weights. The expected ranking (ER) of each region was the ranking averaged across folds.

Validation of the prediction models

Data from 56 participants (VaD =14, VaMCI =23, CH =19) were used to train the machine learning models, and data from the other 23 participants (VaD =7, VaMCI =10, CH =6) were used as separate independent test samples to evaluate the stability and generalizability of the predictive models trained in this study. The prediction models were analyzed using a training, validation, and testing approach with a nested leave-one-subject-out cross-validation procedure to tune the parameters and estimate the performance of our models. The predicted values from each fold were appended together, and cross-validation accuracy was calculated to evaluate the performance of our models on the training set. The overall procedure is shown in *Figure 1*.

Statistical analysis

A chi-square test was used to test for gender-based differences, and one-way analysis of variance (ANOVA) was performed to test for group differences in age and education levels. We used ANCOVA, adjusted for age, education level, and gender, to test for group differences in Fazekas scores and neurological tests. All statistical analyses were performed with the Statistical Package for Social Science (SPSS; 22.0, IBM Corp., Armonk, NY, USA).

Results

Demographic and neuropsychological information

A total of 79 individuals aged between 50 and 85 years with WMLs were selected for this study and grouped as follows: 21 participants with VaD, 33 with VaMCI, and 25 with CH. The demographic characteristics and neuropsychological performance of the participants are summarized in *Table 2*. Significant differences between the groups were found in all the neuropsychological testing scores ($P<0.05$). Participants with CH performed best in the neuropsychological

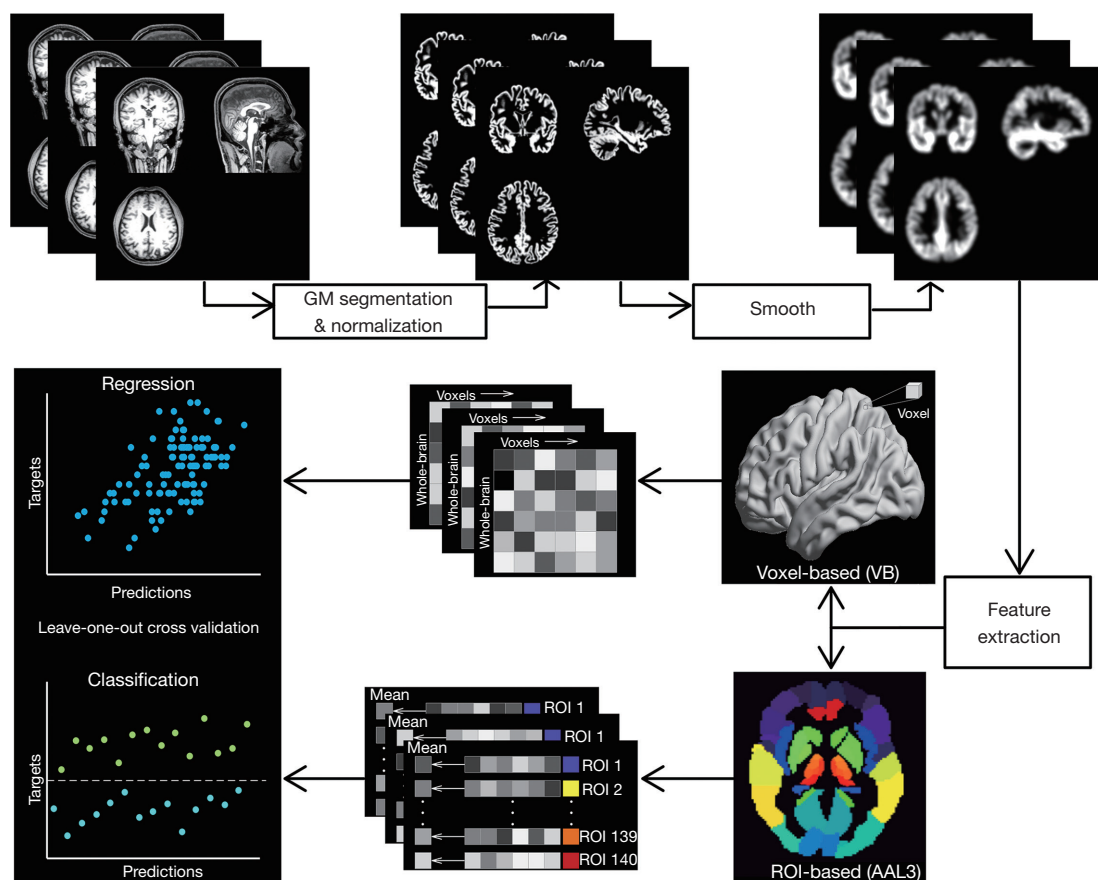


Figure 1 The overall study procedure. Firstly, GM images were segmented for each participant and normalized to the standard MNI space. Then the segmented GM images were smoothed using an 8 mm Gaussian kernel. Next, the input data were derived from the smoothed GM images at voxel-level and region-level. Finally, the computed data were submitted to the prediction models for regression and classification analysis, and the weight distribution of each model was extracted. GM, gray matter; MNI, Montreal Neurological Institute.

testing, followed by participants in the VaMCI group, and finally participants in the VaD group, who had the worst neuropsychological testing scores. The age, gender, and education level of the participants did not produce any significant differences among these groups ($P > 0.05$).

Prediction results of RVR

To investigate whether GM information can predict the severity of cognitive decline in older individuals with WMLs, we used the RVR model to assess GM information from the neuropsychological test scores with a nested leave-one-subject-out cross-validation scheme in model training. We then applied our trained models to the imaging data of 23 participants from the unseen testing dataset. The results of the predictions with VB and AAL3 data derived

from GM images are reported in *Table 3*. Our findings suggest that multi-domain cognitive performances can be accurately predicted based on the pattern of GM atrophy in older individuals with WMLs. In 6 of 9 neuropsychological tests, the clinically measured scores were predicted by RVR models with both VB and AAL3 data used in the training sample ($P < 0.05$), which included scores for MMSE, MoCA, visuospatial/executive, attention, delayed recall, and orientation. The well-trained RVR models were then applied to independent testing datasets and predicted cognitive performance of the older individuals with WMLs with an accuracy comparable to that of the training set.

The RVR models using VB data (RVR-VB) yielded better results in the training sample than models that used AAL3 data (RVR-AAL3). However, in the validation results, the predictive performance of the RVR-AAL3 was

Table 2 Demographics and clinical characteristics of the participants

Variable	Total (Ntr/Nte), 79 (56/23)	VaD, 21 (14/7)	VaMCI, 33 (23/10)	CH, 25 (19/6)	F/χ^2	P	VaD vs. CH	VaMCI vs. CH	VaD vs. VaMCI
Demographics									
Male/Female	38/41	13/8	15/18	10/15	2.35	0.308	2.19	0.17	1.39*
Age	64.0±9.8	65.5±13.0	62.3±9.5	65.0±6.5	0.88	0.421	0.48	-2.71	3.19
Education	11.3 ±3.3	11.3±3.2	11.2±3.5	11.4±3.0	0.003	0.975	-0.11	-0.20	0.09
Fazekas score	1.8±0.8	2.1±0.9	1.7±0.8	1.6±0.6	2.77	0.069	0.53	0.18	0.35
Neuropsychological tests									
CDR	-	1.3±0.7	0.5±0.1	0	72.20	<0.001	1.32***	0.51***	0.81***
MMSE	26.7±3.7	22.5±4.2	27.6±2.2	29.1±1.1	36.91	<0.001	-6.30***	-1.62	-4.63***
MoCA	22.5±5.3	15.6±4.0	23.2±2.6	27.4±1.4	107.7	<0.001	-11.60***	-4.32***	-7.29***
Visuospatial/ executive	3.5±1.5	1.7±0.9	3.6±1.1	4.8±0.5	66.74	<0.001	-3.11***	-1.23***	-1.88***
Naming	2.7±0.6	2.3±0.8	2.8±0.5	2.9±0.4	5.16	0.008	-0.55**	-0.10	-0.46*
Attention	5.0±1.4	3.5±1.4	5.3±1.2	5.8±0.4	31.04	<0.001	-2.20***	-0.66*	-1.54***
Language	1.8±0.9	1.0±0.8	2.0±0.6	2.3±0.7	21.73	<0.001	-1.32***	-0.36	-0.99***
Abstraction	1.5±0.7	1.1±0.8	1.6±0.7	1.6±0.6	4.19	0.017	-0.51*	-0.01	-0.50*
Delayed recall	2.3±1.6	1.2±0.9	1.9±1.4	3.8±1.3	25.85	<0.001	-2.51***	-1.86***	-0.65
Orientation	5.5±1.0	4.6±1.4	5.7±0.6	6.0±0.2	17.81	<0.001	-1.31***	-0.30	-1.01***

Values are presented as mean (\pm standard deviation). * $P<0.05$; ** $P<0.01$; *** $P<0.001$. Ntr, number of subjects for model training; Nte, number of subjects for model test; CH, good cognitive health; VaMCI, vascular mild cognitive impairment; VaD, vascular dementia; CDR, clinical dementia rating; MMSE, Chinese version of the Mini-Mental State Examination; MoCA, Beijing version of the Montreal Cognitive Assessment.

generally superior to the RVR-VB, except for predictions of delayed recall and orientation scores. In the prediction of MoCA scores, the RVR-VB yielded the best predictive performance for the training data (CORR =0.69, $P=0.001$), while RVR-AAL3 yielded a CORR =0.65 ($P=0.001$). In the validation results for the MoCA scores, the RVR-AAL3 model yielded a CORR =0.47 ($P=0.022$) achieving a slightly better predictive performance than the RVR-VB model (CORR =0.46, $P=0.029$). In predicting the MMSE scores, the RVR models using the testing data performed significantly better than the models that used the training data. The corresponding scatter plots for predicted clinical scores versus observed MoCA scores for the RVR-VB and the RVR-AAL3, respectively, are shown in *Figure 2*.

Weight maps of the RVR-VB and RVR-AAL3 had similar spatial distribution, but the rank of brain regions may have had little effect on the prediction of the same clinical test scores. The right nucleus accumbens, the

bilateral temporal pole along the middle temporal gyrus, and subdivisions of thalamic nuclei were the most salient regions for both the RVR-VB and RVR-AAL3 models. In predictions of the MoCA, attention and orientation scores, the right nucleus accumbens had the 5 highest weights in both the RVR-VB and RVR-AAL3 models. For the prediction of the MMSE test scores, this region was ranked 3rd in RVR-VB model and 6th in the RVR-AAL3 model. The weight maps of the RVR-VB- and RVR-AAL3-model predicted scores of MMSE and MoCA are shown in *Figure 3*. The list of the 5 selected regions with highest contribution to the RVR-VB and RVR-AAL3 models are summarized in *Tables S1,S2*.

Classification results of SVM and GPC

Table 4 shows that both SVM and GPC could distinguish between VaD and CH, and VaMCI and CH in the training

Table 3 Prediction of neuropsychological test scores in older individuals with WMLs based on GM microstructure assessed on T1 weighted (T1W) images, using relevance vector regression (RVR) models. $P < 0.05$ was considered significant for the predictive performance of RVR models using the training data

Model	Neuropsychological tests	Data type	Training set			Testing set		
			CORR	R ²	MSE	CORR	R ²	MSE
RVR	MMSE	VB	0.37	0.14	12.99	0.60	0.36	1.76
		AAL3	0.36	0.13	13.60	0.65	0.42	7.13
	MoCA	VB	0.69	0.48	13.79	0.46	0.21	0.77
		AAL3	0.65	0.43	15.46	0.47	0.23	3.02
	Visuospatial/executive	VB	0.44	0.19	1.91	0.35	0.12	4.39
		AAL3	0.43	0.19	1.96	0.37	0.14	4.25
	Attention	VB	0.60	0.36	1.08	0.31	0.10	0.15
		AAL3	0.60	0.36	1.10	0.36	0.13	1.74
	Delayed recall	VB	0.51	0.26	1.76	0.39	0.16	0.55
		AAL3	0.49	0.24	1.86	0.36	0.13	1.68
	Orientation	VB	0.62	0.39	0.58	0.62	0.38	0.77
		AAL3	0.54	0.29	0.70	0.61	0.38	0.75

VB, voxel-based; AAL3, automated anatomical labelling atlas 3; CORR, Pearson's correlation coefficient; R², coefficient of determination; MSE, mean-squared error; MMSE, Chinese version of the Mini-Mental State Examination; MoCA, Beijing version of the Montreal Cognitive Assessment; WMLs, white matter lesions; GM, gray matter; RVR, relevance vector regression; VB, voxel-based; AAL3, automated anatomical labelling atlas 3.

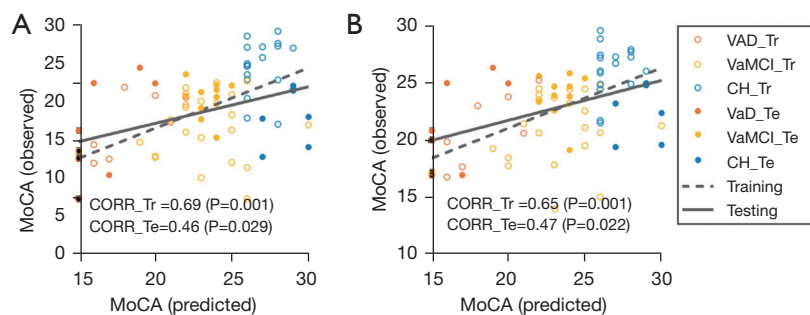


Figure 2 Predicting MoCA using the RVR model based on GM images of older individuals with WMLs. (A) Scatter plot of the predicted MoCA scores *vs.* the observed scores based on voxel-based (VB) data. (B) Scatter plot of predicted MoCA scores *vs.* the observed scores based on AAL3 data. The scores of participants in the CH group are shown in blue, those of participants diagnosed with VaMCI clinically are shown in orange, and those of participants diagnosed with VaD are shown in yellow. The lines indicate the linear correlation between predicted and actual MoCA scores. A solid line shows testing results and a dashed line shows training results. MoCA, Beijing version of the Montreal Cognitive Assessment; WMLs, white matter lesions; RVR, relevance vector regression; GM, gray matter; AAL3, automated anatomical labelling atlas 3; Tr, training; Te, testing; CH, good cognitive health; VaMCI, vascular mild cognitive impairment; VaD, vascular dementia.

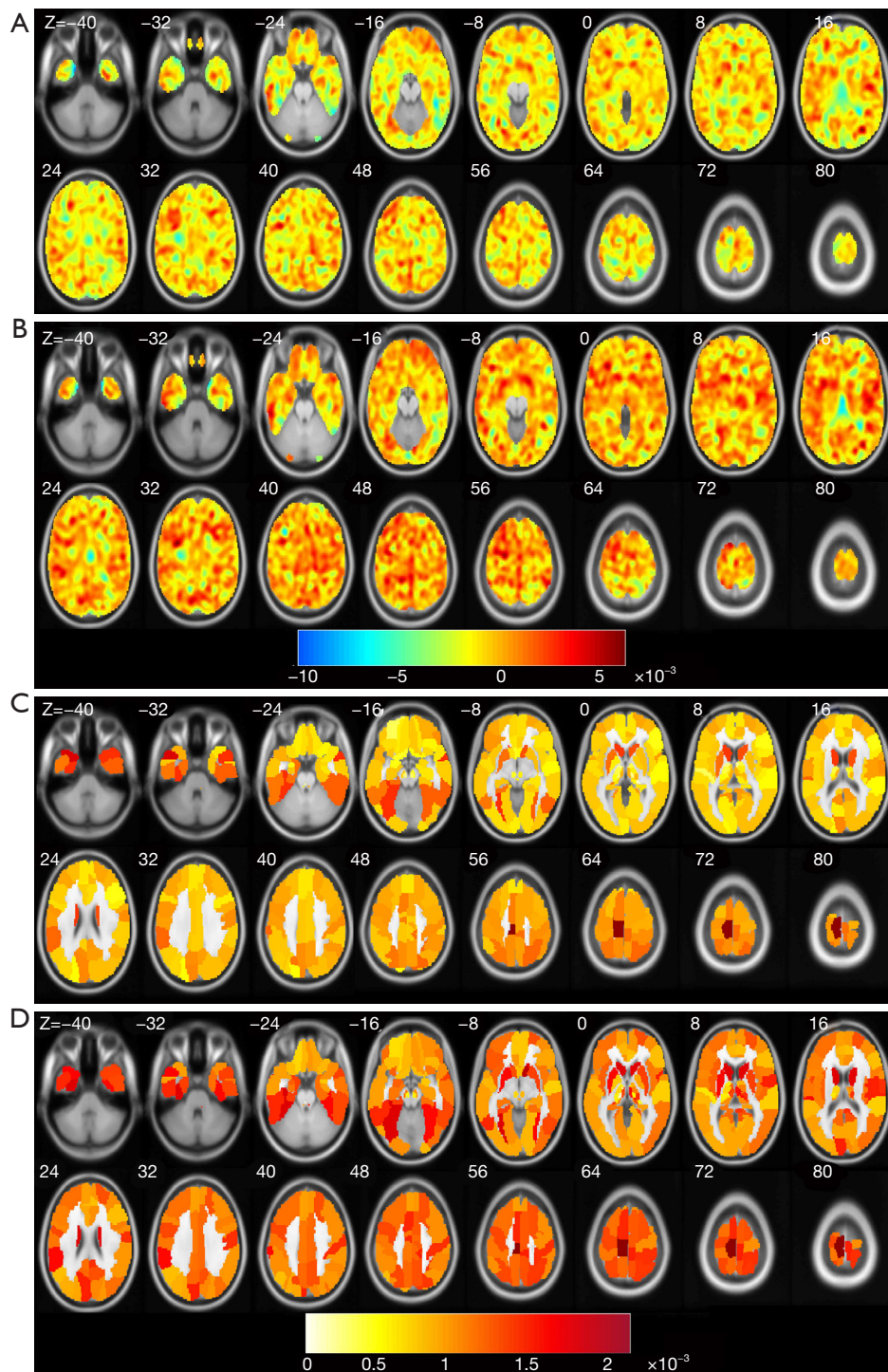


Figure 3 Weight maps of RVR model predictions of MMSE and MoCA scores in older individuals with WMLs, based on the GM atrophy patterns. (A) and (B) are weight maps for RVR-VB predictions of MMSE and MoCA scores, respectively, in which the weights of voxels located in the same regions were averaged across the AAL3. (C) and (D) are weight maps for RVR-AAL3 predictions of MMSE and MoCA scores respectively. RVR, relevance vector regression; WMLs, white matter lesion; MMSE, Chinese version of the Mini-Mental State Examination; MoCA, Beijing version of the Montreal Cognitive Assessment; GM, gray matter; VB, voxel-based; AAL3, the automated anatomical labelling atlas 3.

Table 4 Comparison of classification performance of the SVM and GPC algorithms using VB and AAL3 data

Model	Classification	Data type	ACC% (P)	ACC% for VaD (P)	ACC% for VaMCI (P)	ACC% for CH (P)
SVM	VaD_Tr vs. CH_Tr	VB	93.94 (0.001)	94.74 (0.001)	–	89.47 (0.004)
		AAL3	93.94 (0.001)	100 (0.001)	–	89.47 (0.217)
	VaMCI_Tr vs. CH_Tr	VB	95.24 (0.001)	–	100 (0.001)	89.47 (0.001)
		AAL3	95.24 (0.001)	–	100 (0.002)	89.47 (0.001)
	VaD_Tr vs. VaMCI_Tr	VB	62.16 (0.179)	35.71 (0.236)	78.26 (0.339)	–
		AAL3	67.57 (0.026)	50.00 (0.031)	78.26 (0.889)	–
GPC	VaD_Tr vs. CH_Tr	VB	93.94 (0.001)	100 (0.001)	–	89.47 (0.002)
		AAL3	93.94 (0.001)	100 (0.001)	–	89.47 (0.007)
	VaMCI_Tr vs. CH_Tr	VB	95.24 (0.001)	–	100 (0.001)	89.47 (0.001)
		AAL3	95.24 (0.001)	–	100 (0.001)	89.47 (0.001)
	VaD_Tr vs. VaMCI_Tr	VB	59.46 (0.227)	28.57 (0.221)	78.26 (0.601)	–
		AAL3	56.76 (0.667)	21.43 (0.106)	78.26 (0.969)	–

CH, good cognitive health; VaMCI, vascular mild cognitive impairment; VaD, vascular dementia; VB, voxel-based; AAL3, automated anatomical labelling atlas 3; ACC, accuracy; Tr, training; SVM, support vector machine; GPC, Gaussian process classifier.

set (VaD *vs.* CH ACC =93.94%, P=0.001; VaMCI *vs.* CH ACC =95.24%, P=0.001), but that only SVM-AAL3 could distinguish between VaD and VaMCI (VaD *vs.* VaMCI ACC =67.57%, P=0.026) in the training set. The ROC curves and confusion matrices of these classifiers on the training dataset are presented in *Figure 4*. Overall, the predictive performance of the SVM classifiers was superior to the GPC models, and the trained SVM classification models were applied to participant data from an independent unseen testing sample. As can be seen from *Table 5*, the SVM models stably predicted individual-level subtypes in the testing dataset (VaD *vs.* CH ACC =76.92%, VaMCI *vs.* CH ACC =87.50%, and VaD *vs.* VaMCI =70.59%. Confusion matrices for the SVM classifiers predicting with the testing sample were presented in *Figure S2*. As similar to the results of RVR models, there was no significant difference in the predictive power between model using VB features and the one using AAL3 features.

The SVM and GPC classifiers had similar distribution of weight maps, as did the VB and AAL3 input data. The weight maps for the classifiers of VaD versus CH, VaMCI versus CH, and VaD versus VaMCI are presented in *Figures 5-7* respectively, and the 5 regions with highest weights are reported in *Tables S3,S4*. Classifiers of VaD versus CH and VaMCI versus CH had similar distribution of weight maps, with both SVM and GPC. The bilateral putamen and

thalamic subregions, including the ventral posterolateral and ventral lateral nuclei, and the mediodorsal lateral parvocellular, were the most important ROIs. In the classification of VaD versus VaMCI, the right posterior cingulate gyrus, the bilateral nucleus accumbens, the left inferior temporal gyrus, and the caudate nucleus had higher weights than other brain regions.

Discussion

The goal of this study was to investigate whether GM information could predict the severity of cognitive decline in older individuals with WMLs. The results of multivariate RVR analyses revealed the capacity of GM atrophy patterns to predict the decline of multiple cognitive functions in the study population, including general and domain-specific cognition, including memory, visuospatial/executive function, orientation and attention. The results of the classification models used indicated that GM volume can provide differential diagnostic information that distinguishes between VaD, VaMCI, and CH in older individuals with WMLs using both training and testing datasets.

Previous studies have suggested that WMLs are associated with GM atrophy, which may consequently cause cognitive decline (36,37), and increase the risk of dementia (38), especially in dementia-related regions such as the temporal

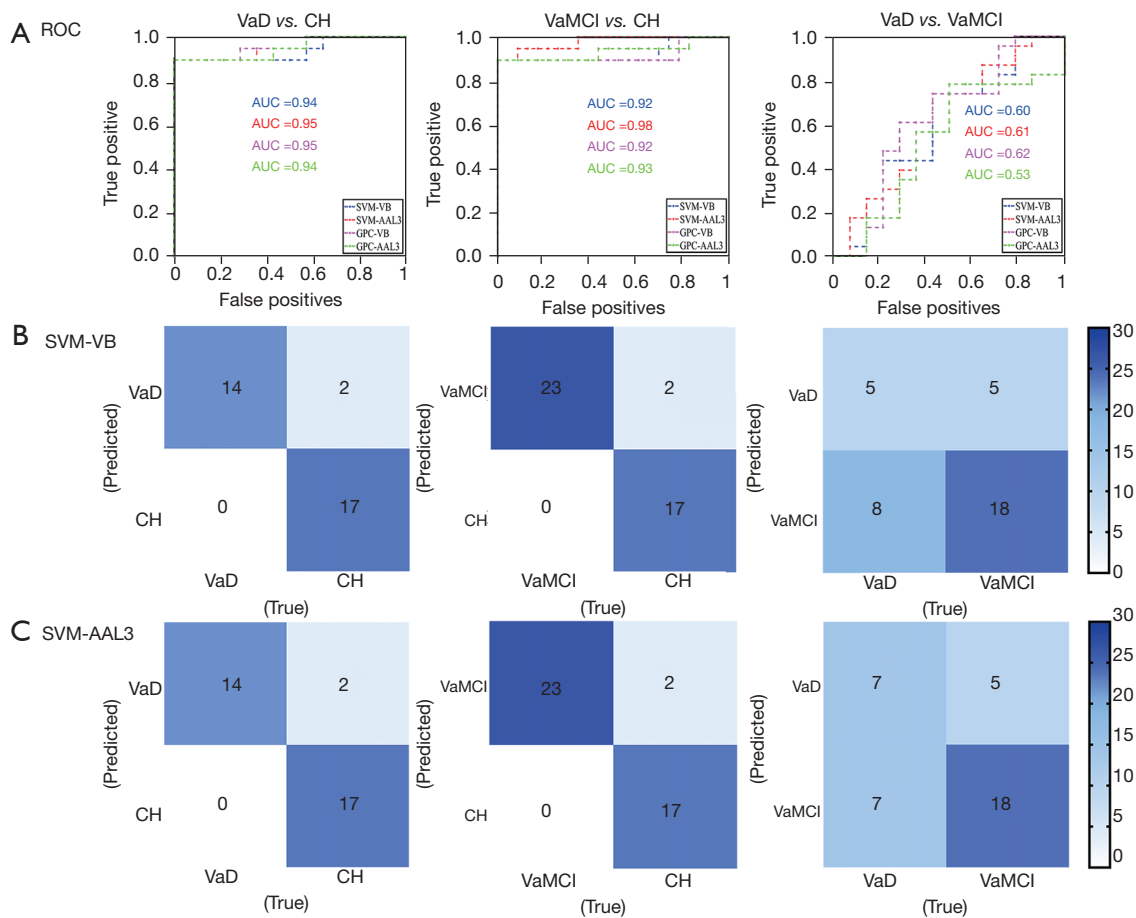


Figure 4 ROC curves and confusion matrices of SVM and GPC. (A) ROC curves of SVM and GPC models for the training dataset; (B) confusion matrices of SVM with VB data; (C) confusion matrices of SVM with AAL3 data. The left column shows VaD vs. CH, the middle column shows VaMCI vs. CH, and the right column shows VaD vs. VaMCI. ROC, receiver operating characteristic; SVM, support vector machine; GPC, Gaussian process classifiers; VB, voxel-based; AAL3, automated anatomical labelling atlas 3; CH, good cognitive health; VaMCI, vascular mild cognitive impairment; VaD, vascular dementia.

Table 5 Classification performance of the SVM algorithm prediction using the testing dataset

Model	Classification	Data type	ACC%	ACC% for VaD	ACC% for VaMCI	ACC% for CH
SVM	VaD_Te vs. CH_Te	VB	69.23	57.14	–	83.33
		AAL3	76.92	57.14	–	100.00
	VaMCI_Te vs. CH_Te	VB	87.50	–	100.00	71.43
		AAL3	68.75	–	63.00	90.00
	VaD_Te vs. VaMCI_Te	VB	70.59	50.00	80.00	–
		AAL3	64.71	50.00	70.00	–

SVM, support vector machine; CH, good cognitive health; VaMCI, vascular mild cognitive impairment; VaD, vascular dementia; VB, voxel-based; AAL3, automated anatomical labelling atlas 3; ACC, accuracy; Te, testing.

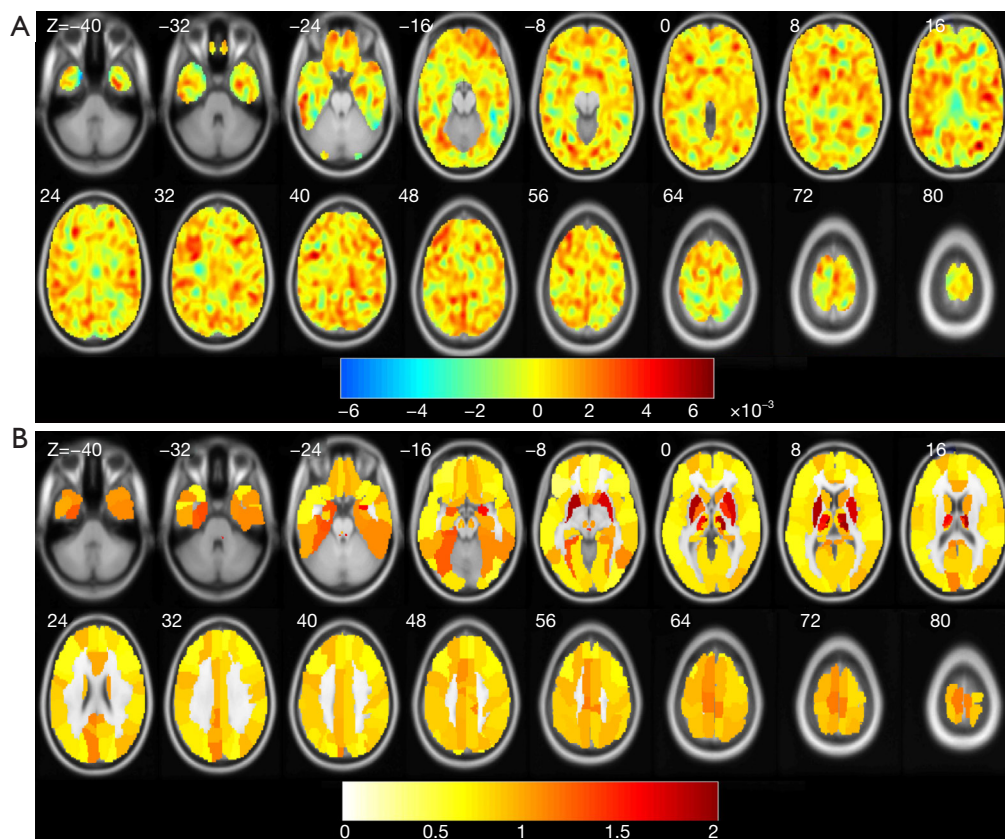


Figure 5 Weight maps for SVM classifications of VaD vs. CH. (A) Voxel-level weight maps of SVM-VB, in which the weights of voxels were averaged across the AAL3 atlas. (B) Atlas-based region-level weight maps of SVM-AAL3. SVM, support vector machine; VB, voxel-based; AAL3, automated anatomical labelling atlas 3; CH, good cognitive health; VaD, vascular dementia.

and frontal cortex. It has been suggested that WMLs are associated with frontal lobe function, especially executive function and processing speed (7,39,40), although some studies have also found a link with the decline of episodic memory (41,42). However, the underlying mechanisms and the effects of WMLs on the development of cognitive impairment are still unclear. Moreover, though significant synergistic interactions have been found between WMLs and brain atrophy in multi-domain cognitive functions across time and the rate of cognitive decline, the underlying mechanisms remain unclear (43,44). Applying multivariate machine learning techniques, such as RVR, to MRI data allows for quantitative prediction of clinical performance with excellent accuracy as machine learning models are more sensitive to the subtle alterations caused by psychiatric or neurodegenerative disorders (45,46). This study found that a few brain regions, including the right nucleus accumbens, the bilateral temporal pole along the

middle temporal gyrus, the left paracentral lobule, and the bilateral thalamic subregions, including the mediodorsal lateral, mediodorsal medial magnocellular, and the lateral pulvinar nuclei contributed most to the accurate prediction of cognition performance in participants with WMLs. Most of these regions are part of the default mode network and limbic system, the dysfunction of which has been implicated in the pathophysiology of dementia and MCI (47-49). Nie *et al.* found that atrophy in the right posterior nucleus accumbens and the medial-ventral area of the right thalamus significantly correlated with worse clinical scores, such as the MMSE and MoCA scores, in the elderly (50). Lee *et al.* reported that regions of the right middle-temporal pole and precentral gyrus might be pivotal neural substrates of cognitive reversal in the aging population (51).

Machine learning techniques like the GPC and SVM algorithms can accurately distinguish between VaD and CH (11). As there is no agreed criteria for the diagnosis

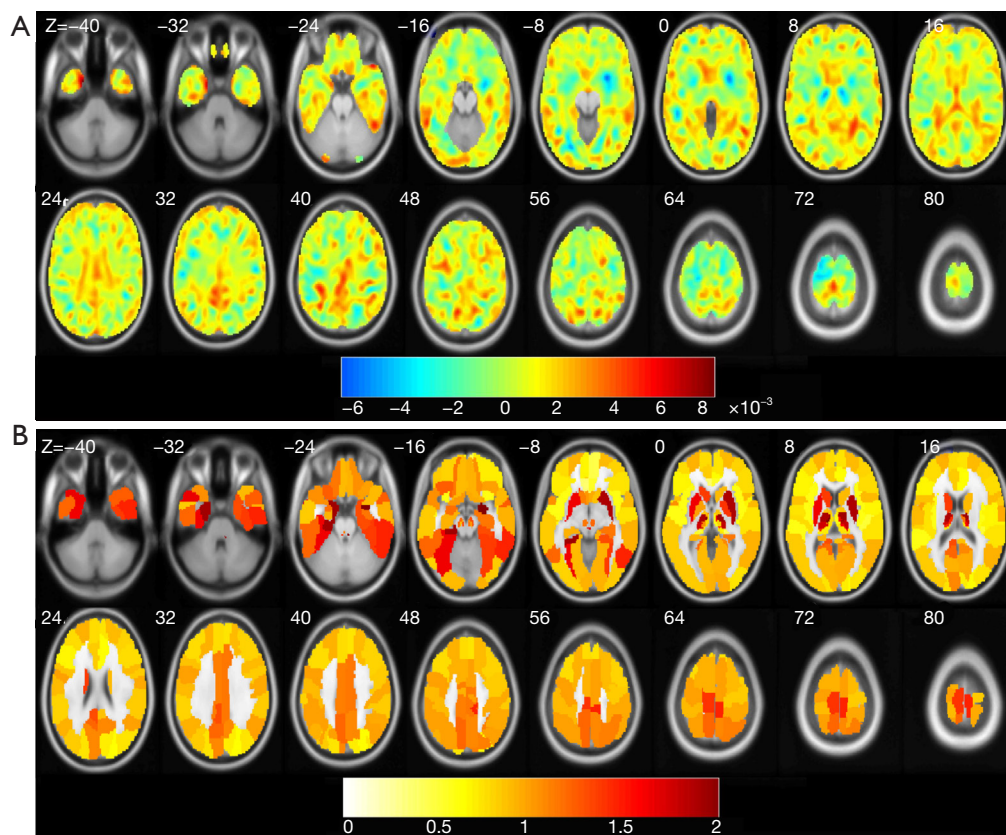


Figure 6 Weight maps for SVM for classifications of VaMCI vs. CH. (A) Voxel-level weight maps of SVM-VB. (B) Atlas-based region-level weight maps of SVM-AAL3. SVM, support vector machine; VB, voxel-based; CH, good cognitive health; VaMCI, vascular mild cognitive impairment; AAL3, automated anatomical labelling atlas 3.

of VaD and VaMCI, using automatic classifiers may help increase diagnostic power on an individual level (52,53). In both the VaD and VaMCI versus CH classification tasks, the 2 algorithms achieved cross-validation accuracies greater than 90% on the training sample, and a testing accuracy of 76.92% for the VaD versus CH and 87.50% for VaMCI versus CH. Furthermore, regions of the bilateral putamen and thalamic subregions shown high predictive contributions for both SVM and GPC classification models. The 5 brain regions that contributed most to predictive accuracy included the bilateral lateral ventral lateral nucleus, the left mediodorsal parvocellular, and the right ventral posterolateral nucleus. This is consistent with the results reported by Skrobot *et al.*, who identified a lateral/medial axis across the thalamus and established associations between anatomical nuclei and GM volume (53). Parnaudeau *et al.* also suggested that there is a causal relationship between abnormalities of

the mediodorsal thalamus-prefrontal cortex and cognitive impairment (54). Our results were consistent with previous studies on widespread subcortical and cortical structural alternations (55). Liu *et al.* found that cortex thinning or volume decline in the cingulate, thalamus, caudate nucleus, and amygdala was associated with cognitive impairment in subcortical ischemic VaD (56). The strong predictive power of data from the putamen in distinguishing between healthy individuals and those with VaD and VaMCI may be inferred from our results. As in the regression models, the medial-ventral parts of the bilateral thalamic nuclei were also found to be important in the classification of VaD and VaMCI versus CH, which has been reported as significantly related to the clinical scores of patients with MCI and AD (50). Plachti *et al.* found that the pattern of structural covariance of the hippocampus voxels in the older population was primarily reduced to the parieto-occipital and frontal-medial brain regions that conform the default

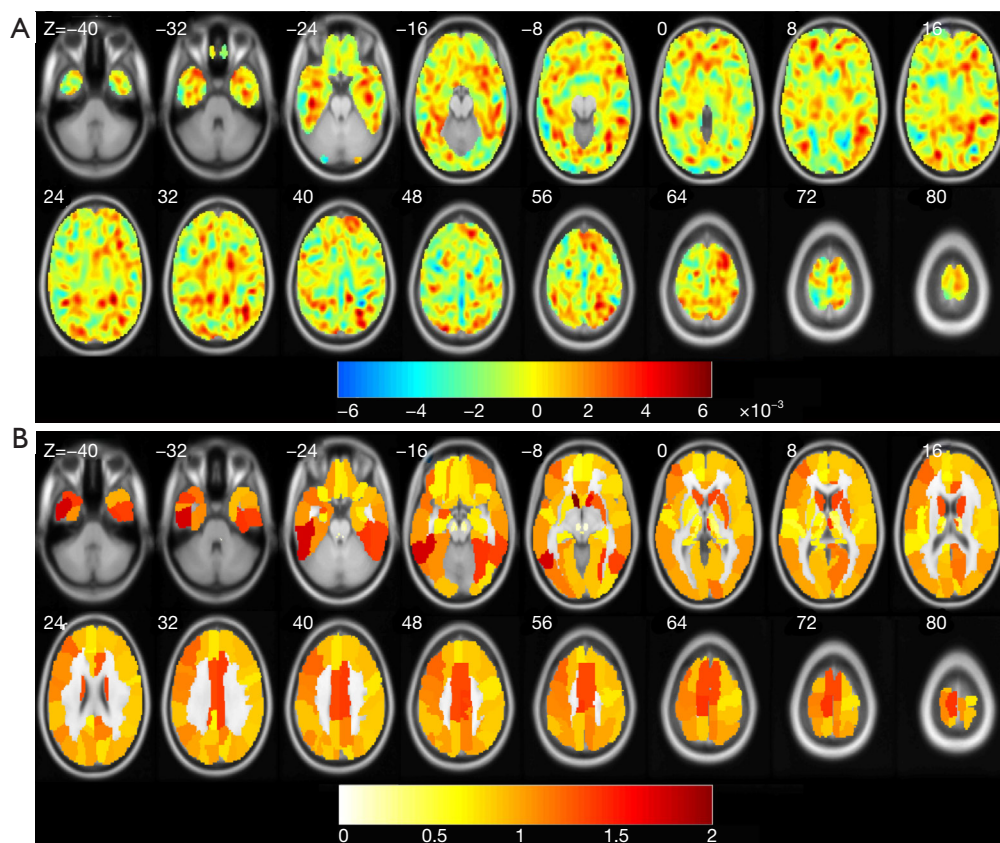


Figure 7 Weight maps for SVM classifications of VaD vs. VaMCI. (A) Voxel-level weight maps of SVM-VB. (B) Atlas-based region-level weight maps of SVM-AAL3. SVM, support vector machine; VaMCI, vascular mild cognitive impairment; VaD, vascular dementia; VB, voxel-based; AAL3, automated anatomical labelling atlas 3.

mode network, such as the posterior cingulate cortex, the precuneus, and the frontal medial cortex (57). In our results, the high predictive power of data from the thalamic nuclei and other subcortical brain regions of the participants with WMLs might be linked to the co-atrophy of these regions.

Much previous research has used GM density maps as data for SVM in dementia classification, achieving promising accuracy ranging from 81% to 95% (21,58). Our results showed that, comparing with the classification between VaD and VaMCI, the ACC was higher in classifying VaD or VaMCI from CH. This is consistent with previous studies that indicated that the challenges in differentiating between MCI and dementia led to lower accuracies of 62.07% to 90% (59). Overall, our results showed that GM density maps could be used to distinguish between a cognitively healthy control group and groups with different levels of cognitive impairment in an older cohort with WMLs. Cross-validation can reduce the

effect of random variations and ensure the generality and robustness of the classification model.

Some limitations should be mentioned. In contrast to previous studies that reported significant results using data from the frontal brain regions, data from the frontal gyrus were rarely found to contribute significantly to the models predicting cognitive impairment in older individuals with WMLs in this study, especially for RVR. A possible reason for this may be the high collinearity between the data from frontal regions and other selected data. As discussed in previous research, multivariate predictive models cannot provide a clear explanation of why a specific region or feature has a higher weight in the decision function of a model (60,61). The predictive performance of the testing dataset showed an overall drop compared to the prediction results of the training dataset, especially for the SVM classifications of VaD and VaMCI versus CH. The decreased predictive performance of the test dataset may be

associated with the limited number of older individuals with WMLs enrolled in this study and the slight differences in the relative numbers of participants from each group in the training and test datasets. Our findings need to be replicated in larger independent test samples to confirm and test the generalizability of our models, as the metrics only reflected the predictive accuracy of these models for our particular datasets. Another limitation of the current study was that although the multivariate regression and classification models capitalized on complex GM atrophy patterns to make predictions of cognitive decline in older individuals with WMLs, we have no specific understanding of how and why the atrophy of specific GM regions drives the models' predictions. Alternative approaches, such as multiple kernel learning models, should also be used to investigate why some regions provide greater predictive data and to test the reproducibility and stability of our findings. Further work is required to obtain a deeper understanding of the effects of WML pathologies and brain atrophy on cognitive impairment in the aging population.

Conclusions

In conclusion, our results provide preliminary evidence that decline in multidomain cognitive functions in older people with WMLs can be predicted using machine learning techniques, and that it is possible to classify patients with VaD, VaMCI, and CH based on their patterns of GM atrophy. These findings may be of use in the early clinical diagnosis of VaMCI and VaD.

Acknowledgments

The abstract of this manuscript was presented at the World Statistics Congress (WSC) 2021 as an e-poster.

Funding: This work was funded with grants from the National Key Technology Research and Development Program of China (Nos. 2018YFC2002300, 2018YFC2002302, 2020YFC2004102), the Scientific Research Project of Beijing Municipal Education Commission (No. KM202110025022), the National Natural Science Foundation of China (NSFC; Nos. 31600933, 81972144, 81972148, 31872785), the Health commission of Hubei Province scientific research project (No. WJ2019H366), and the Wuhan Health Scientific Research Project (No. WX19Q19).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at <https://atm.amegroups.com/article/view/10.21037/atm-21-3571/rc>

Data Sharing Statement: Available at <https://atm.amegroups.com/article/view/10.21037/atm-21-3571/dss>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://atm.amegroups.com/article/view/10.21037/atm-21-3571/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the Beijing Tiantan Hospital, Capital Medical University, China (KYSB2016-023) and informed consent was provided by all participants.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Sun X, Liang Y, Wang J, et al. Early frontal structural and functional changes in mild white matter lesions relevant to cognitive decline. *J Alzheimers Dis* 2014;40:123-34.
2. Prins ND, Scheltens P. White matter hyperintensities, cognitive impairment and dementia: an update. *Nat Rev Neurol* 2015;11:157-65.
3. Habes M, Sotiras A, Erus G, et al. White matter lesions: Spatial heterogeneity, links to risk factors, cognition, genetics, and atrophy. *Neurology* 2018;91:e964-75.
4. Yu D, Hennebelle M, Sahlas DJ, et al. Soluble Epoxide

- Hydrolase-Derived Linoleic Acid Oxylipins in Serum Are Associated with Periventricular White Matter Hyperintensities and Vascular Cognitive Impairment. *Transl Stroke Res* 2019;10:522-33.
5. Brickman AM, Zahodne LB, Guzman VA, et al. Reconsidering harbingers of dementia: progression of parietal lobe white matter hyperintensities predicts Alzheimer's disease incidence. *Neurobiol Aging* 2015;36:27-32.
 6. Aribisala BS, Valdés Hernández MC, Royle NA, et al. Brain atrophy associations with white matter lesions in the ageing brain: the Lothian Birth Cohort 1936. *Eur Radiol* 2013;23:1084-92.
 7. Debette S, Markus HS. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ* 2010;341:c3666.
 8. Caunca MR, Siedlecki K, Cheung YK, et al. Cholinergic White Matter Lesions, AD-Signature Cortical Thickness, and Change in Cognition: The Northern Manhattan Study. *J Gerontol A Biol Sci Med Sci* 2020;75:1508-15.
 9. Tuladhar AM, Reid AT, Shumskaya E, et al. Relationship between white matter hyperintensities, cortical thickness, and cognition. *Stroke* 2015;46:425-32.
 10. Habes M, Erus G, Toledo JB, et al. Regional tract-specific white matter hyperintensities are associated with patterns to aging-related brain atrophy via vascular risk factors, but also independently. *Alzheimers Dement (Amst)* 2018;10:278-84.
 11. Pellegrini E, Ballerini L, Hernandez MDCV, et al. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimers Dement (Amst)* 2018;10:519-35.
 12. Ahmed MR, Zhang Y, Feng Z, et al. Neuroimaging and Machine Learning for Dementia Diagnosis: Recent Advancements and Future Prospects. *IEEE Rev Biomed Eng* 2019;12:19-33.
 13. Habes M, Pomponio R, Shou H, et al. The Brain Chart of Aging: Machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans. *Alzheimers Dement* 2021;17:89-102.
 14. Gong Q, Li L, Du M, et al. Quantitative prediction of individual psychopathology in trauma survivors using resting-state fMRI. *Neuropsychopharmacology* 2014;39:681-7.
 15. Tipping ME. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research* 2001;1:211-44.
 16. Sui J, Jiang R, Bustillo J, et al. Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises. *Biol Psychiatry* 2020;88:818-28.
 17. Tognin S, Pettersson-Yeo W, Valli I, et al. Using structural neuroimaging to make quantitative predictions of symptom progression in individuals at ultra-high risk for psychosis. *Front Psychiatry* 2013;4:187.
 18. Ranlund S, Rosa MJ, de Jong S, et al. Associations between polygenic risk scores for four psychiatric illnesses and brain structure using multivariate pattern recognition. *Neuroimage Clin* 2018;20:1026-36.
 19. Orrù G, Pettersson-Yeo W, Marquand AF, et al. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev* 2012;36:1140-52.
 20. Challis E, Hurley P, Serra L, et al. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. *Neuroimage* 2015;112:232-43.
 21. Rathore S, Habes M, Iftikhar MA, et al. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage* 2017;155:530-48.
 22. Wahlund LO, Barkhof F, Fazekas F, et al. A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke* 2001;32:1318-22.
 23. Wardlaw JM, Smith EE, Biessels GJ, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol* 2013;12:822-38.
 24. Katzman R, Zhang MY, Ouang-Ya-Qu, et al. A Chinese version of the Mini-Mental State Examination; impact of illiteracy in a Shanghai dementia survey. *J Clin Epidemiol* 1988;41:971-8.
 25. Nie K, Zhang Y, Wang L, et al. A pilot study of psychometric properties of the Beijing version of Montreal Cognitive Assessment in patients with idiopathic Parkinson's disease in China. *J Clin Neurosci* 2012;19:1497-500.
 26. Lin L, Xue Y, Duan Q, et al. Microstructural White Matter Abnormalities and Cognitive Dysfunction in Subcortical Ischemic Vascular Disease: an Atlas-Based Diffusion Tensor Analysis Study. *J Mol Neurosci* 2015;56:363-70.
 27. Chen X, Zhang R, Xiao Y, et al. Reliability and Validity of

- the Beijing Version of the Montreal Cognitive Assessment in the Evaluation of Cognitive Function of Adult Patients with OSAHS. *PLoS One* 2015;10:e0132361.
28. Rondina JM, Ferreira LK, de Souza Duran FL, et al. Selecting the most relevant brain regions to discriminate Alzheimer's disease patients from healthy controls using multiple kernel learning: A comparison across functional and structural imaging modalities and atlases. *Neuroimage Clin* 2018;17:628-41.
 29. Wang J, Liang Y, Chen H, et al. Structural changes in white matter lesion patients and their correlation with cognitive impairment. *Neuropsychiatr Dis Treat* 2019;15:1355-63.
 30. Guggenmos M, Scheel M, Sekutowicz M, et al. Decoding diagnosis and lifetime consumption in alcohol dependence from grey-matter pattern information. *Acta Psychiatr Scand* 2018;137:252-62.
 31. Wagner F, Duering M, Gesierich BG, et al. Gray Matter Covariance Networks as Classifiers and Predictors of Cognitive Function in Alzheimer's Disease. *Front Psychiatry* 2020;11:360.
 32. Rolls ET, Huang CC, Lin CP, et al. Automated anatomical labelling atlas 3. *Neuroimage* 2020;206:116189.
 33. Schrouff J, Rosa MJ, Rondina JM, et al. PRoNTo: pattern recognition for neuroimaging toolbox. *Neuroinformatics* 2013;11:319-37.
 34. Samper-González J, Burgos N, Bottani S, et al. Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. *Neuroimage* 2018;183:504-21.
 35. Marquand A, Howard M, Brammer M, et al. Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *Neuroimage* 2010;49:2178-89.
 36. Rizvi B, Lao PJ, Chesebro AG, et al. Regional white matter hyperintensities predict Alzheimer's-like neurodegeneration. *Alzheimer's & Dementia* 2020;16:e044776.
 37. Habes M, Erus G, Toledo JB, et al. White matter hyperintensities and imaging patterns of brain ageing in the general population. *Brain* 2016;139:1164-79.
 38. Wardlaw JM, Valdés Hernández MC, Muñoz-Maniega S. What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. *J Am Heart Assoc* 2015;4:001140.
 39. Veldsman M, Tai XY, Nichols T, et al. Cerebrovascular risk factors impact frontoparietal network integrity and executive function in healthy ageing. *Nat Commun* 2020;11:4340.
 40. Seiler S, Fletcher E, Beiser AS, et al. Structural brain network efficiency and cognitive processing speed in healthy aging. *Alzheimer's & Dementia* 2020;16(S5). doi:10.1002/alz.044563.
 41. Lockhart SN, Mayda AB, Roach AE, et al. Episodic memory function is associated with multiple measures of white matter integrity in cognitive aging. *Front Hum Neurosci* 2012;6:56.
 42. Liang Y, Sun X, Xu S, et al. Preclinical Cerebral Network Connectivity Evidence of Deficits in Mild White Matter Lesions. *Front Aging Neurosci* 2016;8:27.
 43. Heinen R, Groeneveld ON, Barkhof F, et al. Small vessel disease lesion type and brain atrophy: The role of co-occurring amyloid. *Alzheimers Dement (Amst)* 2020;12:e12060.
 44. Jokinen H, Koikkalainen J, Laakso HM, et al. Global Burden of Small Vessel Disease-Related Brain Changes on MRI Predicts Cognitive and Functional Decline. *Stroke* 2020;51:170-8.
 45. Stonnington CM, Chu C, Klöppel S, et al. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *Neuroimage* 2010;51:1405-13.
 46. Cui Z, Gong G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage* 2018;178:622-37.
 47. Morgane PJ, Galler JR, Mokler DJ. A review of systems and networks of the limbic forebrain/limbic midbrain. *Prog Neurobiol* 2005;75:143-60.
 48. Hafkemeijer A, van der Grond J, Rombouts SA. Imaging the default mode network in aging and dementia. *Biochim Biophys Acta* 2012;1822:431-41.
 49. Mather M, Harley CW. The Locus Coeruleus: Essential for Maintaining Cognitive Function and the Aging Brain. *Trends Cogn Sci* 2016;20:214-26.
 50. Nie X, Sun Y, Wan S, et al. Subregional Structural Alterations in Hippocampus and Nucleus Accumbens Correlate with the Clinical Impairment in Patients with Alzheimer's Disease Clinical Spectrum: Parallel Combining Volume and Vertex-Based Approach. *Front Neurol* 2017;8:399.
 51. Lee DH, Lee P, Seo SW, et al. Neural substrates of cognitive reserve in Alzheimer's disease spectrum and normal aging. *Neuroimage* 2019;186:690-702.
 52. Perneczky R, Tene O, Attems J, et al. Is the time ripe for new diagnostic criteria of cognitive impairment due to cerebrovascular disease? Consensus report of the

- International Congress on Vascular Dementia working group. *BMC Med* 2016;14:162.
53. Skrobot OA, Black SE, Chen C, et al. Progress toward standardized diagnosis of vascular cognitive impairment: Guidelines from the Vascular Impairment of Cognition Classification Consensus Study. *Alzheimers Dement* 2018;14:280-92.
 54. Parnaudeau S, Bolkan SS, Kellendonk C. The Mediodorsal Thalamus: An Essential Partner of the Prefrontal Cortex for Cognition. *Biol Psychiatry* 2018;83:648-56.
 55. Alves GS, de Carvalho LA, Sudo FK, et al. A panel of clinical and neuropathological features of cerebrovascular disease through the novel neuroimaging methods. *Dement Neuropsychol* 2017;11:343-55.
 56. Liu C, Li C, Gui L, et al. The pattern of brain gray matter impairments in patients with subcortical vascular dementia. *J Neurol Sci* 2014;341:110-8.
 57. Plachti A, Kharabian S, Eickhoff SB, et al. Hippocampus co-atrophy pattern in dementia deviates from covariance patterns across the lifespan. *Brain* 2020;143:2788-802.
 58. Cuadrado-Godia E, Dwivedi P, Sharma S, et al. Cerebral Small Vessel Disease: A Review Focusing on Pathophysiology, Biomarkers, and Machine Learning Strategies. *J Stroke* 2018;20:302-20.
 59. Schmitter D, Roche A, Maréchal B, et al. An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease. *Neuroimage Clin* 2015;7:7-17.
 60. Schrouff J, Monteiro JM, Portugal L, et al. Embedding Anatomical or Functional Knowledge in Whole-Brain Multiple Kernel Learning Models. *Neuroinformatics* 2018;16:117-43.
 61. Vogel JW, Vachon-Preseu E, Pichet Binette A, et al. Brain properties predict proximity to symptom onset in sporadic Alzheimer's disease. *Brain* 2018;141:1871-83.
- (English Language Editor: L. Roberts and J. Jones)

Cite this article as: Wang J, Zhao C, Wei J, Li C, Zhang X, Liang Y, Zhang Y. Individual prediction and classification of cognitive impairment in patients with white matter lesions based on gray matter volume. *Ann Transl Med* 2022;10(5):246. doi: 10.21037/atm-21-3571

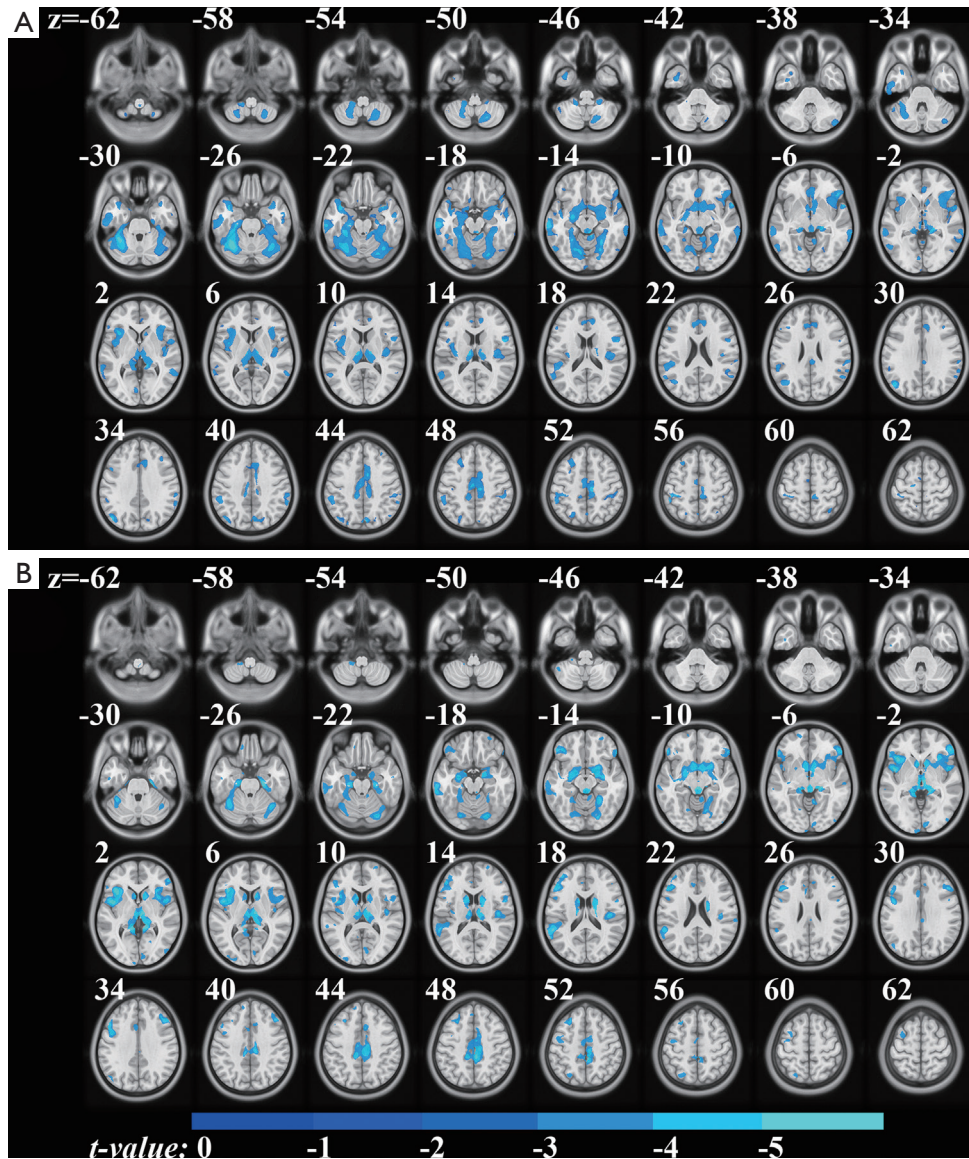


Figure S1 GM atrophy in older individuals with WMLs. (A) GM atrophy in individuals with VaD *vs.* CH. (B) GM atrophy in individuals with VaD *vs.* VaMCI. A one-way ANCOVA test, adjusted for age, gender, and education levels, was used to test differences in the GM images for the 3 groups. A P value of <0.05 was considered significant. Family-discovery rated (FDR) corrected. GM, gray matter; WMLs, white matter lesions; VaD, vascular dementia; VaMCI, vascular mild cognitive impairment; ANCOVA, analysis of covariance; CH, good cognitive health.

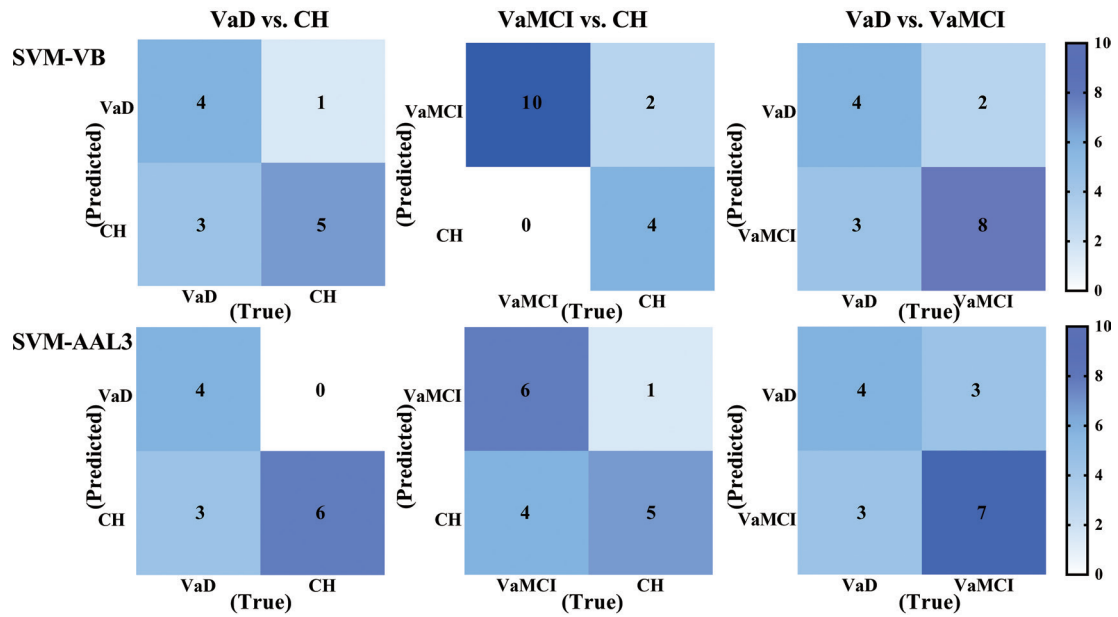


Figure S2 Confusion matrices of SVM predictions in the testing set. The first row shows SVM classifiers using VB data, and the second row is SVM classifiers using AAL3 data. The left column shows SVM models classifying VaD and CH, the middle column shows the classification of VaMCI *vs.* CH, and the right column shows VaD *vs.* VaMCI. SVM, support vector machine; VB, voxel-based; AAL3, automated anatomical labelling atlas 3; VaD, vascular dementia; CH, good cognitive health; VaMCI, vascular mild cognitive impairment.

Table S1 The 5 most relevant regions for the RVR models using on VB data

Method	Neuropsychological Tests	Regions	Importance (%)	ER
RVR_VB	MMSE	Cingulate_Post_R	1.58	2.05
		Thal_PuA_L	1.54	2.52
		N_Acc_R	1.49	3.36
		Thal_PuL_L	1.46	5.91
		Temporal_Pole_Mid_L	1.34	6.29
	MoCA	N_Acc_R	1.68	1.14
		SN_pr_L	1.47	2.75
		Rolandic_Oper_R	1.42	3.77
		Thal_MDI_R	1.33	5.41
		Thal_LGN_L	1.32	6.52
	Visuospatial/Executive	Thal_VPL_R	1.70	1.14
		Heschl_R	1.53	2.36
		Frontal_Inf_Oper_L	1.35	4.82
		Thal_PuL_R	1.33	7.09
		Thal_VL_R	1.33	5.80
	Attention	N_Acc_R	2.10	1.09
		SN_pr_L	1.69	2.41
		Thal_LP_R	1.60	3.14
		Thal_AV_R	1.53	4.48
		Thal_MDI_L	1.40	6.50
	Delayed recall	Thal_PuL_R	2.43	1.02
		Thal_MDI_R	1.96	2.36
		Thal_MDm_R	1.89	3.38
		Thal_PuA_R	1.85	3.70
		Thal_PuM_R	1.82	4.50
	Orientation	Thal_PuA_L	1.51	1.86
		N_Acc_R	1.47	2.68
		Thal_PuM_R	1.44	3.29
Thal_LGN_R		1.43	3.50	
Thal_MGN_R		1.33	5.63	

ER, expected ranking; RVR, relevance vector regression; VB, voxel-based; R, right; L, left; Cingulate_Post, posterior cingulate gyrus; Thal_PuA, pulvinar anterior; N_Acc, nucleus accumbens; Thal_PuL, pulvinar lateral; Temporal_Pole_Mid, temporal pole: middle temporal gyrus; SN_pr, substantia nigra (pars reticulata); Rolandic_Oper, rolandic operculum; Thal_MDI, mediodorsal lateral parvocellular; Thal_LGN, lateral geniculate; Thal_VPL, ventral posterolateral; Heschl, Heschl's gyrus; Frontal_Inf_Oper, inferior frontal gyrus (opercular part); Thal_VL, ventral lateral; Thal_LP, lateral posterior; Thal_AV, Thalamus & Anteroventral Nucleus; Thal_MDm, mediodorsal medial magnocellular; Thal_PuM, pulvinar medial; Thal_MGN, medial geniculate.

Table S2 the 5 most relevant regions for the RVR models using AAL3 data

Method	Neuropsychological Tests	Regions	Importance (%)	ER
RVR_AAL3	MMSE	Paracentral_Lobule_L	2.16	1.27
		Temporal_Pole_Mid_L	1.81	2.16
		Temporal_Pole_Mid_R	1.40	4.21
		Fusiform_L	1.39	4.32
		Caudate_L	1.36	5.36
	MoCA	Paracentral_Lobule_L	1.71	1.96
		N_Acc_R	1.57	2.04
		Caudate_L	1.36	3.75
		Fusiform_L	1.34	4.21
		Caudate_R	1.33	4.93
	Visuospatial/Executive	Temporal_Pole_Mid_L	1.78	1.95
		Paracentral_Lobule_L	1.73	2.57
		Amygdala_L	1.71	2.98
		Paracentral_Lobule_R	1.71	3.32
		Fusiform_L	1.46	5.09
	Attention	Paracentral_Lobule_L	1.74	2.09
		Thal_MDI_L	1.66	2.05
		Thal_MDm_L	1.54	4.34
		N_Acc_R	1.51	4.18
		Thal_PuL_R	1.45	5.04
	Delayed recall	Thal_PuL_R	1.95	1.30
		Thal_MDm_R	1.82	2.11
		Thal_MDI_R	1.74	3.05
		Postcentral_R	1.48	5.38
		Paracentral_Lobule_R	1.46	5.80
	Orientation	Paracentral_Lobule_L	2.22	1.00
		N_Acc_R	1.83	1.95
Caudate_R		1.46	3.09	
Caudate_L		1.36	4.32	
Temporal_Pole_Mid_R		1.29	5.41	

RVR, relevance vector regression; AAL3, automated anatomical labelling atlas 3; Paracentral_Lobule, paracentral lobule; Fusiform, fusiform gyrus; Caudate, caudate nucleus; Postcentral, postcentral gyrus.

Table S3 The 5 most relevant regions for the classifiers using VB data

Classification	SVM			GPC		
	Region	Importance (%)	ER	Region	Importance (%)	ER
VaD vs. CH	Thal_VL_R	1.89	1.27	Thal_MDI_L	2.25	1.03
	Thal_MDI_L	1.84	2.15	Putamen_L	2.10	2.58
	Thal_VPL_R	1.83	2.82	Thal_VL_R	2.08	2.45
	Putamen_L	1.65	4.67	Thal_VPL_R	1.99	4.00
	Thal_MDI_R	1.65	4.42	Putamen_R	1.95	4.82
VaMCI vs. CH	Thal_MDI_L	1.91	1.24	Thal_MDI_L	2.19	1.26
	Thal_VL_R	1.87	2.00	Putamen_R	2.15	1.71
	Thal_VL_L	1.78	3.19	Thal_VL_R	1.98	2.98
	Putamen_R	1.74	4.05	Thal_VPL_R	1.87	4.62
	Thal_VPL_R	1.69	5.45	Thal_VL_L	1.87	4.31
VaD vs. VaMCI	Cingulate_Post_R	1.61	1.22	Cingulate_Post_R	1.62	1.73
	Caudate_L	1.50	2.08	Caudate_L	1.62	1.57
	N_Acc_R	1.28	5.08	N_Acc_R	1.36	4.32
	Temporal_Inf_L	1.25	4.59	Putamen_R	1.34	4.95
	Caudate_R	1.23	6.24	N_Acc_L	1.33	5.35

VB, voxel-based; SVM, support vector machine; GPC, Gaussian process classification; Thal_IL, intralaminar; Putamen, lenticular nucleus & putamen; Temporal_Inf, inferior temporal gyrus; VaD, vascular dementia; VaMCI, vascular mild cognitive impairment; CH, good cognitive health.

Table S4 The 5 most relevant regions for the classifiers using AAL3 data

Classification	SVM			GPC		
	Region	Importance (%)	ER	Region	Importance (%)	ER
VaD vs. CH	Thal_MDI_L	1.97	1.03	Amygdala_R	1.65	2.03
	Putamen_L	1.77	2.27	Thal_MDI_L	1.64	2.58
	Thal_VPL_R	1.74	2.88	Putamen_L	1.63	3.06
	Thal_VL_R	1.69	3.97	Thal_VPL_R	1.60	3.58
	Putamen_R	1.58	5.85	Thal_VL_R	1.59	4.36
VaMCI vs. CH	Thal_MDI_L	1.88	1.12	LC_R	1.84	1.10
	Thal_VL_R	1.71	2.26	Amygdala_R	1.68	2.48
	ParaHippocampal_L	1.66	3.95	Putamen_R	1.62	3.21
	LC_R	1.65	3.90	Thal_MDI_L	1.61	3.67
	Amygdala_R	1.64	4.76	Raphe_D	1.50	5.45
VaD vs. VaMCI	N_Acc_R	1.57	2.24	N_Acc_R	1.55	2.30
	N_Acc_L	1.55	2.32	N_Acc_L	1.44	4.76
	Temporal_Inf_L	1.50	2.65	Cingulate_Post_R	1.43	4.76
	Paracentral_Lobule_L	1.44	4.22	Caudate_L	1.42	5.35
	Temporal_Pole_Mid_L	1.31	6.70	Paracentral_Lobule_L	1.41	4.92

AAL3, automated anatomical labelling atlas 3; SVM, support vector machine; GPC, Gaussian process classification; ParaHippocampal, parahippocampal gyrus; LC, locus coeruleus; Raphe_D, raphe nucleus (dorsal); VaD, vascular dementia; VaMCI, vascular mild cognitive impairment; CH, good cognitive health.