NeuroImage: Clinical 40 (2023) 103533

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Detecting conversion from mild cognitive impairment to Alzheimer's disease using FLAIR MRI biomarkers

Owen Crystal [a,j,*], Pejman J. Maralani [b], Sandra Black [c,d,e,f,g], Corinne Fischer [c,h,i], Alan R. Moody [b], April Khademi [a,b,i,j,k]

[a] Electrical, Computer and Biomedical Engineering, Toronto Metropolitan University, Toronto, ON, Canada
[b] Department of Medical Imaging, University of Toronto, Toronto, ON, Canada
[c] Institute of Medical Science, University of Toronto, Toronto, ON, Canada
[d] Hurvitz Brain Sciences Research Program, Sunnybrook Research Institute, Toronto, ON, Canada
[e] Division of Neurology, Department of Medicine, Sunnybrook Health Sciences Centre, Toronto, ON, Canada
[f] L.C. Campbell Cognitive Neurology Research Unit, Sunnybrook Health Sciences Centre, Toronto, ON, Canada
[g] Department of Neurology, University of Toronto, Toronto, ON, Canada
[h] Department of Psychiatry, St. Michael's Hospital, Toronto, ON, Canada
[i] Keenan Research Center, St. Michael's Hospital, Toronto, ON, Canada
[j] Institute of Biomedical Engineering, Science and Technology (iBEST), Toronto, ON, Canada October 5, 2023
[k] Vector Institute for Artificial Intelligence, Toronto, ON, Canada

ARTICLE INFO

Keywords:
FLAIR MRI
Biomarkers
Alzheimer's disease
Mild cognitive impairment

ABSTRACT

Mild cognitive impairment (MCI) is the prodromal phase of Alzheimer's disease (AD) and while it presents as an imperative intervention window, it is difficult to detect which subjects convert to AD (cMCI) and which ones remain stable (sMCI). The objective of this work was to investigate fluid-attenuated inversion recovery (FLAIR) MRI biomarkers and their ability to differentiate between sMCI and cMCI subjects in cross-sectional and longitudinal data. Three types of biomarkers were investigated: volume, intensity and texture. Volume biomarkers included total brain volume, cerebrospinal fluid volume (CSF), lateral ventricular volume, white matter lesion volume, subarachnoid CSF, and grey matter (GM) and white matter (WM), all normalized to intracranial volume. The mean intensity, kurtosis, and skewness of the GM and WM made up the intensity features. Texture features quantified homogeneity and microstructural tissue changes of GM and WM regions. Composite indices were also considered, which are biomarkers that represent an aggregate sum (z-score normalization and summation) of all biomarkers. The FLAIR MRI biomarkers successfully identified high-risk subjects as significant differences ($p < 0.05$) were found between the means of the sMCI and cMCI groups and the rate of change over time for several individual biomarkers as well as the composite indices for both cross-sectional and longitudinal analyses. Classification accuracy and feature importance analysis showed volume biomarkers to be most predictive, however, best performance was obtained when complimenting the volume biomarkers with the intensity and texture features. Using all the biomarkers, accuracy of 86.2 % and 69.2 % was achieved for normal control-AD and sMCI-cMCI classification respectively. Survival analysis demonstrated that the majority of the biomarkers showed a noticeable impact on the AD conversion probability 4 years prior to conversion. Composite indices were the top performers for all analyses including feature importance, classification, and survival analysis. This demonstrated their ability to summarize various dimensions of disease into single-valued metrics. Significant correlation ($p < 0.05$) with phosphorylated-tau and amyloid-beta CSF biomarkers was found with all the FLAIR biomarkers. The proposed biomarker system is easily attained as FLAIR is routinely acquired, models are not computationally intensive and the results are explainable, thus making this pipeline easily integrated into clinical workflow.

* Corresponding author.
E-mail address: ocrystal@torontomu.ca (O. Crystal).

https://doi.org/10.1016/j.nicl.2023.103533
Received 11 April 2023; Received in revised form 5 October 2023; Accepted 26 October 2023
Available online 3 November 2023

1. Introduction

Alzheimer's disease (AD) is the most common type of dementia (Alzheimer's Association. , 2023) and causes a rapid decrease in cognition and memory over time (Mayeux and Stern, 2012). AD prevalence is expected to increase to 1 in 85 persons across the world by 2050 (Brookmeyer et al., 2007); demonstrating the urgency in identifying high-risk individuals early during the prodromal stage of AD, known as mild cognitive impairment (MCI) (Thung et al., 2018) (Petersen, 2004). MCI is characterized by abnormal changes in cognitive domains that have not reached the severity of AD. >50 % of individuals diagnosed with MCI will further progress to AD within 5 years (Gauthier et al., 2006) and this phase presents an imperative intervention window. MCI subjects that convert to AD are referred to as converting MCI (cMCI) as opposed to stable MCI (sMCI) subjects that have cognitive issues but do not progress to AD. Identifying cMCI individuals could select candidates for early treatment and also for patient stratification in large clinical trials. One of the most common causes of failure within AD treatment trials is the incorrect selection and grouping of subjects (Marinescu et al., 2020). Increasing the homogeneity of the cohort within clinical trials can help address this issue.

For objective and efficient patient monitoring and stratification, imaging biomarkers automatically mea- sured from magnetic resonance imaging (MRI) of the brain can be used. Image analysis and machine learning techniques have huge potential for identifying early signs of AD. Previous studies used biomarkers from T1- weighted MRI such as cortical thickness and hippocampal volume to detect AD subjects as well as cMCI subjects (Thung et al., 2018) (Moscoso et al., 2019) (Tong et al., 2017) (Eskildsen et al., 2013). Sorensen et al. used linear discriminant analysis and T1-weighted MRI biomarkers to classify between sMCI and cMCI with an accuracy of 63 % (Sørensen et al., 2014). Tong et al. used T1-weighted MRI intensity features to achieve an normal control (NC) vs AD classification accuracy of 73 % and an accuracy of 69 % when classifying between sMCI and cMCI (Tong et al., 2017). MRI-based brain atrophy measurements, such as hippocam- pal volume, had the strongest relationship with cognitive decline (Jack et al., 2010) (Vemuri et al., 2009) (Vemuri et al., 2009) and play an important role in predicting conversion to AD (Jack et al., 2010) (Frisoni et al., 2010).

Other methods explored multi-modal inputs for automated classification of dementia groups. Korolev et. al used probabilistic pattern classification with T1-weighted MRI, plasma biomarkers, clinical variables, and cognitive measures to achieve an sMCI-cMCI classification accuracy of 80 % (Korolev et al., 2016). Davatzikos et. al used T1-weighted MRI and cerebrospinal fluid (CSF) biomarkers to classify between sMCI and cMCI and achieved an overall accuracy of 62 % (Davatzikos et al., 2011). Deep learning has also become a popular solution for cognitive classification and disease prediction (Duc et al., 2020) (Jo et al., 2019) (Lee et al., 2019) (Li et al., 2019) (Spasov et al., 2019). For example, Cheng et. al implemented transfer learning using T1-weighted MRI, positron emission tomography (PET), and CSF biomarkers and achieved a classification accuracy of 79 %. PET imaging offers unique insights as it provides information on amyloid-beta that is very useful in diagnosing AD. In the work by Cheng and colleagues, comparatively high classification accuracy was achieved but using multi-modal inputs is very expensive and it is much less likely to have access to these images. CSF blood biomarkers have been used alone for cognitive classification and predicting MCI to AD conversion (Palmqvist et al., 2012) (Caminiti, 2018) (Davatzikos et al., 2011); but studies have shown that while CSF tau, a common CSF biomarker, is accurate in quantifying neuronal injury and neurodegeneration, it does not necessarily predict the onset of AD (Jack et al., 2010) (Hesse et al., 2001) (Schoonenboom et al., 2012). Other studies have also used demographic information (age, sex, education) as well as cognitive scores (e.g., ADAS, MMSE, CDR-SB) to classify between cognitive groups (Korolev et al., 2016) (Gaser et al., 2013) (Inglese et al., 2022). All existing technologies have been designed and tested on a single dataset.

This is the first work that investigates the clinical utility of FLAIR biomarkers for differentiating between sMCI and cMCI in three large multicentre dementia cohorts. Twelve biomarkers related to intensity, tex- ture, and volumes of objects in the brain are extracted from three cross-sectional and longitudinal dementia cohorts using fully-automated and validated algorithms. A composite index is also explored, which is the normalized sum of biomarkers into a single-valued metric. Unlike most previous works that focus on T1- MRI in only cross-sectional data, this work brings new knowledge about FLAIR biomarkers in a large-scale neuroimaging dataset. FLAIR sequences highlight white matter disease and white matter lesions (WML) related to many neurological disorders (DiGregorio et al., 2022) (Khademi et al., 2021) (Rocca et al., 2016); cognitive impairment, age, and CSF-biomarkers (DiGregorio et al., 2022) (Bahsoun et al., 2022) (Crystal et al., 2022) (Chan et al., 2022) (Chan et al., 2022). As a result, FLAIR MRI is routinely acquired in clinical settings and FLAIR biomarkers have high translation potential.

Several types of analyses were conducted to examine the clinical utility of FLAIR biomarkers in differ- entiating between sMCI and cMCI subjects. First, statistical analysis was perform to test for significant differences between biomarkers across cognitive groups were tested in both cross-sectional and longitudinal data using ANOVA and regression. Second, the predictive capacity of FLAIR biomarkers were examined by designed a machine learning classifier to predict cognitive label based on the biomarkers. Feature importance for the sMCI vs cMCI classification was performed using Shapley (SHAP) feature analysis. SHAP feature importance was used to determine the most important features as well as to develop biomarker thresholds between groups. Lastly, survival analysis was conducted to determine if FLAIR biomarkers significantly impact the AD conversion probability.

2. Materials and methods

2.1. Cohorts

Three multi-centre cohorts were used in this work: 1) The Alzheimer's Disease Neuroimaging Initiative (ADNI) is an international dataset with longitudinal imaging for studying AD and dementia disease (n = 978 subjects) (Jack et al., 2008). 2) The Canadian Consortium on Neurodegeneration in Aging (CCNA) is a cross-sectional pan- Canadian study for the investigation of different types of dementia, including Alzheimer's disease, vascular dementia, frontotemporal dementia, and Lewy body dementia (n = 369 subjects) (Chertkow et al., 2019) (Mohaddes et al., 2018). 3) The Ontario Neurodegenerative Disease Research Initiative (ONDRI) dataset is a dementia cohort and subjects with NC, sMCI, cMCI, and AD cognitive labels were used (n = 509 subjects) (Farhan et al., 2017). All subjects with FLAIR imaging and available Montreal Cognitive Assessment (MoCA) scores were used. See Table 1 for patient breakdown.

2.2. MR imaging

FLAIR MRI images were acquired from 77 centres worldwide, on 1.5/3T machines from GE, Siemens and Philips. Pixel spacing is 0.4285 mm-1.2 mm, TR: 2250–11000 ms, TE: 90–200 ms, TI: 2200–16500 ms. Table 1 summarizes the acquisition parameters for each of the datasets.

2.3. Cognitive labeling

MoCA was used to label subjects as NC (MoCA > 26), MCI (19 < MoCA < 25), or AD (MoCA < 18). The MCI group was subdivided into stable MCI (sMCI) and converting MCI (cMCI) based on whether they converted to AD or not. For cross-sectional analyses, an MCI subject was considered cMCI if their MoCA score progressed from the MCI range to the AD range. Only the scan immediately prior (approximately 1 year) to AD conversion was labeled cMCI and used for the analysis. If a subject remained within the MCI MoCA range for all time points, they were

Table 1
FLAIR MRI datasets. All data is acquired at 1.5/3T and 3–5 mm slice thickness.

	Patient Information					
Database	**Disease**	**Volumes**	**Images**	**Patients**	**Centres**	**Age ± SD (years)**
ADNI	Dementia	3072	117523	978	58	73.75 ± 7.35
CCNA	Dementia	361	17802	369	19	72.68 ± 7.31
ONDRI	Dementia	1388	66659	509	17	68.59 ± 7.68
Total	Dementia	4821	201983	1856	77	72.67 ± 7.68
			Acquisition Parameters			
Database	**GE/Philips/Siemens**	**TR (ms)**	**TE (ms)**	**TI (ms)**	**X Spacing (mm)**	**Y Spacing (mm)**
ADNI	769/571/1732	9000–11000	90–154	2250–2500	0.8594	0.8594
CCNA	30/59/272	9000–11000	117–150	2200–2800	0.4285–1	0.4285–1
ONDRI	684/17/687	2250–10000	120–200	2250–16500	0.92–0.94	0.89–0.94
Total	1483/647/2691	2250–11000	90–200	2200–16500	0.4295–1.2	0.4295–1.2

labeled sMCI. For longitudinal analysis, subjects labeled MCI that converted to AD in later scans, were assigned a cMCI label for up to 6 years prior to AD conversion (based on the available data). All scans after progressing to AD were considered AD. The cMCI label was assigned to all scans of subjects before conversion (compared to cross-sectional analysis, which defined cMCI scans as only the scan 1 year before conversion). Time points for longitudinal cMCI scans were normalized about the conversion point, therefore visit number 5 represents 1 year before converting to AD and all preceding time points represent scans further in the past. See Table 2 for a breakdown of the sample sizes for the cross-sectional (N = 1356) and longitudinal (N = 4707) analyses broken down by cognitive group. For training AD/NC models, the number of NC and AD was balanced. To avoid issues with variance and class imbalance, the number of sMCI samples was balanced with the number of cMCI samples through random sampling.

2.4. Biomarker measurement

FLAIR MRI volumes underwent bias field correction and intensity standardization to align intensity ranges of each tissue across datasets (Reiche et al., 2019). Intracranial volume (ICV) segmentation was performed using a Mul- tiResUNet convolutional neural network (CNN) to extract brain tissue (DiGregorio et al., 2021). Total brain volume (TBV) and CSF were segmented by thresholding the intensity standardized volumes (Reiche et al., 2019). A skip connection U-Net CNN was used to segment WMLs (Khademi et al., 2021) and the normal-appearing brain matter (NABM), defined as the WM and GM, was detected by subtracting CSF and WML masks from the ICV. Dataset-specific fine-tuned 2D-UNet CNNs were used to segment the lateral ventricular volume (LVV) (Crystal et al., 2023). Subarachnoid CSF (Sub CSF) was found by calculating the difference between the CSF and LVV masks. Volumes from each of the structures were computed, and normalized by ICV to account for varying head size (Hansen et al., 2015) (Nordenskjöld et al., 2013). In total there are six volume biomarkers and Fig. 1 shows example segmentations. To complement the volume biomarkers, three FLAIR texture features, and three FLAIR intensity features from the NABM region were included. The texture features are previously designed biomarkers called macrostructural damage (MAD), microstructural damage (MID), and microstructural integrity, which were computed based on local binary patterns and spatial correlation of intensities within small regions (Khademi et al., 2009). MAD measures the local structural changes and edge content (Bahsoun et al., 2022). MID summarizes smaller, variations within the NABM. Integrity measures the homogeneity of the NABM (Bahsoun et al., 2022). The NABM intensity features are mean intensity, skewness (lack of symmetry in the intensity distribution), and kurtosis (how peaked the intensity distribution is). Fig. 2 shows example texture features.

To reduce dimensionality, we propose composite indices, which combines biomarkers into a single-valued metric for each subject. Each biomarker was z-score normalized with respect to the mean and standard deviation of the NC group for that biomarker. Biomarkers were then grouped based on their trends with worsening cognitive state/neurodegeneration: increasing or decreasing (Bahsoun et al., 2022) (DiGregorio et al., 2021). The increasing and decreas- ing biomarkers are listed in Table 9. Z-score normalized biomarkers were added together for the increasing and decreasing biomarkers separately, yielding a z-score increasing composite index (ZSI) and a z-score decreasing composite index (ZSD) respectively.Table 10.

Table 2
Dataset for cross-sectional and longitudinal analyses.

Dataset	Cross-Sectional				Longitudinal			
	NC	AD	sMCI	cMCI	NC	sMCI	cMCI	AD
ADNI	376	376	62	62	1059	388	161	1707
ONDRI	165	165	27	27	573	155	27	637
CCNA	38	48	–	–	–	–	–	–
Total	589	589	89	89	1632	543	188	2344

2.5. Statistical analysis

Intensity, texture, and volume biomarkers were extracted from all imaging volumes and two types of statis- tical analyses were performed. First, ANOVA and Tukey's posthoc were used to compare biomarker means between cognitive groups with effect sizes determined by computing Cohen's *d*. This was done using the cross sectional data shown in Table 2. Biomarkers with significantly different means between sMCI and cMCI groups were retained to compute the cross-sectional ZSI (Equation (2) and ZSD (Equation (3) compos- ite indices. The second analysis uses regression to determine differences in longitudinal progression between groups. Biomarkers were computed for all subjects with available longitudinal data (Table 2). The rate of biomarker change over serial scans is modeled using linear regression for each cognitive group using the following equation: $Biomarker \sim time + diagnosis + time * diagnosis$. The interaction term is added to analyze differences across diagnostic groups over time. Slopes and intercepts for each group were compared using ANOVA and Tukey's posthoc analysis as well as Cohen's *d* to examine effect size and practical signif- icance. Biomarkers with significantly different slopes between sMCI and cMCI groups are retained to create longitudinal composite indices, ZSI (Equation (4) and ZSD (Equation (5). Cross-sectional and longitudinal composite indices were used in the classification and survival analysis experiments respectively.

2.6. Classification of cognitive label

To examine the predictive capacity of FLAIR biomarkers, automated classification of the cognitive label is investigated using cross-sectional data (Table 2). Imaging biomarkers from AD and NC subjects from all datasets were used to train a random forest classifier (RFC) to classify between NC vs AD and sMCI vs cMCI. This training setup was motivated by other successful works that described sMCI as similar to NC and cMCI as similar to AD, which gives way to a larger training set (NC/AD) and

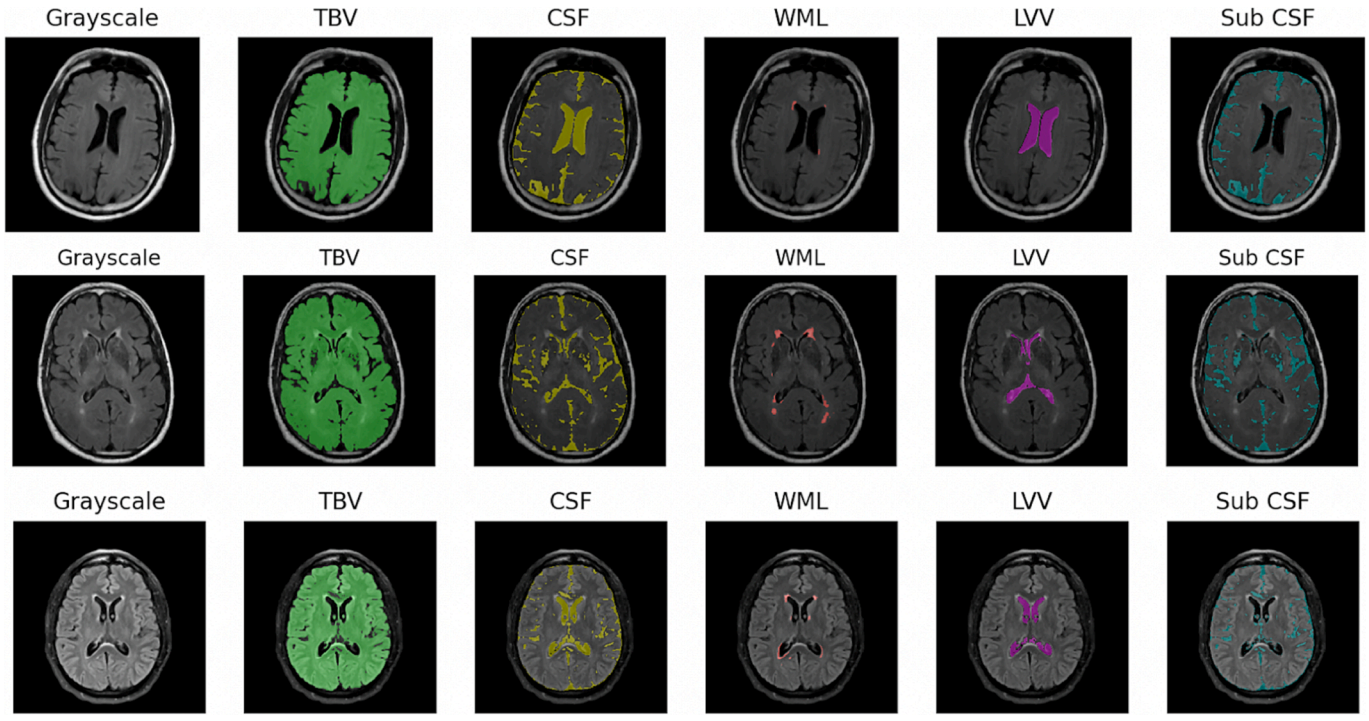

Fig. 1. Example segmentations from three different subjects.

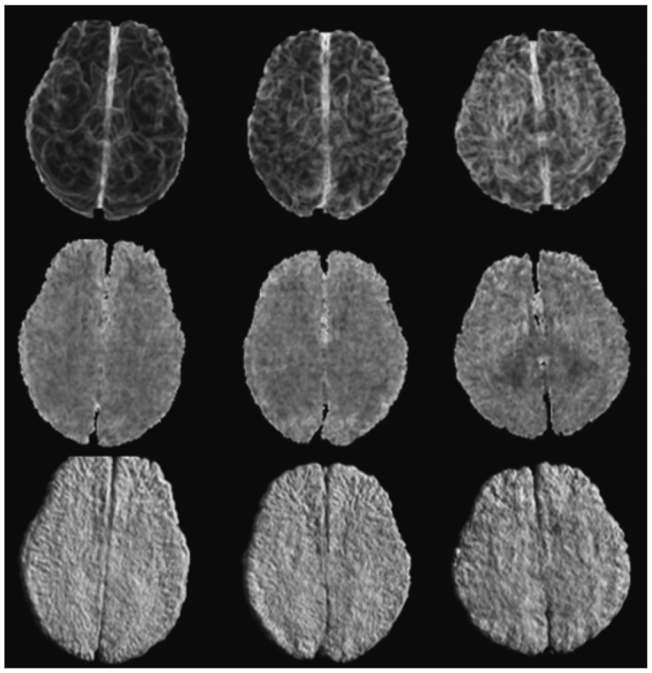

Fig. 2. Texture maps used to compute (each subjects is one column): macrostructural damage (top row), microstructural damage (middle row), microstructural integrity (bottom row).

successful classification between sMCI and cMCI (Tong et al., 2017). An equal number of NC and AD subjects were randomly selected from the entire dataset to avoid class imbalance. For NC vs AD classification, a 70/30 train/test data split was used across all datasets and 5-fold cross-validation was implemented. To examine classification performance between sMCI vs cMCI, all NC and AD data was used to train the final model, which was tested on all sMCI and cMCI subjects. Classification performance was measured using overall classification accuracy (ACC) and area under the receiver operating curve (AUC).

Using the final classification model, SHAP feature analysis is performed to determine which biomarkers were contributing the most to classifying between sMCI and cMCI. Feature importance is determined based on the Shapley value, which is the average of the marginal contributions of each possible feature combination (Nohara et al., 2022). SHAP analysis is known to be the state-of-the-art feature importance analysis in terms of explainability as it is known to be more consistent than ensemble tree methods (Lundberg and Lee, 2017). The SHAP feature importance is used to generate SHAP dependence plots and biomarker thresholds.

2.7. Survival analysis

Survival analysis is used to investigate if FLAIR biomarkers show differences in AD conversion probability between subjects above and below a specified biomarker threshold. Using the previously computed SHAP values for each biomarker, SHAP dependence plots are generated to determine biomarker thresholds (Chan et al., 2023). First, a regression line is used to model the relationship between biomarker value vs SHAP value. The biomarker value (x-value) in which the regression line crosses the $y = 0$ point is used as the threshold between the sMCI and cMCI groups.

Using the biomarker thresholds, and the longitudinal data in Table 2, Kaplan-Meier curves were used to estimate the "survival function" over time of the threshold-determined groups. In this work, survival refers to maintaining an MCI diagnosis over time and not converting to AD. Survival rates of subjects with biomarker values above and below the thresholds were determined via the Cox Proportional Hazard model. Significant differences are tested between groups to determine weather the hazard ratio (HR) is equal to 1 (no difference in survival rate between groups). Since HR yields a percent change in hazard with one unit increase of the variable, each biomarker underwent z-score normalization. Showing significant differences between the hazard rate of biomarkers above and below the thresholds would demonstrate the biomarkers' ability to distinguish between sMCI and cMCI.

2.8. Biomarker correlation

To validate the extracted and computed biomarkers, their correlation with CSF biomarkers and genotypes, that are known as gold standard indicators of AD and cognitive decline, was investigated cross-sectionally using the ADNI dataset. Each biomarker's correlation with phosphorylated-tau (p-tau) and amyloid-β (Aβ) CSF biomarkers and the APOE4 genotype was measured (Andersson et al., 2008) (DeTure and Dickson, 2019). Elevated p-tau and reduced levels of Aβ in the CSF are strong characteristics of AD, while the presence of the APOE4 allele is an established indicator for AD.

3. Results

All 12 vol, texture and intensity FLAIR MRI biomarkers were extracted from the FLAIR MRI dataset and subjects were labeled as NC, sMCI, cMCI or AD as in section 2. Analysis was completed on a NVIDIA GeForce RTX 3090 Ti GPU with 32 GB of RAM. Recall that Table 2 breaks down the number of subjects used for the cross-sectional and longitudinal analyses. These sample sizes, which were by cognitive group for each of the three datasets, were based on the exclusion criterias from the original dataset shown in Table.

1. For example, longitudinal volumes from the same subjects were not used for both testing and training to avoid data leakage, hence the difference in sample sizes between Table 1 and Table 2 (subjects used for training were only used for training). Note that only ADNI was used in the correlation analysis since it was the only dataset that provided the amyloid-beta, phosphorylated-tau, and APOE4 biomarkers.

3.1. Statistical analysis

Distributions of the volume, texture, and intensity biomarkers over cognitive groups for all cross-sectional data can be seen in Fig. 9, Fig. 10, and Fig. 11respectively. These figures show progressively in- creasing and decreasing patterns over different cognitive labels. To compare biomarkers between cognitive labels, ANOVA and Tukey's posthoc analysis was performed. The results of the sMCI-cMCI comparison are shown in Table 3 and all comparisons are in Table 11. All biomarkers except for MID and MAD showed significant differences ($p < 0.05$) between sMCI and cMCI. Since MID and MAD did not differentiate between cMCI and sMCI, these biomarkers were removed and the remaining biomarkers (TBV, CSF, WML, LVV, SubCSF, Integrity, mean Intensity, Kurtosis, Skewness, and NABM/ICV) were used to create ZSI and ZSD. Distributions of the composite indices, over cognitive groups are shown in Fig. 3 showing increas- ing and decreasing trends with worsening cognitive diagnosis for ZSI and ZSD respectively. ANOVA and posthoc analysis showed significant differences ($p < 0.001$) between all groups for both the ZSI and ZSD composite indices. All biomarkers showed large effect size and practical significance ($d > 0.80$) for the NC–AD comparison except WML. LVV showed the highest effect size ($d = 0.29$) for the sMCI-cMCI comparison.

Table 3
ANOVA/posthoc of cross-sectional biomarkers between sMCI and cMCI. Bold is significant.

Group	Biomarker	p-value	Cohen's *d*
Volume	**TBV**	**<0.01**	0.20
	CSF	**<0.01**	0.20
	WML	***0.04***	0.12
	LVV	**<0.01**	0.29
	SubCSF	***0.03***	0.10
	NABM/ICV	**<0.01**	0.22
Texture	**Integrity**	**<0.01**	0.22
	MAD	*0.12*	0.21
	MID	0.79	0.12
Intensity	**Intensity**	***0.04***	0.24
	Kurtosis	**<0.01**	0.18
	Skewness	**< 0.01**	0.23
Composites	**ZSD**	**<0.01**	0.22
	ZSI	**<0.01**	0.23

For longitudinal analysis, linear regression was performed to analyze disease progression over time for each FLAIR biomarker. The biomarker regression plots for the volume, texture, and intensity biomarkers are shown in Fig. 14, Fig. 15, and Fig. 16 respectively. The results show increasing and decreasing slope trends for worsening cognitive diagnosis as expected. ANOVA and posthoc analysis were used for the slopes and intercepts for the sMCI-cMCI regression and are shown in Table 4 (all group comparisons are shown in Table 12). TBV, CSF, SubCSF, MAD, MID, Kurtosis, and Skewness demonstrated significant differences ($p < 0.05$) between the rate of biomarker change between the sMCI and cMCI group. These significant biomarkers were used to compute the longitudinal composite indices ZSI (Equation (4) and ZSD (Equation (5). The ZSI and ZSD biomarkers showed significant differences ($p < 0.05$) between the sMCI and cMCI slopes (Fig. 4). The cMCI slopes for both composite indices showed the fastest rate of change while the sMCI group progressed at a similar rate to the NC group. These biomarkers were retained and used as longitudinal markers for the survival analysis. All biomarkers yielded medium or large effect sizes ($d > 0.50$) when comparing sMCI-cMCI except for WML, MID, intensity, and skewness, while ZSI ($d = 1.62$) and ZSD ($d = 1.51$) had the two largest effect sizes. Table 13 shows results of the ANOVA and Tukey's posthoc comparing the intercepts of the regression lines between cognitive groups for each biomarker. All biomarkers showed significant differences ($p < 0.05$) when comparing the intercepts between NC/sMCI/cMCI and AD. None of the biomarkers showed significant intercept differences ($p < 0.05$) when comparing cMCI to NC or cMCI to sMCI. These results indicate that AD has a significantly greater amount of baseline disease in comparison to the other three groups.

3.2. Classification performance

To determine the predictive capacity of FLAIR biomarkers, cross-sectional biomarkers from AD and NC subjects were used to train an RFC to classify cognitive labels. Fivefold classification was performed for NC vs AD classification which yielded an ACC of 86.2 % (confidence interval of [0.83, 0.89] over five folds using all 12 biomarkers. Other feature (biomarker) combinations were experimented with, namely using only the composite indices (ZSD/ZSI), volume biomarkers only, intensity biomarkers only and texture biomarkers alone. The ACC results for models trained with different biomarkers is shown in Fig. 5 and feature groups in Table 5. The composite indices achieved similar performance as all biomarkers combined, followed by volume then texture and intensity biomarkers.

Using all NC and AD data, the final model is trained using all 12 biomarkers and classified between sMCI and cMCI with an ACC of 69.2 % (confidence interval of [0.62, 0.75]). Models developed using all biomarkers and individual biomarkers are shown in Fig. 6. Similar to the NC-AD classification, using all biomarkers was the best, followed by composite indices which were top performers, followed by volume, then intensity and then texture biomarkers. To demonstrate the tools work equally well across datasets, Fig. 12 and Fig. 13 were included to show the impact of using different training and testing datasets, which resulted in consistent performance across datasets likely due to the intensity standardization.

SHAP values were generated to measure feature importance for sMCI vs cMCI classification and are shown in Fig. 7. The top 5 features were ZSI, Skewness, ZSD, TBV, and NABM/ICV. These biomarkers, along with LVV and CSF, showed a nearly bimodal SHAP value distribution, indicating two distinct groups. Trends of the feature rankings agree with results of the feature group-based classification, where the composite indices and volume biomarkers contribute the most to sMCI vs cMCI classification. Interestingly, intensity skewness is included, and there is a

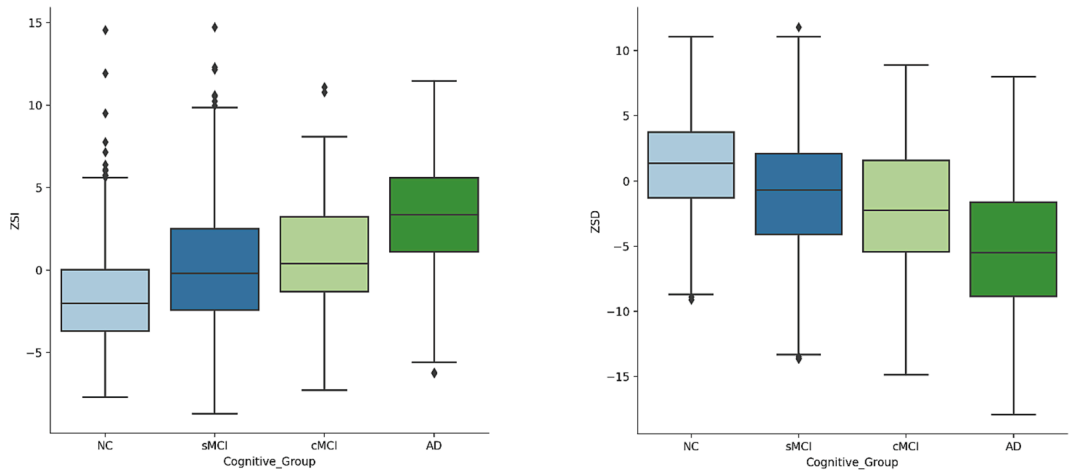

Fig. 3. ZSI (left) and ZSD (right) composite indices by cognitive group. Significant differences ($p < 0.05$) were found in all cognitive group comparisons for both composite indices.

Table 4
ANOVA/posthoc comparing longitudinal biomarker slopes of sMCI-cMCI groups. Bold is significant.

Group	Biomarker	p-value	Cohen's *d*
Volume	**TBV**	***0.02***	0.75
	CSF	***0.02***	0.75
	WML	*0.49*	0.01
	LVV	*0.41*	0.65
	SubCSF	***0.04***	0.67
	NABM/ICV	*0.10*	1.00
Texture	Integrity	*0.13*	0.65
	MAD	***0.03***	0.65
	MID	***0.04***	0.40
Intensity	Intensity	*0.85*	0.16
	Kurtosis	***0.04***	0.81
	Skewness	***< 0.01***	0.17
Composites	**ZSD**	***0.02***	1.51
	ZSI	***0.01***	1.62

long tail in that biomarker.

3.3. Survival analysis

SHAP values and corresponding biomarker thresholds were used to perform survival analysis. Using SHAP value vs Biomarker plots, linear regression was used to determine the x-value (biomarker threshold) in which y = 0 (the point at which the other class becomes more favourable). Threshold values as well as determined hazard ratios are listed in Table 6. Only ZSD, ZSI, LVV, and Kurtosis showed a significant difference ($p < 0.05$) between the hazard rate of subjects above and below their respective threshold. Kaplan-Meier curves of composite indices, volume, texture, and intensity biomarkers are shown in Fig. 8, Fig. 17, Fig. 18, Fig. 19. Recall that "Survival Probability" represents the probability of the subject not developing AD. Therefore, a decreased survival probability indicates a higher likelihood of converting to AD.

3.4. Biomarker correlation

To validate the correlation between the FLAIR biomarkers and composite indices, and AD conversion, a sub-analysis was performed. Pearson's correlation coefficient was computed between biomarkers and

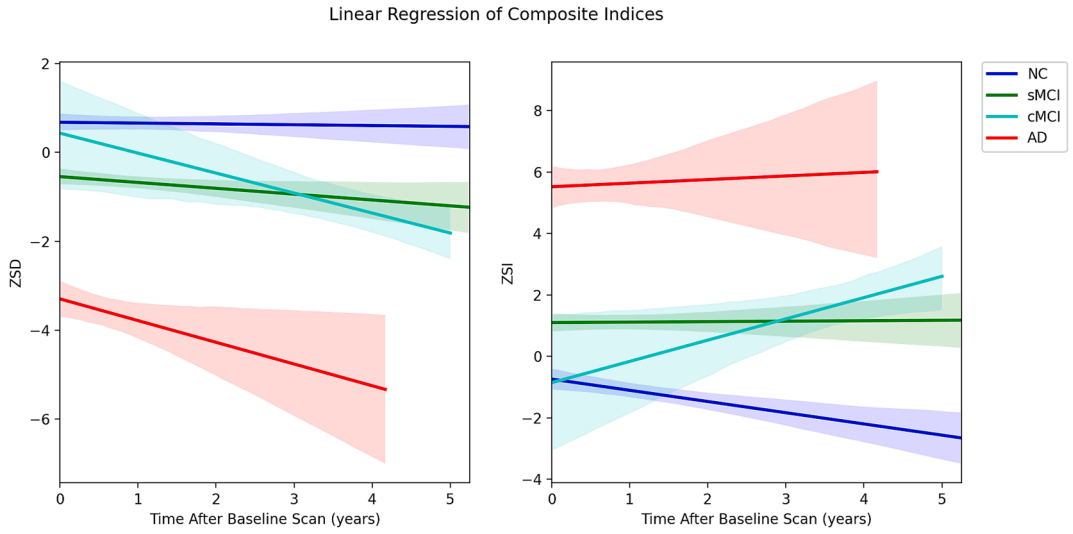

Fig. 4. ZSD (left) and ZSI (right) over time by cognitive group. Significant differences ($p < 0.05$) between the slopes of the sMCI and cMCI groups for both composite indices.

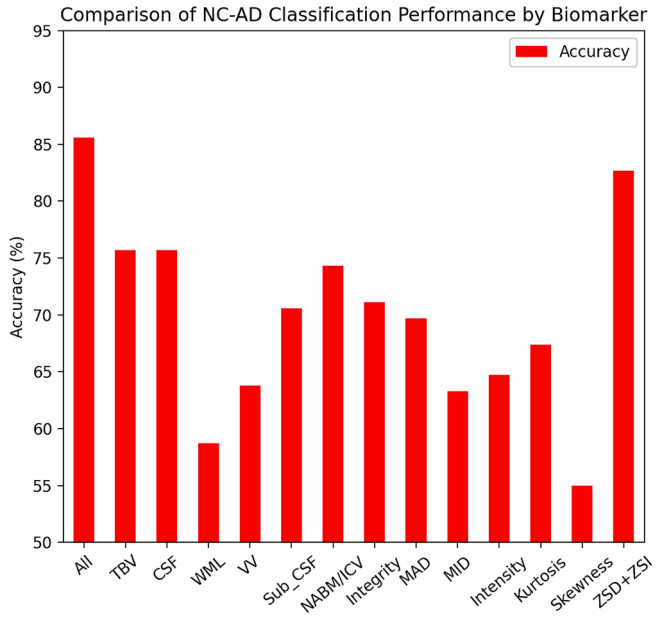

Fig. 5. NC vs AD classification accuracy for each individual biomarker and the composite indices.

Table 5
Classification accuracy on held out NC, sMCI, cMCI and AD datasets over all datasets.

Comparison	Features	Accuracy	AUC
NC-AD	All Biomarkers	86.2 %	87.0 %
	ZSD/ZSI	82.7 %	83.0 %
	Volume	80.8 %	82.0 %
	Intensity	76.1 %	76.0 %
	Texture	76.5 %	77.0 %
sMCI-cMCI	All Biomarkers	69.2 %	69.0 %
	ZSD/ZSI	65.2 %	65.0 %
	Volume	62.3 %	61.0 %
	Intensity	59.7 %	60.0 %
	Texture	55.6 %	57.0 %

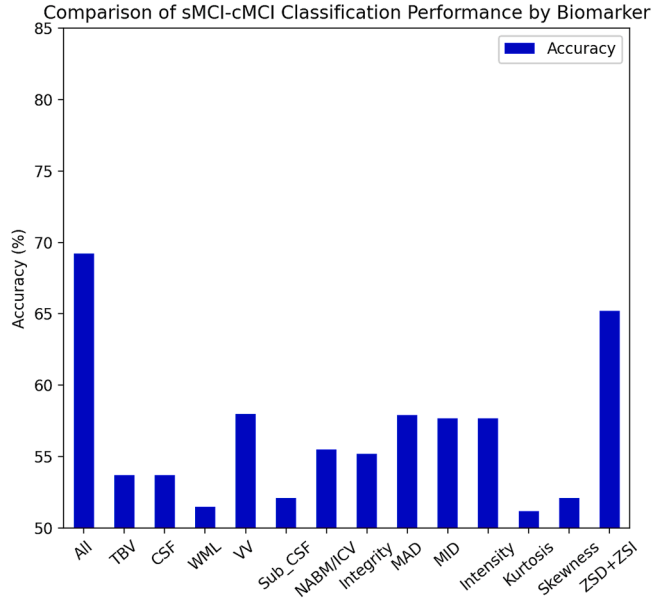

Fig. 6. sMCI vs cMCI classification accuracy for each individual biomarker and the composite indices.

p-tau and amyloid-β (Aβ) CSF biomarkers and APOE4 genotype, which are known as established indicators for MCI to AD conversion. All biomarkers showed a significant correlation ($p < 0.01$) to Aβ. Only skewness did not show a significant correlation to p-tau. LVV, Integrity, Intensity, NABM/ICV, and ZSD showed a significant correlation ($p < 0.001$) to the APOE4 genotype. The correlation shown between the FLAIR i-biomarkers and the gold standard AD conversion indicators further strengthens the results of the cross- sectional and longitudinal analyses performed throughout this paper.

3.5. Summary

A complete summary of the results of each experiment for all 12 biomarkers and the two composite indices, specifically for sMCI-cMCI differentiation is shown in Table 7.

4. Discussion

This is the first work that uses multi-centre, FLAIR-MRI data to identify high-risk MCI subjects, which was done both cross-sectionally and longitudinally using quantitative biomarkers, and composite indices. Previously validated and developed tools were used to accurately extract and measure the biomarkers (Khademi et al., 2021) (Reiche et al., 2019) (DiGregorio et al., 2021) (Hansen et al., 2015) (Nordenskjöld et al., 2013). It was hypothesized that six volume-based (TBV, CSF, WML, LVV, SubCSF, NAB- M/ICV), three texture (MAD, MID, Integrity) and three intensity (mean Intensity, Kurtosis, Skewness) biomarkers could identify cMCI subjects in cross-sectional and longitudinal data using statistical models and machine learning classifiers. Composite indices were designed to compare sMCI and cMCI groups using single-valued metrics that summarize the neurodegeneration.

Longitudinal analysis of biomarkers is valuable as it can be leveraged to analyze disease progression, identify earlier diseases, used for patient monitoring, and help understand disease mechanisms. DiGregorio et al. demonstrated longitudinal FLAIR volume biomarkers are significantly different between cognitive groups, with longitudinal biomarkers computed as the normalized average annual rate of change of the volume (DiGregorio et al., 2022) (Bahsoun et al., 2022). We expanded on this work by focusing on a more traditional longitudinal analysis to provide insight into the serial differences between cognitive groups via linear regression and survival analysis.

Volume-based biomarkers related to brain tissue loss (TBV, CSF, LVV, Sub CSF, and NABM/ICV) showed significant differences between sMCI and cMCI groups cross-sectionally and all but LVV and NAB- M/ICV showed significant differences longitudinally. Medium to large effect sizes were found in all volumes biomarkers except WML when comparing rate of change between sMCI and cMCI groups. Volume biomark- ers yielded higher NC vs AD and sMCI vs cMCI classification performance and were generally ranked higher based on SHAP feature importance in comparison to the texture and intensity biomarkers. This further supports current literature that shows neurodegeneration is strongly associated with a loss of brain tissue and so is the onset of AD (Jack et al., 2010) (Vemuri et al., 2009) (Vemuri et al., 2009) (DiGregorio et al., 2022); which was confirmed by the proposed analysis where TBV and NABM/ICV were significantly lower and CSF, LVV, and SubCSF were significantly higher for worse cognitive diagnosis. It is worth noting that LVV ranked higher in terms of SHAP feature importance in comparison to Sub CSF, but Sub CSF showed significant differences in the rate of change between sMCI and cMCI. This indicates that on a cross-sectional level, the magnitude of tissue loss is more noticeable in periventricular regions, but in terms of the rate of tissue loss, it is occurring significantly faster in the GM. Although WML showed significant differences cross-sectionally between sMCI and cMCI, it was irrelevant in all other analyses. Although in the past, it has been shown that WML is correlated with cognitive impair- ment (Khademi et al., 2021) (Brugulat-Serrat et al., 2020), more recent research has begun to

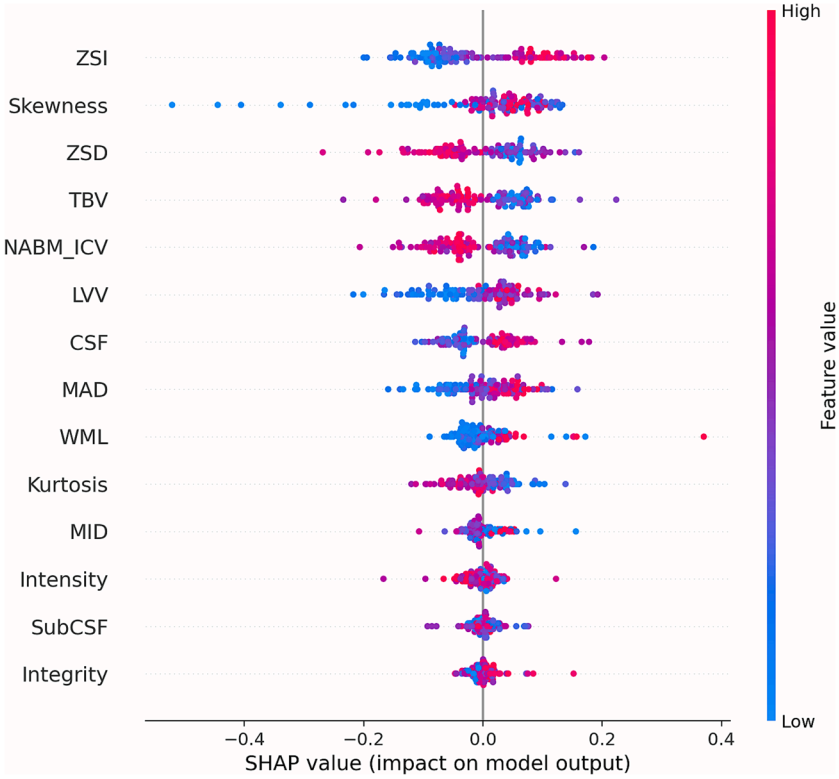

Fig. 7. SHAP feature importance ranking. Red is subjects contributing to a cMCI classification and blues indicates sMCI. Features ordered by feature importance, top is most important. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

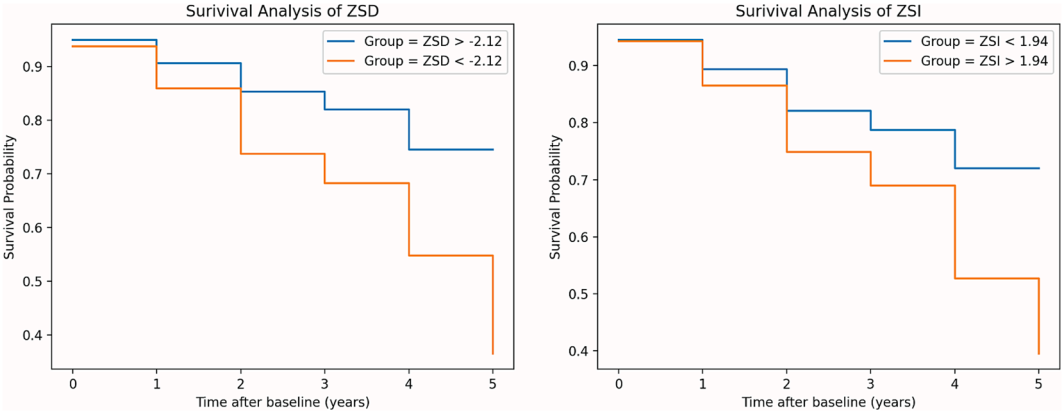

Fig. 8. ZSD (left) and ZSI (right) Kaplan-Meier curves representing the survival probability over time between the SHAP defined groups. Significant differences ($p < 0.05$) found between the hazard ratios of the two groups.

suggest that WML may not be contributing to cognitive decline in a straightforward manner (Chan et al., 2023). Our findings show that cross-sectional volume may differentiate between cMCI and sMCI, but longitudinally, our results suggest that WML volume change may not be a contributing factor for diagnosis of prodromal AD.

Despite yielding lower classification accuracy and SHAP rankings compared to volume biomarkers, adding texture and intensity biomarkers to the feature set resulted in the best performing model for both NC vs AD and sMCI vs cMCI classification. This indicates that in addition to overall tissue loss in the WM and GM based on the results of the volume biomarkers, there are more localized changes occurring within GM and WM regions that is related to prodromal AD. FLAIR texture biomarkers were shown to measure microstructural properties of GM and WM tissue, which is correlated to both mean diffusivity and fractional anisotropy in DTI, as well as cognitive impairment and neurodegeneration (Bahsoun et al., 2022). FLAIR intensity has been shown to be related to vasogenic edema, ischemia and demylination (Black et al., 2009), and was a predictor for future WML development. All the intensity and texture features except Skewness showed significant correlation with the p-tau and Aβ CSF biomarkers, further validating their relationship with the pathophysiological process of AD. Therefore, FLAIR texture and intensity are providing complementary information to the volume biomarkers that can be used further tease out differences between cMCI and sMCI. It is impor- tant to measure various dimensions of the disease since there are individualized patterns in both AD and prodromal (Badhwar et al., 2020) (Noh et al., 2014) (Poulakis et al., 2018); so including volume, texture, and intensity biomarkers encapsulates various types of neurodegenerative processes.

The integrity texture biomarker and intensity features (kurtosis, skewness, mean intensity) were able to distinguish between sMCI and

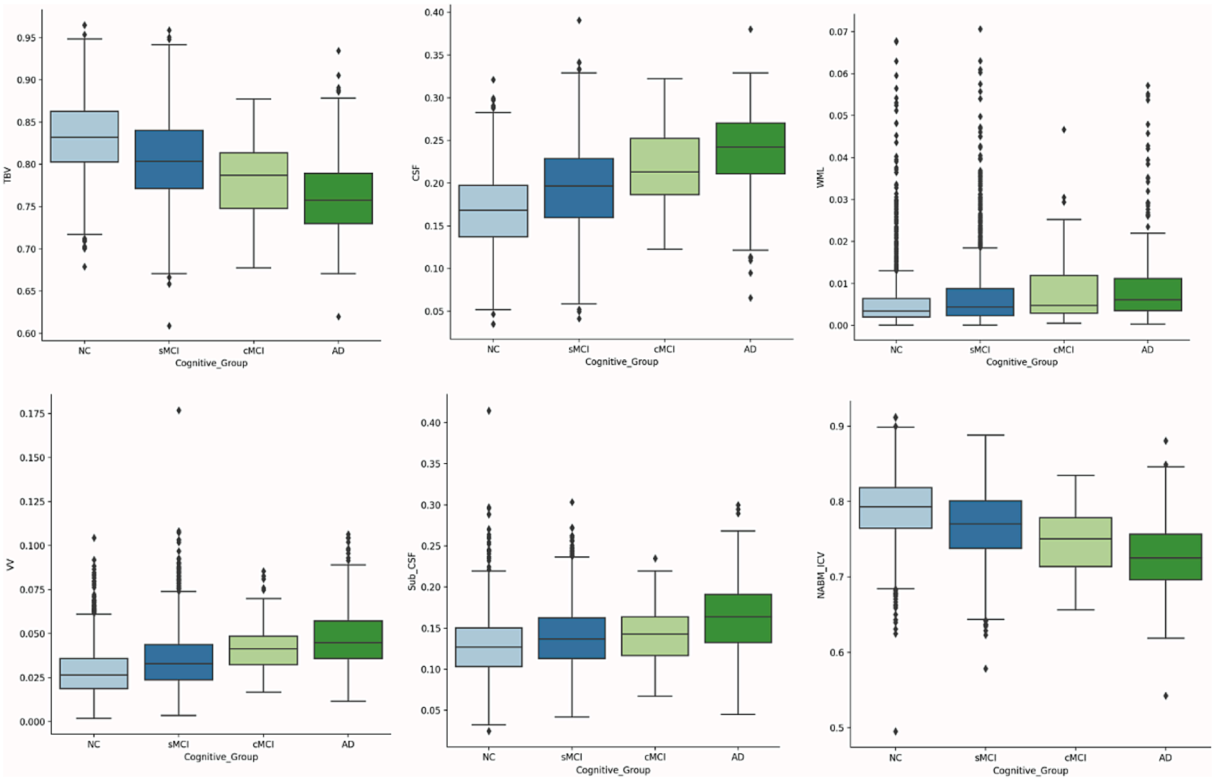

Fig. 9. Comparison of volumes biomarkers between cognitive groups. Significant differences ($p < 0.05$) seen between the sMCI and cMCI groups for all six volume biomarkers.

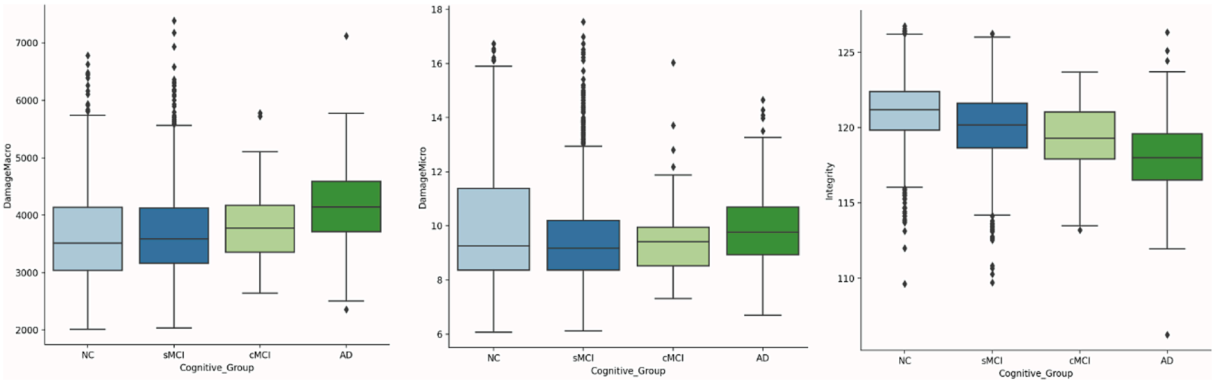

Fig. 10. Comparison of texture biomarkers between cognitive groups. Significant differences ($p < 0.05$) seen between the sMCI and cMCI groups for the integrity feature (right).

cMCI groups with statistical significance cross-sectionally while the texture markers MAD and MID were distinguished between groups in terms of longitudinal progression. This is despite MID and MAD not being significantly different between cMCI and sMCI on the cross-sectional level. In Bahsoun et al., FLAIR texture was found to be related to cognition, where it was hypothesized that GM and WM degeneration are reflected in local texture changes (Bahsoun et al., 2022). In this work, we showed longitudinal FLAIR texture also differentiates between cMCI and sMCI, indicating that serial changes to the tissue mi- crostructure in both WM and GM may play a role in subjects who convert to AD. MAD and MID quantify the roughness of the WM and GM based on local intensity differences (Bahsoun et al., 2022) as larger MAD and MID values indicate signs of potential GM atrophy and WM tract degeneration. Therefore, the longitudinal results suggest the rate of GM atrophy and WM tract degeneration was significantly greater for cMCI subjects than for sMCI subjects. The integrity texture biomarker is related to repeating structural patterns with less sim- ilar patterns reflecting a lower metric (more randomness). Cross-sectionally, integrity differentiated between cMCI and sMCI, but the rate of change of this biomarker is not significant between the groups. This could be related to the sensitivity of the metric, or could indicate the change in repeating patterns over time is not notable for prodomal AD. Intensity features differentiated between cMCI and sMCI using ANOVA on cross sectional and longitudinal differences, despite being ranked relatively low on the SHAP feature importance (except for skewness). Intensity in FLAIR MRI may be related to edema and water content related to tissue degeneration, and therefore, intensity may be a beneficial surrogate marker for identifying high risk subjects. The composite indices (ZSD and ZSI) were computed for cross-sectional and longitudinal analyses. The composite indices were the best performing biomarker group for all analyses as they showed significant differ- ences cross-sectionally and longitudinally with the largest effect sizes, the highest classification performance,

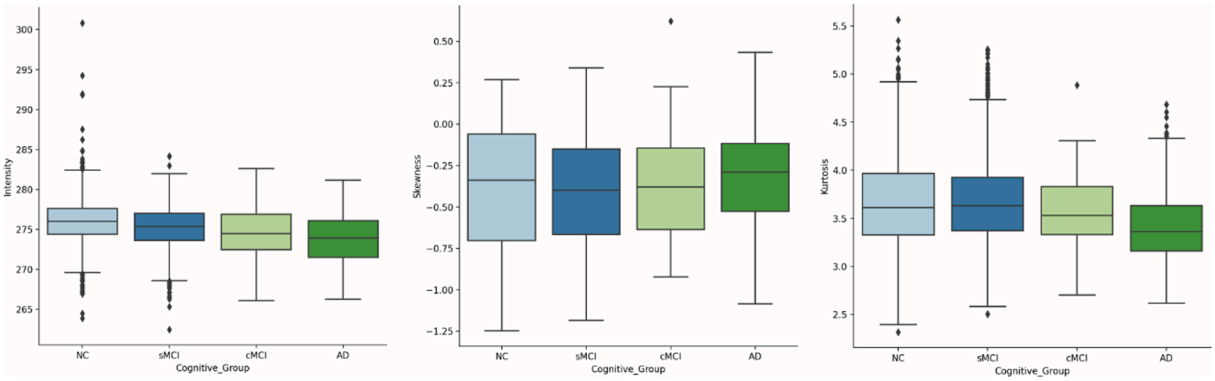

Fig. 11. Comparison of intensity biomarkers between cognitive groups. Significant differences ($p < 0.05$) seen between the sMCI and cMCI groups for all three intensity biomarkers.

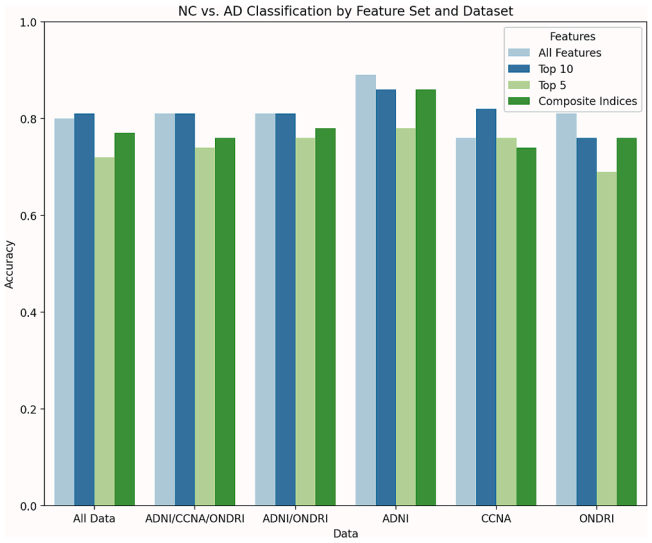

Fig. 12. NC vs AD classification accuracy by dataset and feature group.

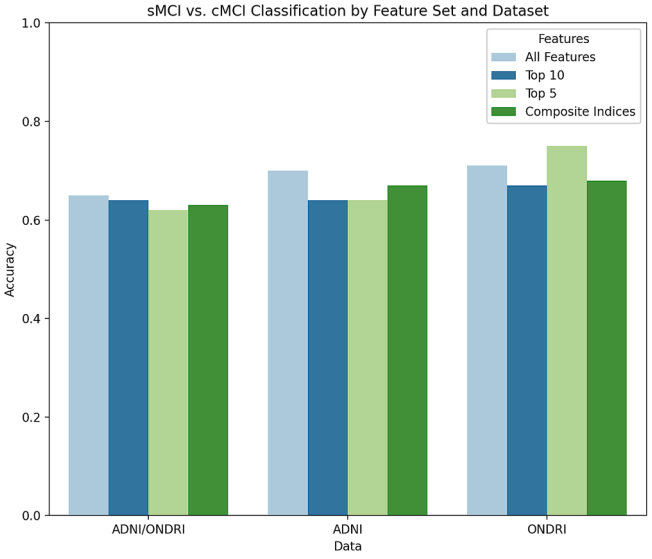

Fig. 13. sMCI vs cMCI classification accuracy by dataset and feature group.

ranked within the top 3 for SHAP feature importance, and both showed a significant impact on the risk of AD conversion via survival analysis. These results indicate the computed composite can summarize the neuroanatomical state into two single-valued metrics to distinguish between high and low-risk MCI subjects. The ZSD and ZSI metrics show great potential in clinical applications as they are easily monitored longi- tudinally. Otherwise, each biomarker would have to be monitored individually, which makes it difficult for clinicians to consolidate the information. They also offer a pre-clinical use as be used for patient strati- fication in applications such as cohort selections in clinical trials. Using the composite indices can allow clinicians to consider a single unified, explainable metric that summarizes all biomarkers which considers many dimensions of disease, and can identify normal aging or high-risk subjects.

A summary of the existing techniques for sMCI-cMCI classification is shown in Table 8. All meth- ods include T1-weighted MRI, while others combine T1 biomarkers with PET, cognitive measures, plasma biomarkers, and CSF biomarkers. The sample sizes vary from in the twenties to into the hundreds. Us-ing multi-modal inputs makes the process more time consuming and expensive. The proposed sMCI-cMCI classification model performance is similar (or better) to existing technologies using T1 only, demonstrating the ability to classify cMCI subjects on a patient-level basis using FLAIR biomarkers. These biomarkers provide explanation for the classification based on known pathological processes (e.g., brain tissue loss). Providing automated diagnosis using intuitive and explainable biomarkers is desired in the medical field (Duc et al., 2020) (Jo et al., 2019) (Lee et al., 2019) (Li et al., 2019) (Spasov et al., 2019). This helps to better understand the pathological process, which can help clinicians pro- vide a more personalized treatment plan.

Although no significant differences were seen in the intercepts of the sMCI and cMCI group in the longi- tudinal analysis, majority of the biomarkers and both composite indices show the cMCI group is declining or progressing at the fastest rate, highlighting the importance of identifying subjects in the prodromal phase. Interestingly, AD subjects did not have the highest slope but did have the highest baseline disease (inter- cept) for all biomarkers. Existing literature has found biomarkers related to AD are significantly different compared to NC and MCI in terms of the baselines, while the rate of change was not as exaggerated (Bahsoun et al., 2022). This could also explain the lack of differences between the NC and sMCI slopes compared to the AD group, while showing significant differences between the intercepts.

All Kaplan-Meier curves except for WML and SubCSF show that after time point 2, the survival proba- bility of the cMCI group begins to drop and is less than the sMCI group. Based on the time-normalization of the cMCI group, this indicates cMCI subjects (determined by biomarker thresholds) show signs of increased AD-conversion probability up to 4 years before AD diagnosis compared to sMCI subjects. Similar results are seen in Shi et al. as they used survival analysis to compare risk of AD

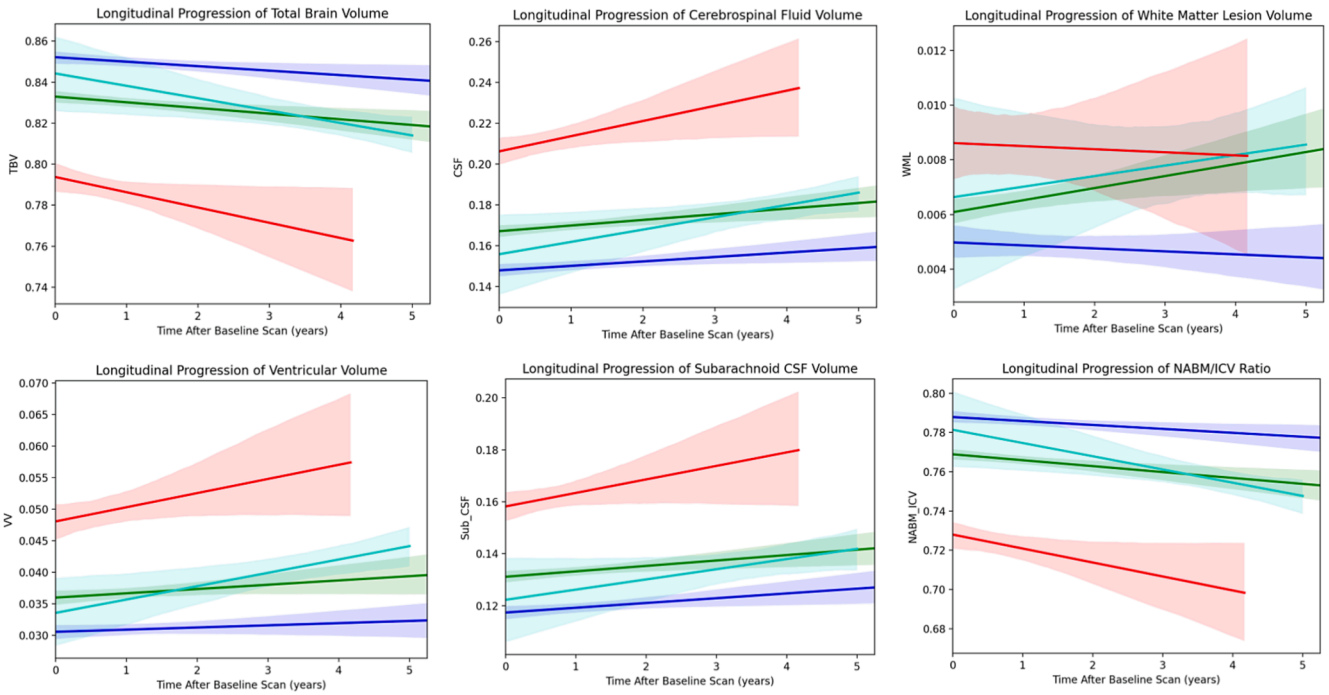

Fig. 14. Linear regression of the longitudinal progression of volume biomarkers over time by cognitive group. Significant differences ($p < 0.05$) seen between the slopes of the sMCI and cMCI groups for TBV, CSF, and Sub CSF.

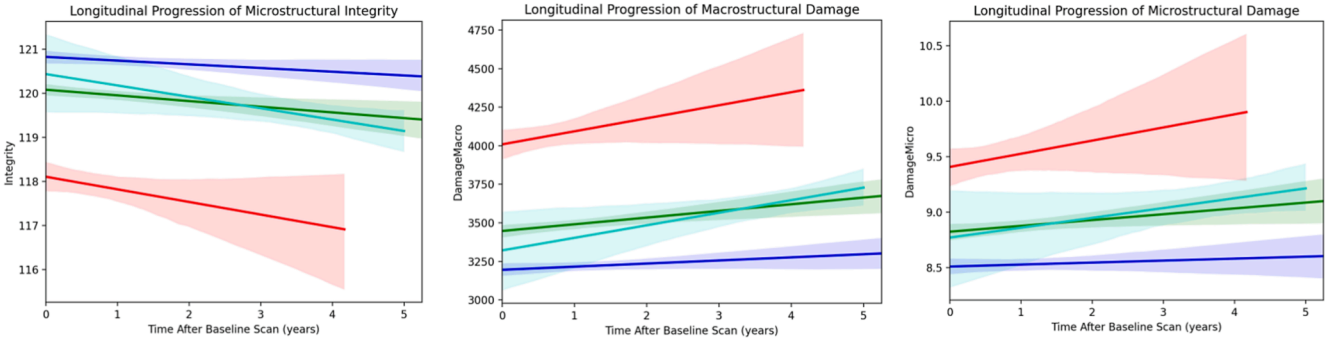

Fig. 15. Linear regression of the longitudinal progression of texture biomarkers over time by cognitive group. Significant differences ($p < 0.05$) seen between the slopes of the sMCI and cMCI groups for the MAD and MID features.

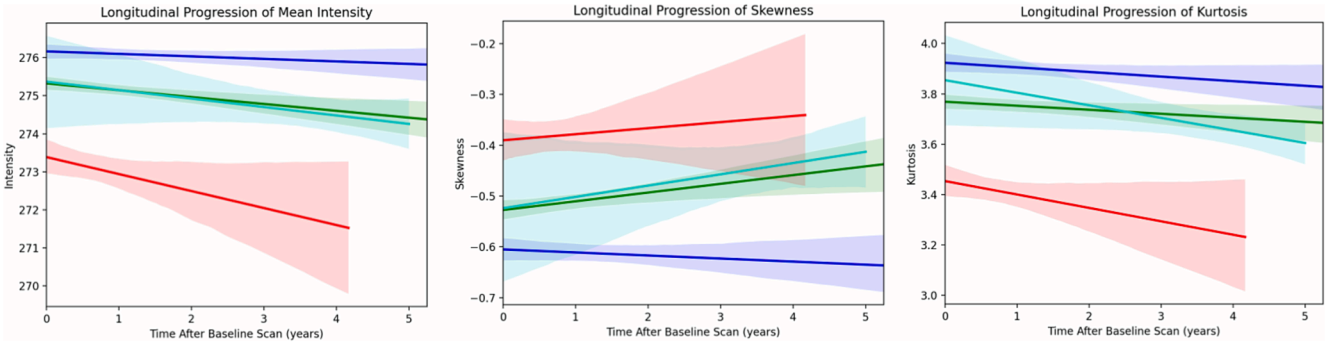

Fig. 16. Linear regression of the longitudinal progression of intensity biomarkers over time by cognitive group. Significant differences ($p < 0.05$) seen between the slopes of the sMCI and cMCI groups for kurtosis and skewness.

conversion between groups using T1-MRI images (Shi et al., 2022). Comparing the results of the proposed work to that of Shi et al., the survival functions have similar shape. However, in (Shi et al., 2022); the high risk group shows an drop in probability (i.e. increased chance of conversion) approximately 4 years after baseline, whereas our biomarkers show a reduction in survival probability just two years after baseline, indicating the FLAIR biomarkers may be more sensitive to detecting conversion earlier. FLAIR MRI may be providing unique

Table 6
Results of SHAP analysis and Cox Proportional Hazard model. P-value is in relation to the null hypothesis that the HR = 1 (no significant difference in risk of AD conversion).

Group	Biomarker	SHAP Rank	Biomarker Threshold	Hazard Ratio	p-value
Volume	TBV	4	0.81	0.94	*0.23*
	CSF	7	0.19	1.07	*0.23*
	WML	9	0.0068	1.02	*0.71*
	LVV	6	0.04	1.12	***0.03***
	Sub CSF	13	0.17	1.03	*0.38*
	NABM/ICV	5	0.76	0.92	*0.12*
Texture	Integrity	14	119.29	0.92	*0.08*
	MAD	8	3614.35	1.09	*0.12*
	MID	11	9.14	1.06	0.26
Intensity	Intensity	12	273.65	0.93	*0.17*
	Kurtosis	10	3.63	0.90	***0.05***
	Skewness	2	-0.49	1.101.1	0.06
Composites	**ZSD**	3	−2.12	0.88	***0.01***
	ZSI	1	1.94	1.11	***0.04***

insight into disease progression for early disease detection compared to more traditional modalities. The difference in survival probability or relative risk of converting to AD seen between groups is significantly impacted by ZSI, ZSD, LVV, and Kurtosis. Hazard ratios indicate how a single normalized unit increase impacts the risk of conversion. Biomarkers yielding hazard ratios>1 have positive correlations with the risk of conversion. Therefore, with every unit increase in ZSI, a subject is 1.11 times more likely to convert from MCI to AD. Biomarkers with HRs<1 (ZSD HR: 0.88) have an inverse relationship with conversion. Therefore, with every normalized unit decrease in ZSD, a subject is 1.14 (1/0.88) times more likely to convert to AD. The results demonstrate the ability to use these biomarkers to stratify subjects into low and high-risk groups. This has the potential to have significant real-world implications for important applications such as clinical trial cohort selection.

The proposed biomarkers not only offer automated processes for extracting volume biomarkers but also unique insights into the level of tissue loosening/degeneration based on the calculated damage and intensity features. When analyzing biomarkers, the proposed pipeline is much more efficient and comparably accurate in comparison to manual annotations. There is also a preclinical application. Previous Alzheimer's treat- ment trials have failed due to poor patient selection and subgrouping. The proposed biomarkers can help objectively group subjects to yield more homogeneous subgroups, which commonly leads

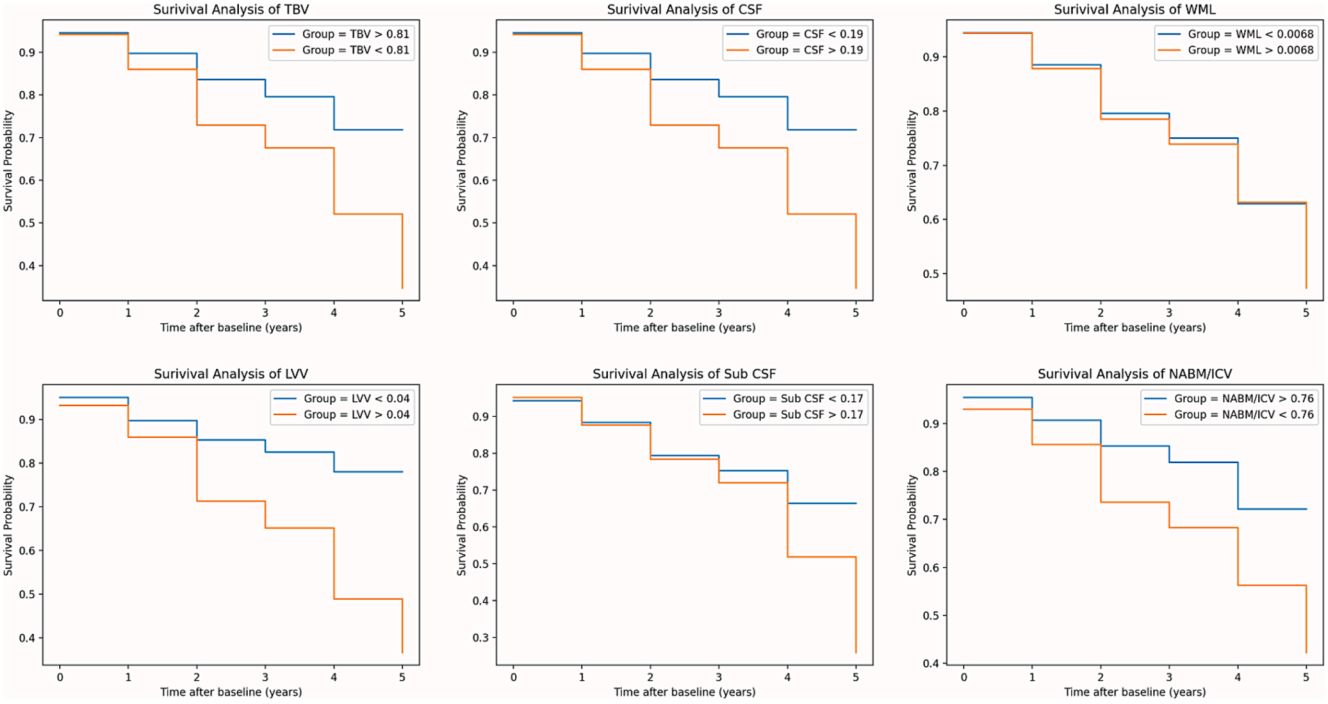

Fig. 17. Kaplan Meier curve of the volume biomarkers.

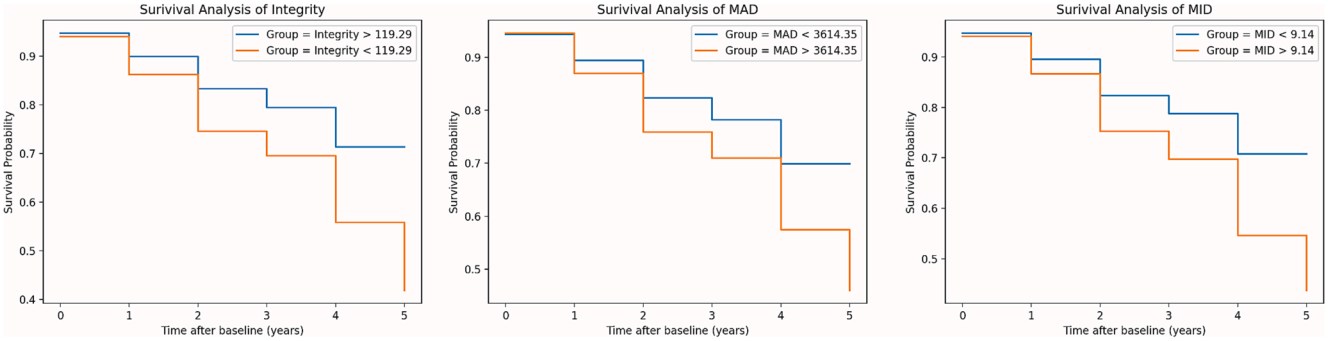

Fig. 18. Kaplan Meier curve of the texture biomarkers.

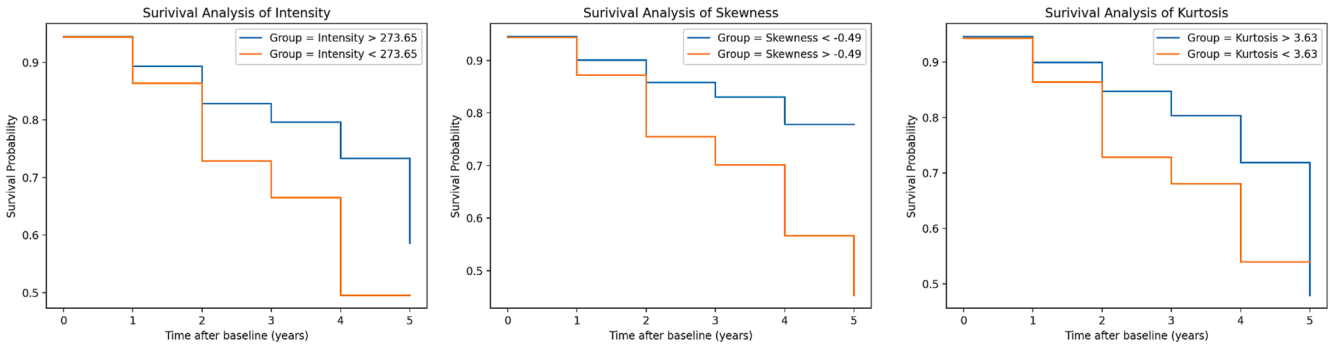

Fig. 19. Kaplan Meier curve of the intensity biomarkers.

Table 7
Summary of results comparing sMCI and cMCI groups. Bold indicates significance.

Group	Biomarker	ANOVA		Classification		Survival		Correlation (p-values)		
		Cross	Long.	SHAP	ACC	p-value	HR	p-tau	Aβ	APOE4
Volume	TBV	***<0.01***	***0.02***	4	53.7 %	*0.23*	0.94	***<0.01***	***<0.01***	*0.06*
	CSF	***<0.01***	***0.02***	7	53.7 %	*0.23*	1.07	***<0.01***	***<0.01***	*0.06*
	WML	***0.04***	*0.49*	9	51.5 %	*0.71*	1.02	***0.04***	***<0.01***	*0.49*
	LVV	***<0.01***	*0.41*	6	58.0 %	***0.03***	1.12	***0.02***	***<0.01***	***<0.01***
	SubCSF	***0.03***	***0.04***	13	52.1 %	*0.38*	1.03	***<0.01***	***<0.01***	*0.32*
	NABM/ICV	***<0.01***	*0.10*	5	55.5 %	*0.12*	0.92	***<0.01***	***<0.01***	***0.03***
	Integrity	***<0.01***	*0.13*	14	55.2 %	*0.08*	0.92	***<0.01***	***<0.01***	***0.03***
Texture										*0.08*
										0.72
	MAD	*0.12*	***0.03***	8	57.9 %	*0.12*	1.09	***<0.01***	***<0.01***	*0.08*
	MID	*0.79*	*0.04*	11	57.7 %	*0.26*	1.06	***0.04***	***<0.01***	0.72
	Intensity	***0.04***	*0.85*	12	57.7 %	*0.14*	0.93	***<0.01***	***<0.01***	***0.04***
Intensity										*0.21*
										0.87
	Kurtosis	***<0.01***	***0.04***	10	51.2 %	***0.05***	0.90	***<0.01***	***<0.01***	*0.21*
	Skewness	***<0.01***	***<0.01***	2	51.2 %	*0.06*	1.10	*0.07*	***<0.01***	*0.87*
Composites	ZSD	***<0.01***	***0.02***	3	65.2 %	***0.01***	0.88	***<0.01***	***<0.01***	***0.02***
	ZSI	***<0.01***	***0.01***	1		***0.04***	1.11	***<0.01***	***<0.01***	0.22

Table 8
Summary of existing sMCI-cMCI classification machine learning technologies.

Author	Dataset	Sample Size (sMCI, cMCI)	Modality	ACC, AUC
Tong et. al (Tong et al., 2017)	ADNI	129, 171	T1-MRI, Cognitive	0.81, –
Davatzikosa et. al (Davatzikos et al., 2011)	ADNI	170, 69	T1-MRI, CSF	–, 0.54–0.63
Xu et. al (Xu et al., 2016)	ADNI	83, 27	T1-MRI, PET	0.83, –
Korolev et. al (Korolev et al., 2016)	ADNI	120, 139	T1-MRI, Cognitive	0.80, –
Chupin et. al (Chupin et al., 2009)	ADNI	134, 76	T1-MRI	0.64, –
Misra et. al (Misra et al., 2009)	ADNI	76, 27	T1-MRI	0.82, –
Wolz et. al (Wolz et al., 2011)	ADNI	238, 167	T1-MRI	0.56–0.68, –
Cho et. al (Cho et al., 2012)	ADNI	65, 37	T1-MRI	0.71, –
Proposed	**ADNI, ONDRI**	**89, 89**	**FLAIR**	**0.69, 0.68**

Table 9
List of features categorized as increasing and decreasing with worsening cognition.

Increasing	Decreasing
LVV	TBV
CSF	NABM/ICV
WML	Intensity
SubCSF	Kurtosis
Skewness	Integrity
MAD	–
MID	–

Table 10
Classification Performance metrics for various machine learning classifiers for sMCI vs cMCI classification problem.

Classifier	Accuracy	Sensitivity	Specificity
RFC	69 %	65 %	71 %
SVM	57 %	29 %	84 %
Logistic Regression	58 %	53 %	63 %
Gaussian Naive Bayes	59 %	51 %	66 %

to more indicative results of treatment trials. More specifically, the composite indices offer the ability to longitudinally monitor cognitively declining individuals using single-valued metrics that summarize the neuroanatomical state of the subject based on various dimensions of disease. The ZSI and ZSD composite indices are easily integrated into clinical workflow and can be used for patient selection in clinical trials.

A limitation of this study is that clinical variables such as lifestyle, socioeconomic status, and ethnicity were not available or included in the analysis. While the sex variable was available, due to sample sizes in the cMCI group, we were unable to further stratify the data. It is also worth noting that other papers (Table 8) used ADNI diagnostic labels to

Table 11
ANOVA and post-hoc analysis comparing the means of all 12 FLAIR MRI biomarkers and the ZSD and ZSI composite indices between cognitive classes. Cohen's *d* reported for each comparison to determine effect size. Biomarkers showing significant differences between sMCI-cMCI are bolded.

iomarker	NC-AD *p* (Cohen's *d*)	cMCI-AD *p* (Cohen's *d*)	sMCI-AD *p* (Cohen's *d*)	cMCI-NC *p* (Cohen's *d*)	sMCI-NC *p* (Cohen's *d*)	sMCI-cMCI *p* (Cohen's *d*)
TBV	<***0.001*** (*1.73*)	<***0.001*** (*0.81*)	<***0.001*** (*1.00*)	<***0.001*** (*0.78*)	<***0.001*** (*0.56*)	<***0.001*** (*0.20*)
CSF	<***0.001*** (*1.73*)	<***0.001*** (*0.81*)	<***0.001*** (*1.00*)	<***0.001*** (*0.78*)	<***0.001*** (*0.56*)	<***0.001*** (*0.20*)
WML	<***0.001*** (*0.60*)	*0.72* (*0.06*)	<***0.001*** (*0.29*)	<***0.001*** (*0.52*)	<***0.001*** (*0.27*)	<***0.05*** (*0.12*)
LVV	<***0.001*** (*1.37*)	<***0.05*** (*0.54*)	<***0.001*** (*0.78*)	<***0.001*** (*0.78*)	<***0.001*** (*0.43*)	<***0.001*** (*0.29*)
SubCSF	<***0.001*** (*1.50*)	<***0.001*** (*0.71*)	<***0.001*** (*0.87*)	<***0.001*** (*0.60*)	<***0.001*** (*0.49*)	<***0.05*** (*0.10*)
Integrity	<***0.001*** (*1.73*)	<***0.001*** (*0.81*)	<***0.001*** (*1.03*)	<***0.001*** (*0.75*)	<***0.001*** (*0.48*)	<***0.001*** (*0.22*)
MAD	<***0.001*** (*1.80*)	<***0.05*** (*0.85*)	<***0.001*** (*0.98*)	*0.07* (*0.84*)	*0.13* (*0.58*)	*0.12* (*0.21*)
MID	*0.59* (*1.13*)	*0.54* (*0.45*)	<***0.01*** (*0.62*)	*0.32* (*0.65*)	<***0.001*** (*0.40*)	*0.79* (*0.21*)
Intensity	<***0.001*** (*1.41*)	<***0.05*** (*0.56*)	<***0.001*** (*0.85*)	*0.06* (*0.71*)	<***0.001*** (*0.44*)	<***0.05*** (*0.24*)
Kurtosis	<***0.001*** (*1.22*)	<***0.01*** (*0.66*)	<***0.001*** (*0.79*)	<***0.001*** (*0.57*)	<***0.001*** (*0.38*)	<***0.01*** (*0.18*)
Skewness	<***0.001*** (*0.91*)	*0.30* (*0.21*)	<***0.001*** (*0.47*)	<***0.001*** (*0.65*)	<***0.001*** (*0.40*)	<***0.01*** (*0.23*)
NABM/ICV	<***0.001*** (*1.83*)	<***0.001*** (*0.86*)	<***0.001*** (*1.05*)	<***0.001*** (*0.85*)	<***0.001*** (*0.58*)	<***0.001*** (*0.22*)
ZSD	<***0.001*** (*1.77*)	<***0.001*** (*0.87*)	<***0.001*** (*1.05*)	<***0.001*** (*0.80*)	<***0.001*** (*0.53*)	<***0.001*** (*0.22*)
ZSI	<***0.001*** (*1.78*)	<***0.001*** (*0.73*)	<***0.001*** (*0.96*)	<***0.001*** (*0.90*)	<***0.001*** (*0.61*)	<***0.001*** (*0.23*)

Table 12
Results of ANOVA and post-hoc analysis comparing the regression slopes of all 12 FLAIR MRI biomarkers, and ZSD and ZSI composite indices. Cohen's *d* computed and shown to determine effect size. Biomarkers showing significant differences between sMCI-cMCI are bolded.

Biomarker	NC-AD *p* (Cohen's *d*)	cMCI-AD *p* (Cohen's *d*)	sMCI-AD *p* (Cohen's *d*)	cMCI-NC *p* (Cohen's *d*)	sMCI-NC *p* (Cohen's *d*)	sMCI-cMCI *p* (Cohen's *d*)
TBV	*0.107* (*1.67*)	*0.138* (*0.25*)	*0.226* (*1.33*)	<***0.01*** (*1.33*)	*0.421* (*1.00*)	<***0.05*** (*0.75*)
CSF	*0.107* (*1.67*)	*0.138* (*0.25*)	*0.226* (*1.33*)	<***0.01*** (*1.33*)	*0.421* (*1.00*)	<***0.05*** (*0.75*)
WML	*0.915* (*0*)	*0.231* (*0.50*)	*0.415* (*0.50*)	*0.113* (*0.50*)	*0.079* (*0.50*)	*0.488* (*0*)
LVV	<***0.01*** (*1.70*)	*0.283* (*0*)	<***0.001*** (*1.30*)	*0.476* (*1.70*)	*0.556* (*0.40*)	*0.411* (*0.65*)
SubCSF	*0.524* (*1.33*)	<***0.05*** (*0.33*)	*0.997* (*1.00*)	<***0.01*** (*1.33*)	*0.227* (*1.00*)	<***0.05*** (*0.67*)
Integrity	*0.08* (*1.38*)	*0.526* (*0.20*)	*0.153* (*0.87*)	<***0.05*** (*1.46*)	*0.502* (*0.77*)	*0.132* (*0.65*)
MAD	<***0.05*** (*1.45*)	*0.192* (*0.06*)	*0.195* (*0.94*)	<***0.001*** (*1.45*)	*0.104* (*1.21*)	<***0.05*** (*0.65*)
MID	<***0.05*** (*1.28*)	*0.122* (*0.32*)	*0.567* (*0.91*)	<***0.001*** (*1.28*)	<***0.01*** (*0.91*)	<***0.05*** (*0.40*)
Intensity	<***0.001*** (*1.96*)	*0.068* (*1.20*)	<***0.001*** (*1.15*)	*0.255* (*0.96*)	<***0.05*** (*1.33*)	*0.849* (*0.16*)
Kurtosis	<***0.001*** (*0.91*)	*0.764* (*0*)	<***0.01*** (*1.06*)	<***0.01*** (*1.11*)	<***0.01*** (*0.29*)	<***0.05*** (*0.81*)
Skewness	<***0.05*** (*0.17*)	*0.055* (*0.42*)	*0.958* (*0.27*)	<***0.001*** (*1.56*)	<***0.001*** (*2.30*)	<***0.05*** (*0.17*)
NABM/ICV	<***0.05*** (*1.67*)	*0.889* (*0*)	<***0.05*** (*1.33*)	<***0.05*** (*1.67*)	*0.834* (*1.00*)	*0.099* (*1.00*)
ZSD	<***0.01*** (*1.51*)	*0.464* (*0.22*)	<***0.05*** (*1.35*)	<***0.01*** (*1.53*)	*0.349* (*0.41*)	<***0.05*** (*1.51*)
ZSI	<***0.05*** (*1.62*)	<***0.05*** (*0.19*)	*0.457* (*1.07*)	<***0.001*** (*1.73*)	<***0.05*** (*1.32*)	<***0.05*** (*1.62*)

find cMCI subjects, whereas in this work, we used MoCA. Diagnostic labels are subjective with continually evolving criteria, and may not be collected all the time. In contrast, MoCA is widely used and could reduce subjectivity compared to a diagnostic label. The number of available cMCI subjects (n = 89) could also be considered a limitation in comparison to existing technologies but the training set used in this work contained data from three datasets (ADNI, CCNA, and ONDRI), which contrasts previous works which focus on a single dataset. Data variability in multicentre datasets can also be a challenge, and could have limited comparison between biomarkers. However, the intensity standardization pipeline helped to minimize this impact allowing for differences to been seen between groups cross-sectionally and longitudinally.

5. Conclusion

This work examines FLAIR-only biomarkers for the differentiation of sMCI and cMCI using cross-sectional, longitudinal and survival-based analysis in a large multicentre dementia cohort. Three types of biomarkers were examined, which included intensity, texture and volume. Cross-sectional biomarkers showed ANOVA differences across sMCI and cMCI groups, and were able to predict cognitive label using machine learning. The most important features were the volume-based features and the composite indices, although intensity and texture were found to improve performance. SHAP analysis also supported the volume biomarkers as being critical for detecting prodomal AD, but the composite indices were top performers. Longitudinally, FLAIR biomarkers modeled disease progression and there were differences in slope and intercepts across the cognitive groups for specific volume, intensity and texture biomarkers The composite indices had that highest effect size. The survival analysis showed that AD conversion can be predicted up to four years in advance, using the composite indices. This FLAIR-only biomarker system is attainable and accessible as FLAIR is an often acquired MRI sequence. The results are easily interpretable, thus making them easily integrated into the clinical workflow.

B Equations

$$ZScore = \frac{Biomar\mathrm{ker}_i - \mu_{NC}}{\sigma NC} \tag{1}$$

$$ZSI_{Cross\mathrm{Sectional}} = ZS_{LVV} + ZS_{CSF} + ZS_{SubCSF} + ZS_{WML} + ZS_{Skewness} \tag{2}$$

$$ZSD_{Cross\mathrm{Sectional}} = ZS_{TBV} + ZS_{NABM\_ICV} + ZS_{Intensity} + ZS_{Integrity} + ZS_{Kurtosis} \tag{3}$$

$$ZSI_{Longitudinal} = ZS_{CSF} + ZS_{SubCSF} + ZS_{Skewness} + ZS_{MAD} + ZS_{MID} \tag{4}$$

$$ZSD_{Longitudinal} = ZS_{TBV} + ZS_{Integrity} + ZS_{Kurtosis} \tag{5}$$

CRediT authorship contribution statement

Owen Crystal: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation, Supervision, Validation, Writing – review & editing, Formal analysis. **Pejman J. Maralani:** Conceptualization, Methodology, Validation. **Sandra Black:** Conceptualization, Methodology, Validation. **Corinne Fischer:** Conceptualization, Methodology, Validation, Writing – review & editing. **Alan R. Moody:** Conceptualization, Methodology, Validation.

Table 13
Results of ANOVA and post-hoc analysis comparing the regression intercepts of all 12 FLAIR MRI biomarkers, and ZSD and ZSI composite indices. Cohen's *d* computed and shown to determine effect size.

Biomarker	NC-AD	cMCI-AD	sMCI-AD	cMCI-NC	sMCI-NC	sMCI-cMCI
	***p* (Cohen's *d*)**	***p* (Cohen's *d*)**	***p* (Cohen's *d*)**	***p* (Cohen's *d*)**	***p* (Cohen's *d*)**	***p* (Cohen's *d*)**
TBV	$<$***0.001*** (20.00)	$<$***0.001*** (12.50)	$<$***0.001*** (13.33)	*0.36* (20)	$<$***0.001*** (3.33)	$<$***0.001*** (2.50)
CSF	$<$***0.001*** (20.00)	$<$***0.001*** (12.50)	$<$***0.001*** (13.33)	*0.36* (20)	$<$***0.001*** (3.33)	*0.25* (2.50)
WML	$<$***0.001*** (4.00)	$<$***0.001*** (2.00)	$<$***0.001*** (3.00)	*0.30* (2.00)	$<$***001*** (1.00)	*0.75* (1.00)
LVV	$<$***0.001*** (17.00)	$<$***0.001*** (14.00)	$<$***0.001*** (12.00)	*0.35* (3.00)	$<$***0.001*** (5.00)	*0.52* (1.00)
SubCSF	$<$***0.001*** (13.67)	$<$***0.001*** (12.00)	$<$***0.001*** (9.00)	*0.49* (2.50)	$<$***0.001*** (14.00)	*0.26* (2.14)
Integrity	$<$***0.001*** (17.89)	$<$***0.001*** (15.53)	$<$***0.001*** (10.76)	*0.33* (3.17)	$<$***0.001*** (11.38)	*0.45* (1.75)
MAD	$<$***0.001*** (18.23)	$<$***0.001*** (12.79)	$<$***0.001*** (12.78)	*0.28* (3.49)	$<$***0.001*** (13.17)	*0.37* (2.14)
MID	$<$***0.001*** (11.53)	$<$***0.001*** (6.80)	$<$***0.001*** (15.32)	*0.20* (4.13)	$<$***0.001*** (8.48)	*0.82* (5.35)
Intensity	$<$***0.001*** (14.32)	$<$***0.001*** (10.42)	$<$***0.001*** (4.53)	*0.14* (5.06)	$<$***0.001*** (46.27)	*0.94* (12.25)
Kurtosis	$<$***0.001*** (14.24)	$<$***0.001*** (7.84)	$<$***0.001*** (10.00)	*0.47* (2.59)	$<$***0.001*** (10.71)	*0.81* (1.90)
Skewness	$<$***0.001*** (10.00)	$<$***0.001*** (5.15)	$<$***0.001*** (6.27)	*0.20* (4.78)	$<$***0.001*** (8.20)	*0.96* (0.14)
NABM/ICV	$<$***0.001*** (19.33)	$<$***0.001*** (12.75)	$<$***0.001*** (12.33)	*0.43* (2.33)	$<$***0.001*** (21.00)	*0.19* (3.50)
ZSD	$<$***0.001*** (19.02)	$<$***0.001*** (13.49)	$<$***0.001*** (13.33)	*0.32* (3.33)	$<$***0.001*** (13.61)	*0.30* (2.42)
ZSI	$<$***0.001*** (18.15)	$<$***0.001*** (12.13)	$<$***0.001*** (12.25)	*0.18* (4.37)	$<$***0.001*** (14.11)	*0.42* (1.93)

April Khademi: Resources, Supervision, Project administration, Funding acquisition, Conceptualization, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC); the Government of Ontario; and the Alzheimer's Society Research Program. We would like to also thank the CCNA and ONDRI participants for the time, consent, and participation in our study, as well as the CCNA and ONDRI investigators that made this work possible.

References

Alzheimer's Association. (2023). Alzheimer's Disease Facts and Figures. *Alzheimers Dement*, 19(4), doi 10.1002/alz.13016.

Andersson, C., Blennow, K., Almkvist, O., Andreasen, N., Engfeldt, P., Johansson, S.E., Lindau, M., Eriksdotter-Jönhagen, M., 2008. Increasing CSF phospho-tau levels during cognitive decline and progression to dementia. Neurobiol. Aging 29 (10), 1466–1473. https://doi.org/10.1016/j.neurobiolaging.2007.03.027.

Badhwar, A., McFall, G. P., Sapkota, S., Black, S. E., Chertkow, H., Duchesne, S., Masellis, M., Li, L., Dixon, R. A., & Bellec, P. (2020). A multiomics approach to heterogeneity in Alzheimer's disease: Focused review and roadmap. *Brain*, 143(5), 1315–1331. https://doi.org/10. 1093/brain/awz384.

Bahsoun, M.A., Khan, M.U., Mitha, S., Ghazvanchahi, A., Khosravani, H., Jabehdar Maralani, P., Tardif, J.C., Moody, A.R., Tyrrell, P.N., Khademi, A., 2022. FLAIR MRI biomarkers of the normal appearing brain matter are related to cognition. Neuroimage Clin. 34, 102955.

Black, S., Gao, F., Bilbao, J., 2009. Understanding white matter disease: imaging-pathological correlations in vascular cognitive impairment. Stroke 40 (3), S48–S52. https://doi.org/10.1161/STROKEAHA.108.537704.

Brookmeyer, R., Johnson, E., Ziegler-Graham, K., Arrighi, H.M., 2007. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement. 3 (3), 186–191.

Brugulat-Serrat, A., Salvado, G., Operto, G., Cacciaglia, R., Sudre, C., Grau-Rivera, O., Suarez-Calvet, M., Falcon, C., Sanchez-Benavides, G., Gramunt, N., et al., 2020. White matter hyperintensities mediate gray matter volume and processing speed relationship in cognitively unimpaired participants. Human Brain Mapping 41, 1309–1322.

Caminiti, S.P., et al., 2018. FDG-PET and CSF biomarker accuracy in prediction of conversion to different dementias in a large multicentre MCI cohort. NeuroImage. Clinical 18, 167–177.

Chan, K., Crystal, O., Khosravani, H., Black, S., Tyrrell, P., Maralani, P., Moody, A.R., Khademi, A., 2022. Texture Biomarkers of the Normal Appearing Brain Matter in FLAIR are Related to Cognition. American Society of Neuroradiology (ASNR) Annual Meeting.

Chan, K., Fischer, C., Khosravani, H., Black, S.E., Tyrrell, P., Jabehdar Maralani, P., Moody, A.R., & Khademi, A. (2022). FLAIR MRI Biomarkers are related to APOE-4 status and Aβ42 and Tau CSF markers. *Alzheimer's Association International Conference (AAIC)*.

Chan, K., Fischer, C., Maralani, P., Black, S., Moody, A.R., Khademi, A., 2023. 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.

Chertkow, H., Borrie, M., Whitehead, V., Black, S.E., Feldman, H.H., Gauthier, S., Hogan, D.B., Masellis, M., McGilton, K., Rockwood, K., 2019. The comprehensive assessment of neurodegeneration and dementia: canadian cohort study. Canadian Journal of Neurological Sciences 46 (5), 499–511.

Cho, Y., Seong, J.K., Jeong, Y., Shin, S.Y., 2012. Alzheimer's disease neuroimaging initiative. individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage 59 (3), 2217–2230.

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., Benali, H., Garnero, L., Colliot, O., 2009. Alzheimer's disease neuroimaging initiative. fully automatic hippocampus seg- mentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus 19 (6), 579–587.

Crystal, O., Chan, K., Khosravani, H., Black, S., Tyrrell, P., Maralani, P., Moody, A.R., Khademi, A., 2022. Characterizing Dementia using Cross-Sectional Biomarkers from Multicenter FLAIR MRI. American Society of Neuroradiology (ASNR) Annual Meeting.

Crystal, O., Maralani, P., Black, S., Moody, A.R., & Khademi, A. (2023). Domain Adaptation using Silver Standard Masks for Lateral Ventricle Segmentation in FLAIR MRI. *Proceedings of Machine Learning Research*, Accepted.

Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q., 2011. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol Aging 32(12):2322, e19–e27.

Davatzikos, C., Bhatt, P., ShawbcKayhan, L.M., Batmanghelich, K.N., Trojanowski, J.Q., 2011. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiology of Aging 32 (12), 2322.e19–2322.e27.

Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., & Trojanowski, J.Q. (2011). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging*, 32(12), 2322.e19-27.

DeTure, M.A., Dickson, D.W., 2019. The neuropathological diagnosis of Alzheimer's disease. Mol Neurodegeneration 14, 32. https://doi.org/10.1186/s13024-019-0333-5.

DiGregorio, J., Arezza, G., Gibicar, A., Moody, A.R., Tyrrell, P.N., Khademi, A., 2021. Intracranial volume segmentation for neurodegenerative populations using multicentre flair mri. Neuroimage: Reports 1 (1), 100006.

DiGregorio, J., Gibicar, A., Khosravani, Jabehdar Maralani, P., Tardif, J.-C., Tyrrell, P.N., A.R.Moody, A.R., & Khademi, A. (2022). Cross-sectional and longitudinal Biomarker extraction and analysis for multicentre FLAIR brain MRI, *Neuroimage: Reports*, 2(2), 100091.

Duc, N.T., Ryu, S., Qureshi, M.N.I., Choi, M., Lee, K.H., Lee, B., 2020. 3D-deep learning based automatic diagnosis of Alzheimer's disease with Joint MMSE prediction using resting-state fMRI. Neuroinformatics. 71–86.

Eskildsen, S.F., Coupé, P., García-Lorenzo, D., Fonov, V., Pruessner, J.C., Collins, D.L., 2013. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. Neuroimage 65, 511–521.

Farhan, S., Bartha, R., Black, S., Corbett, D., Finger, E., Freedman, M., Strong, M., 2017. The ontario neu- rodegenerative disease research initiative (ondri). Canadian Journal of Neurological Sciences 44 (2), 196–202.

Frisoni, G.B., Fox, N.C., Jack Jr, C.R., Scheltens, P., Thompson, P.M., 2010. The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 6 (2), 67–77.

Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H., 2013. Alzheimer's disease neuroimag- ing initiative. brainage in mild cognitive impaired patients: predicting the conversion to alzheimer's disease. PLoS One 8 (6), e67346.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., et al., 2006. Mild cognitive impairment. Lancet 367 (9518), 1262–1270.

Hansen, T., Brezova, V., Eikenes, L., Håberg, A., Vangberg, T., 2015. How does the accuracy of intracranial volume measurements affect normalized brain volumes? sample size estimates based on 966 subjects from the HUNT MRI cohort. American Journal of Neuroradiology 36 (8), 1450–1456.

Hesse, C., Rosengren, L., Andreasen, N., Davidsson, P., Vanderstichele, H., Vanmechelen, E., Blennow, K., 2001. Transient increase in total tau but not phospho-tau in human cerebrospinal fluid after acute stroke. Neurosci Lett 297 (3), 187–190.

Inglese, M., Patel, N., Linton-Reid, K., Loreto, F., Win, Z., Perry, R.J., Carswell, C., Grech-Sollars, M., Crum, W.R., Lu, H., Malhotra, P.A., 2022. Alzheimer's disease neuroimaging initiative, Aboagye EO. a predictive model using the mesoscopic architecture of the living brain to detect Alzheimer's disease. Commun Med (Lond) 2 (70), 1–16.

Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L. Whitwell, J., & Ward, C. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4), 685–691.

Jack Jr, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 9 (1), 119–128.

Jo, T., Nho, K., Saykin, A.J., 2019. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Front Aging Neurosci 11, 220.

Khademi, A., Hosseinzadeh, D., Venetsanopoulos, A., & Moody, A. (2009) Nonparametric statistical tests for exploration of correlation and nonstationarity in images. *16th International Conference on Digital Signal Processing*, 1–6.

Khademi, A., Gibicar, A., Arezza, G., DiGregorio, J., Tyrrell, P.N., Moody, A., 2021. Segmentation of white matter lesions in multicentre FLAIR MRI. Neuroimage: Reports 1 (4), 1–29.

Korolev, I.O., Symonds, L.L., Bozoki, A.C., 2016. Alzheimer's Disease Neuroimaging Initiative. Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and plasma biomarkers via probabilistic pattern classification. PLoS One 11 (2), e0138866.

Lee, G., Nho, K., Kang, B., Sohn, K.A., Kim, D. (2019). for Alzheimer's Disease Neuroimaging Initiative. Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Sci Rep*. 9(1), 1952.

Li, H., Habes, M., Wolk, D.A., Fan, Y., 2019. Alzheimer's disease neuroimaging initiative and the australian imaging biomarkers and lifestyle study of aging. a deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. Alzheimers Dement. 15 (8), 1059–1070.

Lundberg, S., & Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. number: arXiv:1705.07874 arXiv:1705.07874 [cs, stat]. [Online].

Marinescu, R.V., Bron, E.E., Oxtoby, N.P., Young, A.L., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Golland, P., Klein, S., Alexander, D.C., 2020. Predicting Alzheimer's disease progression: Results from the TADPOLE Challenge. Alzheimer's Dement. 16, e039538.

Mayeux, R., Stern, Y., 2012. Epidemiology of Alzheimer disease. Cold Spring Harbor Perspectives in Medicine 2 (8), 1–18.

Misra, C., Fan, Y., Davatzikos, C., 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. Neuroimage 15, 44(4), 1415–1422.

Mohaddes, Z., Das, S., Abou-Haidar, R., Safi-Harab, M., Blader, D., Callegaro, J., Henri-Bellemare, C., Tunteng, J.-F., Evans, L., Campbell, T., 2018. National neuroinformatics framework for canadian consortium on neurodegeneration in aging (CCNA). Frontiers in Neuroinformatics 12, 85.

Moscoso, A., Silva-Rodríguez, J., Aldrey, J.M., Cortés, J., Fernández-Ferreiro, A., Gómez-Lado, N., Ruibal, Á., Aguiar, P., 2019. Prediction of Alzheimer's disease dementia with MRI beyond the short- term: Implications for the design of predictive models. Neuroimage Clin 23, 1–9.

Noh, Y., Jeon, S., Lee, J.M., Seo, S.W., Kim, G.H., Cho, H., Ye, B.S., Yoon, C.W., Kim, H.J., Chin, J., Park, K.H., 2014. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. Neurology 83 (21), 1936–1944. https://doi.org/10.1212/wnl.0000000000001003.

Nohara, Y., Matsumoto, K., Soejima, H., Nakashima, N., 2022. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Computer Methods and Programs in Biomedicine 214, 106584.

Nordenskjöld, R., Malmberg, F., Larsson, E.-M., Simmons, A., Brooks, S.J., Lind, L., Ahlström, H., Johansson, L., Kullberg, J., 2013. Intracranial volume estimated with commonly used methods could introduce bias in studies including brain volume measurements. Neuroimage 83, 355–360.

Palmqvist, S., Hertze, J., Minthon, L., Wattmo, C., Zetterberg, H., Blennow, K., et al., 2012. Com- parison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: six-year follow-up Study. PLoS ONE 7 (6), e38639.

Petersen, R., 2004. Mild cognitive impairment as a diagnostic entity. J. Intern. Med. 256, 183–194.

Poulakis, K., Pereira, J.B., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soinin-nen, H., Lovestone, S., Simmons, A., Wahlund, L.O., Westman, E., 2018. Heterogeneous pat- terns of brain atrophy in Alzheimer's disease. Neurobiology of Aging 65, 98–108. https://doi.org/10.1016/j.neurobiolaging.2018.01.009.

Reiche, B., Moody, A., Khademi, A., 2019. Pathology-preserving intensity standardization frame- work for multi-institutional FLAIR MRI datasets. Magnetic Resonance Imaging 62, 59–69.

Rocca, M., Bataaglini, M., Benedict, R., De Stefano, N., Geurts, J., Henry, R., Horsfield, M., Jenkinson, M., Pagani, E., Filippi, M., 2016. Brain MRI atrophy quantification in MS. Neurology 403–413.

Schoonenboom, N.S., Reesink, F.E., Verwey, N.A., Kester, M.I., Teunissen, C.E., van de Ven, P.M., Pijnenburg, Y.A., Blankenstein, M.A., Rozemuller, A.J., Scheltens, P., van der Flier, W.M., 2012. Cerebrospinal fluid markers for differential dementia diagnosis in a large memory clinic cohort. Neurology 78 (1), 47–54.

Shi, R., Zhang, L., Han, Y., Sheng, C., Jin, S., Zhang, Q., Zhang, S., Ding, C., Wang, L., Jiehui Jiang, J., 2022. Generative adversarial network constrained multiple loss autoencoder: a deep learning- based individual atrophy detection for Alzheimer's disease and mild cognitive impairment. Hum Brain Mapp 44 (3), 1129–1146. https://doi.org/10.1002/hbm.26146.

Sørensen, L., Pai, A., Anker, C., Balas, I., Lillholm, M., Igel, C., Nielsen, M., 2014. Dementia diag- nosis using MRI cortical thickness, shape, texture, and volumetry. In Proc MICCAI workshop challenge on computer-aided diagnosis of dementia based on structural MRI data 111–118.

Spasov, S., Passamonti, L., Duggento, A., Liò, P., Toschi, N., 2019. Alzheimer's disease neuroimag- ing initiative.a parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. Neuroimage 189, 276–287.

Thung, K., Yap, P., Adeli, E., Lee, S., Shen, D., 2018. Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion. Medical Image Analysis 45, 68–82.

Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D., 2017. A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's Disease. IEEE Trans Biomed Eng. 64 (1), 155–165.

Vemuri, P., Wiste, H.J., Weigand, S.D., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Knopman, D.S., Petersen, R.C., Jack Jr, C.R., 2009. Alzheimer's disease neuroimaging initiative. MRI and CSF biomarkers in normal, MCI, and AD subjects: diagnostic discrimination and cognitive correlations. Neurology. 73 (4), 287–293.

Vemuri, P., Wiste, H.J., Weigand, S.D., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Knopman, D.S., Petersen, R.C., Jack Jr, C.R., 2009. Alzheimer's disease neuroimaging initiative. MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. Neurology. 73 (4), 294–301.

Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D.P., Rueckert, D., Soininen, H., & Lötjönen, J. (2011). Alzheimer's Disease Neuroimaging Initiative. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. *PLoS One*, 2011, 6(10), e25446.

Xu, L., Wu, X., Li, R., Chen, K., Long, Z., Zhang, J., Guo, X., Yao, L., 2016. Alzheimer's disease neuroimaging initiative. prediction of progressive mild cognitive impairment by multi-modal neuroimag- ing biomarkers. J Alzheimers Dis 51 (4), 1045–1056.