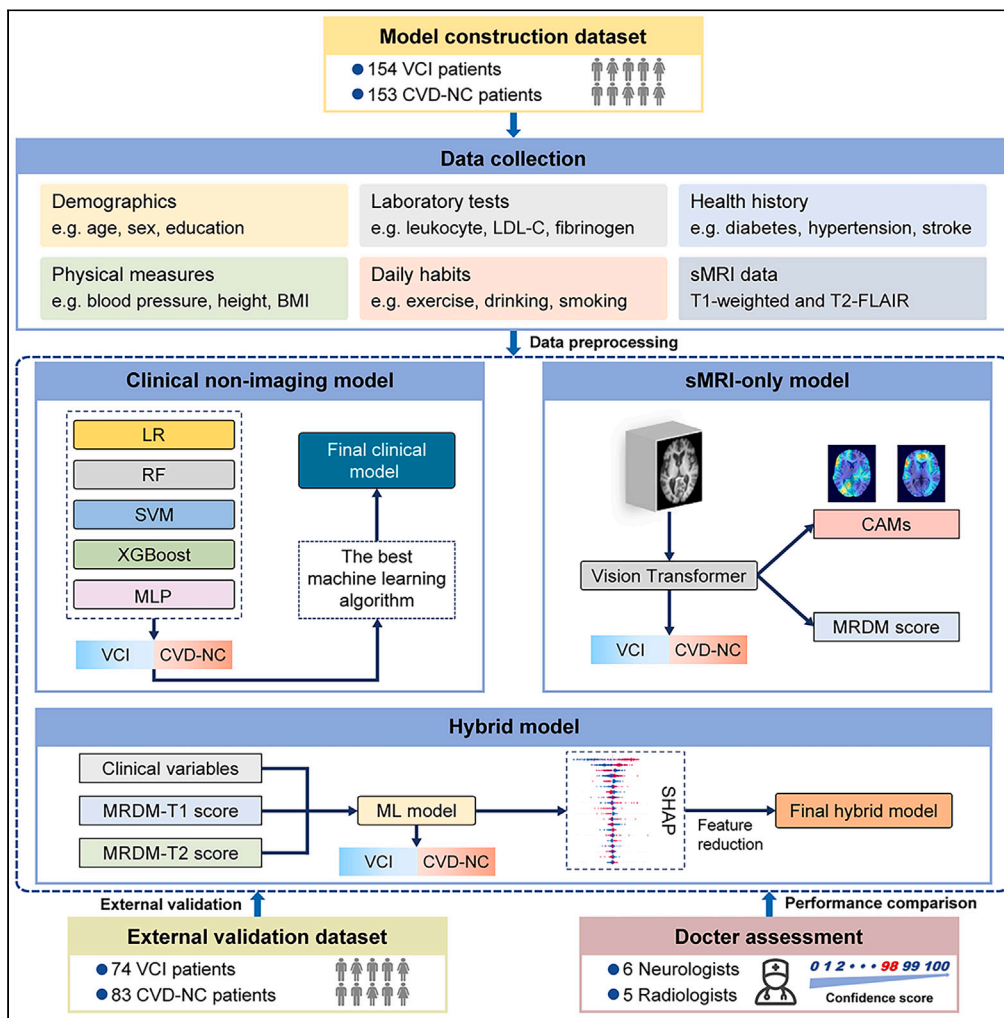CellPress
OPEN ACCESS

Article

Development and validation of a multimodal deep learning framework for vascular cognitive impairment diagnosis

Fan Fan, Hao Song, Jiu Jiang, ..., Ming Lei, Chu He, Junjian Zhang

chuhe@whu.edu.cn (C.H.)
zhangjj@whu.edu.cn (J.Z.)

Highlights
Our model provides clinicians with a clinical decision tool for diagnosing VCI

Our model performed excellently in both internal and external validation

The diagnostic performance of the model can be on par with that of clinical experts

Our model identifies key brain regions and clinical features related to VCI

Article

Development and validation of a multimodal deep learning framework for vascular cognitive impairment diagnosis

Fan Fan,[1,5] Hao Song,[1,5] Jiu Jiang,[2] Haoying He,[1] Dong Sun,[1] Zhipeng Xu,[1] Sisi Peng,[1] Ran Zhang,[1] Tian Li,[1] Jing Cao,[1] Juan Xu,[1] Xiaoxiang Peng,[3] Ming Lei,[4] Chu He,[2,*] and Junjian Zhang[1,6,*]

SUMMARY

Cerebrovascular disease (CVD) is the second leading cause of dementia worldwide. The accurate detection of vascular cognitive impairment (VCI) in CVD patients remains an unresolved challenge. We collected the clinical non-imaging data and neuroimaging data from 307 subjects with CVD. Using these data, we developed a multimodal deep learning framework that combined the vision transformer and extreme gradient boosting algorithms. The final hybrid model within the framework included only two neuroimaging features and six clinical features, demonstrating robust performance across both internal and external datasets. Furthermore, the diagnostic performance of our model on a specific dataset was demonstrated to be comparable to that of expert clinicians. Notably, our model can identify the brain regions and clinical features that significantly contribute to the VCI diagnosis, thereby enhancing transparency and interpretability. We developed an accurate and explainable clinical decision support tool to identify the presence of VCI in patients with CVD.

INTRODUCTION

Vascular cognitive impairment (VCI) is a recognized disease entity that encompasses a varying severity of cognitive impairments predominantly caused by cerebrovascular pathologies, spanning from mild cognitive impairment (MCI) to dementia.[1] Epidemiological findings indicate that vascular dementia (VD) constitutes approximately 15%–30% of dementia cases, ranking second only to Alzheimer's disease (AD).[2] In China, MCI associated with vascular factors prevails, constituting around 42% of all MCI cases.[3] With growing evidence suggesting a critical role for vascular interventions in dementia management,[4] early and accurate VCI diagnosis becomes imperative for crafting prevention and therapeutic strategies in cerebrovascular disease (CVD).

Diagnosing VCI is complex, hinging on the outcomes of clinical evaluation, structural magnetic resonance imaging (sMRI), and neuropsychological tests.[5] Nevertheless, achieving an accurate diagnosis of VCI based on these assessments remains challenging even for experienced neurologists. Currently, due to the lack of reliable VCI biomarkers, clinical evaluation typically necessitates a comprehensive medical history, neurological physical examination, and laboratory tests, further complicating clinical diagnosis. Additionally, sMRI plays a pivotal role in the VCI diagnosis. However, the relationship between neuroimaging markers of vascular brain injury, such as white matter hyperintensity (WMH), infarcts, and enlarged perivascular spaces (EPVSs), and the cognitive decline remains not entirely elucidated.[6] Besides, such MRI-visible lesions exclusively reflect the late and irreversible stages of cerebrovascular pathology. The neuropsychological tests of VCI patients content typically encompass both global and five domain-specific cognitive functions,[1] the complexity of which demands specially trained professionals. However, due to the unequal distribution of health resources, primary care institutions and economically underdeveloped areas often lack such professionals.[7] Moreover, subjective bias on the part of assessors can substantially impact the outcomes of neuropsychological assessments.[8] Hence, an efficient and easily generalizable model is crucial for the assessment of cognitive impairment in patients with CVD.

Artificial intelligence (AI) technology can develop disease-related assessment models by analyzing interactions and complex associations within large datasets.[9] It has shown significant potential in dementia-related classification tasks, enhancing the accuracy and efficiency of dementia diagnosis and providing reliable support for clinical neurologists.[10,11] Currently, most research has concentrated on employing machine learning (ML) approaches to distinguish between individuals with normal cognition and AD patients (including those with MCI and

[1]Department of Neurology, Zhongnan Hospital of Wuhan University, Wuhan, Hubei province, China
[2]Electronic Information School, Wuhan University, Wuhan, Hubei province, China
[3]Department of Neurology, Third People's Hospital of Hubei Province, Wuhan, Hubei province, China
[4]Department of Neurology, General Hospital of the Yangtze River Shipping, Wuhan, Hubei province, China
[5]These authors contributed equally
[6]Lead contact
*Correspondence: chuhe@whu.edu.cn (C.H.), zhangjj@whu.edu.cn (J.Z.)
https://doi.org/10.1016/j.isci.2024.110945

dementia), due to its prevalence in dementia diagnosis.[12–14] However, its utilization in the realm of VCI research is still limited. Given the significance of neuroimaging in the VCI diagnosis, certain recent studies have used conventional ML methods with neuroimaging data to diagnose VCI.[15–17] For instance, one of the studies developed a support vector machine (SVM) model leveraging the sMRI and quantitative electroencephalography features of subjects to discriminate VCI with no dementia from normal controls, achieving a classification accuracy with an area under the receiver operating characteristic curves (AUC) of 0.72. Notably, feature extraction of images and classifier construction are two separate steps in conventional ML,[18,19] demanding considerable time and workforce. Deep learning (DL) is a branch of ML that unifies feature extraction and classifier construction, enabling the automatic acquisition of high-throughput image features in a task-oriented manner without requiring experts to participate in the process.[20] With the swift advancement of DL techniques, the transformer algorithm, proposed in recent years, has also gained widespread application in image-related tasks. In comparison to convolutional neural network, a DL method commonly used in computer vision tasks, the self-attention mechanism in transformer possesses the advantage of efficiently capturing global information of the input image.[21] Several recent studies have shown that transformer has exhibited exceptional performance in neuroimaging-based AD classification tasks.[22,23] However, there is currently a lack of research on applying the transformer algorithm to the diagnosis of VCI.

Recent advancements indicated that incorporating multimodal data into DL models can enhance the performance of dementia diagnosis.[14,24] Furthermore, a study using an unsupervised ML method for predicting subcortical vascular cognitive impairment (SVCI) demonstrated that the model achieved optimal performance when combining diffusion-weighted imaging (DTI) and resting-state functional MRI (rs-fMRI) data.[25] However, an efficient model developed utilizing routine assessment data collected during CVD patient visits is still lacking. Hence, we developed a multimodal DL framework based on the T1-weighted MRI, T2-fluid-attenuated inversion recovery (FLAIR) MRI, and routine clinical assessment data. Our strategy offers evidence that an automated AI-based method could function as a clinical tool for the accurate diagnosis of VCI.

RESULTS

Demographics

To build the multimodal DL framework, we recruited a total of 307 subjects, of which 40.1% were female and the median age was 61 [56–67] years. To further validate the model performance, additional 157 patients were recruited as an external dataset, with females accounting for 41.4% and the median age of 67 [62–70] years. Moreover, to facilitate group analysis, we defined the subjects with VCI as the VCI group, and the CVD patients with normal cognition as the CVD-NC group. Detailed demographic information of the VCI and CVD-NC groups is shown in Table 1. In both datasets, compared to the CVD-NC group, the VCI group had higher age and lower education years, Mini-Mental State Examination (MMSE) scores, and Montreal Cognitive Assessment (MOCA) scores, with statistically significant differences. There were no significant differences in sex and body mass index between the two groups.

Performance of the multimodal DL model

Three separate models were employed for diagnosing VCI, including the sMRI-only model, clinical non-imaging model, and hybrid model. The sMRI-only model, constructed based on the vision transformer (ViT) network, demonstrated robust performance in diagnosing VCI (Table S1). Interestingly, we observed that the model performed better when T1 images were used as input compared to T2-FLAIR images (Figures 1A and 1B), with the AUCs of 0.905 [0.834–0.963] and 0.861 [0.769–0.938], respectively. Similarly, the area under the precision-recall curve (AP) exhibited comparable values, with 0.901 [0.819–0.969] and 0.828 [0.708–0.940], respectively.

In the clinical non-imaging model, we utilized five ML algorithms: extreme gradient boosting (XGBoost), logistic regression (LR), multi-layer perceptron (MLP), SVM, and random forest (RF). The AUCs and APs of these five models trained for the task of diagnosing VCI were as follows: 0.927 [0.874–0.974]/0.936 [0.886–0.976], 0.880 [0.808–0.940]/0.900 [0.829–0.953], 0.891 [0.821–0.953]/0.899 [0.821–0.963], 0.850 [0.768–0.921]/0.881 [0.806–0.942], and 0.865 [0.780–0.941]/0.859 [0.763–0.948]. XGBoost demonstrated superior performance, prompting us to select it for constructing the hybrid model. Figures 1C and 1D show the receiver operating characteristic (ROC) and precision-recall (PR) curves of the five ML models, and all performance metrics of these five models are provided in Figure S1.

The final hybrid model was constructed by combining the ViT and XGBoost models. The ROC curve and PR curve are shown in Figures 1E and 1F, and all performance metrics are provided in Table S1. We found that when fusing the T1 image features with clinical features and inputting them to the XGBoost model, it showed superior performance in diagnosing VCI compared to the sMRI-only model with input of the T1 image and the clinical non-imaging model. The AUC and AP were 0.965 [0.923–0.993] and 0.972 [0.940–0.994], respectively. Furthermore, when fusing the T2-FLAIR image features and clinical features simultaneously, the model still demonstrated excellent diagnostic performance, with the AUC and AP of 0.953 [0.903–0.992] and 0.964 [0.926–0.992], respectively. Notably, the fusion of T1, T2-FLAIR, and clinical features showed the best model performance, with the AUC and AP reaching 0.972 [0.934–0.997]/0.979 [0.951–0.997].

Utilizing Shapley additive explanations (SHAP) analysis, we ranked all imaging and non-imaging features in descending order of their contribution to the VCI diagnosis (Figure S2). Consecutive models were then constructed by gradually increasing the number of features based on the order of feature ranking. Figure 2A shows the trends of AUC and AP values for the consecutive models. It can be observed that the model performance increased steeply when the first several features were sequentially added and gradually stabilized with slight fluctuations when more features were added subsequently. Ultimately, we selected the top eight features as the inputs for constructing the hybrid model. As shown in Figure 2B, all performance metrics of the final model achieved a comparable level to the hybrid model with all input features, where the AUC and AP were 0.964 [0.920–0.998] and 0.963 [0.920–0.997], respectively.

Table 1. Demographic information

Characteristics	Model construction dataset			External validation dataset		
	VCI group (n = 154)	NC group (n = 153)	p value	VCI group (n = 74)	NC group (n = 83)	p value
Age (years)	64 [58–68]	59 [55–63.5]	<0.001	68.5 [66–71]	63 [60–69]	<0.001
Female, n (%)	58 (37.7%)	65 (42.5%)	0.389	30 (40.5%)	35 (42.2%)	0.837
Education (years)	9 [7–10]	12 [10–14]	<0.001	10 [9–12]	12 [10–14]	<0.001
BMI (kg/m^2)	23.92 ± 2.67	24.07 ± 2.63	0.626	24.94 ± 2.84	25.61 ± 2.90	0.147
MOCA (scores)	18 [13–20.25]	26 [24–27]	<0.001	22 [20–23]	25 [24–27]	<0.001
MMSE (scores)	25 [20–27]	29 [28–30]	<0.001	26 [25–28]	28 [27–29]	<0.001

Note: Continuous variables were expressed as mean ± standard deviation or median (interquartile range) and compared using independent-sample t test or Mann-Whitney U test between two groups. Categorical variables were expressed as counts (%), and chi-squared test was used for frequency comparison between two groups.
Abbreviations: VCI, vascular cognitive impairment; NC, cerebrovascular disease with normal cognition; BMI, body mass index; MOCA, Montreal Cognitive Assessment; MMSE, Mini-Mental State Examination.

Interpretability of the multimodal DL model

Our developed model possesses the capability to automatically identify the important brain regions associated with VCI from whole-brain sMRI and visualized them to generate class activation maps (CAMs). In Figure 3, we presented the mean CAMs generated by the ViT network constructed for the VCI diagnosis, which were displayed in two-dimensional form from three different perspectives (i.e., coronal, axial, and sagittal planes). Notably, it was observed that the highlighted brain regions in CAMs generated based on T1 image features were not identical to those generated based on T2-FLAIR image features. Furthermore, we quantified the contribution of T1 and T2-FLAIR image features to the VCI diagnosis within each brain region as two sets of importance scores. The top 10 anatomical regions with the highest contribution of T1 and T2-FLAIR image features to the VCI diagnosis were listed in Figures 4A and 4B, respectively, accounting more than 40% of the total importance score of all brain regions. Specifically, these regions comprise cerebellar vermis, bilateral cerebellar hemispheres, bilateral anterior cingulate gyrus, bilateral orbital frontal lobe, right superior parietal gyrus, and right amygdala. The importance scores of 116 anatomical regions in AAL atlas were listed in Table S2. Additionally, although the model identified the important brain regions associated with VCI, they were not entirely identical across different subjects. The CAMs of partial VCI subjects are shown in Figure S3.

To provide an interpretable diagnostic model for clinical neurologists, we utilized the SHAP method to explain the output of the hybrid model after feature reduction by calculating the contribution of each feature to the VCI diagnosis. As shown in Figure 5, the contribution of sMRI-derived features and various non-imaging features to the final hybrid model was evaluated utilizing the average SHAP values in descending order. It was found that MRI diagnostic model (MRDM)-T1 score, MRDM-T2 score, and use of antiplatelet drugs were the three most important features for diagnosing VCI, followed by education years, carotid artery stenosis, leukocyte counts, age, and diastolic blood pressure.

Furthermore, to explore the relationship between the sMRI-derived features generated by the ViT model and the neuroimaging markers potentially associated with VCI, we analyzed the association between MRDM scores and the neuroimaging markers of small vessel disease (SVD) using the spearman rank correlation test. As shown in Figure S4, both MRDM-T1 and MRDM-T2 scores have positive correlations with Fazekas scores, number of lacunes, basal ganglia (BG)-EPVS scores, and centrum semiovale (CSO)-EPVS scores. Among them, Fazekas scores exhibited the strongest correlations with MRDM-T1 ($r = 0.564$, $p < 0.001$) and MRDM-T2 scores ($r = 0.584$, $p < 0.001$), while CSO-EPVS scores showed weak correlations with them ($r = 0.382$, $p < 0.001$; $r = 0.308$, $p < 0.001$).

External validation of the final hybrid model

To validate the performance of the final hybrid model after feature reduction, 157 subjects were recruited as an external dataset. For external validation, the final model demonstrated an AUC and AP of 0.902 [0.859–0.941] and 0.900 [0.848–0.943], respectively, which were marginally below those observed in the internal dataset. However, this discrepancy did not reach statistical significance (ΔAUC = 0.061, $p = 0.074$), indicating that the final hybrid model performed excellently in both internal and external validations.

Diagnostic performance of the model and clinicians

To assess whether the performance of our models can be on par with the diagnostic performance of expert clinicians for VCI, we normalized the clinicians' average confidence scores for randomly selected cases from an external dataset to a scale of 0–1, aligning them with the model-predicted probabilities. Based on this, we calculated the performance metrics for clinicians diagnosing VCI. As shown in Figures 6A and 6B, the performance of the hybrid model was slightly higher than that of the neurologists (AUC: 0.893 [0.812–0.960] vs. 0.845 [0.750–0.921], $p = 0.437$; AP: 0.891 [0.801–0.959] vs. 0.762 [0.624–0.898]). Similarly, the diagnostic performance of the sMRI-only model using T1/T2-FLAIR inputs was higher than that of the radiologists (AUC: 0.841 [0.738–0.926]/0.843 [0.749–0.930] vs. 0.757 [0.642–0.854], $p = 0.252/0.186$; AP: 0.774 [0.620–0.914]/0.773 [0.636–0.921] vs. 0.699 [0.544–0.839]) (Figures 6C and 6D). These results indicate that our model at least matched
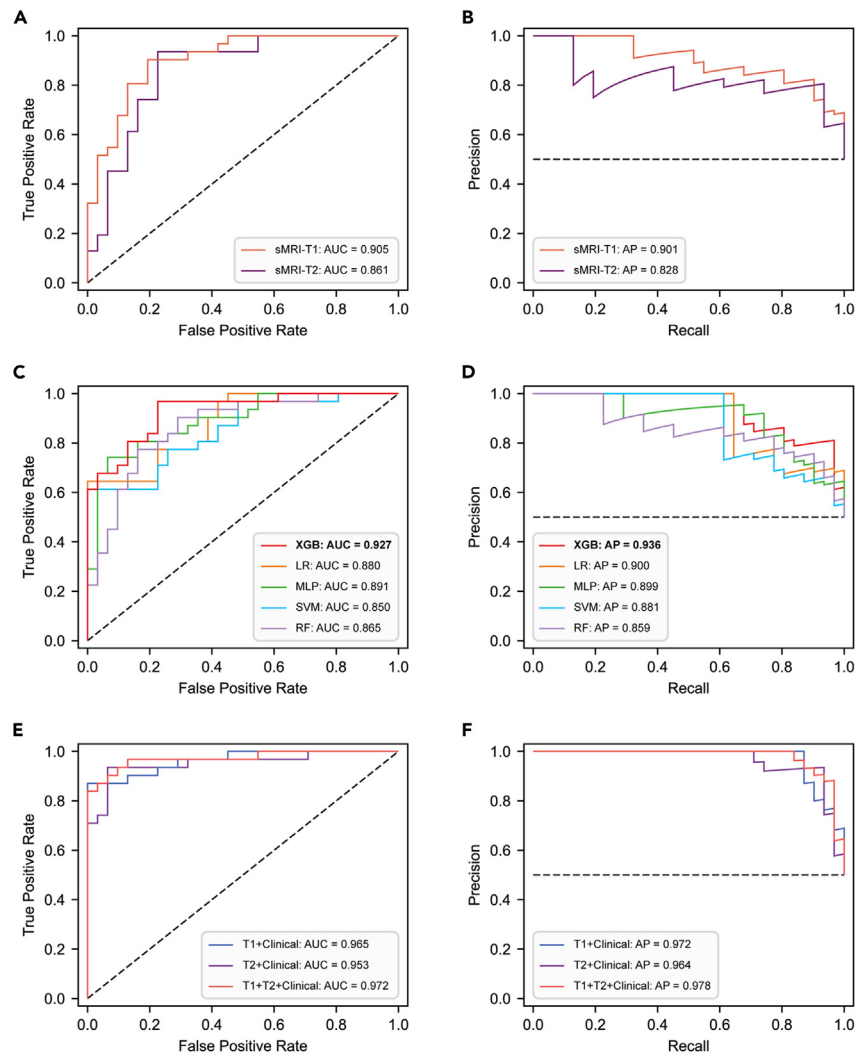
Figure 1. Performance of the multimodal deep learning model

(A and B) The ROC and PR curves of the sMRI-only model. The AUC values and AP values were utilized to evaluate the performance of the sMRI-only model with the T1 and T2-FLAIR images as inputs.

(C and D) The ROC and PR curves of the clinical non-imaging model. To determine the optimal clinical non-imaging model, the AUC and AP values of five machine learning models were compared.

(E and F) The ROC and PR curves of the hybrid model. Different combinations of sMRI-derived features and clinical variables were input into the XGBoost model, and model performance was evaluated by calculating AUC and AP values. sMRI, structural magnetic resonance imaging; AUC, area under the receiver operating characteristic curve; AP, area under the precision-recall curve; XGB, extreme gradient boosting; LR, logistic regression; MLP, multi-layer perceptron; SVM, support vector machine; RF, random forest; ROC, receiver operating characteristic; PR, precision-recall; FLAIR, fluid-attenuated inversion recovery; VCI, vascular cognitive impairment.

the diagnostic ability of the clinicians. Additionally, we assessed the inter-rater agreement among the clinicians' confidence scores (Figure S5). We found strong consistency among the neurologists' assessments ($r = 0.69 \pm 0.02$), whereas only moderate consistency was found among the radiologists ($r = 0.55 \pm 0.05$). This suggests that clinicians' diagnoses vary more when less clinical information is available.

DISCUSSION

In this study, we developed a hybrid model based on multimodal inputs, building upon a sMRI-only model and a clinical non-imaging model. After feature reduction based on feature importance ranking, the final hybrid model inputted only two sMRI-derived features and eight non-imaging features, showing excellent performance in the VCI diagnosis across both internal and external datasets. To augment the interpretability of the models, we identified the brain regions and clinical variables that contribute most to the VCI diagnosis. Moreover, in a randomly selected subset of cases, the diagnostic performance of our models is on par with that of expert clinicians.
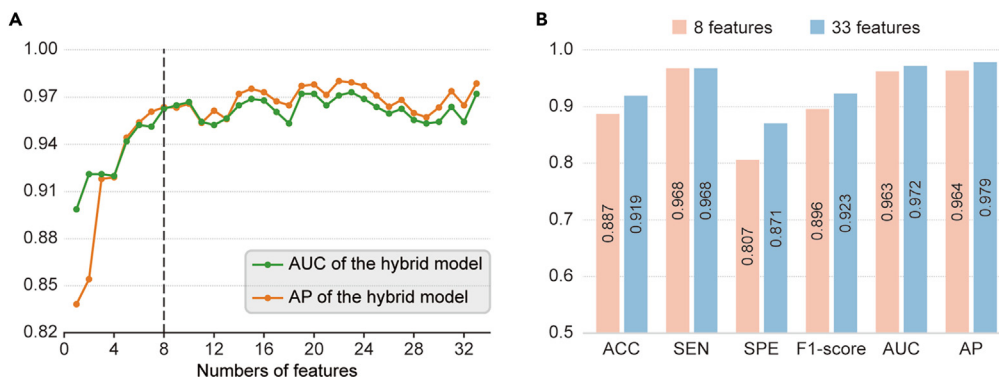
Figure 2. Feature reduction of the hybrid model

(A) The AUC and AP values of the XGBoost model with various numbers of features. According to the contribution of each feature to the VCI diagnosis in descending order, sequential forward selection was performed from 33 pre-selected features.

(B) Accuracy, sensitivity, specificity, F1-score, AUC, and AP of the hybrid model with 8 features and 33 features. AUC, area under the receiver operating characteristic curve; AP, area under the precision-recall curve; ACC, accuracy; SEN, sensitivity; SPE, specificity; XGBoost, extreme gradient boosting; VCI, vascular cognitive impairment.

Neuroimaging is now an indispensable aspect of VCI clinical assessment, particularly with various advanced MRI sequences such as DTI, rs-fMRI, and perfusion-weighted imaging.[26] However, the expensive and time-consuming advanced MRI scanning significantly increases the financial burden on patients, making it challenging to promote their use in medical institutions at all levels. Therefore, our model only utilized sMRI data, aiming to substantially reduce MRI examination time and alleviate the stress on patients caused by prolonged examinations. Previous studies usually inputted only sMRI T1 images to AD diagnostic models.[13,27] This might be attributed to the consideration of hippocampal and temporal lobe atrophy as imaging markers of AD, with these features being more easily identifiable in T1 imaging.[28] However,
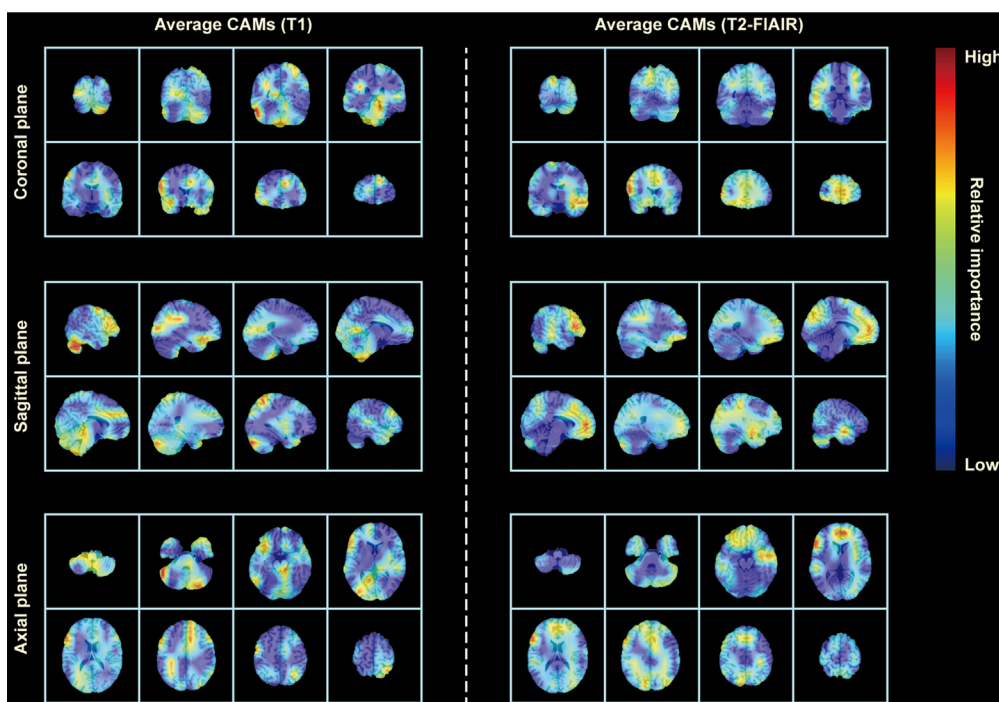
Figure 3. Average CAM generated on test subjects by ViT

CAMs generated by the ViT model on colin27 average brain template consistently highlighted high-risk brain regions associated with VCI, when T1 and T2-FLAIR images were input to the model, respectively. The color red indicates a higher contribution of the highlighted brain regions to the diagnosis of VCI, whereas the color blue indicates a comparatively lower contribution to the VCI diagnosis. CAM, class activation map; ViT, vision transformer; VCI, vascular cognitive impairment; FLAIR, fluid-attenuated inversion recovery.
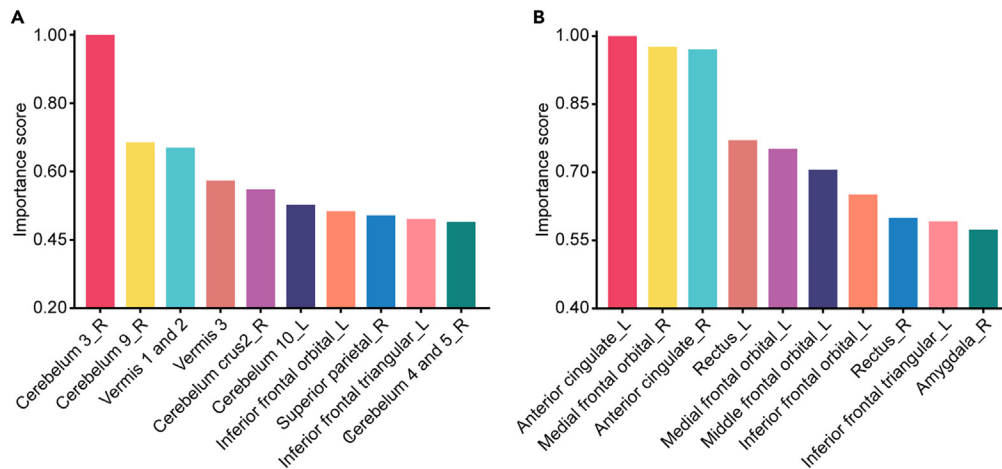
Figure 4. Top 10 anatomical brain regions ranked by their contribution to the VCI diagnosis
The contribution of T1 (A) and T2-FLAIR (B) image features to the VCI diagnosis within each brain region was quantified by calculating the importance scores, respectively. The top 10 brain regions collectively represented more than 40% of the total importance score of all brain regions. VCI, vascular cognitive impairment; FLAIR, fluid-attenuated inversion recovery.

neuroimaging features related to VCI, such as WMH, were more easily identifiable in T2-FLAIR imaging.[29] Therefore, we incorporated both the T1 and T2-FLAIR sequence data in this model, aiming to extract comprehensive high-level image features through DL networks, thereby achieving higher diagnostic accuracy. This study employed the ViT algorithm to construct a DL model based on image data, which could effectively isolate VCI in CVD patients, suggesting that ViT may have advantages in image-based VCI classification task. VCI is a highly heterogeneous syndrome, characterized by various neuroimaging features in its sMRI. These features potentially present in any region of the brain, emphasizing the significance of considering the interrelationships among different brain regions in neuroimaging. ViT can utilize the self-attention mechanism to capture long-range spatial dependencies in the images,[30] thus effectively extracting the image features related to VCI. Furthermore, considering that sMRI data consist of many slices with continuous spatial positional relationships, we used 3D images as the input to the ViT network instead of 2D images, a choice that has been proven to be effective in previous studies.[31]

Due to the diverse and complex pathophysiology and influencing factors of VCI,[4] a combination of imaging and other clinical variables (e.g., age, vascular risk factors, etc.) is required to enhance the diagnostic accuracy of VCI. Therefore, our hybrid model, which combined sMRI data with clinical non-imaging data, demonstrated higher classification accuracy in diagnostic task compared to the sMRI-only model. Nevertheless, in contrast to several dementia diagnostic models using complex and subjective neuropsychological tests or expensive whole-genome sequencing,[24,32] our multimodal VCI diagnostic model relied exclusively on routine clinical variables collected during CVD patient visits; these variables can be collected through rapid questionnaires, physical measures, and simple blood tests. Moreover, in the absence of established guidelines or consensus on feature selection for the VCI diagnostic model, determining the optimal number of features is challenging. Although more features may provide comprehensive information for the classification model, an excessive inclusion may compromise the model's clinical applicability, and non-causal associations between features and outcomes may reduce the diagnostic accuracy.[33] To
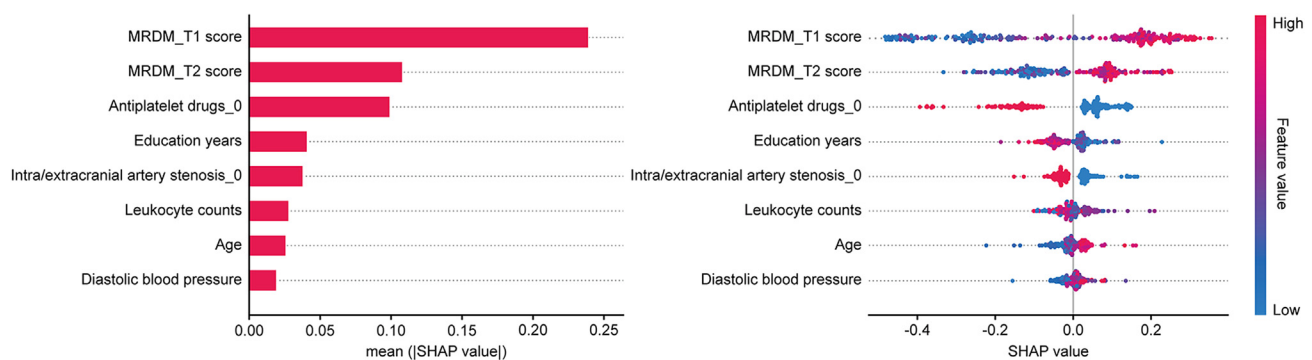
Figure 5. Final model explanation by the SHAP method
SHAP analysis was utilized to rank the contribution of eight imaging and non-imaging features in the final hybrid model to the diagnosis of VCI. The left plot demonstrates the mean absolute SHAP values, and the right plot illustrates the distribution of SHAP values. All features are sorted in descending order of the mean absolute SHAP values. SHAP, Shapley additive explanations; VCI, vascular cognitive impairment.
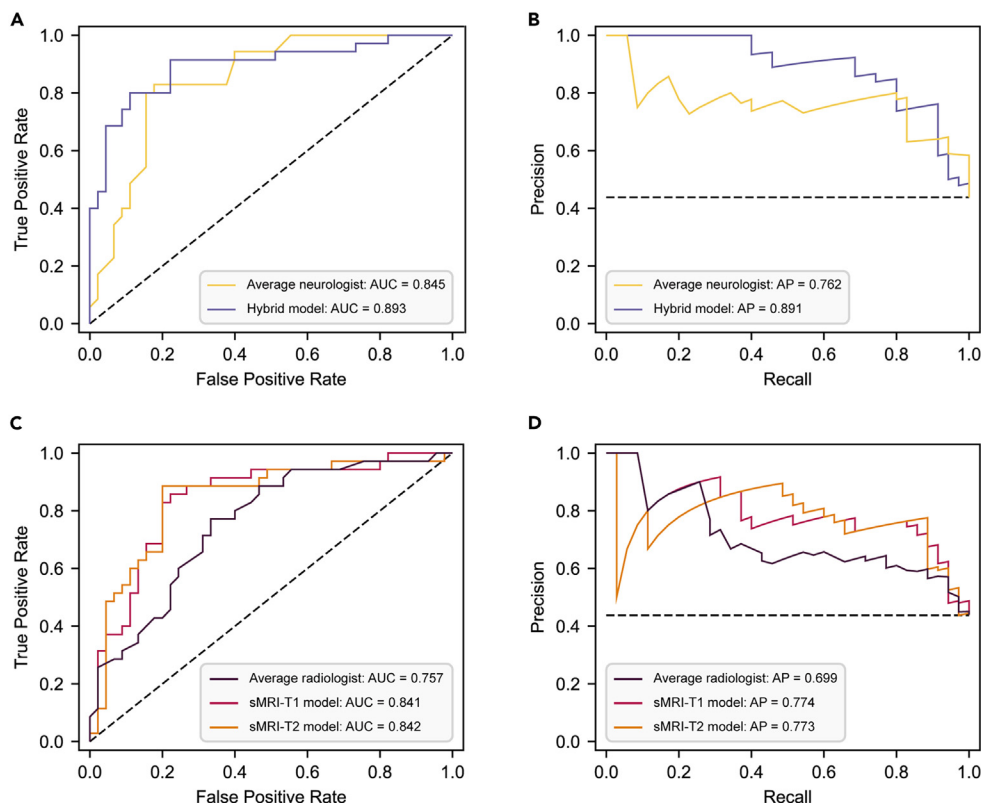
Figure 6. Diagnostic performance of clinicians and models

Six neurologists and five radiologists performed clinical assessments of 80 randomly selected cases, providing confidence scores (0–100) for each case, where higher scores indicated a greater probability of the VCI diagnosis. The confidence scores from neurologists and radiologists were averaged separately to produce a single consensus confidence score for each case. By averaging the confidence scores, we calculated the diagnostic performance of the average neurologist and radiologist.

(A and B) The ROC and PR curves for the average neurologist and the hybrid model.

(C and D) The ROC and PR curves for the average radiologist and the sMRI-only model. The AUC and AP values were utilized to compare the diagnostic performance of the average neurologist with the hybrid model and the average radiologist with the sMRI-only model, respectively. AUC, area under the receiver operating characteristic curve; AP, area under the precision-recall curve; sMRI, structural magnetic resonance imaging; VCI, vascular cognitive impairment; ROC, receiver operating characteristic; PR, precision-recall.

address these limitations, the least absolute shrinkage and selection operator and SHAP methods were utilized to assist feature reduction. Our final model only included eight features, performed well both in internal and external validation, and might be easily generalized to medical institutions at all levels.

The ML approach was usually characterized as a "black box," lacking the ability to provide direct explanations for the outputs of the models.[34] Such lack of transparency may foster uncertainty and doubt among clinicians about the reliability of model outcomes, consequently impacting clinical decision-making processes. In this study, the gradient-weighted CAM and SHAP methods were employed to interpret the sMRI-only model and the final hybrid model, respectively.

Our study explained the validity of DL models by highlighting the brain regions related to the VCI diagnosis by leveraging CAMs, offering potential evidence for the underlying pathophysiology of VCI. Cognitive deficits in VCI patients manifest variably due to the involvement of various brain regions.[35] Early studies revealed the correlation between executive dysfunction and cortical thinning as well as reduced gray matter volume in the prefrontal lobe of VCI patients.[36,37] In addition, a variety of studies utilizing DTI and rs-fMRI have detected white matter microstructural damages and functional brain abnormalities in the prefrontal lobe of SVCI patients, which interrupt the prefrontal-subcortical loop that results in specific cognitive impairments.[38] Our findings indicated that the prefrontal lobe harbors several brain regions crucial for VCI diagnosis, suggesting that DL networks might be able to effectively identify VCI by extracting image features related to prefrontal lobe structural damage from sMRI. Emerging evidence suggested that the cerebellum, apart from its established role in in sensorimotor control, plays a pivotal role in a wide spectrum of cognitive and affective functions.[39] A recent study demonstrated that mild VCI is associated with anatomical atrophy of specific cerebellar regions involved in cognitive functions and reduced functional connectivity with the striatum.[40] Moreover, an autopsy-based study found that the incidence of cerebellar microbleeds and microinfarcts significantly increased in patients with VD.[41] Another experimental study further explored the mechanisms of cerebellar damage following chronic cerebral hypoperfusion.[42]

Consistent with the aforementioned evidence, several anatomical regions of the cerebellum showed pronounced contribution to the VCI diagnosis in our report, further emphasizing the role of the cerebellum in cognitive processes of CVD patients. Notably, our model also highlighted the importance of the amygdala and partial regions of the parietal lobe in the VCI diagnosis, a finding supported by limited prior research. An earlier study showed that the parietal cortex in moderate to severe VCI patients was thinner compared to mild VCI patients.[43] Furthermore, it was observed that the best radiomic features associated with the diagnosis of SVCI without dementia were in four anatomical regions, including the right amygdala.[15] In general, our findings might provide an expanded perspective for further exploring the structural and functional changes of the brain regions related to VCI. It is noteworthy that the discriminative brain regions identified by CAMs varied among different subjects, potentially related to the presence of different regions of pathological brain damage in different subjects. This suggested that our developed model could achieve individualized localization of VCI-related brain regions, providing auxiliary guidance for clinicians in formulating individualized diagnosis and treatment plans. Furthermore, the sMRI-derived features generated by our model showed a positive correlation with multiple neuroimaging markers of SVD, supporting the notion that SVD is one of the most prevalent pathological mechanisms underlying VCI.[44]

In the final hybrid model, SHAP analysis showed that MRDM-T1 and MRDM-T2 scores exerted a greater impact on diagnosing VCI than traditional demographic and vascular risk factors, underscoring the value of neuroimaging in VCI diagnosis. Notably, the use of antiplatelet drugs demonstrated a diagnostic relevance akin to MRDM-T2 scores. Currently, there are controversial results on the role of antiplatelet drugs in VCI prevention and management. A systematic review found that antiplatelet therapy might serve as a preventive strategy for patients with a risk of VD.[45] However, Kwan et al. indicated there is insufficient evidence supporting the antiplatelet therapy for preventing cognitive decline in non-demented patients with cerebral SVD.[46] Our findings found that patients taking antiplatelet drugs have a higher likelihood of being diagnosed with VCI, presumably related to their comorbid CVD and other vascular risk factors, which are key etiologies of VCI. Age and education level significantly contribute to the VCI diagnosis, aligning with established risk factors for dementia. Additionally, diastolic blood pressure emerged as an important feature captured by the model. A large sample study found a cross-sectional association between higher diastolic blood pressure and impaired cognitive status, supporting our findings.[47] Interestingly, leukocyte counts contribute to the VCI diagnosis. Inflammation may be one of the underlying pathophysiological mechanisms of VCI.[48] A recent meta-analysis indicated that inflammatory markers can be used to differentiate between VD and AD patients, with higher levels of inflammatory markers associated with a higher risk of VD.[49] As a peripheral inflammatory marker, leukocyte counts may be associated with the presence of VCI, and our results provide clinical evidence supporting this hypothesis. Further research is needed to validate this association.

In conclusion, we have developed a multimodal DL framework to identify VCI from CVD patients based on routine neuroimaging and clinical non-imaging data collected during patient visits. Our model demonstrated excellent ability to diagnose VCI in both internal and external validations and was further validated against the assessment results from expert clinicians. Furthermore, the model is explainable, with the ViT network highlighting the brain regions associated with VCI, and SHAP analysis identified features that contribute more significantly to the VCI diagnosis, providing additional clinical evidence for further exploration of risk factors and brain pathological changes related to VCI.

Limitations of the study

Our study had several limitations. First, while ML modeling necessitates extensive datasets, no standard is currently available for calculating the sample sizes needed to development of ML models. Despite constructing this model with a modest dataset size of 307 subjects, we mitigated overfitting risks through well-performed internal cross-validation and external validation, suggesting an appropriate sample size may provide sufficient power for exploring the VCI diagnostic model. Second, the model was developed based on a population from the central China, and its generalizability to the entire Chinese population or even the global population was unclear. Third, according to the diagnostic criteria for VCI, cognitive impairments-mixed AD or other pathological changes still existed among the included VCI subjects. Therefore, our model might default to diagnosing these types of patients as VCI. Since mixed dementia is quite common, future work may establish a multi-label classification model based on this study, which can identify different pathological types of dementia within the same individual.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Junjian Zhang (e-mail: zhangjj@whu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data reported in this paper will be shared by the lead contact upon request.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

F.F., H.S., C.H., and J.Z. contributed to the conception and design of the study. J.J. and C.H. contributed to the construction of the deep learning model. F.F., H.S., and H.H. contributed to the analysis of machine learning and statistical analysis. F.F., H.S., H.H., D.S., Z.X., S.P., R.Z., T.L., X.P., and M.L. contributed to the data collection and establishment of the database. J.C., J.X., and R.Z. contributed to the neuropsychological and clinical assessments. F.F., J.J., and H.S. contributed to drafting the text and preparing the figures. C.H. and J.Z. contributed to reviewing and revising the manuscript. All authors have read the final manuscript and accept responsibility for the submission for publication.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
  - External validation
  - Ethical statement
- METHOD DETAILS
  - Neuropsychological assessments
  - MRI protocol
  - Assessment of SVD neuroimaging markers
  - Data collection and preprocessing
  - Model development
  - sMRI-only model
  - Clinical non-imaging model
  - Hybrid model
  - Model interpretability and feature reduction
  - Expert clinicians validation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis
  - Hardware and software for model development

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110945.

REFERENCES

1. van der Flier, W.M., Skoog, I., Schneider, J.A., Pantoni, L., Mok, V., Chen, C.L.H., and Scheltens, P. (2018). Vascular cognitive impairment. Nat. Rev. Dis. Primers 4, 18003. https://doi.org/10.1038/nrdp.2018.3.

2. Wolters, F.J., and Ikram, M.A. (2019). Epidemiology of Vascular Dementia. Arterioscler. Thromb. Vasc. Biol. 39, 1542–1549. https://doi.org/10.1161/atvbaha.119.311908.

3. Jia, L., Quan, M., Fu, Y., Zhao, T., Li, Y., Wei, C., Tang, Y., Qin, Q., Wang, F., Qiao, Y., et al. (2020). Dementia in China: epidemiology, clinical management, and research advances. Lancet Neurol. 19, 81–92. https://doi.org/10.1016/s1474-4422(19)30290-x.

4. Rundek, T., Tolea, M., Ariko, T., Fagerli, E.A., and Camargo, C.J. (2022). Vascular Cognitive Impairment (VCI). Neurotherapeutics 19, 68–88. https://doi.org/10.1007/s13311-021-01170-y.

5. Skrobot, O.A., Black, S.E., Chen, C., DeCarli, C., Erkinjuntti, T., Ford, G.A., Kalaria, R.N., O'Brien, J., Pantoni, L., Pasquier, F., et al. (2018). Progress toward standardized diagnosis of vascular cognitive impairment: Guidelines from the Vascular Impairment of Cognition Classification Consensus Study. Alzheimers Dement. 14, 280–292. https://doi.org/10.1016/j.jalz.2017.09.007.

6. Debette, S., Schilling, S., Duperron, M.G., Larsson, S.C., and Markus, H.S. (2019). Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury: A Systematic Review and Meta-analysis. JAMA Neurol. 76, 81–94. https://doi.org/10.1001/jamaneurol.2018.3122.

7. Zhang, T., Xu, Y., Ren, J., Sun, L., and Liu, C. (2017). Inequality in the distribution of health resources and health services in China: hospitals versus primary care institutions. Int. J. Equity Health 16, 42. https://doi.org/10.1186/s12939-017-0543-9.

8. Jia, X., Wang, Z., Huang, F., Su, C., Du, W., Jiang, H., Wang, H., Wang, J., Wang, F., Su, W., et al. (2021). 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 Psychiatr. 21, 485. https://doi.org/10.1186/s12888-021-03495-6.

9. Noorbakhsh-Sabet, N., Zand, R., Zhang, Y., and Abedi, V. (2019). Artificial Intelligence Transforms the Future of Health Care. Am. J. Med. 132, 795–801. https://doi.org/10.1016/j.amjmed.2019.01.017.

10. Martin, S.A., Townend, F.J., Barkhof, F., and Cole, J.H. (2023). Interpretable machine learning for dementia: A systematic review. Alzheimers Dement. 19, 2135–2149. https://doi.org/10.1002/alz.12948.

11. Borchert, R.J., Azevedo, T., Badhwar, A., Bernal, J., Betts, M., Bruffaerts, R., Burkhart, M.C., Dewachter, I., Gellersen, H.M., Low, A., et al. (2023). Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: A systematic review. Alzheimers Dement. 19, 5885–5904. https://doi.org/10.1002/alz.13412.

12. Ahmed, M.R., Zhang, Y., Feng, Z., Lo, B., Inan, O.T., and Liao, H. (2019). Neuroimaging and Machine Learning for Dementia Diagnosis: Recent Advancements and Future Prospects. IEEE Rev. Biomed. Eng. 12, 19–33. https://doi.org/10.1109/RBME.2018.2886237.

13. Qiu, S., Joshi, P.S., Miller, M.I., Xue, C., Zhou, X., Karjadi, C., Chang, G.H., Joshi, A.S., Dwyer, B., Zhu, S., et al. (2020). Development

and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain 143, 1920–1933. https://doi.org/10.1093/brain/awaa137.

14. Golovanevsky, M., Eickhoff, C., and Singh, R. (2022). Multimodal attention-based deep learning for Alzheimer's disease diagnosis. J. Am. Med. Inform. Assoc. 29, 2014–2022. https://doi.org/10.1093/jamia/ocac168.

15. Liu, B., Meng, S., Cheng, J., Zeng, Y., Zhou, D., Deng, X., Kuang, L., Wu, X., Tang, L., Wang, H., et al. (2022). Diagnosis of Subcortical Ischemic Vascular Cognitive Impairment With No Dementia Using Radiomics of Cerebral Cortex and Subcortical Nuclei in High-Resolution T1-Weighted MR Imaging. Front. Oncol. 12, 852726. https://doi.org/10.3389/fonc.2022.852726.

16. Wang, Y., Lu, P., Zhan, Y., Wu, X., Qiu, Y., Wang, Z., Xu, Q., and Zhou, Y. (2021). The Contribution of White Matter Diffusion and Cortical Perfusion Pathology to Vascular Cognitive Impairment: A Multimode Imaging-Based Machine Learning Study. Front. Aging Neurosci. 13, 687001. https://doi.org/10.3389/fnagi.2021.687001.

17. Li, Z., Wu, M., Yin, C., Wang, Z., Wang, J., Chen, L., and Zhao, W. (2024). Machine learning based on the EEG and structural MRI can predict different stages of vascular cognitive impairment. Front. Aging Neurosci. 16, 1364808. https://doi.org/10.3389/fnagi.2024.1364808.

18. Sørensen, L., Igel, C., Liv Hansen, N., Osler, M., Lauritzen, M., Rostrup, E., and Nielsen, M.; Alzheimer's Disease Neuroimaging Initiative and the Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing (2016). Early detection of Alzheimer's disease using MRI hippocampal texture. Hum. Brain Mapp. 37, 1148–1161. https://doi.org/10.1002/hbm.23091.

19. Rathore, S., Habes, M., Iftikhar, M.A., Shacklett, A., and Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. Neuroimage 155, 530–548. https://doi.org/10.1016/j.neuroimage.2017.03.057.

20. Lian, C., Liu, M., Pan, Y., and Shen, D. (2022). Attention-Guided Hybrid Network for Dementia Diagnosis With Structural MR Images. IEEE Trans. Cybern. 52, 1992–2003. https://doi.org/10.1109/tcyb.2020.3005859.

21. Matsoukas, C., Fredin Haslum, J., Söderberg, M., and Smith, K. (2021). Is it Time to Replace CNNs with Transformers for Medical Images?. Preprint at arXiv. https://doi.org/10.48550/arXiv.2108.09038.

22. Sarraf, S., Sarraf, A., DeSouza, D.D., Anderson, J.A.E., and Kabia, M.; The Alzheimer's Disease Neuroimaging Initiative (2023). OViTAD: Optimized Vision Transformer to Predict Various Stages of Alzheimer's Disease Using Resting-State fMRI and Structural MRI Data. Brain Sci. 13, 260. https://doi.org/10.3390/brainsci13020260.

23. Hoang, G.M., Kim, U.H., and Kim, J.G. (2023). Vision transformers for the prediction of mild cognitive impairment to Alzheimer's disease progression using mid-sagittal sMRI. Front. Aging Neurosci. 15, 1102869. https://doi.org/10.3389/fnagi.2023.1102869.

24. Qiu, S., Miller, M.I., Joshi, P.S., Lee, J.C., Xue, C., Ni, Y., Wang, Y., De Anda-Duran, I., Hwang, P.H., Cramer, J.A., et al. (2022). Multimodal deep learning for Alzheimer's

disease dementia assessment. Nat. Commun. 13, 3404. https://doi.org/10.1038/s41467-022-31037-5.

25. Qin, Q., Qu, J., Yin, Y., Liang, Y., Wang, Y., Xie, B., Liu, Q., Wang, X., Xia, X., Wang, M., et al. (2023). Unsupervised machine learning model to predict cognitive impairment in subcortical ischemic vascular disease. Alzheimers Dement. 19, 3327–3338. https://doi.org/10.1002/alz.12971.

26. Heiss, W.D., Rosenberg, G.A., Thiel, A., Berlot, R., and de Reuck, J. (2016). Neuroimaging in vascular cognitive impairment: a state-of-the-art review. BMC Med. 14, 174. https://doi.org/10.1186/s12916-016-0725-0.

27. Lian, C., Liu, M., Zhang, J., and Shen, D. (2020). Hierarchical Fully Convolutional Network for Joint Atrophy Localization and Alzheimer's Disease Diagnosis Using Structural MRI. IEEE Trans. Pattern Anal. Mach. Intell. 42, 880–893. https://doi.org/10.1109/tpami.2018.2889096.

28. Frisoni, G.B., Fox, N.C., Jack, C.R., Jr., Scheltens, P., and Thompson, P.M. (2010). The clinical use of structural MRI in Alzheimer disease. Nat. Rev. Neurol. 6, 67–77. https://doi.org/10.1038/nrneurol.2009.215.

29. Duering, M., Biessels, G.J., Brodtmann, A., Chen, C., Cordonnier, C., de Leeuw, F.E., Debette, S., Frayne, R., Jouvent, E., Rost, N.S., et al. (2023). Neuroimaging standards for research into small vessel disease-advances since 2013. Lancet Neurol. 22, 602–618. https://doi.org/10.1016/s1474-4422(23)00131-x.

30. He, K., Gan, C., Li, Z., Rekik, I., Yin, Z., Ji, W., Gao, Y., Wang, Q., Zhang, J., and Shen, D. (2022). Transformers in Medical Image Analysis: A Review. Preprint at arXiv. https://doi.org/10.48550/arXiv.2202.12165.

31. Dhinagar, N.J., Thomopoulos, S.I., Laltoo, E., and Thompson, P.M. (2023). Efficiently Training Vision Transformers on Structural MRI Scans for Alzheimer's Disease Detection. Preprint at arXiv. https://doi.org/10.48550/arXiv.2303.08216.

32. Yaqub, A., Mens, M.M.J., Klap, J.M., Weverling, G.J., Klatser, P., Brakenhoff, J.P.J., Roshchupkin, G.V., Ikram, M.K., Ghanbari, M., and Ikram, M.A. (2023). Genome-wide profiling of circulatory microRNAs associated with cognition and dementia. Alzheimers Dement. 19, 1194–1203. https://doi.org/10.1002/alz.12752.

33. Hu, J., Xu, J., Li, M., Jiang, Z., Mao, J., Feng, L., Miao, K., Li, H., Chen, J., Bai, Z., et al. (2024). Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study. EClinicalMedicine 68, 102409. https://doi.org/10.1016/j.eclinm.2023.102409.

34. Azodi, C.B., Tang, J., and Shiu, S.H. (2020). Opening the Black Box: Interpretable Machine Learning for Geneticists. Trends Genet. 36, 442–455. https://doi.org/10.1016/j.tig.2020.03.005.

35. Moorhouse, P., and Rockwood, K. (2008). Vascular cognitive impairment: current concepts and clinical developments. Lancet Neurol. 7, 246–255. https://doi.org/10.1016/s1474-4422(08)70040-1.

36. Kim, H.J., Ye, B.S., Yoon, C.W., Noh, Y., Kim, G.H., Cho, H., Jeon, S., Lee, J.M., Kim, J.H., Seong, J.K., et al. (2014). Cortical thickness and hippocampal shape in pure vascular mild cognitive impairment and dementia of

subcortical type. Eur. J. Neurol. 21, 744–751. https://doi.org/10.1111/ene.12376.

37. Liu, C., Li, C., Gui, L., Zhao, L., Evans, A.C., Xie, B., Zhang, J., Wei, L., Zhou, D., Wang, J., and Yin, X. (2014). The pattern of brain gray matter impairments in patients with subcortical vascular dementia. J. Neurol. Sci. 341, 110–118. https://doi.org/10.1016/j.jns.2014.04.017.

38. Ye, Q., and Bai, F. (2018). Contribution of diffusion, perfusion and functional MRI to the disconnection hypothesis in subcortical vascular cognitive impairment. Stroke Vasc. Neurol. 3, 131–139. https://doi.org/10.1136/svn-2017-000080.

39. Jacobi, H., Faber, J., Timmann, D., and Klockgether, T. (2021). Update cerebellum and cognition. J. Neurol. 268, 3921–3925. https://doi.org/10.1007/s00415-021-10486-w.

40. Acharya, A., Ren, P., Yi, L., Tian, W., and Liang, X. (2022). Structural atrophy and functional dysconnectivity patterns in the cerebellum relate to cerebral networks in svMCI. Front. Neurosci. 16, 1006231. https://doi.org/10.3389/fnins.2022.1006231.

41. De Reuck, J.L., Deramecourt, V., Auger, F., Durieux, N., Cordonnier, C., Devos, D., Defebvre, L., Moreau, C., Capparos-Lefebvre, D., Pasquier, F., et al. (2015). The significance of cortical cerebellar microbleeds and microinfarcts in neurodegenerative and cerebrovascular diseases. A post-mortem 7.0-tesla magnetic resonance study with neuropathological correlates. Cerebrovasc. Dis. 39, 138–143. https://doi.org/10.1159/000371488.

42. Poh, L., Razak, S.M.B.A., Lim, H.M., Lai, M.K.P., Chen, C.L.H., Lim, L.H.K., Arumugam, T.V., and Fann, D.Y. (2021). AIM2 inflammasome mediates apoptotic and pyroptotic death in the cerebellum following chronic hypoperfusion. Exp. Neurol. 346, 113856. https://doi.org/10.1016/j.expneurol.2021.113856.

43. Thong, J.Y.J., Du, J., Ratnarajah, N., Dong, Y., Soon, H.W., Saini, M., Tan, M.Z., Ta, A.T., Chen, C., and Qiu, A. (2014). Abnormalities of cortical thickness, subcortical shapes, and white matter integrity in subcortical vascular cognitive impairment. Hum. Brain Mapp. 35, 2320–2332. https://doi.org/10.1002/hbm.22330.

44. Yang, Q., Wei, X., Deng, B., Chang, Z., Jin, D., Huang, Y., Zhang, J.H., Yenari, M.A., Jin, K., and Wang, Q. (2022). Cerebral small vessel disease alters neurovascular unit regulation of microcirculation integrity involved in vascular cognitive impairment. Neurobiol. Dis. 170, 105750. https://doi.org/10.1016/j.nbd.2022.105750.

45. Alexander, P., Visagan, S., Jawhar, S., Kare, A., Issa, N., Issa, R., Jawhar, A., Thomas, S., and Gorantla, V. (2022). Antiplatelets and Vascular Dementia: A Systematic Review. J. Aging Res. 2022, 9780067. https://doi.org/10.1155/2022/9780067.

46. Kwan, J., Hafdi, M., Chiang, L.L.W., Myint, P.K., Wong, L.S., and Quinn, T.J. (2022). Antithrombotic therapy to prevent cognitive decline in people with small vessel disease on neuroimaging but without dementia. Cochrane Database Syst. Rev. 7, Cd012269. https://doi.org/10.1002/14651858.CD012269.pub2.

47. Tsivgoulis, G., Alexandrov, A.V., Wadley, V.G., Unverzagt, F.W., Go, R.C.P., Moy, C.S., Kissela, B., and Howard, G. (2009). Association of higher diastolic blood

pressure levels with cognitive impairment. Neurology *73*, 589–595. https://doi.org/10.1212/WNL.0b013e3181b38969.

48. Hosoki, S., Hansra, G.K., Jayasena, T., Poljak, A., Mather, K.A., Catts, V.S., Rust, R., Sagare, A., Kovacic, J.C., Brodtmann, A., et al. (2023). Molecular biomarkers for vascular cognitive impairment and dementia. Nat. Rev. Neurol. *19*, 737–753. https://doi.org/10.1038/s41582-023-00884-1.

49. Custodero, C., Ciavarella, A., Panza, F., Gnocchi, D., Lenato, G.M., Lee, J., Mazzocca, A., Sabbà, C., and Solfrizzi, V. (2022). Role of inflammatory markers in the diagnosis of vascular contributions to cognitive impairment and dementia: a systematic review and meta-analysis. GeroScience *44*, 1373–1392. https://doi.org/10.1007/s11357-022-00556-w.

50. Sachdev, P., Kalaria, R., O'Brien, J., Skoog, I., Alladi, S., Black, S.E., Blacker, D., Blazer, D.G., Chen, C., Chui, H., et al. (2014). Diagnostic criteria for vascular cognitive disorders: a VASCOG statement. Alzheimer Dis. Assoc. Disord. *28*, 206–218. https://doi.org/10.1097/wad.0000000000000034.

51. Ghafar, M.Z.A.A., Miptah, H.N., and O'Caoimh, R. (2019). Cognitive screening instruments to identify vascular cognitive impairment: A systematic review. Int. J. Geriatr. Psychiatry *34*, 1114–1127. https://doi.org/10.1002/gps.5136.

52. Wardlaw, J.M., Smith, E.E., Biessels, G.J., Cordonnier, C., Fazekas, F., Frayne, R., Lindley, R.I., O'Brien, J.T., Barkhof, F., Benavente, O.R., et al. (2013). Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. Lancet Neurol. *12*, 822–838. https://doi.org/10.1016/S1474-4422(13)70124-8.

53. Zeng, W., Chen, Y., Zhu, Z., Gao, S., Xia, J., Chen, X., Jia, J., and Zhang, Z. (2020). Severity of white matter hyperintensities: Lesion patterns, cognition, and microstructural changes. J. Cereb. Blood Flow Metab. *40*, 2454–2463. https://doi.org/10.1177/0271678x19893600.

54. Romero, J.R., Pinheiro, A., Aparicio, H.J., DeCarli, C.S., Demissie, S., and Seshadri, S. (2022). MRI Visible Perivascular Spaces and Risk of Incident Dementia: The Framingham Heart Study. Neurology *99*, e2561–e2571. https://doi.org/10.1212/wnl.0000000000201293.

55. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B Methodol. *58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

56. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2016). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1610.02391.

57. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage *15*, 273–289. https://doi.org/10.1006/nimg.2001.0978.

58. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.I. (2020). From Local Explanations to Global Understanding with Explainable AI for Trees. Nat. Mach. Intell. *2*, 56–67. https://doi.org/10.1038/s42256-019-0138-9.

59. You, J., Zhang, Y.R., Wang, H.F., Yang, M., Feng, J.F., Yu, J.T., and Cheng, W. (2022). Development of a novel dementia risk prediction model in the general population: A large, longitudinal, population-based machine-learning study. EClinicalMedicine *53*, 101665. https://doi.org/10.1016/j.eclinm.2022.101665.

60. DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics *44*, 837–845.

61. de Raadt, A., Warrens, M.J., Bosker, R.J., and Kiers, H.A.L. (2021). A Comparison of Reliability Coefficients for Ordinal Rating Scales. J. Classif. *38*, 519–543. https://doi.org/10.1007/s00357-021-09386-5.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
MRIConvert software	Version 2.0.8	https://idoimaging.com/
MIPAV software	Version11.0.7	https://mipav.cit.nih.gov
ANTs	Version 2.4.3	http://stnava.github.io/ANTs
FSL software	Version 6.0.3	http://www.fmrib.ox.ac.uk/fsl/
MedCalc software	Version 20	https://www.medcalc.org
GraphPad Prism software	Version 8.0.2	https://www.graphpad.com/
Python	Version 3.8.0	https://www.python.org/
PyTorch	Version 1.8.0	https://pytorch.org/
Matplotlib	Version 3.7.1	https://matplotlib.org/
numpy	Version 1.24.3	https://numpy.org/
XGBoost	Version 2.0.3	https://xgboost.readthedocs.io/
scikitlearn	Version 1.0.2	https://scikit-learn.org/
Other		
Origin code	This study	https://doi.org/10.5281/zenodo.13151114

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

In this study, we recruited 154 patients who met the inclusion criteria for VCI and 153 CVD patients with normal cognition from the Department of Neurology of Zhongnan Hospital of Wuhan University and Shuiguohu Street Community Health Service Center between August 2019 and April 2023. All subjects underwent clinical assessment and imaging examinations in Zhongnan Hospital of Wuhan University.

The diagnosis of VCI was established by a panel of three experienced neurologists based on the following inclusion criteria: (1) age 50-75 years; (2) education of 6 years or more; (3) neuropsychological assessments revealing a Montreal Cognitive Assessment (MoCA) score < 24 points or impairment in any one of the five major cognitive domains (including executive function, attention, memory, language, and visuo-spatial function), with or without compromised activities of daily living; (4) MRI findings meeting one of the following[50]: (i) one or more large vessel infarcts, single infarct in an extensive or key brain region (thalamic or basal ganglia); (ii) multiple lacunar infarcts outside the brainstem, 1-2 lacunes combined with extensive white matter lesions or located in key brain regions, or extensive and confluent white matter lesions (Fazekas score > 1); (iii) intracerebral hemorrhage in key brain regions, two or more intracerebral hemorrhages; (5) exclusion of other specific causes of hemorrhage, cortical and watershed infarction, hydrocephalus, and white matter lesions, such as multiple sclerosis. In addition, the inclusion criteria for CVD patients with normal cognition were as follows: (1) age 50-75 years; (2) education of 6 years or more; (3) neuropsychological assessments indicating global cognitive function and all five major cognitive domains were normal; (4) MRI results providing evidence of CVD, such as any severity of SVD and stroke.

Subjects meeting the following criteria were excluded: (1) history of stroke within 3 months; (2) other neurological disorders that cause cognitive impairment, such as AD, dementia with Lewy bodies, Parkinson's disease, post-traumatic dementia, brain tumors, or central nervous system infections; (3) epilepsy, depression, or psychiatric disorders; (4) conditions that would interfere the completion of neuropsychological tests, such as severe visual impairment, hearing impairment , aphasia, or physical dysfunction; (5) inability to cooperate in completing the MRI examination.

External validation

A dataset consisting of 74 VCI subjects and 83 CVD subjects with normal cognition recruited from General Hospital of the Yangtze River Shipping and Third People's Hospital of Hubei Province between August 2022 and April 2024 was utilized for the external validation. The inclusion and exclusion criteria were identical to those of the model construction dataset.

Ethical statement

The study was approved by the Ethics Committee (Approval number: 2020185) and has been registered with ClinicalTrials.gov (Registration number: NCT04999813). All subjects signed informed consent form in accordance with the Declaration of Helsinki.

METHOD DETAILS

Neuropsychological assessments

All subjects underwent comprehensive neuropsychological tests, administered by the same senior neuropsychologist, with the assessment results quality-controlled by two additional neuropsychologists. The Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MOCA) were employed to evaluate global cognitive function. Considering the MOCA's higher accuracy in diagnosing vascular mild cognitive impairment than the MMSE,[51] we relied on MOCA results as one of the criteria for diagnosing VCI. The Hopkins Verbal Learning Test, Verbal Fluency Test, and Clock Drawing Test were employed to measure subjects' memory, language, and visuospatial functions, respectively. Attention was assessed by the Symbol Digit Patterns Test and Trajectory Making Test-A (TMT-A), while executive functioning was appraised through the Victoria Stroop Test and TMT-B. Additionally, the Hamilton Depression Rating Scale (HAMD) was employed to evaluate the severity of depression, and subjects with an HAMD score > 17 were excluded from participation in this study. The Activities of Daily Living (ADL)/Instrumental Activities of Daily Living (IADL) scale was employed to assess the independence in daily living of all subjects.

MRI protocol

Neuroimages were acquired in this study using a 3.0T magnetic resonance imaging (MRI) scanner with a 32-channel array coil (Siemens Healthcare, Erlangen, Germany). Padding was used to limit the subject's head movement during the scanning process.

T1-weighted images were acquired with a sagittal three-dimensional magnetization prepared rapid gradient echo sequence, the parameters were: reception time (TR) = 6.7 ms, echo time (TE) = 2.26 ms, inversion time (TI) = 900 ms, flip angle = 9°, field of view (FOV) = 224 × 256 mm, voxel size = 1 × 1 × 1 mm, sagittal slices number = 176.

T2-FLAIR images were acquired using the inversion recovery MATRIX sequence with the following parameters: TR = 6,000 ms, TE = 388 ms, TI = 2200 ms, flip angle = 120°, echo sequence length = 848, bandwidth = 781 Hz/pixel, FOV = 512 × 512 mm, voxel size = 0.5 × 0.5 × 1 mm, sagittal slices number = 160.

Assessment of SVD neuroimaging markers

Neuroimaging markers of small vessel disease (SVD) included recent small subcortical infarct, whit matter hyperintensity (WMH), lacune, enlarged perivascular spaces (EPVS), cerebral microbleed, and brain atrophy.[52] WMH, lacune, and EPVS were assessed by two well-trained neuroradiologists who were blinded to this study. Evaluation of these three biomarkers was performed on T1 and T2-FLAIR imaging. WMH was defined as the white matter region with high signal on T2-FLAIR imaging, which could appear as isointense or hypointense in T1 imaging. The degree of WMH was visually assessed on T2-FLAIR imaging using the Fazekas scores, scored separately for periventricular WMH and deep WMH. The total Fazekas score was the sum of periventricular WMH and deep WMH, ranging from 0 to 6.[53] Lacune was defined as a subcortical round or oval fluid-filled cavity, with a signal similar to cerebrospinal fluid (CSF), presenting a hyperdense rim on T2-FLAIR imaging, and ranging from 3 to 15 mm in diameter. The number of lacunes was recorded. EPVS were defined as a region with signal intensity identical to that of CSF, consistent with the course of penetrating vessels, and exhibiting a linear or round/ovoid shape, typically with a diameter smaller than 3 mm. The number of EPVS was recorded and graded. EPVS were categorized into centrum semiovale (CSO) and basal ganglia (BG) based on the brain topography. EPVS was graded according to the EPVS counts: 0 = no EPVS, 1 = 1~10 EPVS, 2 = 11~20 EPVS, 3 = 21~40 EPVS, and 4 = >40 EPVS.[54] Microbleed and recent small subcortical infarct were not evaluated in this study because the subjects did not undergo scanning with DWI sequence and SWI sequence. Furthermore, as cerebral atrophy was not a specific biomarker of SVD,[29] it was likewise not evaluated in this study.

Data collection and preprocessing

All subjects underwent standardized collection of the clinical non-imaging data and sMRI data. As shown in Table S3, the clinical non-imaging data encompassed demographic, medical history, personal and family history, daily habits, physical measures, and laboratory tests. Moreover, the sMRI data consisting of T1 and T2-FLAIR images. All types of data were batch pre-processed by professional computer data analysts.

For the non-imaging data, we initially included all data from subjects in the baseline information collection, encompassing continuous, categorical, and hierarchical data types. Subsequently, we conducted an initial screening of all data, manually removing redundant variables with clinical meaninglessness (e.g., name, ID number, time of testing, etc.) with the participation of clinicians in this study. Variables with more than 20% missing values were excluded to minimize bias resulting from missing data. The remaining variables with missing values were imputed with the median (for continuous variables) or mode (for categorical variables). To make the textual data applicable to the ML algorithms, we employed Z score normalization for continuous variables and utilized one-hot encoding to create new binary labels for the categorical variables. We then utilized the least absolute shrinkage and selection operator (LASSO) algorithm with ten-fold cross-validation to restrict the number of variables.[55] Ultimately, the diagnostic model included 31 clinical non-imaging variables.

Additionally, all sMRI data underwent preprocessing using a standardized pipeline, which included image format conversion, anterior commissure (AC)-posterior commissure (PC) correction, image registration, skull stripping, image resampling, and resolution adjustment. To facilitate program script processing, we initially converted the images of the T1 and T2-FLAIR sequences from the original Dicom format to the NIFTI format using MRIConvert software (https://idoimaging.com/). AC-PC correction was then applied to all images using MIPAV software (https://mipav.cit.nih.gov), thereby eliminating positional bias present in the subject scans during the experiment.[20] The T1 images were aligned to the Colin27 template using Advanced Normalization Tools (ANTs) (http://stnava.github.io/ANTs), and the T2-FLAIR images were

aligned to the already aligned T1 images to correct for global linearity error. Furthermore, Brain Extraction Tool (BET) in the FSL software package (http://www.fmrib.ox.ac.uk/fsl/) was employed to remove the skull from the T1 images, and the generated masks were utilized to remove the skull from the T2-FLAIR images. Finally, all images were resampled to standardize voxel spacing to 1 mm x 1 mm x 1 mm, and images were resized to achieve identical spatial voxel resolution of 181 x 217 x 181. Throughout the image cropping process, only the meaningless zero-valued background regions were removed, preserving the complete structural information of the brain.

Model development

We utilized DL and conventional ML algorithms to develop a set of diagnostic models for VCI based on multimodal data. These models perform a classification task to determine whether a subject belongs to class 0 (CVD with normal cognition) or class 1 (VCI). The multimodal framework comprises three separate models: an sMRI-only model, a clinical non-imaging model, and a hybrid model. To input the sMRI features into the hybrid model in textual form, we denoted the probability that a subject was diagnosed with VCI through the sMRI-only model as the MRI Diagnostic Model (MRDM) score. Furthermore, the data from the model construction dataset, which comprises the Zhongnan Hospital and the community health service center, were randomly divided, with 80% utilizing for training and remaining 20% forming the test set. The random seeds were consistent across all models. To reduce overfitting to a specific dataset and construct a more generalized model, we partitioned the training set randomly into five equal-sized subsets and performed a five-fold cross-validation. The grid search technique was employed to search for the optimal hyperparameters. Subsequently, the test set data were input into the model with the optimal hyperparameters to evaluate the model's performance. Additionally, an external dataset was used for model testing (external validation).

sMRI-only model

We employed pre-processed whole-brain sMRI as input to develop a classification model based on the Vision Transformer (ViT). The architecture of the ViT backbone is shown in Figure S6. Specifically, the input images were initially split into identical patches, with $N=HWD/P^2$ denoting the number of produced patches, where H, W, and D represented the height, width, and depth of each input image, respectively, and $P \times P \times P$ represented the resolution of each patch. The dimensions of the patches were reduced to E dimensions by a linear projection layer, resulting in transformed linear features represented in the form of a matrix as $X \in R^{N \times E}$. The features of each patch were subsequently combined with positional encoding to obtain spatial information. Furthermore, the features with spatial information were input to the Transformer encoder to generate the deep features of networks. Transformer encoder served as the core structure of the ViT, comprising alternating Multi-head Self-Attention (MSA) layer and Multi-Layer Perceptron (MLP). Layer normalization was applied to the input features before each MSA layer and MLP. The MSA layers were employed to capture interactions between the features of all normalized patches. The MSA consists of multiple Self-Attention (SA) modules which were connected in series by channels. Each SA had three learnable weight matrices, which were the query (W^Q), key (W^K), and value (W^V). We projected the input features X onto the above weight matrices to obtain three matrices for Q, K, and V ($Q = XW^Q$, $K = XW^K$, and $V = XW^V$), the features from each SA module could be expressed as:

$$Z = \text{softmax}\left(\frac{QK^T}{\sqrt{E^q}}\right)V.$$

The feature matrices formed by each SA were summed and linearly processed to generate the features output from the MSA. These features underwent dimensionality reduction by the MLP. The operations within the encoder were then repeated, and this process was iterated L times. The final MLP conducted the classification of the corresponding subjects based on the deep features output from the Transformer encoder. It should be noted that we inputted the T1 and T2-FLAIR images as different variables to the ViT model, and finally generated two sets of MRDM scores, namely MRDM-T1 scores and MRDM-T2 scores.

Clinical non-imaging model

To delineate the contribution of clinical information other than neuroimaging in the VCI diagnosis, we developed a conventional ML classification model utilizing clinical non-imaging data. To simulate the clinical setting of a CVD patient visit, the input variables in the clinical non-imaging model include only routine tests other than neuropsychological tests, owing to their accessibility and objectivity in clinical assessment. Although the AD biomarkers such as apolipoprotein E ε4 allele, β-amyloid, and tau proteins have been proposed as potential correlates with VCI in several studies, these correlations lack validation in a broader body of research. Moreover, these tests are costly and impractical for routine use, thus, we excluded them from the model development. Five conventional ML algorithms were utilized to construct the clinical non-imaging model, comprising logistic regression (LR), SVM, random forest (RF), MLP, and eXtreme gradient boosting (XGBoost). Ultimately, the performance of all models was computed, and the model exhibiting the best performance was chosen as the foundation for the subsequent model construction.

Hybrid model

In addition to the two independent models described earlier, we integrated sMRI and clinical non-imaging features into a comprehensive hybrid model. Feature fusion was achieved by combining the MRDM scores computed from the sMRI-only model with other clinical variables. The concatenated feature vectors were then input into a conventional ML classifier. The ML classifier was selected from the algorithm with the best performance among the clinical non-imaging models. Consequently, the final hybrid model was formed by combining the ViT and a

conventional ML model. Moreover, to investigate the extent to which the features of different sMRI sequences influenced the hybrid model performance, the following combinations of features were input to the hybrid model: (1) sMRI T1 features and clinical variables; (2) sMRI T2-FLAIR features and clinical variables; and (3) sMRI T1 features, sMRI T2-FLAIR features, and clinical variables, respectively.

Model interpretability and feature reduction

To validate the interpretability of DL model, we utilized the Gradient-weighted Class Activation Mapping (Grad-CAM) method to conduct brain regions visualization analysis on the features extracted from the ViT network.[56] Specifically, the ViT network performed feature extraction and category prediction on input images. The associations between disease classification and image regions were then established by back-propagating gradients. This process generated class activation maps (CAMs) for each subject, highlighting the significant brain regions associated with VCI. To further analyze, the CAMs from all subjects were averaged to obtain an average CAM. The anatomical automatic labeling (AAL) atlas[57] was overlaid on the average CAM and extracted voxel values (ranging from 0-1) from 116 brain regions. The importance score was calculated by determining the ratio of the sum of voxel values to the number of voxels in each brain region, and then normalized to a range between 0 and 1 for consistency. Finally, two sets of importance scores for the 116 brain regions were generated based on T1 and T2-FLAIR image features, respectively.

Moreover, to analyze the interpretability of the ML model, we utilized the Shapley Additive exPlanations (SHAP) algorithm. SHAP, a game-theory based approach that can be applied to interpret any ML model, which achieved this by computing the Shapley value for each individual feature of a given training sample.[58] Utilizing SHAP, a more comprehensive understanding of the underlying factors driving the predictions of model could be obtained, thereby enhancing the interpretability of the ML model. To identify the number of features in the final model, we firstly ranked a total of 33 features, including MRDM-T1 scores, MRDM-T2 scores, and 31 clinical variables, inputted to the hybrid model in descending order based on their Shapley values. Next, a sequential forward selection strategy was employed to reduce the number of features.[59] Specifically, consecutive ML models were constructed with sequentially added features based on their contribution to the model, and the performance of each model were evaluated. In this process, the stopping point was determined when there was no significant improvement in model performance with further increase in the number of features. Afterwards, the final hybrid model was constructed based on the filtered features, and SHAP analysis was utilized to re-rank the contribution of each feature to the model.

Expert clinicians validation

In clinical practice, the diagnosis of VCI is conducted through a comprehensive assessment by practicing clinicians based on the patient's clinical data. Consequently, to further validate the performance of the model, we invited a team of experts comprising six neurologists and five radiologists to participate in the diagnostic task for a subset of cases from an external dataset. The expert clinicians were presented with 80 randomly selected cases from the external dataset, comprising 35 VCI and 45 CVD with normal cognition cases. The neurologists were provided with data for these cases, including imaging data (T1 and T2-FLAIR images) and non-imaging data (demographics, health history, physical measures, daily habits, laboratory tests, and neurological tests). The radiologists were provided with only the imaging data (T1 and T2-FLAIR images), age, sex, and education years. They were required to provide confidence scores for each case, on a scale from 0 to 100, where higher scores indicated a greater probability of the VCI diagnosis. To assess the overall diagnostic performance of the clinicians, the confidence scores from the six neurologists and five radiologists were averaged, respectively. We then calculated the diagnostic performance of the neurologists and radiologists separately using the average confidence scores.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

In the descriptive analysis of textual data, continuous variables with normal distribution were expressed as mean $\pm$ standard deviation, those with skewed distribution as median [interquartile range], and categorical variables as counts (percentage). For continuous variables, independent sample t-test or Mann-Whitney U test was employed to compare the differences of baseline characteristics between VCI group and control group. Chi-square test was employed to compare the categorical variables in the two groups regarding baseline characteristics. Spearman's rank correlation was employed to analyze the relationship between neuroimaging markers of SVD and MRDM scores predicted by the ViT model. To evaluate model performance, the accuracy, sensitivity, specificity, and F1-score (harmonic mean of recall and precision) were calculated for each model prediction. The formulas for accuracy, sensitivity, specificity and F1 score were $\frac{TP+TN}{TP+TN+FP+FN}$, $\frac{TP}{TP+FN}$, $\frac{TN}{FP+TN}$, and $\frac{2*TP}{2*TP+FP+FN}$, respectively. Additionally, the receiver operating characteristic curves (ROC) and precision-recall curves (PR) were generated, and the AUC and the area under the precision-recall curves (AP) were calculated for each ROC and PR curve. Among these metrics, AUC and AP served as the primary indicators for assessing model performance. The 95% confidence intervals (CIs) for model evaluation metrics were generated with 1000 bootstrap sets. The difference between AUCs was analyzed utilizing the non-parametric method of Delong[60] in MedCalc software, version 20 (https://www.medcalc.org). To evaluate the inter-rater agreement among clinicians' confidence scores for a specific set of cases, the clinicians' scores were conducted pairwise Pearson correlation analysis.[61] We also calculated the average pairwise correlation coefficient and its 95% confidence interval. The significance level for all statistical analyses were 0.05.

Hardware and software for model development

All sMRI and non-imaging data processing, as well as the proposed transformer-based multimodal DL framework were implemented on an Ubuntu workstation with a single NVIDIA GTX TITAN GPU. The ViT model was developed using PyTorch (version 1.8.0) in Python (version 3.8.0), the matplotlib library (version 3.7.1) and numpy (version 1.24.3) in Python were used for generating graphs of the analyzed results and vectorized numerical computation, respectively. Each ML model was developed using the scikitlearn library (version 1.0.2) and in python (version 3.8.0). Additionally, MedCalc (version 20) was performed to complete other elementary statistical analyses.