26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Discovery of MRI Biomarker for Early Alzheimer's Disease Integrating Regional Atrophy and Ventricular Expansion

I. Aberathne[a], D. Kulasiri[a], S. Samarasinghe[a], and J.Zhang[b]

[a] Centre for Advanced Computational Solutions (C-fACS), Lincoln University, Christchurch, New Zealand
[b] College of Life Sciences, Peking University, Beijing, China
Email: iroshan.aberathne@lincolnuni.ac.nz

Abstract: Timely detection of neurodegeneration in Alzheimer's disease, particularly during the early and late mild cognitive impairment stages (EMCI and LMCI, respectively), is crucial for effective clinical intervention and disease management. Despite widespread use of absolute and normalised brain or ventricular volumes as imaging biomarkers, the relative expansion of ventricles per unit of brain tissue remains underexplored. We introduce the Ventricle-to-Tissue Ratio (VTR), and intuitive biomarker capturing this ratio, and define its global (VTR_G) and estimated (VTR_E) forms. Through machine learning analyses, we demonstrate that VTR_E robustly approximates VTR_G, accurately tracks age and group related atrophic changes, and exhibits superior discriminative performance in distinguishing EMCI from LMCI when compared to alternative composite features. Moreover, explainable AI techniques confirm VTR_E's dominant and interpretable role in model predictions. Together, these findings establish VTR_E as a sensitive, scalable, and clinically interpretable biomarker for early neurodegeneration, offering significant translational value for Alzheimer's disease diagnostics and research.

Keywords: *Early Alzheimer's disease, biomarker, machine learning, explainable ai, brain MRI*

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder marked by structural brain changes that begin in its prodromal stages, such as early and late mild cognitive impairment (EMCI, LMCI) (Wu et al., 2021). Volumetric magnetic resonance imaging (MRI) biomarkers have become essential for detecting and tracking these changes, with ventricular enlargement and regional brain tissue atrophy among the earliest and most consistent features observed (Cho et al., 2012). These affected regions are critical to memory and cognition, and their deterioration contributes directly to the clinical symptoms of cognitive decline. Despite the utility of absolute brain and ventricular volume measurements, their interpretability in early disease detection is limited by individual differences in brain and head size, which can obscure subtle yet clinically significant changes. To address this limitation, ratio-based measures, particularly those normalizing ventricular volume to brain or intracranial volume offer anatomically sensitive, individualized metrics that better control for inter-subject variability (Colliot et al., 2008). These normalized ratios have shown stronger correlations with cognitive impairment and AD pathology than absolute volumes, underscoring their promise as more accurate imaging biomarkers that capture both structural degeneration and its functional consequences (Carmichael et al., 2007).

In pursuit of more refined biomarkers for neurodegeneration, we introduce the Ventricle-to-Tissue Ratio (VTR) metric designed to quantify brain atrophy by balancing regional ventricular expansion against regional brain tissue loss. Our findings reveal that VTR values increase with age and disease progression, with consistently higher ratios observed in individuals with LMCI compared to those with EMCI. Building on this, we developed an estimated VTR (VTR_E) using a minimal, yet critical feature set identified through robust algorithmic feature selection. This anatomically grounded measure incorporates volumes from key ventricular structures such as the lateral-ventricle, inferior-lateral ventricle, and 3rd-ventricle alongside brain regions vulnerable to early AD neurodegeneration, notably the hippocampus and amygdala. VTR_E demonstrated superior predictive accuracy for estimating the global VTR (VTR_G) and significantly enhanced group discrimination between EMCI and LMCI. Explainability analyses using Local Interpretable Model-agnostic eXplaination (LIME) and SHaply Additive exPlanation (SHAP) (Sathyan et al., 2022) further confirmed VTR_E as an interpretable and dominant contributor to VTR_G prediction. Together, these results position VTR_E as a sensitive, scalable, and biologically meaningful biomarker with strong potential for improving early detection and monitoring of neurodegenerative progression in mild cognitive impairment (MCI).

The motivation and rationale behind this study are outlined in Section 2. The data, methodology, and implementation details for developing and validating the VTR are provided in Section 3. Experimental evaluations and key results are discussed in Section 4, followed by concluding remarks and implications in Section 5.

2. BACKGROUND AND MOTIVATION

MRI biomarkers are pivotal in detecting and characterizing structural brain changes associated with AD and its early stages. Hallmark features like ventricular enlargement and gray matter atrophy emerge early and intensify as the disease progresses, with ventricular expansion largely reflecting passive anatomical changes due to localized tissue loss (Carmichael et al., 2007, Colliot et al., 2008). These changes follow a region specific, sequential pattern, with ventricular volume increasing from normal aging through EMCI to AD. While volumetric MRI measures are sensitive indicators of neurodegeneration, their clinical interpretation especially in early stages is complicated by individual variability in brain anatomy and head size (Schuff et al., 2009). This variability means identical absolute volumes may reflect different levels of atrophy or disease severity across individuals, potentially masking subtle changes crucial for early diagnosis and reducing the sensitivity and specificity of these biomarkers in diverse populations.

To overcome these limitations and enhance the early identification of AD, there is strong motivation to adopt ratio-based volumetric biomarkers. By normalizing ventricular volume to the whole brain or intracranial volume through ventricle-to-brain or ventricle-to-intracranial volume ratios (Bartos et al., 2019, Colliot et al., 2008), these metrics provide a personalized reference that accounts for anatomical variability. This approach allows for more accurate detection of disease-related brain changes in the context of each individual's unique cranial structure, even when absolute volume losses are still modest. Moreover, research has shown that these ratio-based measures correlate more robustly with cognitive impairment and AD pathology than absolute measures alone, making them particularly valuable for distinguishing between early clinical stages and for tracking disease progression (Bartos et al., 2019, Colliot et al., 2008, Todd et al., 2018). Therefore, ratio-based volumetric biomarkers arise from the need to improve the precision and sensitivity of early AD detection, enabling clinicians and researchers to reliably differentiate pathological brain changes from normal variation in brain size thereby strengthening the role of MRI markers in early diagnosis and intervention.

3. MATERIALS AND METHODS

The primary objective of this study is to identify anatomically specific and computationally reliable volumetric biomarkers for estimating the whole-brain/global Ventricular-to-Tissue Ratio (VTR_G) in individuals with MCI. To this end, we introduce and evaluate a regionally grounded metric VTR_E which captures the relative expansion of ventricular volume to the decline in brain tissue volume. A central aim is to assess how accurately this anatomically guided VTR_E can approximate the global VTR_G measure, thereby determining its potential utility as a sensitive biomarker for whole-brain atrophy.

3.1. Ventricle-to-Tissue-Ratio (VTR)

We define a ventricle-to-tissue-ratio (VTR) to quantify the degree of brain atrophy in (1) where v is total ventricular volume and x is total brain volume.

$$VTR = \frac{v}{x} \tag{1}$$

A near zero implies a small ventricular volume relative to brain tissue, typically representing a healthy brain, where $v \to 0$ and x remains high. Conversely, as brain atrophy progresses, tissue volume x decreases while ventricular volume v increases. In advanced stages, where $x \to 0$, the VTR becomes large, indicating severe atrophy as defined in (2).

$$\lim_{x \to 0} \frac{v}{x} \to \infty \tag{2}$$

At early stages (EMCI and LMCI), the brain tissue volume x begins to decline, but not drastically. The resulting moderate increase in v leads to a finite, gradually increasing VTR, making it a sensitive maker for early atrophy changes. To incorporate anatomical specificity, we redefine v and x as the sum of regional volumes as $v = \sum_{i=1}^{n} \alpha_i$, and $x = \sum_{j=1}^{n} \beta_j$ where α_i and β_j are prominent ventricular expansion and atrophic brain region respectively.

3.2. Data Preparation

Data were obtained from the ADNI2/GO cohort of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu), which focuses on identifying early biomarkers of cognitive decline (Diogo et al., 2022). We selected participants aged 65–75 and used T1-weighted MRI scans to evaluate structural brain changes. The dataset finally comprised 112 subjects: 61 with EMCI and 51 with LMCI, including 58 males and 54 females. MRI images were processed using FreeSurfer (v7.2.0-centos7_x86_64; https://surfer.nmr.mgh.harvard.edu), on the New Zealand eScience Infrastructure (NeSI; https://www.nesi.org.nz). Brain region volumes were extracted in mm^3.

To validate the VTR metric, we first evaluated candidate measurements for x and v over VTR_E. The VTR_G can be computed directly from two existing measures: BrainSegVol and BrainSegVolNotVent, as in (3). Here, BrainSegVol represents the total brain volume including ventricle, while BrainSegVolNotVent excludes ventricular regions, reflecting only the brain tissue.

$$VTR_G = \frac{BrainSegVol - BrainSegNotVent}{BrainSegNotVent} \tag{3}$$

However, identifying appropriate candidates for VTR_E is less straightforward, particularly for selecting the x. Therefore, we considered the following brain tissue volume candidates for x : TotalGrayVol, and CerebralWhiteMatterVol. Each of these captures distinct anatomical structures and may reflect different aspects of age-related brain atrophy. The TotalGrayVol consists of eight subcortical structures: *Thalamus*, *Caudate*, *Putamen*, *Pallidum*, *Hippocampus*, *Amygdala*, *Accumbens-area*, and *VentralDC*. The CerebralWhiteMatterVol measure encompasses regions such as the *Optic-Chiasm*, Corpus-Callosum subdivisions *(CC_Posterior, CC_Mid_Posterior, CC_Central, CC_Mid_Anterior, CC_Anterior)*, while the BrainSegVolNotVent includes all regions from TotalGrayVol and CerebralWhiteMatterVol, in addition to *Cerebellum-White-Matter, Cerebellum-Cortex,* and *Brain-Stem*. The ventricular volume corresponds to six anatomically established ventricular structures: *Lateral-Ventricle*, *Inferior-Lateral-Ventricle*, *3rd-Ventricle*, *4th-Ventricle*, *5th-Ventricle*, and *Cerebrospinal Fluid (CSF)*.

To systematically identify the most informative regions for VTR_E, we applied Prominent Feature Selection (PFS) algorithm (Aberathne et al., 2025), optimizing its hyperparameters for robust feature selection from initial pool of features which included all TotalGrayVol, CerebralWhiteMatterVol, and additional brain regions from BrainSegVolNotVent for brain tissue and all ventricular volumes. Among these, five features have been identified by PFS as prominent including three ventricular (lateral-ventricle, inferior-lateral-ventricle, and 3rd-ventricle) and two brain tissue (hippocampus, and amygdala) regions. To further assess the relevance of PFS selected features,

we analyse volume change rate for different brain tissue measures between the ages 65 and 75, as summarised in Table 1.

Table 1. Brain tissue volume at age 65 and 75 in mm^3. Gray and white matter volume decrease with age, with gray matter showing the greatest decline indicating more pronounced atrophy.

Tissue Volume	Mean Volume at age 65	Mean Volume at age 75	Rate of Change
TotalGrayVol	0.3864	0.3691	4.47%
CerebralWhiteMatterVol	0.2939	0.2898	1.39%
TotalGrayVol + CerebralWhiteMatterVol	0.6804	0.6590	3.14%

The combined TotalGrayVol and CerebralWhiteMatterVol exhibit a moderate decline (3.14%), closely aligning with the 4.47% reduction observed in gray matter alone. This finding supports the relevance of features selected by PFS, particularly those from TotalGrayVol, indicating that these regions are well-suited for capturing meaningful structural brain changes. Therefore, the feature vector is thus composed of a combination of these selected features for VTR_E as characterised in (4).

$$VTR_E = \frac{Lateral\ Ventricle + Inferior\ Lateral\ Ventricle + 3rd\ Ventricle}{Hippocampus + Amygdala} \tag{4}$$

4. EXPERIMENTAL RESULTS AND FINDINGS

To better understand brain tissue-to-ventricular relationships in early cognitive decline, we first examined the behaviour of VTR_G across EMCI and LMCI groups as illustrated in Figure 1.

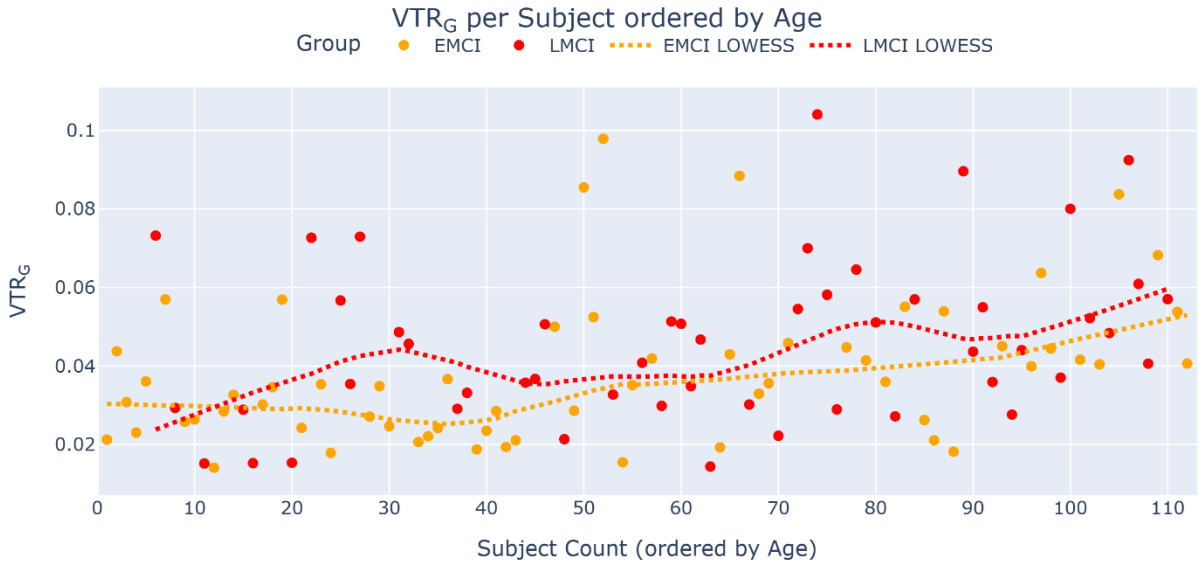

Figure 1. Variation of VTR_G with age in EMCI and LMCI groups. VTR_G values are shown for 112 subjects aged 65–75, with individuals arranged left to right by increasing age along the x-axis. LOWESS lines for each group depict smoothed age-related trends in VTR_G, highlighting atrophy progression in either group.

The VTR_G demonstrates strong potential as shown in Figure 1, where an upward trend is evident with increasing age for both EMCI and LMCI groups. Notably, VTR_G values are consistently higher in the LMCI group compared to EMCI, which served as the primary motivation for exploring VTR_E as a proxy by leveraging a minimal set of features which could efficiently estimate VTR_G.

4.1. Performance Evaluation of VTR_E

To assess the predictive utility of VTR_E, we performed a regression analysis using the Random Forest Regressor (RFR). RFR was selected due to its proven ability to model complex, non-linear relationships and effectively capture interactions between features without requiring extensive hyperparameter tuning. Its ensemble-based structure also provides robustness against overfitting, which is particularly beneficial for small datasets.

We constructed four alternative composite features in addition to VTR_E, by recombining the ventricular and brain tissue regions that were excluded from VTR_E. The excluded ventricular regions are defined as $\alpha_E = CSF +$

$4th\ Ventricle + 5th\ Ventricle$. The excluded brain tissue regions are grouped into three categories: gray matter → $\beta_{GE} = Caudate + Putamen + Pallidum + Thalamus + Accumbens\ area + VentralDC$, white matter → $\beta_{WE} = Optic\ Chiasm + CC\ Posterior + CC\ Mid\ Posterior + CC\ Mid\ Anterior$, and BrainSegNotVent → $\beta_{BNVE} = \beta_{GE} + \beta_{WE} + Brain\ Stem + Cerebellum\ White\ Matter + Cerebellum\ Cortex$. These groupings formed the basis for the four alternative composite features as $VTR_{GM} = \frac{\alpha_E}{\beta_{GE}}$, $VTR_{WM} = \frac{\alpha_E}{\beta_{WE}}$, $VTR_{GWM} = \frac{\alpha_E}{\beta_{GE} + \beta_{WE}}$, and $VTR_{BNV} = \frac{\alpha_E}{\beta_{BNVE}}$. The prediction results presented in Table 2 indicate that the VTR_E achieved the highest predictive accuracy across all groups. In comparison, the four alternative feature sets demonstrated significantly lower performance. These findings indicate the robustness and effectiveness of VTR_E for predicting VTR_G.

Table 2. Comparison of predictive power of VTR_E. VTR_E consistently shows the highest values indicating superior sensitivity and reliability as a biomarker. Its higher performance across groups suggests VTR_E better captures disease-related changes compared to other features.

Feature set	EMCI (%)	LMCI (%)	EMCI + LMCI (%)
VTR_E	65.37	53.78	62.34
VTR_{GM}	34.24	26.28	33.44
VTR_{WM}	39.09	16.03	40.67
VTR_{GWM}	35.72	27.43	35.77
VTR_{BNV}	32.16	37.58	45.88

4.2. Explainability Power of VTR_E

To assess the explainability of VTR_E in predicting VTR_G, we employed two model-agnostic interpretation methods: LIME and SHAP. LIME was used to analyse local model behaviour, revealing how variations in VTR_E influence individual predictions of VTR_G by approximating the model with a locally linear surrogate. In contrast, SHAP provided a global perspective by computing feature contribution scores across the entire dataset, quantifying how much VTR_E shifts the predicted values relative to the model's baseline. We created another feature set containing VTR_E and the next highest correlated feature towards the VTR_G from ventricular regions, gray matter, white matter and BrainSegNotVent regions and performed the LIME and SHAP evaluation.

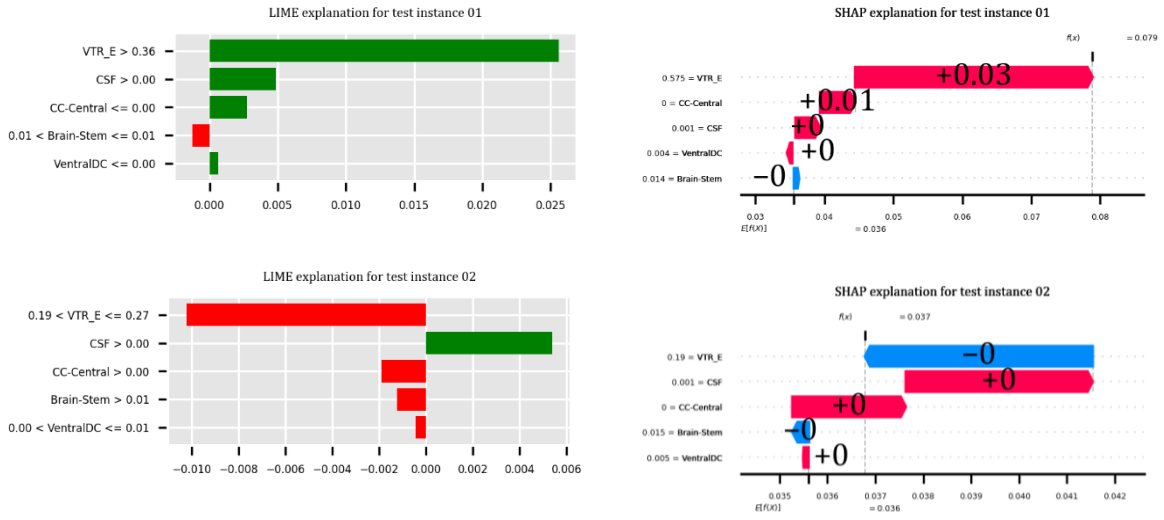

Figure 2. Quantifying VTR_E's contribution to model prediction in EMCI subjects. LIME and SHAP highlight VTR_E as the dominant feature in predicting an EMCI subject. LIME offers intuitive local explanations, while SHAP provides a detailed quantitative breakdown, together emphasizing VTR_E's key role in enhancing robust model interpretability.

Figure 2, and Figure 3 illustrate the feature attribution results for selected test instances using LIME and SHAP. In both EMCI and LMCI groups, VTR_E emerged as the dominant contributor, consistently showing the highest influence on the model's prediction across explanations. These local interpretability methods independently highlight the significance of VTR_E. The consistency across methods and diagnostic groups reinforces the reliability of VTR_E as an interpretable and meaningful biomarker at the early stage of AD.

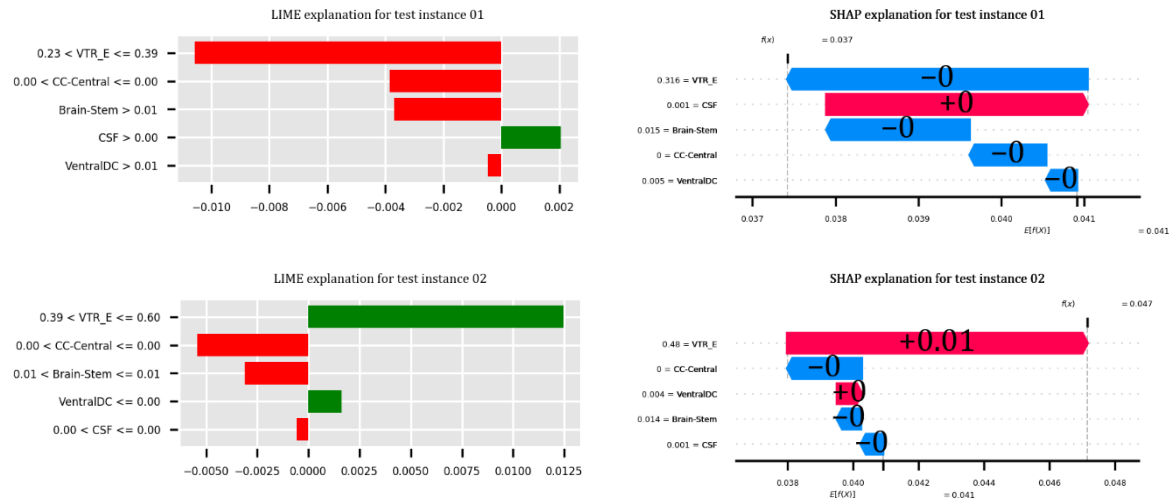

Figure 3. Quantifying VTR_E's contribution to model prediction in LMCI subjects. Even in LMCI subjects VTR_E dominates the model prediction explainability. This consistency across stages highlights VTR_E's robust importance.

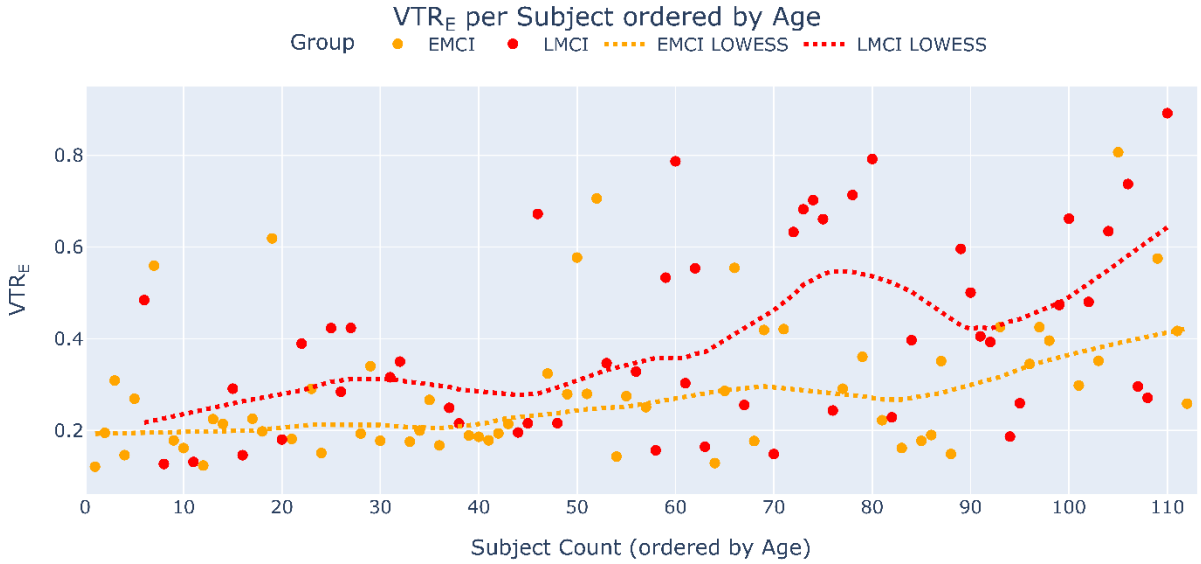

Figure 4. Evaluating VTR_E as a surrogate estimator for VTR_G. VTR_E values for 112 subjects aged 65-75, ordered by age on the x-axis. LOWESS trend lines highlight group-specific patterns, revealing clear differences in VTR_E variation and progression with age.

As illustrated in Figure 4, VTR_E not only captures the same rising trend of VTR_G for both EMCI and LMCI, but also provides enhanced discriminative power evidenced by the greater separation between the EMCI and LMCI trend lines relative to VTR_G seen in Figure 1. This suggests that VTR_E may serve as a more robust and practical marker for distinguishing between the groups and progress estimation.

5. CONCLUSION

This study validates the Ventricle-to-Tissue Ratio (VTR) as a sensitive and reliable biomarker at MCI. VTR_E emerged as a particularly efficient and interpretable proxy. Derived from a minimal yet anatomically meaningful set of features, VTR_E maintains strong predictive power and group discrimination while enhancing model explainability. These findings support VTR_E's potential as a scalable and clinically useful biomarker for early detection and monitoring of neurodegeneration in the context of Alzheimer's disease.

ACKNOWLEDGEMENT

This project was supported by the Alzheimer's Disease Neuroimaging Initiative (ADNI) for MRI data and by the New Zealand eScience Infrastructure (NeSI) for high-performance computing resources.

REFERECES

Aberathne, I., Kulasiri, D., Samarasinghe, S., 2025. Vulnerable brain region identification in early stage of Alzheimer's disease (AD) using magnetic resonance images (MRI): a novel unsupervised feature selection algorithm. Proceedings of ISBM 2025, (in press).

Bartos, A., Gregus, D., Ibrahim, I., Tintěra, J., 2019. Brain volumes and their ratios in Alzheimer's disease on magnetic resonance imaging segmented using Freesurfer 6.0. Psychiatry Research: Neuroimaging 287, 70–74. DOI: https://doi.org/10.1016/j.pscychresns.2019.01.014.

Carmichael, O.T., Kuller, L.H., Lopez, O.L., Thompson, P.M., Dutton, R.A., Lu, A., Becker, J.T., 2007. Ventricular volume and dementia progression in the Cardiovascular Health Study. Neurobiology of Aging 28(3), 389–397. DOI: https://doi.org/10.1016/j.neurobiolaging.2006.01.006.

Cho, Y., Seong, J.K., Jeong, Y., Shin, S.Y., Alzheimer's Disease Neuroimaging Initiative, 2012. 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. DOI: https://doi.org/10.1016/j.neuroimage.2011.09.085

Colliot, O., Chételat, G., Chupin, M., Desgranges, B., Magnin, B., Benali, H., Lehéricy, S., 2008. Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. Radiology 248(1), 194–201. DOI: https://doi.org/10.1148/radiol.2481070876.

Diogo, V.S., Ferreira, H.A., Prata, D., Alzheimer's Disease Neuroimaging Initiative, 2022. Early diagnosis of Alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach. Alzheimer's Research & Therapy 14(1), 107. DOI: https://doi.org/10.1186/s13195-022-01047-y.

Sathyan, A., Weinberg, A.I., Cohen, K., 2022. Interpretable AI for bio-medical applications. Complex Engineering Systems 2(4), 18. DOI: https://doi.org/10.20517/ces.2022.41

Schuff, N., Woerner, N., Boreta, L., Kornfield, T., Shaw, L.M., Trojanowski, J.Q., Alzheimer's Disease Neuroimaging Initiative, 2009. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain 132(4), 1067–1077. DOI: https://doi.org/10.1093/brain/awp007.

Todd, K.L., Brighton, T., Norton, E.S., Schick, S., Elkins, W., Pletnikova, O., Alzheimer's Disease Neuroimaging Initiative, 2018. Ventricular and periventricular anomalies in the aging and cognitively impaired brain. Frontiers in Aging Neuroscience 9, 445. DOI: https://doi.org/10.3389/fnagi.2017.00445.

Wu, Z., Peng, Y., Hong, M., Zhang, Y., 2021. Gray matter deterioration pattern during Alzheimer's disease progression: a regions-of-interest based surface morphometry study. Frontiers in Aging Neuroscience 13, 593898. DOI: https://doi.org/10.3389/fnagi.2021.593898.

