

<https://doi.org/10.1038/s44303-024-00063-x>

Stratifying vascular disease patients into homogeneous subgroups using machine learning and FLAIR MRI biomarkers

Check for updates

Karissa Chan^{1,2} ✉, Corinne Fischer³, Pejman Jabehtar Maralani⁴, Sandra E. Black^{5,6,7,8}, Alan R. Moody⁴ & April Khademi^{1,2,3,4,5,9}

This study proposes a framework to stratify vascular disease patients based on brain health and cerebrovascular disease (CVD) risk using regional FLAIR biomarkers. Intensity and texture biomarkers were extracted from FLAIR volumes of 379 atherosclerosis patients. K-Means clustering identified five homogeneous subgroups. The 15 most important biomarkers for subgroup differentiation, identified via Random Forest classification, were used to generate biomarker profiles. ANOVA tests showed age and white matter lesion volume were significantly ($p < 0.05$) different across subgroups, while Fisher's tests revealed significant ($p < 0.05$) differences in the prevalence of several vascular risk factors across subgroup. Based on biomarker and clinical profiles, Subgroup 4 was characterized with neurodegeneration unrelated to CVD, Subgroup 3 identified patients with high CVD risk requiring aggressive intervention, and Subgroups 1, 2, and 5 identified patients with varying levels of moderate risk, suitable for long-term lifestyle interventions. This study supports personalized treatment and risk stratification based on FLAIR biomarkers.

Cerebrovascular disease (CVD) is defined as neurological deficits caused by arterial insufficiency or occlusion, venous occlusive disease, or hemorrhage and can manifest in a person as an acute nonfatal event, or fatal event, with stroke being the primary disease type¹. There are approximately 795,000 acute strokes every year in the US which carries an annual healthcare cost of \$17.9 billion². As a result, CVD is a leading cause of serious long-term disability, and the second leading cause of death worldwide, posing a significant global health challenge.

There are many risk factors for CVD, some of the most common being high blood pressure, high BMI, diabetes, smoking, age, carotid artery disease (stenosis and atherosclerosis), previous CVD events and more^{3–5}. Despite improved management of these risk factors, there is still a large public health burden of CVD, creating urgency to develop new methods in characterizing disease earlier and more accurately to prevent fatal and non-fatal events attributed to CVD.

Recent research has proposed brain-based subtyping methods in neuroimaging as a step towards precision medicine in diagnostics and

therapeutics⁶. Subtyping methods use biomarkers to categorize subjects into homogeneous subgroups, which in the context of CVD, can be used to categorize subjects into risk factor groups for targeted therapy and clinical trials. Additionally, novel brain subtypes can be used for early detection, finding distinct disease phenotypes, and to learn more about mechanisms of disease. In Drysdale et al., the authors use brain subtyping methods with fMRI biomarkers to uncover novel signatures for different disease courses in brain disorders such as depression⁷. As many neurological disorders are comorbid and often present overlapping clinical symptoms, subtyping can also be used to identify distinguishing features of disease^{7,8}.

MR imaging provides insight into CVD manifestation in the brain, with existing studies showing that in addition to major events such as stroke, MRI can visualize other pathologies caused by CVD such as lacunes, infarcts and white matter lesions (WML)^{9,10}. The T2 Fluid Attenuation Inversion Recovery (FLAIR) MRI sequence is commonly used to identify CVD pathology as the suppressed cerebrospinal fluid (CSF) signal enhances contrast between healthy tissue and high signal due to ischemia. While

¹Department of Electrical, Computer and Biomedical Engineering, Toronto Metropolitan University, Toronto, ON, Canada. ²Institute for Biomedical Engineering, Science Tech (iBEST), a Partnership between St. Michael's Hospital and Toronto Metropolitan University, Toronto, ON, Canada. ³Keenan Research Center for Biomedical Science, St. Michael's Hospital, Unity Health Network, Toronto, ON, Canada. ⁴Department of Medical Imaging, University of Toronto, Toronto, ON, Canada. ⁵Institute of Medical Science, University of Toronto, Toronto, ON, Canada. ⁶Horvitz Brain Sciences Research Program, Sunnybrook Research Institute, Toronto, ON, Canada. ⁷Division of Neurology, Department of Medicine, Sunnybrook Health Sciences Centre, Toronto, ON, Canada. ⁸L.C. Campbell Cognitive Neurology Research Unit, Sunnybrook Health Sciences Centre, Toronto, ON, Canada. ⁹Vector Institute for Artificial Intelligence, Toronto, ON, Canada.

✉ e-mail: karissa.chan@torontomu.ca

WML in FLAIR MRI have been heavily studied^{11,12}, there is growing interest in other features from FLAIR that may be related to neurodegeneration and CVD. Previous research suggests FLAIR intensity is related to myelination and water content^{13,14}, while FLAIR texture was found to be related to microstructural tissue integrity and organization^{14,15}. FLAIR texture biomarkers in WM tracts and WML penumbra (boundary) regions differentiated patients with mixed disease (vascular and dementia) and subcortical vascular disease, respectively, from healthy and demented non-vascular disease patients¹⁵. As a routinely acquired imaging sequence, FLAIR holds clinical promise as an attractive and cost-effective modality for stratification of subjects with CVD. In this work, we aim to uncover clinically distinct subgroups of CVD patients through unsupervised clustering of explainable FLAIR biomarkers and to analyze clinical information to further characterize the subgroup phenotypes.

We hypothesize regional FLAIR texture and intensity biomarkers can discern changes in brain health linked to WM disease and CVD risk levels. Low, moderate and high CVD risk levels are considered based on the burden of vascular risk factors, existence of infarcts and high WML volume¹⁶, and subsequent occurrence of ischemic events. The characterization of homogeneous subgroups can aid in better understanding of CVD and may further facilitate personalized treatment decisions through stratification.

Methods

Data

This study was approved by the local institutional review board of a Canadian university (Toronto Metropolitan University). Due to the retrospective nature of the study, informed consent was waived by the local research ethics board (2021-430-3). The Canadian Atherosclerosis Imaging Network (CAIN) dataset is a multicenter pan-Canadian study containing 379 baseline FLAIR imaging volumes of patients with atherosclerotic disease pertaining to carotid artery disease¹⁷. The inclusion criteria of the study were the following: 1) male and female patients over the age of 18 years, 2) patients with mild to severe carotid artery disease (carotid stenosis $\geq 30\%$). Data on the occurrence of ischemic events including stroke and transient ischemic attacks were acquired during yearly follow-up imaging. FLAIR images were acquired using GE, Philips, and SIEMENS scanners from 8 different imaging centers with magnetic field strength of 3 T and acquisition parameters TR, TE, TI of 9000–11000 ms, 117–141 ms, and 2200–2500 ms respectively. FLAIR images had voxel sizes of 0.4286–1 mm \times 0.4286–1 mm \times 3 mm. Carotid artery imaging was completed with a volumetric high resolution T1 fat-saturated gradient Echo image for IPH identification and MRI angiography for stenosis measurements (MRIPH)¹⁷. On baseline images, established lacunar and territorial infarcts were identified by a neuroradiologist (P.M.). Cohort demographics are shown in Table 1. Additionally, the FLAIR Brainerd atlas¹⁸ and blood supply territory (BST) atlas developed in Chan et al.¹⁵ were used.

Image pre-processing

FLAIR volumes were intensity standardized¹⁹, brain extracted²⁰ and registered to atlas space¹⁵ using the Advanced Normalization Tools (ANTs) symmetric normalization^{20,21}, resulting in dimensions of 256x256x55 for all volumes. To extract WM tract regions, a generative adversarial network (GAN) model was trained as in Chan et al.²² which demonstrated high accuracies for diffusion tensor imaging (DTI) fractional anisotropy (FA) volumes particularly in the WM region.

The registered FLAIR volumes were used to generate corresponding FA volumes for each patient. The unsupervised tract segmentation method in Chan et al.¹⁵ was then used to acquire FLAIR tract masks for all patients. This involved using K-means clustering to segment tract regions from the FA volumes and morphological operations to clean the resulting binary tract masks. The Dice Similarity Coefficient (DSC) metric was used to compare synthetic tract masks to those segmented from the 107 acquired real FA volumes, along with additional evaluation metrics shown in Supplementary Table 1^{23–25}.

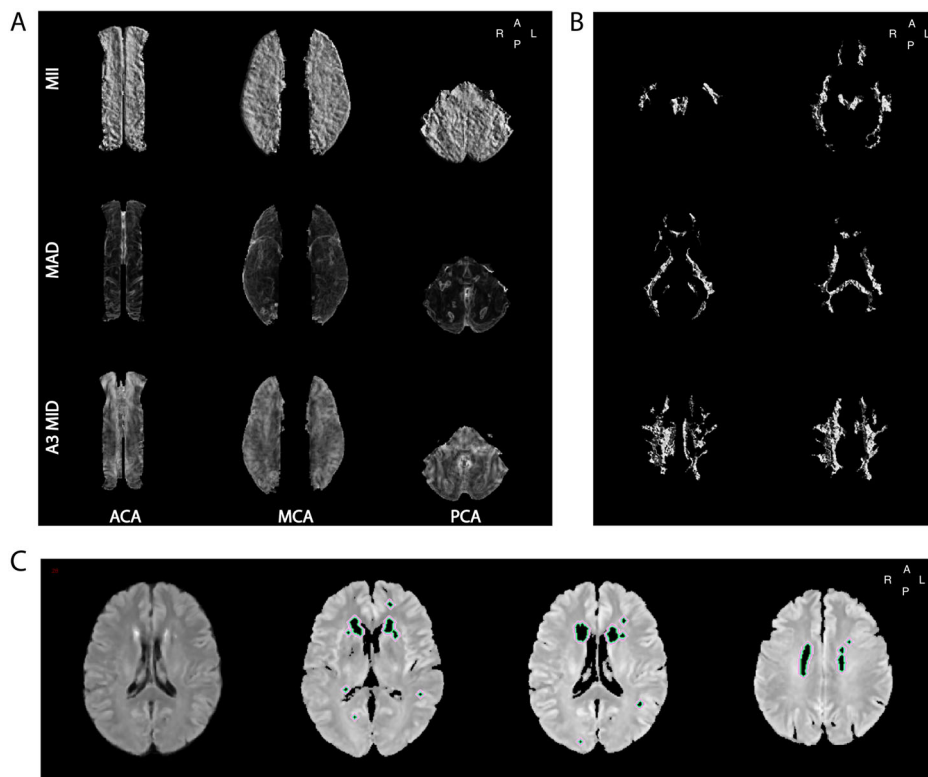
Table 1 | Summary of CAIN clinical data

| All patients, N = 379 | |
|---|----------------|
| Age, mean (SD) | 70.33 (7.97) |
| Sex, male (%) | 62.5% |
| MOCA, mean (SD) | 24.98 (5.32) |
| WML volume (mL), mean (SD) | 19.82 (19.28) |
| Left stenosis (%/100), mean (SD) | 0.331 (0.288) |
| Right stenosis (%/100), mean (SD) | 0.340 (0.264) |
| Left IPH volume (mL), mean (SD) | 0.024 (0.083) |
| Right IPH volume (mL), mean (SD) | 0.023 (0.079) |
| Systolic blood pressure (mmHg), mean (SD) | 138.37 (19.36) |
| Diastolic blood pressure (mmHg), mean (SD) | 73.45 (9.43) |
| Weight (kg), mean (SD) | 78.61 (14.78) |
| BMI (kg/m ²), mean (SD) | 27.59 (4.92) |
| Waist circumference (cm), mean (SD) | 95.81 (15.64) |
| Hypertension, N/% | 28/ 7.38% |
| Diabetes mellitus, N/% | 8/2.11% |
| Hyperlipidemia, N/% | 26/6.86% |
| Smoking, N/% | 15/3.96% |
| Coronary artery disease, N/% | 12/3.17% |
| Myocardial infarction, N/% | 5/1.31% |
| Peripheral vascular disease, N/% | 6/1.58% |
| Atrial fibrillation, N/% | 3/0.79% |
| Existing infarcts (observed at baseline imaging), N/% | 41/10.8% |
| Future ischemic event occurrence, N/% | 30/7.92% |

To extract white matter (WM) tract regions, a generative adversarial network (GAN) model was trained following the method in Chan et al.²², which demonstrated high accuracy specifically for diffusion tensor imaging (DTI) fractional anisotropy (FA) volumes, particularly in WM regions. The motivation for generating FA volumes from registered FLAIR images stems from the limitations of FLAIR images in distinguishing fine white matter tracts, as this level of precision is necessary for studies of brain connectivity and tract-specific pathology. While FLAIR imaging is effective for identifying lesions and gross WM abnormalities, it does not provide sufficient contrast for isolating individual tracts, which requires the finer detail provided by FA maps derived from DTI. By synthesizing FA volumes, we aimed to leverage the higher WM contrast from FA without acquiring DTI for all patients. The synthetic FA maps were generated by training the GAN to model the mapping between FLAIR and FA, ensuring tract delineation could be performed in a manner similar to tractography from actual DTI datasets. The GAN model was trained and evaluated using 420 FLAIR and DTI volumes (11,957 images), including volumes from the CAIN dataset alongside a separate multi-center cohort of dementia and vascular disease patients²².

Once the FA maps were generated, the unsupervised tract segmentation method from Chan et al.¹⁵ was applied. This involved using K-means clustering to identify the WM tract regions from the FA volumes, followed by morphological operations to refine the binary tract masks. The synthetic tract masks were then compared against 107 real FA volumes, acquired through DTI from the same patients, using the DSC to assess the accuracy of the segmentation. The resulting mean DSC of 0.644 demonstrates comparable performance to existing unsupervised brain tissue segmentation techniques^{23–26}, while providing a more accessible method for WM tract segmentation in datasets where

Fig. 1 | Visuals of the ROIs in this work. **A** Sample mean texture maps in ACA, MCA, and PCA BST regions. **B** Sample microstructural integrity maps of WM tract regions from samples slices of a volume, lower to higher slices from left to right and top to bottom. **C** Middle slice of original FLAIR volume (left), and WML penumbra regions P1 (green) to P5 (light pink) delineated on FLAIR NABM slices.



DTI is not readily available. Additional evaluation metrics of the WM tract segmentations are shown in Supplementary Table 1.

Regions of interest

Three major regions of interest are included in this work: the WM tracts, WML penumbra, and BST regions (Fig. 1). The WM tract region is analyzed as an entire region of all tracts combined. WM tract biomarkers were found in an existing study to be optimal in differentiating patients with Mixed (vascular and dementia) disease from other disease groups¹⁵. This is expected as the tracts are at the distal borders of BSTs, where they are susceptible to vascular insufficiencies. The WML penumbra, which is the boundary region directly surrounding WMLs, was found in previous studies to carry signs of abnormal diffusion and cerebral blood flow alterations related to vascular disease²⁷⁻²⁹. Additionally, studies have found biomarkers extracted from penumbra regions to be optimal for identifying subcortical vascular MCI patients from other disease groups¹⁵, and for observing progression of WM injury²⁷. WML penumbra sub-regions were segmented into five regions P1 (adjacent to WML) to P5 (most distal) using the methods outlined in Chan et al.¹⁵. Each penumbra region is a voxel (0.86 mm) further from the WML than the previous. Lastly, the BST atlas was employed to segment FLAIR volumes into regions supplied by the middle cerebral artery (MCA), posterior cerebral artery (PCA), and anterior cerebral artery (ACA), as the BSTs are directly related to cerebral vascularization and are likely to be affected by vascular disease. The FLAIR Brainerd atlas with the masks of the MCA, PCA, and ACA territories are shown in Supplementary Fig. 1. This resulted in a total of 9 regions of interest.

Biomarker extraction

Three FLAIR texture markers - damage, integrity and wavelet biomarkers - and FLAIR intensity were computed to provide a comprehensive assessment of WM alterations, which are critical markers for assessing CVD and neurodegeneration. These biomarkers were specifically chosen based on their strong correlations found in previous studies^{14,15} with fractional anisotropy (FA) and mean diffusivity (MD), two widely recognized diffusion

metrics used to characterize microstructural integrity and tissue changes in the brain. This resulted in four biomarkers from each of the nine regions for a total of 36 biomarkers per subject.

Damage biomarker. Measures fluctuations in intensity in a local window of an image, thus describing heterogeneity of tissue intensity. Higher damage values indicate more roughness in the tissue which is associated with lower cognitive scores and increased water diffusion described by DTI MD values^{14,15} (Eq. 1).

$$Damage = \sum_{i=1}^N \sum_{j=1}^N W_{ij} U_{ij} \quad (1)$$

where W_{ij} is the distance between pixels s_i and s_j and U_{ij} is the absolute difference of their intensities¹⁴.

Integrity biomarker. Measures the repetition of local texture patterns, in which a higher number of similar repeating structured patterns indicates more tissue integrity which is correlated with better cognition and lower MD/higher FA values^{14,15} (Eq. 2).

$$Integrity = \sum_{p=0}^{P-1} s(I_p - I_c) 2^p, \text{ where } s(x) = \{1, \text{ when } x > 0, 0 \text{ otherwise} \quad (2)$$

where P is the number of neighbours, I_p is the intensity of the neighbouring pixel, and I_c is the intensity of the central pixel¹⁴.

A3 mean (wavelet) biomarker. Computed as the mean of the approximation coefficients from the three-level decomposed FLAIR volumes using a Haar wavelet, and describes the homogeneity of tissue on a microstructural level¹⁵. Lower A3 Mean values are associated with

higher tissue integrity and structural organization as shown by correlations with MD and FA¹⁵.

Intensity biomarker. Computed as the median intensity value of masked regions.

Following the methods as in Chan et al.¹⁵ and Bahsoun et al.¹⁴, textures were computed exclusively from the normal-appearing brain matter (NABM) volume to ensure that only non-lesioned, normal tissue was analyzed. NABM volumes were created by removing WML and CSF from each patient's FLAIR image. Damage and integrity texture maps are computed on a slice-wise basis for each NABM volume, resulting in 3D texture volumes. The regions of interest were then masked from the texture volumes and voxel-wise averaged across slices to create 2D texture maps per region. The median of the 2D texture maps were taken as the final biomarker values for damage and integrity. For intensity, regions were masked from the FLAIR NABM volumes and the median intensity value was computed as the intensity biomarker. For the A3 Mean biomarker, regions were masked from the wavelet-decomposed approximation volumes and the mean coefficient value was computed as the final biomarker.

From previous studies, negative correlations found between DTI MD values and A3 Mean and integrity in all regions¹⁵ identify these markers as "integrity" markers, which are lower for low levels of neurodegeneration. Similar "integrity" marker trends were found for intensity in only the BST regions¹⁵. Conversely, damage in all regions, and intensity in the WML penumbra and tract regions, are identified as "damage" markers in previous studies where higher values indicate more neurodegeneration.

Brain subtyping

Unsupervised K-means clustering was performed to identify patient subtypes using FLAIR biomarkers. K-means clustering involves iteratively assigning data points with similar features to a cluster by minimizing the total distance between data points and the subgroup centroid. The FLAIR biomarkers were z-score standardized to have a mean of 0 and variance of 1. Principal component analysis (PCA) was then employed to reduce the feature dimensionality to two principal components describing 95% of the variance in the data. The optimal number of clusters, *k*, was determined using the elbow method. The resulting clusters/subgroups describe homogeneous brain signatures.

FLAIR biomarker signatures

To examine the important biomarkers for differentiating subgroups found using K-means clustering, a multi-class Random Forest classifier (RFC) was trained using the biomarkers to classify subjects into each cluster. Evaluation was completed using all the FLAIR biomarkers with 5-fold cross validation. Feature importance was determined based on mean decrease in impurity, a common method for determining important features in decision tree-based models³⁰. FLAIR biomarker signatures are then constructed for each subgroup using the 15 most important biomarkers contributing to the classification task. To analyze the global disease burden per subgroup, composite integrity and damage biomarkers are computed by taking the mean of all z-scored regional damage/integrity biomarkers per subgroup.

Subgroup characterization

To characterize the patient subgroups, demographic and clinical variables including age, MOCA score, WML volume, degree of left and right carotid stenosis, left and right IPH volume were used. Further, CVD risk factors were included as categorical variables, with specific thresholds for conversion to binary categories. These CVD risk factors included: hypertension, diabetes, hyperlipidemia, smoking, high BMI (>25)³, high waist circumference (>88.9 cm for females, >101.6 cm for males)³, coronary artery disease (CAD), peripheral vascular disease (PVD), myocardial infarction (MI), atrial fibrillation, high systolic blood pressure (>140 mmHg)³, high diastolic blood pressure (>90 mmHg)³, and the presence of prior infarcts. Each risk factor was converted into a binary categorical variable, where the presence of the condition or risk (e.g., BMI >25, systolic blood

pressure >140 mmHg) was assigned a value of 1, and the absence was assigned a value of 0. Similarly, the occurrence of ischemic events (including TIAs and strokes) after baseline imaging was also treated as a binary variable (event = 1, no event = 0).

To create a compound CVD risk score, these CVD risk factors are aggregated into a single measure based on the prevalence of the risks for each subgroup. The compound CVD risk score per subgroup was estimated using the prevalence of each risk factor in each subgroup, subtracted by the population prevalence of each risk factor, and summed across risk factors. All factors were given equal weighting. As a result, the compound risk score quantifies the level of CVD risk in each subgroup relative to the entire population where a risk score of 0 is the population risk. "Low-" and "high-risk" thresholds were quantified as the mean plus/minus one standard deviation of the risk scores. The final characterization of subgroups considers: 1) neurodegenerative brain biomarker signatures, 2) significant clinical variables, 3) compound CVD risk, and 4) future ischemic events.

Statistical analysis

Pearson's correlation tests were performed to investigate the associations between clinical variables and FLAIR biomarkers. ANOVA or Kruskal-Wallis and their respective post-hocs with multiple comparison corrections (Tukey's HSD and Bonferroni, respectively) were performed to compare clinical variables and FLAIR biomarkers between the patient subgroups. Fisher's Exact Tests were also performed to investigate associations between Subgroup and the prevalence of each vascular risk factor as well as future ischemic events. Fisher's Exact Test is a statistical test used for the analysis of categorical variables and small sample sizes, and calculates the exact probability of obtaining the observed samples in the data. Significance in the Fisher's test indicates there are differences in risk factor prevalence across the patient subgroups but does not identify specific pairwise differences between subgroups. Statistical significance was defined as $p < 0.05$ for all tests.

Results

In total, 36 total regional texture and intensity biomarkers were extracted from each of the 379 patients. The biomarkers were used to cluster the patients into homogeneous subgroups using unsupervised K-Means clustering, and statistical tests were employed to investigate the clinical characteristics of each subgroup.

Brain subtyping and FLAIR biomarker signatures

Using the regional FLAIR biomarkers extracted from the dataset, the optimal number of clusters was determined to be $K = 5$ using the elbow method. The resulting cluster sizes ranged from 27–108 patients. The FLAIR biomarkers were then used to train a multi-class RFC to determine the 15 features to include in the biomarker profiles. The resulting classifier had mean classification accuracy, recall, precision, and F1 score of 0.842, 0.853, 0.842, and 0.84 respectively - indicating the biomarkers are robustly clustering the subgroups. The 15 most important biomarkers identified from the classification and results of ANOVA and Kruskal-Wallis tests are shown in Table 2, with significant differences ($p < 0.001$) across the five subgroups for every FLAIR biomarker. Post-hoc tests showed majority of the biomarkers were significantly different across all subgroups (Supplementary Fig. 2).

Subgroup biomarker profiles are shown in Fig. 2A and composite integrity and damage biomarkers are shown in Fig. 2B. Composite biomarkers are computed by taking the mean of all z-scored regional damage/integrity biomarkers by subgroup. Subgroup 4 exhibits the most neurodegeneration, represented by significantly higher intensity and damage biomarkers ($p < 0.001$) and significantly lower integrity biomarkers ($p < 0.001$) than all other subgroups, with the highest composite damage and lowest composite integrity. Subgroup 1 has the lowest composite damage with a slight positive composite integrity which indicates this group has better brain health. This is also confirmed on a regional level, where Subgroup 1 had significantly lower WML penumbra intensity ($p < 0.001$), lower BST

Table 2 | Mean (SD) values of FLAIR biomarkers by subgroup, along with *p*-value results of ANOVA/Kruskal-Wallis tests for significant differences between subgroups

| Subgroup | 1 | 2 | 3 | 4 | 5 | <i>p</i> -value |
|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|
| N | 92 | 53 | 108 | 27 | 99 | - |
| MCA A3 Mean | 173.71 (6.38) | 158.01 (7.26) | 163.55 (7.49) | 140.28 (13.73) | 178.31 (6.98) | <0.001 |
| ACA A3 Mean | 174.26 (7.10) | 160.14 (7.90) | 164.98 (6.97) | 141.20 (13.17) | 177.47 (7.40) | <0.001 |
| P2 A3 Mean | 151.02 (10.56) | 139.43 (8.88) | 151.45 (10.06) | 136.76 (15.23) | 165.44 (8.90) | <0.001 |
| P3 A3 Mean | 155.35 (10.06) | 144.77 (8.32) | 156.68 (9.16) | 141.67 (15.25) | 170.10 (8.28) | <0.001 |
| P4 A3 Mean | 159.21 (9.66) | 149.31 (8.05) | 161.09 (8.46) | 145.57 (15.38) | 174.20 (7.77) | <0.001 |
| P5 A3 Mean | 162.91 (9.37) | 153.42 (7.89) | 165.03 (7.90) | 148.81 (15.53) | 178.00 (7.35) | <0.001 |
| MCA Integrity | 122.45 (1.09) | 119.72 (1.67) | 120.49 (1.65) | 117.43 (2.95) | 122.72 (1.19) | <0.001 |
| ACA Intensity | 271.98 (3.32) | 272.85 (2.69) | 276.06 (3.14) | 274.80 (5.71) | 275.62 (3.13) | <0.001 |
| PCA MAD | 2551.39 (382.70) | 3355.38 (491.66) | 3096.83 (447.93) | 4105.12 (822.73) | 2413.48 (335.23) | <0.001 |
| MCA MAD | 2795.46 (358.46) | 3631.78 (475.92) | 3518.92 (512.87) | 4507.16 (857.23) | 2675.46 (382.08) | <0.001 |
| Tract Intensity | 271.16 (5.66) | 276.42 (5.13) | 280.89 (4.69) | 286.58 (7.40) | 276.38 (5.17) | <0.001 |
| P2 Intensity | 284.83 (5.64) | 291.11 (4.89) | 297.17 (5.51) | 307.11 (10.25) | 291.14 (5.15) | <0.001 |
| P3 Intensity | 278.57 (5.00) | 283.50 (4.47) | 288.99 (4.38) | 296.27 (9.02) | 284.46 (4.09) | <0.001 |
| P4 Intensity | 275.02 (4.91) | 279.28 (4.53) | 284.34 (3.84) | 289.56 (8.24) | 280.74 (3.66) | <0.001 |
| P5 Intensity | 272.86 (4.89) | 276.69 (4.54) | 281.30 (3.57) | 285.12 (7.84) | 278.46 (3.51) | <0.001 |

Bold values identify statistical significance (*p* < 0.05).

damage (*p* < 0.001) and higher BST integrity (*p* < 0.001) than all subgroups except Subgroup 5. Subgroup 5 exhibits low composite damage and highest composite integrity. MCA and PCA damage are not significantly different from Subgroup 1 while the integrity biomarkers are significantly higher (*p* < 0.001) than all other groups. As shown by the composite damage marker, which is the mean of all z-scored regional damage biomarkers, Subgroup 2 exhibits slight positive composite damage which reflects the increased regional damage in the MCA and PCA regions only. Subgroup 3 has different trends, with second highest composite damage due to the significantly higher damage biomarkers across all regions in the regional biomarker profile. Subgroup 2 has the second lowest composite integrity due to the decreased integrity biomarkers across all regions, whereas Subgroup 3 only has a slight negative composite integrity due to decreased integrity biomarkers in the MCA region.

Clinical profiles

Summary statistics of clinical variables are shown in Table 3, and the z-scored subgroup profiles are visualized in Fig. 3A. Statistically significant differences (*p* < 0.05) were found for age, WML volume, and left IPH volume, while MOCA score did not reach statistical significance (*p* = 0.051), as shown in Table 3.

Post-hoc tests (Supplementary Fig. 3) revealed significant differences in age (*p* < 0.05) between most subgroups, except for Subgroups 2 and 3, and Subgroups 1 and 5, which were not significantly different. For WML volume, significant differences were observed between all subgroups except for Subgroups 2 and 3, and importantly, Subgroups 3 and 4, which exhibited similar WML volumes despite their other clinical differences. Regarding left IPH volume, Subgroup 4 was significantly lower than all other subgroups (*p* < 0.05).

Although MOCA score approached significance (*p* = 0.051), the differences between Subgroups 3 and 4 were not statistically significant. This is an important distinction, as while Subgroups 3 and 4 have similar WML volumes and MOCA scores, they exhibit markedly different profiles in terms of age and left IPH volume, with both showing statistically significant differences between the two subgroups (*p* < 0.05).

Pearson’s correlations between clinical variables and the 15 important FLAIR biomarkers are shown in Fig. 4. Age and WML volume are significantly correlated with most biomarkers and regions, while MOCA shows a significant positive correlation with the wavelet biomarkers in the

BST and WML penumbra regions. Right stenosis and right IPH volume are significantly correlated only with WM tract intensity, while left stenosis is significantly associated with both WM tract intensity and MAD in the PCA and MCA.

CVD risk profiles

Results of the Fisher’s Exact test for each risk factor are shown in Supplementary Table 2, with hypertension (*p* < 0.01), diabetes (*p* < 0.05), hyperlipidemia (*p* < 0.05), smoking (*p* < 0.05), high BMI (*p* < 0.01), high sBP (*p* < 0.05), and baseline infarcts (*p* < 0.001) showing significant associations with Subgroup. For patients with ischemic events occurring in the future, Fisher’s Exact Tests demonstrated a significant association between Subgroup and future event occurrence (*p* < 0.05). The percentage of risk factors and events by subgroup is shown in Supplementary Fig. 4. Subgroup 3 had the largest prevalence of hypertension, hyperlipidemia, smoking, high BMI, sBP, and dBP. Comparatively, Subgroup 4 has the lowest prevalence of CVD risk factors and future events.

The compound CVD risk scores computed using the vascular risk factors and events are shown in Fig. 3B. Subgroup 4 exhibits the lowest risk score of -1.05, followed by Subgroup 2, 1, and 5 with risk scores of 0.29, 0.79, and 1.14 respectively, and finally Subgroup 3 with the highest risk score of 2.28. Using the thresholds, Subgroup 4 is lower than the mean and would be considered low CVD risk, Subgroups 1, 2, 5 are medium CVD risk, and Subgroup 3 is considered high risk CVD. Subgroup 3 also had the highest number of strokes in the future, followed by Subgroup 1 and Subgroup 5. The lowest number of future ischemic events is in Subgroups 2 and 4.

Subgroup characterization

The subgroup characteristics are defined in Table 4 using the compound integrity and damage markers, MoCA, clinical variables that were significant in the statistical analyses, compound CVD risk scores, and future event prevalence. Subgroup differences in values are indicated by a ↑ if the value was significantly higher than other subgroups and population mean, ↓ for a value significantly lower, and -- for values that were not significantly different for a subgroup compared to the other groups and were similar to the population mean. While there were no statistically significant differences in MoCA scores across

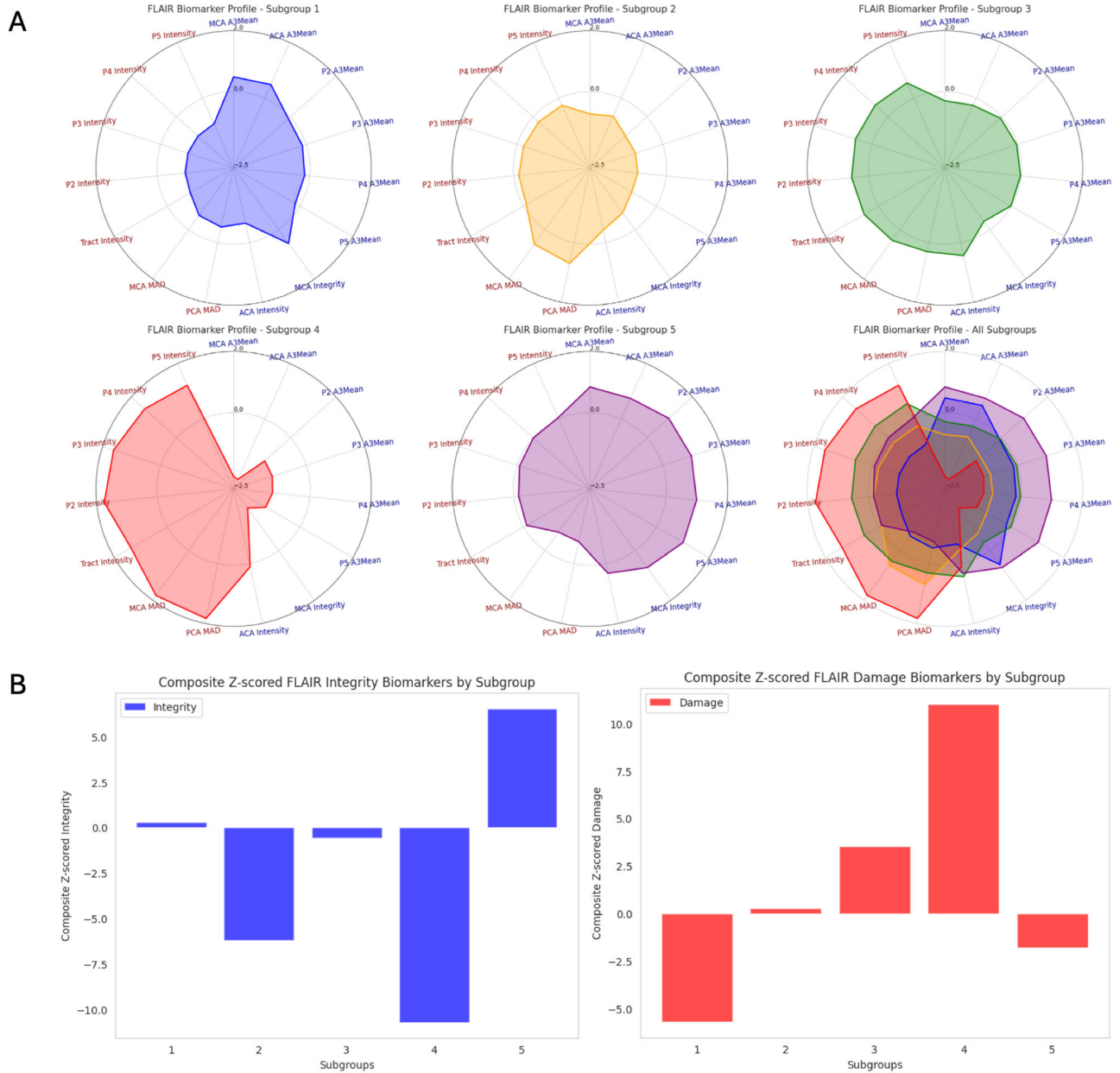


Fig. 2 | Regional and composite FLAIR biomarker profiles for clustered subgroups. **A** Z-scored FLAIR biomarker profiles for each subgroup of top 15 important biomarkers. Integrity biomarkers are shown in blue and Damage biomarkers are shown in red. **B** Composite Z-scored damage (right) and integrity (left) biomarkers by subgroup.

Table 3 | Summary statistics of patient subgroups with p-values of ANOVA t-tests. Bolded values indicate significant differences.

| Cluster | 1 | 2 | 3 | 4 | 5 | p-value |
|------------------|--------------|--------------|--------------|--------------|--------------|------------------|
| N | 92 | 53 | 108 | 27 | 99 | - |
| Gender (M/F, %F) | 59/33, 35.9% | 35/18, 34% | 77/31, 28.7% | 18/9, 33.3% | 48/51, 51.5% | - |
| Age | 68.04(7.46) | 71.87(6.54) | 71.94(6.6) | 75.59(9.43) | 68.45(8.91) | <0.001 |
| MOCA | 24.7(5.71) | 25.7(2.97) | 24.73(5.36) | 22.15(8.11) | 25.91(4.69) | 0.051 |
| WML volume | 11.19(9.47) | 24.43(24.01) | 24.63(19.1) | 41.28(32.46) | 14.27(10.12) | <0.001 |
| Right stenosis | 0.3(0.3) | 0.38(0.31) | 0.33(0.24) | 0.32(0.25) | 0.37(0.23) | 0.428 |
| Left stenosis | 0.27(0.32) | 0.37(0.28) | 0.36(0.27) | 0.39(0.3) | 0.33(0.27) | 0.407 |
| Right IPH volume | 0.018(0.068) | 0.012(0.03) | 0.022(0.061) | 0.024(0.056) | 0.035(0.12) | 0.836 |
| Left IPH volume | 0.028(0.092) | 0.02(0.058) | 0.036(0.11) | 8e-6 (4e-5) | 0.017(0.06) | <0.05 |

Bold values identify statistical significance ($p < 0.05$).

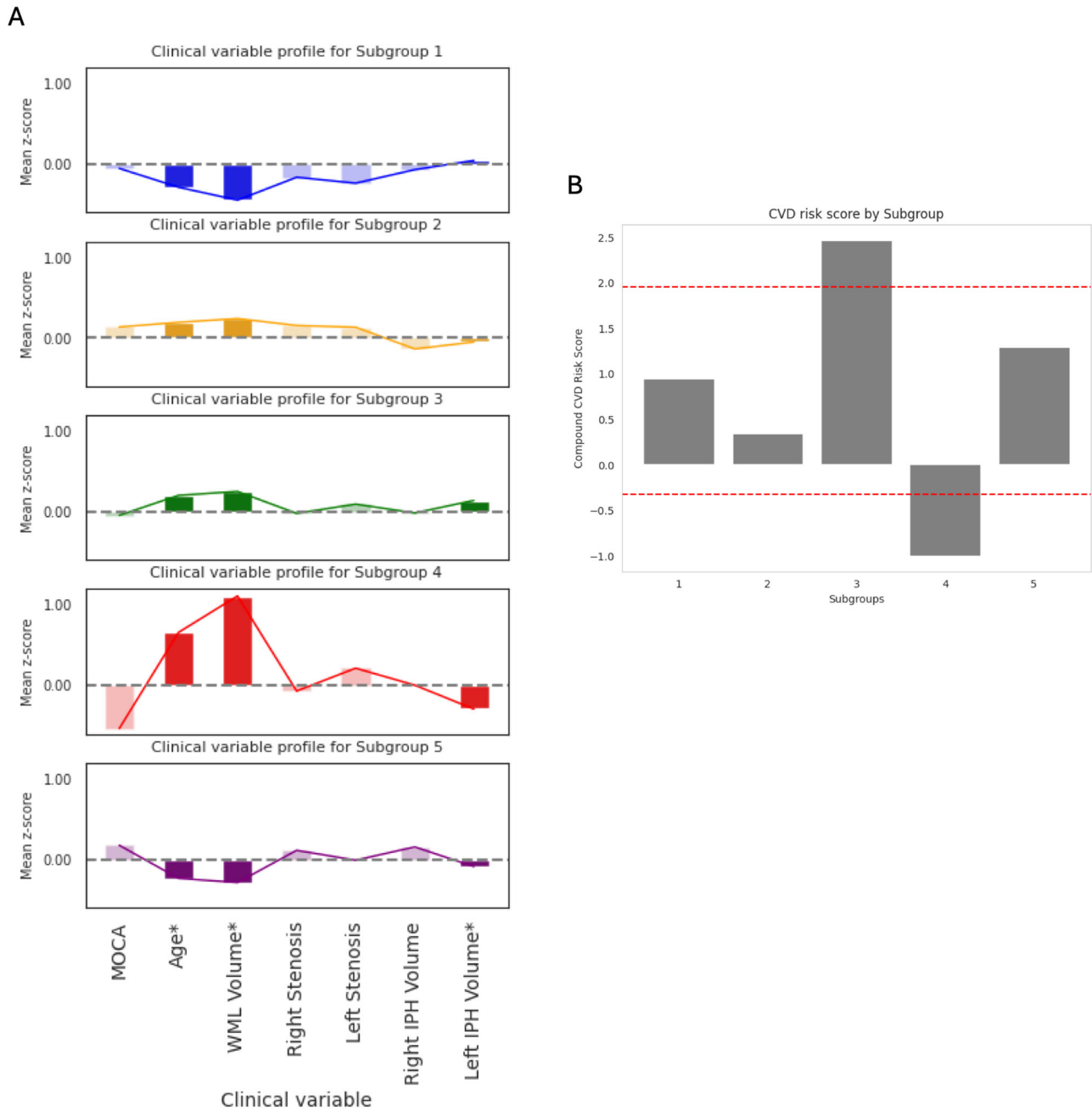


Fig. 3 | Characterization of subgroups using z-scored demographic/clinical variables and CVD risk scores. A Z-scored profiles of each subgroup.* indicate variables with significant differences ($p < 0.05$) between groups. **B** Compound CVD risk score shown by subgroup. High/low score thresholds are shown as dashed lines.

subgroups, we observed numerically large variations in these scores, which we present for context and to acknowledge potential trends.

Subgroup 3 (second highest damage and lower integrity compared to the mean) shows the highest CVD risk score, with a higher number of future strokes, as well as significantly higher patient age and WML volume compared to the mean. While this group has a high CVD risk score and significant damage accumulation in the brain, the future event prevalence is similar to that of Subgroups 1 and 5, based on the risk percentage.

Subgroup 4 (highest damage and lowest integrity) is the oldest patient group, with the highest WML volume and significantly lower left IPH volume. Although this group has a numerically lower MoCA score compared to other subgroups, the difference is not statistically significant. Interestingly, Subgroup 4 has the lowest compound CVD risk score among all subgroups, despite its higher WML burden and lower biomarkers of

integrity. This group also shows a relatively low number of future strokes, though its association with neurodegeneration remains unclear based on the available data.

Subgroups 1, 2, and 5 represent moderate CVD risk groups. Subgroup 2 (mean damage and second lowest integrity biomarkers) has significantly higher age and WML volume than Subgroups 1 and 5, yet its risk score is lower than both. Subgroup 2 also exhibits the lowest number of future strokes. Subgroup 1 (lowest damage and mean integrity biomarkers) is the youngest group with the lowest WML volume and a moderate CVD risk score, slightly lower than Subgroup 5. Despite having a large number of future ischemic events, its CVD risk is comparable to other subgroups. Subgroup 5 (second lowest damage and highest integrity biomarkers) has the second lowest WML volume (significantly higher than Subgroup 1) and is among the youngest groups. It has the highest CVD risk score of the moderate-risk groups and shows a large number of future strokes.

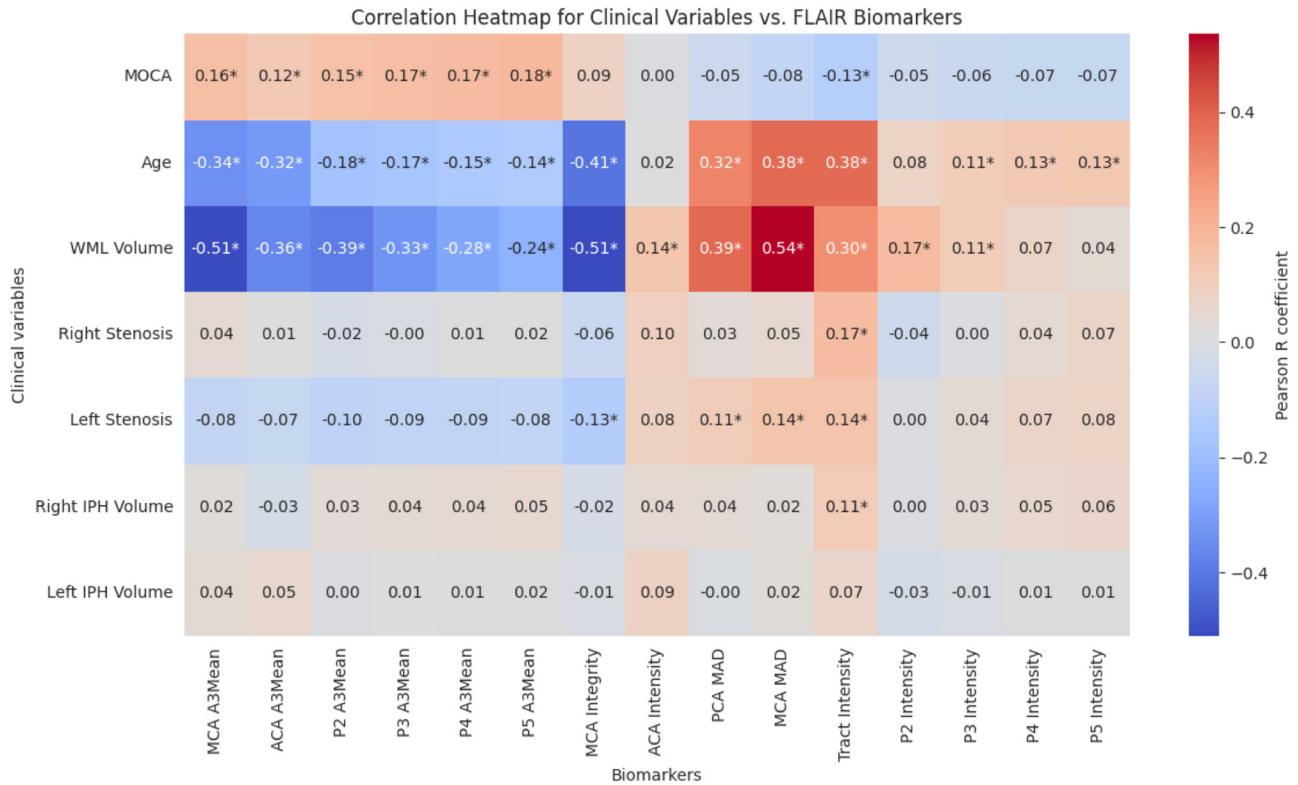


Fig. 4 | Heatmap of R correlation coefficients between clinical variable and FLAIR biomarker. * indicate significant correlations (p < 0.05).

Table 4 | Subgroup characterization.

| Subgroup | Composite Biomarkers | | Significant clinical variables | | | | MoCA | CVD Risk | Future stroke |
|----------|----------------------|-----------|--------------------------------|-----|-------|----|------|----------|---------------|
| | Damage | Integrity | Age | WML | L IPH | | | | |
| 1 | ↓ | -- | ↓ | ↓ | -- | -- | MED | 23.3% | |
| 2 | -- | ↓ | ↑ | ↑ | -- | ↑ | MED | 13.3% | |
| 3 | ↑ | -- | ↑ | ↑ | -- | -- | HIGH | 26.7% | |
| 4 | ↑↑ | ↓↓ | ↑↑ | ↑↑ | ↓ | ↓↓ | LOW | 13.3% | |
| 5 | ↓ | ↑ | ↓ | ↓ | -- | ↑ | MED | 23.3% | |

Discussion

This work uses FLAIR texture and intensity biomarkers to detect WM changes and offers novel stratification of subjects with different levels of CVD risk. Biomarker and clinical differences suggest different underlying processes across subgroups. Automated stratification offers a novel, non-invasive approach for early detection and stratification which can be used to tailor prevention and treatment strategies to avert future ischemic events. Existing studies utilize clinical factors and clustering methods to stratify stroke patients into subgroups with and without subsequent events, for the prediction of vascular outcomes^{31–33}. Sperber et al.³⁴ used clustering methods to stratify stroke patients with cerebral small vessel disease based on lesion type. However, these studies did not include non-stroke patients or imaging biomarkers. Other regression studies often focus on associations between imaging biomarkers, vascular risk factors, and events, to observe relationships between CVD risk and changes in tissue microstructure or event recurrence^{3,35}. While promising, these frameworks do not provide comprehensive methodologies for risk stratification. As such, the novelty of our work is three-fold. First, we demonstrate that explainable FLAIR biomarkers can distinguish patients with varying CVD risk and clinical profiles, from an atherosclerotic cohort. These findings are valuable due to the incorporation

of imaging biomarkers which could be used to monitor patients over time or early disease detection. Secondly, the regional FLAIR biomarkers are correlated with various clinical variables, demonstrating they are quantifying structural changes in the brain related to disease factors. Lastly, this work considers prior, existing, and future CVD-related factors for each patient in the cohort. This provides a more comprehensive profile of CVD burden and risk, addressing the gaps left by previous methodologies.

Among the subgroups, there was one group with low CVD risk (Subgroup 4), one group with high CVD risk (Subgroup 3) and three subgroups with moderate CVD risk (Subgroups 1, 2, 5). A key finding is Subgroup 4, with advanced neurodegeneration, cognitive impairment, highest age and largest WML loads, accompanied with minimal CVD risk factors. This may suggest neurodegeneration is being driven by processes related to accelerated aging or AD, rather than primarily CVD. While WML burden is often associated with CVD, several studies^{9,36} postulated that both vascular and AD processes contribute to WML development. This could perhaps explain the high WML load in Subgroup 4, given its lower CVD risk, although it is not significantly different from Subgroup 3. Uncovering this homogeneous group within a vascular disease cohort underscores the utility of the framework in potentially differentiating underlying disease mechanisms

which could help to choose optimal candidates for therapy or to learn more about disease³⁷. To analyze the relationship to AD further would require other variables such as blood biomarkers, PET imaging and spatial WML patterns³⁸.

Subgroup 3 exhibits accumulated brain damage and this subgroup is likely at an advanced stage of vascular disease and CVD⁵. Contributing factors to the severity of Subgroup 3 may include a markedly high prevalence of MI, though it did not reach statistical significance due to sample size. Witt et al.³⁹ observed a three-fold increase in stroke risk during the first 3 years after incident MI, suggesting the FLAIR biomarker profiles, particularly high WML penumbra and tract intensity may be capturing structural differences in patients with MI. These patients could be automatically identified for aggressive management of lifestyle and cardiovascular risk factors to avoid future negative outcomes such as stroke and death.

The proposed framework offers a nuanced understanding of brain health patterns in patients with moderate CVD risk, despite the fact that stratifying groups with less severe disease levels is more challenging¹⁰. The moderate CVD group with the lowest risk (Subgroup 2), characterized by low integrity and minimal damage, presents clinical features that do not clearly support accelerated aging. Rather, this subgroup's low CVD burden and previous infarcts suggest that its patients may benefit from monitoring for signs of further CVD risk. Subgroup 2 also had the largest proportion of prior infarcts and patients may have received treatment after the stroke, which could have aided in reducing the CVD burden in the brain. Zhang et al.³⁵ found that drug adherence after stroke was a significant factor in predicting stroke recurrence within a 3-year period. However, this would have to be confirmed using a dataset with treatment information in the future.

Subgroups 1 and 5 have less neurodegeneration and share similar biomarker characteristics and risk factors (i.e., large number of future events, lower age, low WML volume). However, they are also unique with different levels of damage and integrity between groups, and Subgroup 5 had the highest MoCA scores in the cohort, suggesting cognitive resilience despite moderate CVD risk. Though these differences in MoCA scores were not statistically significant, we note them here as potential areas for further study. Subjects from these groups had a high number of future stroke events, making it important to identify them, and apply any appropriate therapy early. Subjects are relatively young in these cohorts and could benefit from short- and long-term health benefits afforded by making lifestyle changes (i.e., physical activity, nutrition, weight management, avoidance of tobacco and management of cardiovascular health-related factors such as cholesterol, blood pressure, and glucose)⁴⁰. The automated biomarker system can be used to automatically identify these subjects for lifestyle intervention, to preserve brain health through minimizing modifiable risk factors⁴¹.

Interestingly, the moderate CVD risk groups Subgroups 1 and 5 display notable prevalence of patients with AF and PVD respectively, despite not reaching statistical significance due to small sample sizes of patients with these risk factors. AF is associated with the presence of infarcts¹⁰ corresponding to the larger proportion of baseline infarcts in Subgroup 1, while PVD is associated with increased WML volume⁴², which supports the higher WML volume in Subgroup 5. These differences, noted without statistical significance, suggest AF and PVD as potential contributors to the biomarker and clinical differences between these subgroups that warrant further investigation. Most notably, Subgroup 5 is the only group with more women (51.5%) than men, while the proportion of females in all other subgroups was between 28–35%. Bonberg et al.⁴ found a larger impact of hypertension and smoking on WM damage in women, aligning with the high prevalence of these risk factors observed in Subgroup 5. The brain biomarkers for the two subgroups indicate minimal accumulated brain damage with moderate CVD risk.

A few limitations exist in our study. While the study cohort inclusion criteria was $\geq 30\%$ stenosis, the mean stenosis of the entire population was

$\sim 33\%$ indicating the majority of the cohort had mild stenosis. Future work should include a larger sample size of subjects with severe carotid artery disease, which may allow effects of carotid atherosclerosis to be better distinguished between the subgroups. To further investigate sex differences in the clusters, the study population should also include more female subjects in the future. However, as CVD is more common in men (though the difference decreases with age), the cohort used in this work is a natural sampling. Further, the analysis did not consider the effects of treatments or medications, which should be considered as factors within each subgroup for any future analyses. Lastly, the analysis is done cross-sectionally, making it difficult to draw conclusions about the dynamic processes of CVD and AD-related pathology. Longitudinal experiments as well as validation and refinement of subtypes using integrated diffusion or ASL biomarkers could be performed in future work to further explore pathological mechanisms. Additionally, single-subject studies analyzing new subjects with respect to the subgroups should be investigated in the future to optimize translation.

This study proposes a framework for utilizing regional FLAIR texture and intensity biomarkers to stratify patients with atherosclerosis into homogeneous disease subgroups. By leveraging FLAIR MRI, clinicians can effectively differentiate between various pathological mechanisms, disease stages, and risk factors associated with CVD and its interactions with neurodegeneration. The identified subgroups, namely Subgroup 4 characterized by non-vascular related pathology, Subgroups 1, 2 and 5 representing moderate risk CVD cohorts, and Subgroup 3 with high CVD risk, provide valuable insights into disease manifestations. These subgroups not only shed light on the differential effects of CVD on brain health but also highlight the potential for personalized treatment decisions and risk stratification.

Data availability

The data that support the findings of this study are available from the authors but restrictions apply to the availability of these data, which were used under license from the Canadian Atherosclerosis Imaging Network (CAIN) for the current study, and so are not publicly available. Data may, however, be available from the authors upon reasonable request and with permission from CAIN.

Received: 3 August 2024; Accepted: 3 December 2024;

Published online: 31 December 2024

References

- Li, X. et al. Advances in differential diagnosis of cerebrovascular diseases in magnetic resonance imaging: a narrative review. *Quant. Imaging Med. Surg.* **13**, 2712734–2712734, <https://doi.org/10.21037/qims-22-750> (2023).
- The Burden of Cerebrovascular Disease in the United States. Accessed May 11, 2024. https://www.cdc.gov/pcd/issues/2019/18_0411.htm
- Trofimova, O. et al. Topography of associations between cardiovascular risk factors and myelin loss in the ageing human brain. *Commun. Biol.* **6**, 1–14, <https://doi.org/10.1038/s42003-023-04741-1> (2023).
- Bonberg, N., Wulms, N., Dehghan-Nayyeri, M., Berger, K. & Minnerup, H. Sex-Specific Causes and Consequences of White Matter Damage in a Middle-Aged Cohort. *Front. Aging Neurosci.* **14**, 810296 (2022). Accessed January 29, 2024. <https://www.frontiersin.org/articles/10.3389/fnagi.2022.810296>.
- Farnsworth von Cederwald, B., Josefsson, M., Wählin, A., Nyberg, L. & Karalija, N. Association of Cardiovascular Risk Trajectory With Cognitive Decline and Incident Dementia. *Neurology* **98**, e2013–e2022, <https://doi.org/10.1212/WNL.000000000200255> (2022).
- Woo, C. W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nat. Neurosci.* **20**, 365–377, <https://doi.org/10.1038/nn.4478> (2017).

7. Drysdale, A. T. et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **23**, 28–38, <https://doi.org/10.1038/nm.4246> (2017).
8. Tang, C. C. et al. Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis. *Lancet Neurol.* **9**, 149–158, [https://doi.org/10.1016/S1474-4422\(10\)70002-8](https://doi.org/10.1016/S1474-4422(10)70002-8) (2010).
9. Keith, J. et al. Collagenosis of the Deep Medullary Veins: An Underrecognized Pathologic Correlate of White Matter Hyperintensities and Periventricular Infarction. *J. Neuropathol. Exp. Neurol.* **76**, 299–312, <https://doi.org/10.1093/jnen/nlx009> (2017).
10. Rydén, L. et al. Atrial Fibrillation, Stroke, and Silent Cerebrovascular Disease. *Neurology* **97**, e1608–e1619, <https://doi.org/10.1212/WNL.0000000000012675> (2021).
11. Saba, L. et al. Association Between the Volume of Carotid Artery Plaque and Its Subcomponents and the Volume of White Matter Lesions in Patients Selected for Endarterectomy. *Am. J. Roentgenol.* **201**, W747–W752, <https://doi.org/10.2214/AJR.12.10217> (2013).
12. de Groot, M. et al. Changes in Normal-Appearing White Matter Precede Development of White Matter Lesions. *Stroke* **44**, 1037–1042, <https://doi.org/10.1161/STROKEAHA.112.680223> (2013).
13. Chan, K. et al. Brain Maturation Patterns on Normalized FLAIR MR Imaging in Children and Adolescents. *Am. J. Neuroradiol.* <https://doi.org/10.3174/ajnr.A7966> (2023).
14. Bahsoun, M. A. et al. FLAIR MRI biomarkers of the normal appearing brain matter are related to cognition. *NeuroImage. Clin.* **34**, 102955, <https://doi.org/10.1016/j.nicl.2022.102955> (2022).
15. Chan, K. et al. Alzheimer's and vascular disease classification using regional texture biomarkers in FLAIR MRI. *NeuroImage Clin.* **38**, 103385, <https://doi.org/10.1016/j.nicl.2023.103385> (2023).
16. Yassi, N. et al. Influence of Comorbidity of Cerebrovascular Disease and Amyloid- β on Alzheimer's Disease. *Handb. Prev. Alzheimer's Dis.* 381–394 <https://doi.org/10.3233/AIAD230036> (2024).
17. Tardif, J. C. et al. Atherosclerosis imaging and the Canadian Atherosclerosis Imaging Network. *Can. J. Cardiol.* **29**, 297–303, <https://doi.org/10.1016/j.cjca.2012.09.017> (2013).
18. Winkler, A., Kochunov, P. & Glahn, D. FLAIR templates. *Brainder*. Accessed January 28, 2024. <https://brainder.org/>
19. Reiche, B., Moody, A. R. & Khademi, A. Pathology-preserving intensity standardization framework for multi-institutional FLAIR MRI datasets. *Magn. Reson. Imaging* **62**, 59–69, <https://doi.org/10.1016/j.mri.2019.05.001> (2019).
20. Khademi, A., Reiche, B., DiGregorio, J., Arezza, G. & Moody, A. R. Whole volume brain extraction for multi-centre, multi-disease FLAIR MRI datasets. *Magn. Reson. Imaging* **66**, 116–130, <https://doi.org/10.1016/j.mri.2019.08.022> (2020).
21. Avants, B., Epstein, C., Grossman, M. & Gee, J. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41, <https://doi.org/10.1016/j.media.2007.06.004> (2008).
22. Chan, K., Maralani, P. J., Moody, A. R. & Khademi, A. Synthesis of diffusion-weighted MRI scalar maps from FLAIR volumes using generative adversarial networks. *Front. Neuroinform.* **17**, (2023). <https://www.frontiersin.org/articles/10.3389/fninf.2023.1197330>.
23. Ortiz, A., Górriz, J. M., Ramírez, J. & Salas-González, D. Two fully-supervised methods for MR brain image segmentation using SOM-based strategies. *Appl. Soft Comput.* **13**, 2668–2682, <https://doi.org/10.1016/j.asoc.2012.11.020> (2013).
24. Ortiz, A., Górriz, J. M., Ramírez, J. & Salas-González, D. MRI Brain Image Segmentation with Supervised SOM and Probability-Based Clustering Method. In: Ferrández, J. M., Álvarez Sánchez, J. R., De La Paz, F. & Toledo, F. J. (eds.) *New Challenges on Bioinspired Applications*. 6687. (Springer Berlin Heidelberg, 2011).
25. Mishro, P. K., Agrawal, S., Panda, R. & Abraham, A. A Novel Type-2 Fuzzy C-Means Clustering for Brain MR Image Segmentation. *IEEE Trans. Cybern.* **51**, 3901–3912, <https://doi.org/10.1109/TCYB.2020.2994235> (2021).
26. Vishnuvarthanan, G., Rajasekaran, M. P., Subbaraj, P. & Vishnuvarthanan, A. An unsupervised learning method with a clustering approach for tumor identification and tissue segmentation in magnetic resonance brain images. *Appl. Soft Comput.* **38**, 190–212, <https://doi.org/10.1016/j.asoc.2015.09.016> (2016).
27. Maillard, P. et al. FLAIR and Diffusion MRI Signals Are Independent Predictors of White Matter Hyperintensities. *AJNR Am. J. Neuroradiol.* **34**, 54–61, <https://doi.org/10.3174/ajnr.A3146> (2013).
28. Nasrallah, I. M. et al. White Matter Lesion Penumbra Shows Abnormalities on Structural and Physiologic MRIs in the Coronary Artery Risk Development in Young Adults Cohort. *AJNR. Am. J. Neuroradiol.* **40**, 1291–1298, <https://doi.org/10.3174/ajnr.A6119> (2019).
29. Promjunyakul, N. et al. Characterizing the white matter hyperintensity penumbra with cerebral blood flow measures. *NeuroImage Clin.* **8**, 224–229 (2015). <https://doi.org/10.1016/j.nicl.2015.04.012>
30. Louppe, G., Wehenkel, L., Sutter, A. & Geurts, P. Understanding variable importances in forests of randomized trees. In: *Advances in Neural Information Processing Systems*. 26. Curran Associates, Inc.; (2013). Accessed January 29, 2024. <https://proceedings.neurips.cc/paper/2013/hash/e3796ae838835da0b6f6ea37bcf8bcb7-Abstract.html>
31. Shin, S. et al. Clustering and prediction of long-term functional recovery patterns in first-time stroke patients. *Front Neurol.* **14**, 1130236, <https://doi.org/10.3389/fneur.2023.1130236> (2023).
32. Kabir, A., Ruiz, C., Alvarez, S., Riaz, N. & Moonis, M. Model-based Clustering of Ischemic Stroke Patients. In: *Proceedings of the International Conference on Health Informatics*. SCITEPRESS - Science and Technology Publications; 172–181 <https://doi.org/10.5220/0005278101720181> (2015).
33. Kim, J. T. et al. Neural network-based clustering model of ischemic stroke patients with a maximally distinct distribution of 1-year vascular outcomes. *Sci. Rep.* **12**, 9420, <https://doi.org/10.1038/s41598-022-13636-w> (2022).
34. Sperber, C. et al. A typology of cerebral small vessel disease based on imaging markers. *J. Neurol.* <https://doi.org/10.1007/s00415-023-11831-x> (2023).
35. Zhang, J. et al. Time to recurrence after first-ever ischaemic stroke within 3 years and its risk factors in Chinese population: a prospective cohort study. *BMJ Open* **9**, e032087, <https://doi.org/10.1136/bmjopen-2019-032087> (2019).
36. Black, S., Gao, F. & Bilbao, J. Understanding White Matter Disease: Imaging-Pathological Correlations in Vascular Cognitive Impairment. *Stroke* **40**, S48–S52, <https://doi.org/10.1161/STROKEAHA.108.537704> (2009).
37. Sperling, R. A. et al. Amyloid Related Imaging Abnormalities (ARIA) in Amyloid Modifying Therapeutic Trials: Recommendations from the Alzheimer's Association Research Roundtable Workgroup. *Alzheimers Dement* **7**, 367–385, <https://doi.org/10.1016/j.jalz.2011.05.2351> (2011).
38. Garnier-Crussard, A. et al. White matter hyperintensities in Alzheimer's disease: Beyond vascular contribution. *Alzheimers Dement* **19**, 3738–3748, <https://doi.org/10.1002/alz.13057> (2023).
39. Witt, B. J. et al. A Community-Based Study of Stroke Incidence after Myocardial Infarction. *Ann. Intern. Med.* **143**, 785–792 (2005).
40. Rippe, J. M. Lifestyle Strategies for Risk Factor Reduction, Prevention, and Treatment of Cardiovascular Disease. *Am. J. Lifestyle Med.* **13**, 204–212, <https://doi.org/10.1177/1559827618812395> (2018).

41. Livingston, G. et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* **396**, 413–446, [https://doi.org/10.1016/S0140-6736\(20\)30367-6](https://doi.org/10.1016/S0140-6736(20)30367-6) (2020).
42. Phillips, N. A. & Mate-Kole, C. C. Cognitive Deficits in Peripheral Vascular Disease. *Stroke* **28**, 777–784, <https://doi.org/10.1161/01.STR.28.4.777> (1997).

Acknowledgements

This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (Dr. April Khademi).

Author contributions

K.C. and A.K. wrote the main manuscript text and K.C. prepared figures and supplementary material. K.C. is responsible for the integrity of the data and the accuracy of the data analysis. Additionally, all authors have reviewed, read and approved the manuscript, and they have agreed to the conditions of authorship as set forth by the journal.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44303-024-00063-x>.

Correspondence and requests for materials should be addressed to Karissa Chan.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024