Received: 29 November 2022 | Revised: 6 January 2023 | Accepted: 9 January 2023

DOI: 10.1002/alz.12971

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Unsupervised machine learning model to predict cognitive impairment in subcortical ischemic vascular disease

Qi Qin[1] | Junda Qu[2,3] | Yunsi Yin[1] | Ying Liang[2,3] | Yan Wang[1] | Bingxin Xie[1] | Qingqing Liu[4] | Xuan Wang[5] | Xinyi Xia[1] | Meng Wang[1] | Xu Zhang[2,3] | Jianping Jia[1,6,7,8,9] | Yi Xing[1] | Chunlin Li[2,3] | Yi Tang[1,6]

[1]Department of Neurology & Innovation Center for Neurological Disorders, Xuanwu Hospital, Capital Medical University, National Center for Neurological Disorders, Beijing, China

[2]School of Biomedical Engineering, Capital Medical University, Beijing, China

[3]Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing, China

[4]Department of Neurology, The First Affiliated Hospital of Harbin Medical University, Harbin, China

[5]Department of Endocrinology, The Second People's Hospital of Mudanjiang, Mudanjiang, China

[6]Key Laboratory of Neurodegenerative Diseases, Ministry of Education of the People's Republic of China, Beijing, China

[7]Center of Alzheimer's Disease, Beijing Institute for Brain Disorders, Beijing, China

[8]Beijing Key Laboratory of Geriatric Cognitive Disorders, Beijing, China

[9]National Clinical Research Center for Geriatric Disorders, Beijing, China

Correspondence
Yi Xing and Yi Tang, Department of Neurology & Innovation Center for Neurological Disorders, Xuanwu Hospital, Capital Medical University, 45 Changchun Street, Beijing 100053, China.
Email: 13269627589@163.com and tangyi@xwhosp.org

Chunlin Li, School of Biomedical Engineering, Capital Medical University, 10 XiToutiao outside the YouAnmen, Beijing, 100069, China.
Email: lichunlin1981@163.com

Abstract

INTRODUCTION: It is challenging to predict which patients who meet criteria for subcortical ischemic vascular disease (SIVD) will ultimately progress to subcortical vascular cognitive impairment (SVCI).

METHODS: We collected clinical information, neuropsychological assessments, T1 imaging, diffusion tensor imaging, and resting-state functional magnetic resonance imaging from 83 patients with SVCI and 53 age-matched patients with SIVD without cognitive impairment. We built an unsupervised machine learning model to isolate patients with SVCI. The model was validated using multimodal data from an external cohort comprising 45 patients with SVCI and 32 patients with SIVD without cognitive impairment.

RESULTS: The accuracy, sensitivity, and specificity of the unsupervised machine learning model were 86.03%, 79.52%, and 96.23% and 80.52%, 71.11%, and 93.75% for internal and external cohort, respectively.

DISCUSSION: We developed an accurate and accessible clinical tool which requires only data from routine imaging to predict patients at risk of progressing from SIVD to SVCI.

KEYWORDS
diffusion tensor imaging, resting-state functional magnetic resonance imaging, subcortical ischemic vascular disease, subcortical vascular cognitive impairment, unsupervised machine learning model

Highlights

- Our unsupervised machine learning model provides an accurate and accessible clinical tool to predict patients at risk of progressing from subcortical ischemic vascular disease (SIVD) to subcortical vascular cognitive impairment (SVCI) and requires only data from imaging routinely used during the diagnosis of suspected SVCI.

Qi Qin, Junda Qu and Yunsi Yin contributed equally to this study.

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 82220108009, 81970996, 82201568, 62171300, 81701044; National Key R&D Program of China, Grant/Award Number: 2022YFC3602600; Beijing Natural Science Foundation, Grant/Award Number: JQ19024; Young Elite Scientists Sponsorship Program, Grant/Award Number: 2021QNRC001; Beijing Nova Program, Grant/Award Number: Z211100002121051; Capital's Funds for Health Improvement and Research, Grant/Award Number: CFH 2020-4-1033; STI2030-Major Projects, Grant/Award Number: 2021ZD0201801

- The model yields good accuracy, sensitivity, and specificity and is portable to other cohorts and to clinical practice to distinguish patients with SIVD at risk for progressing to SVCI.
- The model combines assessment of diffusion tensor imaging and functional magnetic resonance imaging measures in patients with SVCI to analyze whether the "disconnection hypothesis" contributes to functional and structural changes and to the clinical presentation of SVCI.

1 | INTRODUCTION

Subcortical ischemic vascular disease (SIVD) is a common form of small vessel disease characterized by asymptomatic lacunes infarcts and white matter hyperintensities (WMHs).[1] The prevalence of SIVD increases with age, affecting approximately 5% of people aged 50 years to almost 100% of people older than 90 years.[2] Half of patients with SIVD will present with cognitive deterioration and eventually progress to subcortical vascular cognitive impairment (SVCI).[3,4] However, in the other half of patients with SIVD, cognitive impairment never manifests, and there are currently no reliable methods to isolate which patients with SIVD will progress to cognitive impairment.[2]

According to the latest Vascular Impairment of Cognition Classification Consensus Study (VICCCS) guidelines, the diagnosis of SVCI generally requires evaluation of clinical symptoms, structural magnetic resonance imaging (MRI), and neuropsychological assessments.[5] For the neuropsychological assessments, the current criteria for SVCI requires analysis of five core cognitive domains,[6] which is assessed using a "60-minute" protocol.[5] However, this detailed neuropsychological testing is neither feasible nor meaningful in certain settings,[2] for example, in patients with severe depression or severe dementia or who are illiterate or have aphasia. Even in the most practicable settings, these neuropsychological assessments are difficult to administer and must be performed by well-trained professional physicians; otherwise, there is a risk of misdiagnosis resulting from improper assessment. Indeed, it has been demonstrated that the outcomes of neuropsychological assessments such as the Montreal Cognitive Assessment (MoCA) are highly variable owing to investigator bias.[7]

In patients with SVCI, structural MRI shows WMHs and lacunes on T2-weighted and fluid-attenuated inversion recovery images.[8] However, these imaging markers alone are not sufficient to diagnose cognitive impairment, because WMHs are common in older populations: prevalence rates for WMHs rise from approximately 50% in adults 45 years of age to approximately 95% in adults 80 years of age.[9,10] Therefore, whether WMH severity correlates with the severity of cognitive decline is uncertain,[11–13] and other imaging approaches are needed to predict and diagnose cognitive decline. Potential approaches that may be better suited for this purpose include diffusion tensor imaging (DTI), which is a sensitive MRI-based technique to identify alterations in the structure of white matter (WM),[14] and functional MRI (fMRI), which can be used to evaluate the functional connectivity (FC) of SVCI.[15,16] Studies that directly compare or combine DTI and fMRI in patients with SVCI are lacking, and future research applying multimodal MRI images to diagnose SVCI or to predict progression to SVCI in patients with SIVD are needed.

Recent advancements in machine learning algorithms can adequately model complex relationships or interactions among high-dimensional, multimodal measures, which creates opportunities to build predictive models to inform clinical practice.[17] Inspired by recent applications of machine learning in Alzheimer's disease (AD) and other neurological diseases,[15,18–21] here we pursue development of a machine learning model to predict development of cognitive impairment in patients with SIVD. We assumed specific WM tract damage that may result in functional changes in particular brain regions and eventually cause cognitive impairment in SIVD patients. Therefore, we created an unsupervised machine learning model based on extensive T1-weighted MRI, DTI, and resting-state fMRI (rsfMRI), as well as clinical data from a multi-center cohort of patients with SIVD. Our unsupervised machine learning model accurately isolated patients with SVCI from patients with SIVD without cognitive impairment and can be used as a clinical tool to predict patients with SIVD at risk of progressing to SVCI.

2 | MATERIALS AND METHODS

2.1 | Participants

Data from 83 patients who met the inclusion criteria for SVCI and 53 age-matched patients with SIVD without cognitive impairment were included to build the unsupervised machine learning model. These patients were recruited from Xuanwu Hospital of Capital Medical University, the Second People's Hospital of Mudanjiang, and the First Affiliated Hospital of Harbin Medical University between 2015 and 2019 (Figure S1 in supporting information). The ethics

RESEARCH IN CONTEXT

1. **Systematic review**: Subcortical ischemic vascular disease (SIVD) is a common form of small vessel disease, affecting approximately 5% of people aged 50 years to almost 100% of people older than 90 years. Half of patients with SIVD will present with cognitive deterioration and eventually progress to subcortical vascular cognitive impairment (SVCI). However, in the other half of patients with SIVD, cognitive impairment never manifests. Currently, there is no accessible and validated methodology to predict which patients who meet criteria for SIVD will ultimately progress to SVCI.
2. **Interpretation**: Our unsupervised machine learning model provides good accuracy, sensitivity, and specificity and is an accessible clinical tool to predict patients at risk of progressing from SIVD to SVCI, requiring only data from imaging routinely used during the diagnosis of suspected SVCI.
3. **Future directions**: The article proposes a method for the generation of new predictive models based on machine learning for other clinical applications and, in particular, for diseases that use multimodal magnetic resonance imaging (MRI) for diagnosis. In addition, our model for SVCI highlights the need for future studies to: (1) assess the ability of diffusion tensor imaging and resting-state functional MRI to follow cognitive impairment in patients with SIVD longitudinally, (2) identify potential therapeutic interventions based on functional and structural changes detected by multimodal MRI in the clinical setting, and (3) determine the relevance of the "disconnection hypothesis" in SIVD-associated cognitive impairment.

committee approved this study. Written informed consent according to the Declaration of Helsinki was provided by all participants.

The diagnosis of SVCI was made by three senior neurologists, and each patient met the following inclusion criteria: (1) age 50 to 80 ; (2) complaint and/or informant report of cognitive impairment at least 3 months; (3) met the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition criteria for cognitive impairment, including a Clinical Dementia Rating (CDR) of ≥0.5 on at least one domain and a global score ≥0.5; and (4) MRI with (i) moderate to severe WM lesions (score ≥ 2, according to the Fazekas rating scale[11]), or multiple (≥3) small supratentorial subcortical infarcts (3-20 mm in diameter),[22] or small infarcts strategically located in the caudate nucleus, globus pallidus, or thalamus; (ii) absence of hemorrhages, cortical and watershed infarcts, hydrocephalus, and WM lesions from specific causes; and (iii) no hippocampal or entorhinal cortex atrophy (score = 0, according to the medical temporal lobe atrophy scale of Scheltens[23]).

The inclusion criteria for patients with SIVD without cognitive impairment were: (1) age 50 to 80; (2) evidence of relevant subcortical vascular ischemia by MRI, including (i) moderate to severe WM lesions (score ≥2, according to the Fazekas rating scale[11]); multiple (≥3) small supratentorial subcortical infarcts (3–20 mm in diameter);[22] or small infarcts strategically located in the caudate nucleus, globus pallidus, or thalamus; (ii) absence of hemorrhages, cortical and watershed infarcts, hydrocephalus, and WM lesions from specific causes; and (iii) no hippocampal or entorhinal cortex atrophy (score = 0, according to the medical temporal lobe atrophy scale of Scheltens[23]); (3) no complaint and/or informant report of cognitive impairment involving memory and/or other cognitive domains; and (4) CDR score = 0 for every domain.

Patients with the following conditions were excluded: (1) other cognitive impairments or neurological disease; (2) any condition, including illiteracy, hearing disorder, or visual impairment, that would preclude completion of neuropsychological assessment testing; and (3) diagnosis of AD by plasma amyloid beta 1-42[24] and/or phosphorylated tau 181[25] concentrations after 2-year follow-up.

For the external validation, 45 patients who met the inclusion criteria for SVCI and 32 age-matched patients with SIVD without cognitive impairment were included by the same inclusion criteria.

2.2 | Measures

2.2.1 | Demographic measure

The demographic measures collected included sex, age, body mass index (BMI), education, medical history, Fazekas score, and apolipoprotein E genotype (Method 1 in supporting information), which were analyzed as clinical information in the unsupervised model.

2.2.2 | Neuropsychological measures

The neuropsychological measures collected included the Mini-Mental State Examination (MMSE), MoCA, Boston Naming Test, Hachinski Ischemic Scale (HIS), Geriatric Depression Scale (GDS), Digit Span, CDR, Trail-Making Test A (TMT-A), Trail-Making Test B (TMT-B), Neuropsychiatric Inventory (NPI), and the Activities of Daily Living (ADL).

2.2.3 | MRI measures

Multimodal MRI images were collected, and measures of the brain regions from T1 structural imaging, DTI, and rsfMRI were extracted for the unsupervised machine learning model. The extraction pipeline of multimodal measures is shown in Figure 1. We used voxel-based morphometry and surface-based morphometry analysis in parallel based on the standard pipeline in Computational Anatomy Toolbox 12 (CAT 12, http://dbm.neuro.uni-jena.de/cat/)[26] in the statistical

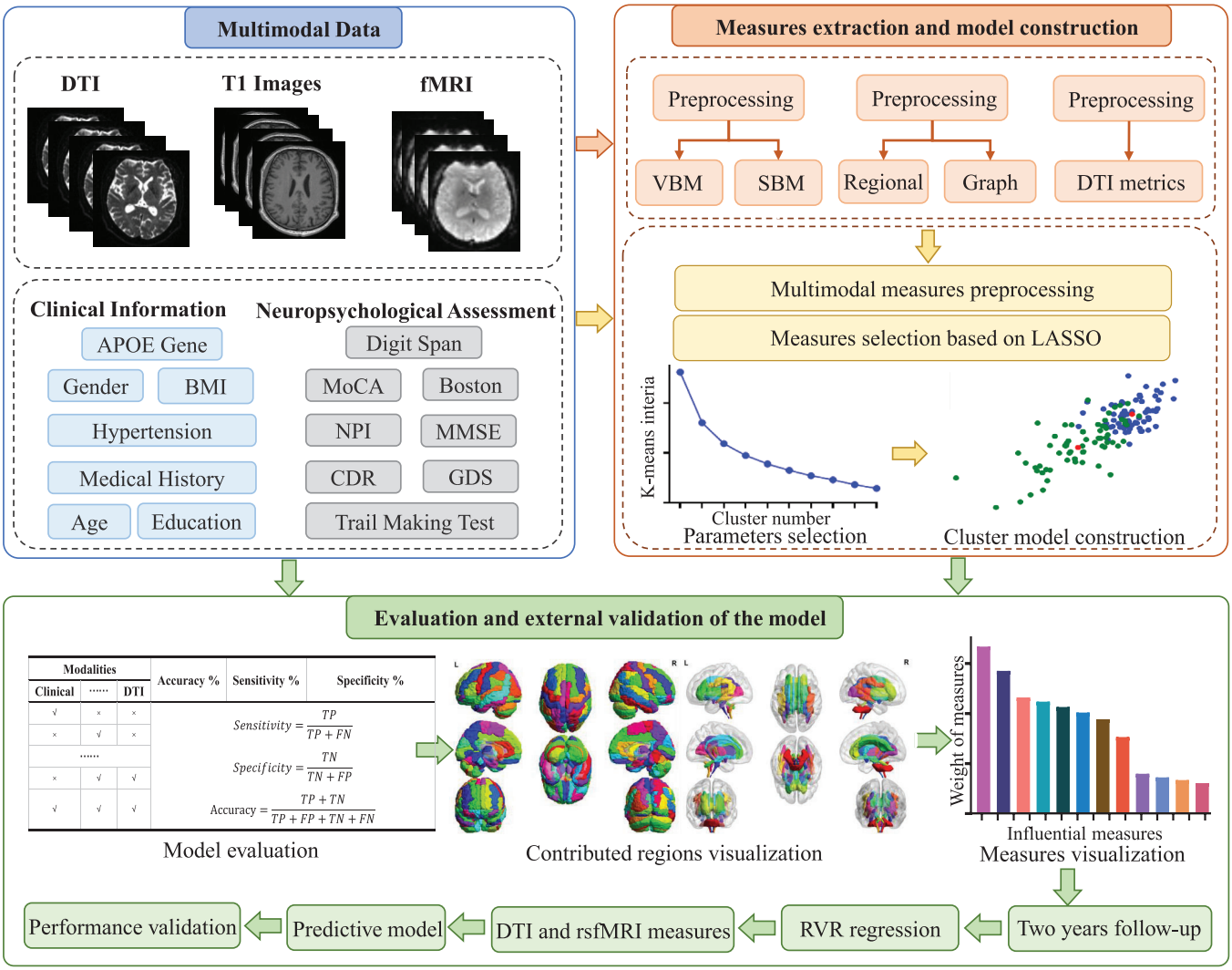

FIGURE 1 Framework for measures processing, model construction, model evaluation, and model validation. Multimodal image measures were extracted from T1, rsfMRI, and DTI using CAT 12 Toolbox, CONN toolbox, RESTPlus toolkit, GRETNA toolkit, and PANDA toolkit. Clinical information and multimodal image measures were preprocessed and selected with LASSO. The elbow method was combined with grid search to determine the cluster number and threshold for the measure of each coefficient. After determining the parameters, unsupervised machine learning models were created, and interpreted to identify the influential measures and corresponding brain regions that contributed to the model. RVR was used to estimate the association between the influential measures and neuropsychological assessments. Model performance was validated in an external cohort. *APOE*, apolipoprotein E; BMI, body mass index; CDR, Clinical Dementia Rating; DTI, diffusion tensor imaging; GDS, Geriatric Depression Scale; LASSO, least absolute shrinkage and selection operator; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; NPI, Neuropsychiatric Inventory; rsfMRI, resting-state functional magnetic resonance imaging; RVR, relevance vector regression; SBM, surface-based morphometry; VBM, voxel-based morphometry.

parametric mapping suit (SPM 12, http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) to obtain the T1 measures: total intracranial volumes, gray matter (GM) volumes, WM volumes, cerebrospinal fluid (CSF) volume, cortical thickness, sulcus depth, fractal dimension,[27] and cortical gyrification.[28] DTI were processed using the automatic pipeline in Pipeline for Analyzing braiN Diffusion imAges (PANDA, www.nitrc.org/projects/panda)[29] to obtain the DTI measures: fractional anisotropy (FA),[30] mean diffusivity,[31] axial diffusivity,[32] radial diffusivity,[32] and local diffusion homogeneity.[33] For rsfMRI, we first applied the CONN toolbox (http://www.nitrc.org/projects/conn)[34] to preprocess the data, and subsequently applied the RESTPlus toolkit (http://restfmri.net/forum/restplus)[35] to extract rsfMRI measures: amplitude of low-frequency fluctuations (ALFF), fractional amplitude of low-frequency fluctuations (fALFF), percent amplitude of fluctuation (PerAF), and regional homogeneity of Kendall's consistency coefficient (KCC-ReHo). In parallel, we also performed graph theory[36] using GRETNA toolbox (http://www.nitrc.org/projects/gretna)[37] to obtain the graph theory measures: assortativity, betweenness centrality, degree centrality, network efficiency, nodal clustering coefficient, nodal efficiency, nodal local efficiency, nodal shortest path length,

rich club, and small world. Combined with the Anatomical Automatic Labeling (AAL) atlas,[38] the Desikan–Killiany (DK 40) atlas,[39] and hand-segmented WM map[40] provided by PANDA, multimodal measures in brain regions of interest (ROIs) were extracted. The details of the multimodal MRI acquisition and images processing are shown in Method 2 in supporting information.

3 | CONSTRUCTION OF THE UNSUPERVISED MACHINE LEARNING MODEL

The pipeline for construction and fitting of the unsupervised machine learning model is shown in Figure 1. We removed measures with a proportion of missing values ≥20%[41] in each cohort or each measure, and imputed other missing values using the mode of measures. One-hot encoding and Z-score transformation were adopted for categorical measures and continuous measures, respectively. The selection of measures was performed in a supervised manner. A two-independent samples test was performed to compare measures from patients with SVCI to patients with SIVD without cognitive impairment using the Scipy package.[42] *P* values were obtained using a Student's *t* test, if equal population variances were met, or a Welch *t* test otherwise. Measures with $P < 0.01$ were selected. We subsequently applied the least absolute shrinkage and selection operator (LASSO) algorithm[43] to further constrain the number of measures. The penalty term of LASSO was determined using the 10-fold cross-validation with the scikit-learn toolkit[44] within range of 0.001 to 10.

The scikit-learn toolkit[44] was used to build the unsupervised model. To estimate the feasibility of distinguishing patients with SVCI, we generated t-distributed stochastic neighbor embedding (t-SNE)[45] plots. Because the supervised algorithm required most of the data for training, *k*-means cluster[46] of unsupervised machine learning was chosen. We also built a supervised random forest model for comparison to the unsupervised machine learning model, and the construction of random forest is shown in Method 3 in supporting information. *K*-means cluster comprised three steps. The first step chose the initial subcluster center. Subsequently, each sample was assigned to its nearest center according to the sum of squared Euclidean distance. Finally, the new center was computed and updated based on the mean value of all subcluster samples.[47] The algorithm kept repeating the last two steps until the distance between the old and new centers was less than the tolerance. The tolerance was set to 0.0001 according to the default value of toolkit. We chose the *k*-means ++[48] to replace the traditional initialization of *k*-means.

Two parameters needed to be determined for the unsupervised machine learning model: the number of clusters *k* and the threshold of the measures' coefficient. We combined the elbow method[49] and grid search[50] to determine clusters *k* and the threshold. The cluster *k* ranged from 1 to 11. The threshold ranged 0 to $\left(scale \times \max\left[\left|\left\lceil \frac{min(coef)}{scale} \right\rceil - 1\right|, \left\lfloor \frac{max(coef)}{scale} \right\rfloor\right]\right)$, where *coef* and *scale*, respectively, represented the coefficient of measures and the scaling parameter. $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$, respectively, represented the round up and round down; *min* and *max* were the minimum value and maximum value. Measures with absolute coefficients above the threshold were identified as influential measures and entered in the unsupervised machine learning model.

To find the most helpful combination of multimodal measures, the following combinations of multimodal measures were input to the unsupervised machine learning model: (1) clinical information; (2) T1 measures; (3) rsfMRI measures; (4) DTI measures; (5) clinical information and T1 measures; (6) clinical information and rsfMRI measures; (7) clinical information and DTI measures; (8) T1 measures and rsfMRI measures; (9) T1 measures and DTI measures; (10) rsfMRI measures and DTI measures; (11) clinical information, T1 measures, and rsfMRI measures; (12) clinical information, T1 measures, and DTI measures; (13) clinical information, rsfMRI measures, and DTI measures; (14) T1 measures, rsfMRI measures, and DTI measures; and (15) clinical information, T1 measures, rsfMRI measures, and DTI measures.

3.1 | Interpretability and evaluation of the unsupervised machine learning model

We used KMeansInterp (https://github.com/YousefGh/kmeans-feature-importance) to gain insight into the influential measures. To visualize the location of brain regions, BrainNet Viewer toolkit (https://www.nitrc.org/projects/bnv/)[51] was performed by mapping the brain regions in the AAL and WM atlases to the smoothed brain surface. The performance of the model was evaluated in terms of accuracy, sensitivity, specificity, and precision. To evaluate the performance of the unsupervised machine learning model, it was necessary to determine which labels the unsupervised machine learning model represented. We input the influential measures only corresponding to patients with SVCI into the trained model. If there were more labels predicted as one than zero, cluster label one was considered representative of patients with SVCI.

We used the relevance vector regression (RVR) model to examine the association between the influential measures and the neuropsychological assessments. The RVR was built using the scikit-rvm toolkit. We used 70% of the data for training and 30% for testing. We subsequently input the influential measures of the test set into the trained RVR to obtain the predicted scores. Pearson correlation coefficient and linear correlation plots were used to examine associations between predicted scores and neuropsychological assessments.

3.2 | External validation

We validated the unsupervised machine learning model using data collected from 45 patients with SVCI and 32 patients with SIVD without cognitive impairment. We extracted the multimodal measures of rsfMRI and DTI using the aforementioned imaging processing pipeline. Influential measures were chosen from multimodal measures and input into the trained unsupervised machine learning model. To

TABLE 1 Demographics.

Demographic variable	Construction cohort number (%) or mean (SD, score range)		Validation cohort number (%) or mean (SD, score range)	
	Patients with SVCI	Patients with SIVD without cognitive impairment	Patients with SVCI	Patients with SIVD without cognitive impairment
Age (years)	65.9 (7.8, 48-79)	63.4 (7.9, 47-77)	66.0 (7.4, 48-79)	64.0 (8.2, 47-77)
Education (years)	10.6 (3.0, 6-18)	11.6 (2.5, 6-17)	10.7 (3.3, 6-18)	11.7 (2.5, 6-16)
BMI (kg/m^2)	23.51 ± 3.13	23.50 ± 2.75	24.72 ± 2.96	24.17 ± 3.03
Female, n (%)	31 (37.35)	21 (39.62)	17 (27.78)	13 (40.63)
Hypertension, *n* (%)	43 (51.81)	27 (50.94)	23 (52.11)	16 (50.00)
Diabetes history, n (%)	20 (24.09)	12 (22.64)	12 (26.66)	8 (25.00)
LDL (mmol/L)	2.77 ± 0.83	2.76 ± 0.89	2.90 ± 0.97	2.96 ± 0.71
HDL (mmol/L)	1.37 ± 0.31	1.38 ± 0.26	1.43 ± 0.33	1.49 ± 0.36
TC (mmol/L)	4.68 ± 1.00	4.47 ± 0.95	4.97 ± 0.92	4.91 ± 0.75
TG (mmol/L)	1.46 ± 0.82	1.53 ± 0.45	1.73 ± 0.96	1.75 ± 0.75
MMSE***	26.5 (2.3, 18-30)	29.0 (0.9, 27-30)	27.3 (2.7, 18-30)	30.0 (0.9, 27-30)
MoCA***	21.1 (3.6, 13-29)	26.8 (1.6, 24-30)	22.8 (3.7, 14-28)	26.4 (0.9, 24-30)
Fazekas score	1.97 ± 0.42	1.89 ± 0.51	2.01 ± 0.4	1.92 ± 0.52
APOE ε4 carrier, *n* (%)	18 (21.69)	7 (13.21)	11 (24.44)	5 (15.63)

Note: There were minor cases of missing values in the demographics, which were filled in by the mean value. Comparisons of patients were performed using the chi-squared test for categorical variables. If continuous variables satisfied the normal distribution, comparisons of continuous variables were performed using Student's *t* test, otherwise using Mann–Whitney *U* test.

Abbreviations: *APOE*, apolipoprotein E; BMI, body mass index; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; SD, standard deviation; SIVD, subcortical ischemic vascular disease; SVCI, subcortical vascular cognitive impairment; TC, total cholesterol; TG, triglycerides.

***$P < 0.001$.

observe the impact of influential measures in the external cohort, we performed two independent samples tests of influential measures. If a *P* value of influential measures was greater than 0.01, it indicated that the influential measure was not reliable for identifying patients with SVCI.

3.3 | Statistical analysis

Demographics were described using number (proportion) for categorical variables and mean (standard deviation, score range) for continuous variables. Comparisons of patients were performed using the chi-squared test for categorical variables. Comparisons of continuous variables were performed using the Student's *t* test or Mann–Whitney *U* test. The performance of the unsupervised machine learning model was measured regarding accuracy, sensitivity, specificity, and precision. The formulas for accuracy, sensitivity, specificity, and precision were defined respectively as $\frac{TP+TN}{TP+FP+TN+FN}$, $\frac{TP}{TP+FN}$, $\frac{TN}{TN+FP}$, and $\frac{TP}{TP+FP}$, where *TP* was both model and actual for patients with SVCI. *FP* was patients who were mistaken by the model as patients with SVCI but who were actually patients with SIVD without cognitive impairment. *FN* referred to those patients who were mistaken by the model as patients with SIVD without cognitive impairment but were actually patients with SVCI. *TN* was both model and actual for patients with SIVD without cognitive impairment.

4 | RESULTS

4.1 | Demographics

To build the unsupervised cluster model, we recruited a cohort of 83 patients with SVCI and 53 age-matched patients with SIVD without cognitive impairment. To evaluate model performance, we recruited a second cohort comprising 45 patients with SVCI and 32 age-matched patients with SIVD without cognitive impairment. Detailed demographics data are presented in Table 1. For both cohorts, there were no significant differences between patients with SVCI and patients with SIVD without cognitive impairment in terms of age, education, BMI, sex, hypertension history, diabetes history, low-density lipoprotein, high-density lipoprotein, total cholesterol, triglycerides, or Fazekas score. There were significant differences for MMSE and MoCA. The MMSE and MoCA values of patients with SVCI were lower than patients with SIVD without cognitive impairment. We also readministered the MMSE and MoCA to patients in the construction cohorts at a 2-year follow-up visit.

TABLE 2 Performance comparison of different modal combinations.

Modalities							
Clinical	T1	rsfMRI	DTI	Accuracy %	Precision %	Sensitivity %	Specificity %
×	×	√	√	**86.0**	**97.0**	**79.5**	**96.2**
×	√	×	√	81.6	92.7	75.9	90.6
√	×	√	√	81.6	98.3	71.1	98.1
√	√	×	√	80.9	92.5	74.7	90.0
√	√	×	×	79.4	91.0	73.5	88.7
×	×	√	×	78.7	77.0	92.8	56.6
×	√	√	√	77.9	100	63.9	100
×	√	×	×	77.2	93.3	67.5	92.5
√	×	√	×	76.5	75.8	90.4	54.7
×	√	√	×	73.5	98.0	57.8	98.1
√	√	√	√	73.5	100	56.6	100
×	×	×	√	72.8	94.2	59.0	94.3
√	×	×	√	71.3	94.0	56.6	94.3
√	√	√	×	69.9	100	50.6	100
√	×	×	×	68.4	76.3	69.9	66.0

Note: The check mark and the cross mark indicate whether the data of the modality is selected to build the model. The bolded part of the table represents the best overall performance of the unsupervised machine learning model, which can balance sensitivity and specificity well.
Abbreviations: DTI, diffusion tensor imaging; rsfMRI, resting-state functional magnetic resonance imaging.

4.2 | Performance of the unsupervised machine learning model

The t-SNE plot indicated that there was little overlap between the patients with SVCI and patients with SIVD without cognitive impairment (Figure S2 in supporting information). Because supervised machine learning required training data, we chose *k*-means to build the unsupervised machine learning model. The elbow method[49] and grid search[50] were used to determine the number of clusters *k* and the threshold of the measures' coefficients (Figure S3 in supporting information). The number of clusters *k* was set to two.

The performance of the 15 models was evaluated for accuracy, sensitivity, specificity, and precision (Table 2). The combination of rsfMRI + DTI performed best with accuracy, sensitivity, and specificity 86.0%, 79.5%, and 96.2%, respectively. By comparison, a previous study that used a sparse logistic regression to isolate patients with vascular cognitive impairment from patients with small vessel disease only achieved accuracy, sensitivity, and specificity of 72.6%, 77.0%, and 64.1%,[52] respectively, which is comparable to results we obtained using other combinations such as T1 + DTI (81.6%, 75.9%, and 90.6%, respectively) and clinical + T1 + DTI (80.9%,74.7%, and 90.0%, respectively). We found that adding clinical measures slightly degraded the performance of the model, which may be due to confounding effects caused by the redundant measures. When focusing on one modality, we found only rsfMRI was responsible for improving the sensitivity of the model. T1 and DTI improved the specificity of the model, with DTI contributing more to improve specificity than T1 (94.3% versus 92.5%, respectively). Thus, the model with the best performance included the two multimodal measures that contributed most to sensitivity (rsfMRI) and specificity (DTI).

In contrast to the unsupervised machine learning model, the supervised random forest model achieved similar performance with accuracy, sensitivity, and specificity of 85.71%, 83.33%, and 90.00% for rsfMRI + DTI. The receiver operating characteristic curve is shown in Figure S4 in supporting information. The supervised and unsupervised model had similar ability, whereas the unsupervised model had the advantage of using the data more efficiently, which suggested that the unsupervised machine learning model was more suitable to isolate the patients with SVCI from patients with SIVD without cognitive impairment.

4.3 | Interpretability and evaluation of the unsupervised machine learning model

After determining that DTI + rsfMRI achieved the best performance compared to other combinations, we adopted the KMeansInterp and BrainNet Viewer toolkit to explore the contributions of measures and their associated brain regions to model performance (Figure 2 and Figure 3). Overall, DTI contributed more to the model than rsfMRI. Among the DTI measures, six brain regions contributed to the model: external capsule, anterior limb of internal capsule, sagittal stratum, tapetum, posterior limb of internal capsule, and splenium of corpus callosum. Patients with SVCI were more likely to have lesions in the anterior or posterior limb of internal capsule and external capsule. These regions affect the connecting fibers of multiple brain regions

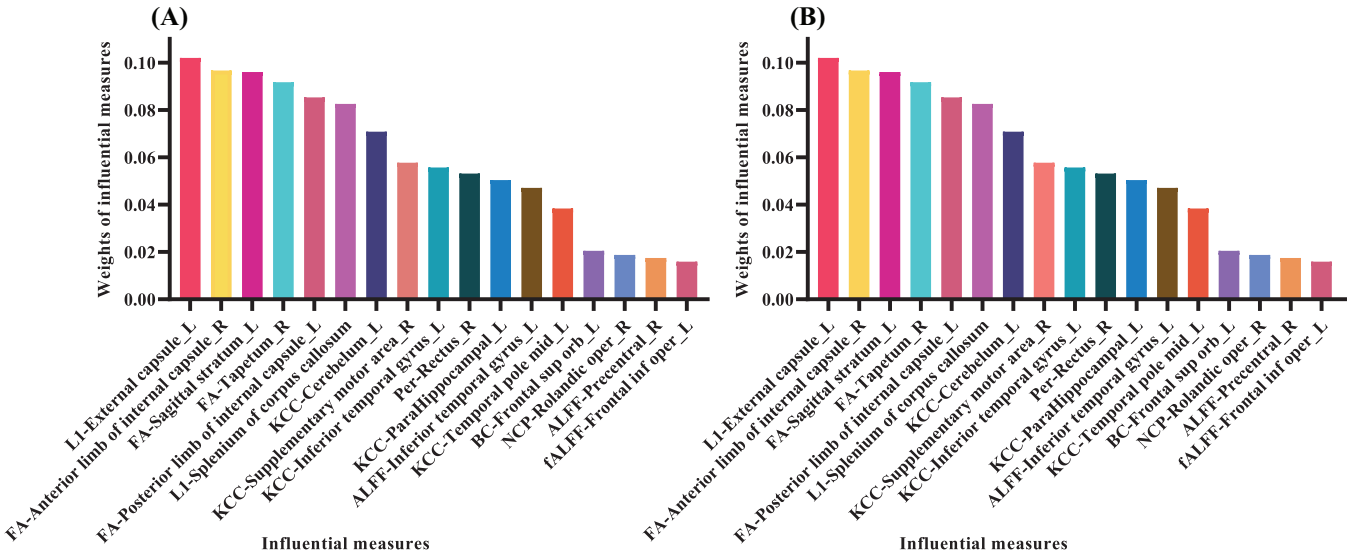

FIGURE 2 Importance of influential measures in unsupervised predictive models. A, Contribution of measures from age-matched patients with SIVD without cognitive impairment: rsfMRI + DTI. B, Contribution of measures from patients with SVCI: rsfMRI + DTI. We processed the weights of influential measures so that the sum of all weights was 1 without changing the distribution of weights to increase the readability. The main contributor to the unsupervised machine learning model was DTI, followed by rsfMRI. ALFF, amplitude of low-frequency fluctuations; BC, betweenness centrality; DTI, diffusion tensor imaging; FA, fractional anisotropy; Frontal inf oper, inferior frontal gyrus, opercular part; Frontal sup orb, superior frontal gyrus, orbital part; KCC, Kendall's consistency coefficient; L1, axial diffusivity; L, left brain; NCP, nodal cluster coefficient; Per, PerAF; PerAF, percent amplitude of fluctuation; R, right brain; Rolandic oper, Rolandic operculum; rsfMRI, resting-state functional magnetic resonance imaging; Sagittal stratum, sagittal stratum includes inferior longitudinal fasciculus and inferior fronto-occipital fasciculus; SIVD, subcortical ischemic vascular disease; SVCI, subcortical vascular cognitive impairment; Temporal pole mid, temporal pole, middle temporal gyrus.

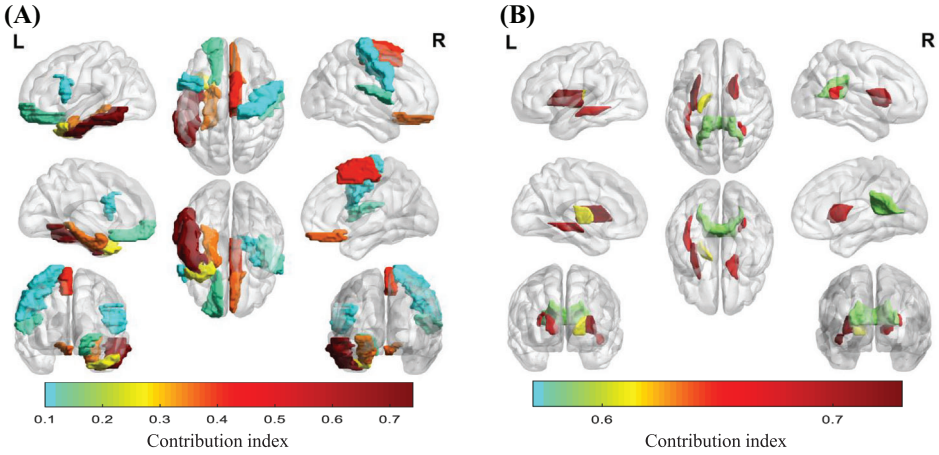

FIGURE 3 Visualization of brain regions of influential measures extracted from rsfMRI and DTI. A, Visualization of brain regions where influential rsfMRI measures were located. B, Visualization of brain regions where influential DTI measures were located. Six and ten brain regions from DTI and rsfMRI contributed to the predictive model, respectively. Two different measures, KCC-ReHo and ALFF, were located in the inferior temporal gyrus; we prioritized this region for further analysis. ALFF, amplitude of low-frequency fluctuations; DTI, diffusion tensor imaging; KCC-ReHo, regional homogeneity of Kendall's consistency coefficient; rsfMRI, resting-state functional magnetic resonance imaging.

through the front and rear limbs, such as the frontal lobe (executive function) and temporal lobe (global cognitive function). Lesions in the corpus callosum, which is an important structure connecting the bilateral cerebral hemispheres, were more likely to manifest in patients with SVCI compared to patients with SIVD without cognitive impairment, which could affect the interaction of neural information between the left and right brain. Damage to the sagittal stratum affects the processing of auditory and visual signals, and lesions in the sagittal stratum may be more likely to occur in patients with SVCI compared to patients with SIVD without cognitive impairment.

RsfMRI measures quantified functional abnormalities and identified 10 brain regions that contributed to the model: cerebellum, supplementary motor area, inferior temporal gyrus, rectus, parahippocampus, temporal pole (middle temporal gyrus), superior frontal gyrus (orbital

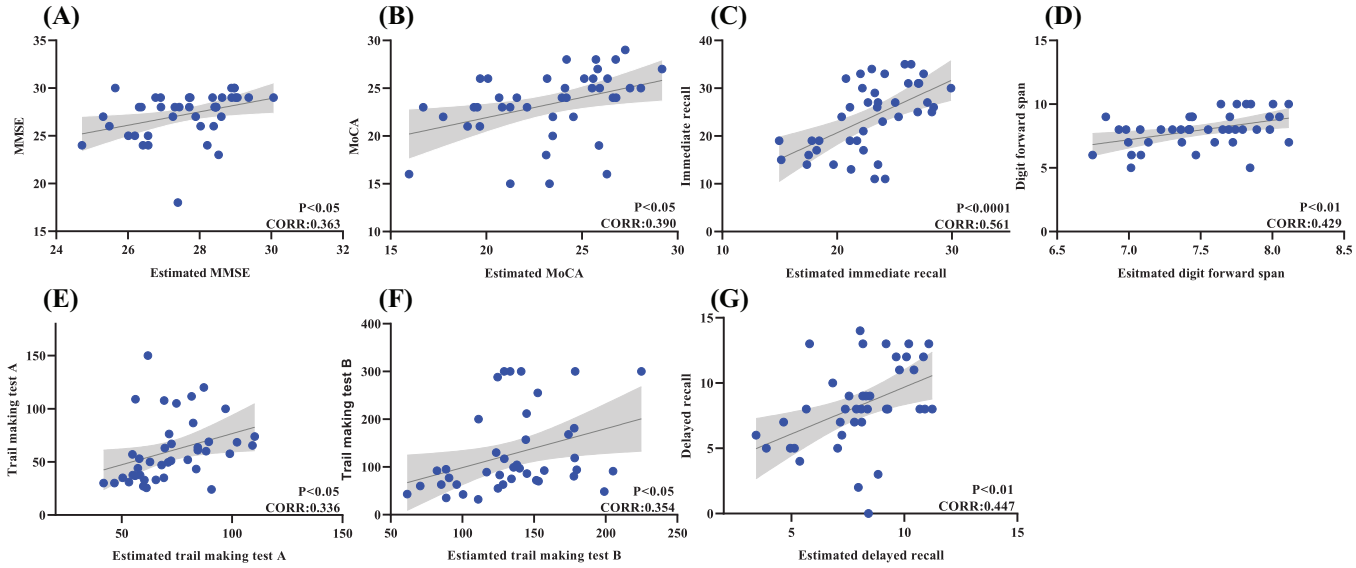

FIGURE 4 RVR-based regression of influential measures and neuropsychological assessments. A, Regression with MMSE. B, Regression with MoCA. C, Regression with immediate recall. D, Regression with digit span forward. E, Regression with TMT-A. F, Regression with TMT-B. G, Regression with delayed recall. The gray line is the fitted line, and the shaded part represents the 95% confidence interval. CORR, Pearson correlation coefficient; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; RVR, relevance vector regression; TMT-A, Trail-Making Test A; TMT-B, Trail-Making Test B.

part), rolandic operculum, precentral, and inferior frontal gyrus (opercular part). Because two different measures, KCC-ReHo and ALFF, were located in the inferior temporal gyrus, we prioritized this region for further analysis. We found most of the abnormal brain regions were in the default mode network (DMN) and frontoparietal control network (FPCN). The DMN is closely related to memory retrieval, cognition, and emotional processing, and brain dysfunction in patients with vascular cognitive impairment occurs mainly in DMN-related regions. FPCN is involved in cognitively demanding task and executive functions, and disruption of the FPCN has been found in neurological disorders.

We trained RVR to assess associations between the influential measures and the neuropsychological assessments (Figure 4). Seven scales in neuropsychological assessments were closely associated with influential measures in the corresponding brain regions. The memory functional domains, immediate recall (Pearson: 0.561 with $P < 0.0001$), delayed recall (Pearson: 0.447 with $P < 0.01$), and digit forward span (Pearson: 0.429 with $P < 0.01$), were the most highly correlated with influential measures. The holistic cognitive domain represented by MoCA (Pearson: 0.390 with $P < 0.05$) and MMSE (Pearson: 0.363 with $P < 0.05$) also correlated with the influential measures. Finally, we also observed a correlation between the executive function domain (TMT-B; Pearson: 0.354 with $P < 0.05$) and attention processing domain (TMT-A; Pearson: 0.336 with $P < 0.05$). The correlations between these neuropsychological assessments and the influential measures suggest that these influential measures may be useful as imaging markers to monitor cognitive performance and assess cognitive decline.

All patients were followed for 2 years to ensure that the diagnoses were accurate (Table S1 in supporting information). In patients with SIVD without cognitive impairment, there were no significant differences in MMSE and MoCA scores between baseline and the 2-year follow-up data. We observed a slight decline in MMSE and MoCA scores in patients with SVCI after 2 years, but the differences were not statistically significant, nor were they unexpected in a cohort with a progressive neurological disease.

4.4 | External validation

To validate model performance, we recruited a second cohort comprising 45 patients with SVCI and 32 patients with SIVD without cognitive impairment. Under the condition of unknown the group of SVCI and SIVD without cognitive impairment, the accuracy, sensitivity, specificity, and precision of the unsupervised machine learning model were 80.52%, 71.11%, 93.75%, and 94.12%, respectively. We found that some of the influential measures extracted from rsfMRI differed significantly between patients with SVCI and patients with SIVD without cognitive impairment within the internal cohort, but these measures were not different between patient groups in the external cohort. Some influential measures of rsfMRI were not as reliable as DTI measures. Thus, we observed slightly lower performance of the unsupervised machine learning model in the external cohort compared to the internal cohort, but such a mild drop in performance is typical upon external validation of machine learning.

5 | DISCUSSION

We aimed to develop a useful and accessible method to address an unmet clinical need: the ability to identify patients with SIVD at risk

of progressing to SVCI. To this end, we collected an extensive and diverse dataset including clinical and neuropsychological assessments and multimodal MRI data from a cohort of patients with SVCI and age-matched patients with SIVD without cognitive impairment who were followed for 2 years. Using an unsupervised machine learning approach to enable data-driven model building, we found that combining the multimodal measures DTI and rsfMRI yielded the best performance among 15 models tested. Therefore, our model only requires input from imaging protocols routinely used in the diagnosis of suspected SVCI based on VICCCS guidelines.[53] During validation using an external cohort of patients, the model demonstrated good accuracy, which indicates that our model is portable to other cohorts and is relevant for the context of clinical practice to distinguish patients with SIVD at risk for progressing to SVCI.

DTI measures the changes of WM microstructure connecting distinct brain regions. Although DTI makes it possible to visualize structural connectivity and lays a solid foundation for functional integration, it cannot be used to study the dynamic FC in brain networks and therefore cannot fully interpret the spatiotemporal dynamics process of the brain. By comparison, fMRI detects synchronous functional activities of related brain regions. To date, no study has directly compared or combined DTI and fMRI findings in patients with SVCI to analyze the mechanisms driving functional and structural changes. Our study combined computational modeling with both structural and functional imaging to elucidate how changes in structural connectivity as well as intrinsic dynamics alter functional networks. This approach identified particular measures and ROIs that contributed to the performance of the model, suggesting that there may be discrete measures that can be extracted from multimodal MRI that can provide information on disease state and prognosis in patients with SIVD or SVCI.

The so-called "disconnection hypothesis" has been put forward as the cause of the cognitive deficits in patients with SVCI.[53] The disconnection hypothesis posits that disconnection of specific WM tracts connecting gray matter regions with the distribution of neuronal networks results in cognitive impairment.[54] Evidence for the disconnection hypothesis comes from numerous experimental investigations of behavioral, functional, structural, and effective connectivity change in SVCI. Our study further supported the "disconnection hypothesis" and highlighted specific tract damage that may result in a functional change of particular brain regions and eventually cause cognitive impairment.

Currently, the most commonly used and most sensitive technique to study the structural changes in WM tracts in vivo is DTI, which quantifies the diffusivity of water molecules in brain tissue.[55] The typical pattern in SVCI is an increase in the extent of water diffusion and a reduction diffusion directionality. Compared to other types of dementia, patients with SVCI have WM alterations mainly in the frontal cortical regions, the genu of the corpus callosum, and periventricular regions.[56] Another study using tractography analysis illustrated that reduced FA in the transcallosal prefrontal tracts was the most effective imaging biomarker for SVCI.[57] Further, it has been demonstrated that these WM abnormalities are associated with impaired neurocognitive performance, which we also observed in our RVR model. In another study, changes in the corpus callosum detected by DTI were closely related to global cognitive function, executive function, psychomotor speed, and concept shifting in patients with SIVD.[58] However, it was not determined whether these anatomical changes were predictive of developing cognitive impairment in patients with SIVD. Notably, our findings showed that structural changes observed by DTI largely coincided with the anterior part of the brain associated with abnormal FC changes in patients with SVCI. Thus, our study identifies the predictive value of DTI measures for SVCI and supports the hypothesis that the distribution of WM tract damage that characterizes SVCI interrupts the prefrontal–subcortical loop, resulting in cognitive impairment.

The rsfMRI features that contributed to the performance of our model were mostly located in cortical regions, which is consistent with previous reports that cognitive functions emerge from communication between cortical and subcortical brain regions.[59] Liu et al. found negative correlation between the ALFF values of the left insula and the MoCA and MMSE scores of patients with SIVD,[60] and Peng et al. have shown that ReHo of related brain regions including the superior frontal gyrus and inferior frontal gyrus was significantly decreased in patients with SVCI.[61] Another study found that the decreased ReHo value in the anterior brain regions was significantly correlated with MMSE and MoCA scores in patients with SIVD, whereas increased ReHo was found in the posterior brain regions.[62] In patients with SVCI, FC changes are significant between different brain regions, and the global topological organization is extensively damaged.[63] In general, our findings support that resting-state local brain function and network connectivity can serve as a predictive biomarker for developing cognitive impairment in patients with SIVD and help delineate the functional changes that occur during progression to SVCI.

Our study developed and applied an unsupervised machine learning model to the comprehensive analysis of multimodal measures by designing a predictive task to constrain the model to extract influential measures. Further regression analysis showed that the influential measures we identified were correlated with neuropsychological assessments. Previous similar studies have usually resorted to a supervised algorithm to achieve prediction. Supervised machine learning requires most of the data to be used in the training set, which can leave insufficient data for validation. By using unsupervised machine learning, we were able to validate the performance of our model and confirm its portability to other cohorts. Our unsupervised machine learning model also has the advantage of easy tuning and interpretability, achieving promising performance while fully analyzing multimodal measures to isolate patients with SVCI.

The SVCI cohorts in this multi-center study were well characterized and followed for 2 years in a memory clinic setting. However, we had a limited dataset for our external cohort, which could result in the failure or success of the model in isolating individual patients with SVCI, significantly influencing changes in metrics. To address this limitation, we will continuously enroll patients for both cohorts, and we are willing to share our data with other researchers. Also, we found that influential fMRI measures were less robust than DTI measures, and it is possible that the current rsfMRI measures are not yet able to fully characterize functional brain activity. In future studies, we will try to

design measures that better represent brain activity or use deep learning technology to fully characterize changes in brain activity to improve the performance of the model. A final limitation of the model that must be considered is selection bias. Because our study excluded patients who were not able to complete neuropsychological assessment testing, our methodology may have selected a patient population with a higher education level compared to the general population by excluding illiteracy. Thus, caution is warranted regarding the generalizability of our findings to the general population.

In conclusion, our study established an unsupervised machine learning model for distinguishing patients with SIVD likely to progress to SVCI. Our model provides an accurate means to stratify patients according to risk of SVCI using multimodal imaging routinely used during the diagnosis of suspected SVCI. Moreover, our study suggests that structural and functional abnormalities can be used to inform on the entirety of the neurobiological pathways that contribute to SVCI.

ACKNOWLEDGMENTS

The authors have nothing to report.

CONFLICTS OF INTEREST STATEMENT

All authors have no actual or potential competing interests to declares. Author disclosures are available in the supporting information.

REFERENCES

1. Carey CL, Kramer JH, Josephson SA. et al. Subcortical lacunes are associated with executive dysfunction in cognitively normal elderly. *Stroke*. 2008;39:397-402.
2. Cannistraro RJ, Badi M, Eidelman BH, Dickson DW, Middlebrooks EH, Meschia JF. CNS small vessel disease: a clinical review. *Neurology*. 2019;92:1146-1156.
3. Grau-Olivares M, Arboix A, Bartrés-Faz D, Junqué C. Neuropsychological abnormalities associated with lacunar infarction. *J Neurol Sci*. 2007;257:160-165.
4. Jacova C, Pearce LA, Costello R, et al. Cognitive impairment in lacunar strokes: the SPS3 trial. *Ann Neurol*. 2012;72:351-362.
5. 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;14:280-292.
6. Stephens S, Kenny RA, Rowan E, et al. Neuropsychological characteristics of mild vascular cognitive impairment and dementia after stroke. *Int J Geriatr Psychiatry*. 2004;19:1053-1057.
7. Jia X, Wang Z, Huang F, et al. A comparison of the Mini-Mental State Examination (MMSE) with the Montreal Cognitive Assessment (MoCA) for mild cognitive impairment screening in Chinese middle-aged and older population: a cross-sectional study. *BMC Psychiatry*. 2021;21:485.
8. van de Pol LA, Korf ESC, van der Flier WM, et al. Magnetic resonance imaging predictors of cognition in mild cognitive impairment. *Arch Neurol*. 2007;64:1023-1028.
9. Wen W, Sachdev PS, Li JJ, Chen X, Anstey KJ. White matter hyperintensities in the forties: their prevalence and topography in an epidemiological sample aged 44-48. *Hum Brain Mapp*. 2009;30:1155-1167.
10. Wardlaw JM. Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study: the Rotterdam Scan Study. *J Neurol Neurosurg Psychiatry*. 2001;70:2-3.
11. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *AJR Am J Roentgenol*. 1987;149:351-356.
12. de Vocht F. Health complaints and cognitive effects caused by exposure to MRI scanner magnetic fields. *Tijdschr Diergeneeskd*. 2007;132:46-47.
13. Debette S, Bombois S, Bruandet A, et al. Subcortical hyperintensities are associated with cognitive decline in patients with mild cognitive impairment. *Stroke*. 2007;38:2924-2930.
14. Qin Q, Tang Y, Dou X, et al. Default mode network integrity changes contribute to cognitive deficits in subcortical vascular cognitive impairment, no dementia. *Brain Imaging Behav*. 2021;15:255-265.
15. Franzmeier N, Koutsouleris N, Benzinger T, et al. Predicting sporadic Alzheimer's disease progression via inherited Alzheimer's disease-informed machine-learning. *Alzheimers Dement*. 2020;16:501-511.
16. Gold G, Bouras C, Canuto A, et al. Clinicopathological validation study of four sets of clinical criteria for vascular dementia. *Am J Psychiatry*. 2002;159:82-87.
17. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health*. 2018;8:20303.
18. 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.
19. Muraoka S, DeLeo AM, Sethi MK, et al. Proteomic and biological profiling of extracellular vesicles from Alzheimer's disease human brain tissues. *Alzheimers Dement*. 2020;16:896-907.
20. Wang Z, Tang Z, Zhu Y, et al. AD risk score for the early phases of disease based on unsupervised machine learning. *Alzheimers Dement*. 2020;16:1524-1533.
21. Dong N, Fu C, Li R, et al. Machine learning decomposition of the anatomy of neuropsychological deficit in Alzheimer's disease and mild cognitive impairment. *Front Aging Neurosci*. 2022;14:854733.
22. Román GC, Erkinjuntti T, Wallin A, Pantoni L, Chui HC. Subcortical ischaemic vascular dementia. *Lancet Neurol*. 2002;1:426-436.
23. Scheltens P, van de Pol L. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *J Neurol Neurosurg & Psychiatry*. 2012;83:1038-1040.
24. Nakamura A, Kaneko N, Villemagne VL, et al. High performance plasma amyloid-β biomarkers for Alzheimer's disease. *Nature*. 2018;554:249-254.
25. Frank B, Ally M, Brekke B, et al. Plasma p-tau(181) shows stronger network association to Alzheimer's disease dementia than neurofilament light and total tau. *Alzheimers Dement*. 2022;18:1523-1536.
26. Gaser C, Dahnke R, Thompson P, Kurth F, Luders E. CAT - A Computational Anatomy Toolbox for the Analysis of Structural MRI Data 2022.
27. Yotter RA, Nenadic I, Ziegler G, Thompson PM, Gaser C. Local cortical surface complexity maps from spherical harmonic reconstructions. *Neuroimage*. 2011;56:961-973.
28. Luders E, Thompson PM, Narr KL, Toga AW, Jancke L, Gaser C. A curvature-based approach to estimate local gyrification on the cortical surface. *Neuroimage*. 2006;29:1224-1230.
29. Cui Z, Zhong S, Xu P, He Y, Gong G. PANDA: a pipeline toolbox for analyzing brain diffusion images. *Front Hum Neurosci*. 2013;7:1-16.
30. Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J Magn Reson*. 2011;213:560-570.
31. Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Di Chiro G. Diffusion tensor MR imaging of the human brain. *Radiology*. 1996;201:637-648.
32. Song S-K, Sun S-W, Ramsbottom MJ, Chang C, Russell J, Cross AH. Dysmyelination revealed through MRI as Increased radial (but unchanged axial) diffusion of water. *Neuroimage*. 2002;17:1429-1436.

33. Gong G. Local diffusion homogeneity (LDH): an inter-voxel diffusion MRI metric for assessing inter-subject white matter variability. *PLoS One*. 2013;8:e66366.
34. Whitfield-Gabrieli S, Conn Nieto-CastanonA. A functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect*. 2012;2:125-141.
35. Jia XZ, Wang J, Sun HY, et al. RESTplus: an improved toolkit for resting-state functional magnetic resonance imaging data processing. *Sci Bull*. 2019;64:953-954.
36. Wang JH, Zuo XN, Gohel S, Milham MP, Biswal BB, He Y. Graph theoretical analysis of functional brain networks: test-retest evaluation on short- and long-term resting-state functional MRI data. *PLoS One*. 2011;6(7):e21976.
37. Wang J, Wang X, Xia M, Liao X, Evans A, He Y. GRETNA: a graph theoretical network analysis toolbox for imaging connectomics. *Front Hum Neurosci*. 2015;9:386.
38. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002;15:273-289.
39. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006;31:968-980.
40. Hua K, Zhang J, Wakana S, et al. Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage*. 2008;39:336-347.
41. Cikes M, Sanchez-Martinez S, Claggett B, et al. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail*. 2019;21:74-85.
42. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods*. 2020;17:261-272.
43. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B*. 1996;58:267-288.
44. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825-2830.
45. van der Maaten LJP, Hinton GE. Visualizing High-dimensional data using t-SNE. *J Mach Learn Res*. 2008;9:2579-2605.
46. MacQueen J. Some methods for classification and analysis of multivariate observations. 1967.
47. Stoitsas K, Bahulikar S, de Munter L, et al. Clustering of trauma patients based on longitudinal data and the application of machine learning to predict recovery. *Sci Rep*. 2022;12:16990.
48. Arthur D, Vassilvitskii S. K-Means++: the advantages of careful seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. 2007;1027-1035.
49. Sharifi H, Lai YK, Guo H, et al. Machine Learning algorithms to differentiate among pulmonary complications after hematopoietic cell transplant. *Chest*. 2020;158:1090-1103.
50. Nguyen L, Van Hoeck A, Cuppen E. Machine learning-based tissue of origin classification for cancer of unknown primary diagnostics using genome-wide mutation features. *Nat Commun*. 2022;13:1-12.
51. Xia M, Wang J, He Y. BrainNet viewer: a network visualization tool for human brain connectomics. *PLoS One*. 2013;8:e68910.
52. Wang Y, Lu P, Zhan Y, et al. The contribution of white matter diffusion and cortical perfusion pathology to vascular cognitive impairment: a multimode imaging-based machine learning study. *Front Aging Neurosci*. 2021;13:1-11.
53. 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;12:822-838.
54. Baykara E, Gesierich B, Adam R, et al. A novel imaging marker for small vessel disease based on skeletonization of white matter tracts and diffusion histograms. *Ann Neurol*. 2016;80:581-592.
55. Pasi M, van Uden IWM, Tuladhar AM, de Leeuw F-E, Pantoni L. White matter microstructural damage on diffusion tensor imaging in cerebral small vessel disease: clinical consequences. *Stroke*. 2016;47:1679-1684.
56. Chen T-F, Lin C-C, Chen Y-F, et al. Diffusion tensor changes in patients with amnesic mild cognitive impairment and various dementias. *Psychiatry Res*. 2009;173:15-21.
57. Zarei M, Damoiseaux JS, Morgese C, et al. Regional white matter integrity differentiates between vascular dementia and Alzheimer disease. *Stroke*. 2009;40:773-779.
58. METACOHORTS for the study of vascular disease and its contribution to cognitive decline and neurodegeneration: an initiative of the joint programme for neurodegenerative disease research. *Alzheimers Dement*. 2016;12:1235-1249.
59. Park H-J, Friston K. Structural and functional brain networks: from connections to cognition. *Science*. 2013;342:1238411.
60. Liu C, Li C, Yin X, et al. Abnormal intrinsic brain activity patterns in patients with subcortical ischemic vascular dementia. *PLoS One*. 2014;9:e87880.
61. Peng C-Y, Chen Y-C, Cui Y, et al. regional coherence alterations revealed by resting-state FMRI in post-stroke patients with cognitive dysfunction. *PLoS One*. 2016;11:e0159574.
62. Ding X, Ding J, Hua B, et al. Abnormal cortical functional activity in patients with ischemic white matter lesions: a resting-state functional magnetic resonance imaging study. *Neurosci Lett*. 2017;644:10-17.
63. Jiang L, Geng W, Chen H, et al. Decreased functional connectivity within the default-mode network in acute brainstem ischemic stroke. *Eur J Radiol*. 2018;105:221-226.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Qin Q, Qu J, Yin Y, et al. Unsupervised machine learning model to predict cognitive impairment in subcortical ischemic vascular disease. *Alzheimer's Dement*. 2023;19:3327–3338. https://doi.org/10.1002/alz.12971