Computers in Biology and Medicine 196 (2025) 110829

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

SMAS: Structural MRI-based AD Score using Bayesian supervised VAE

A. Nemali [1,2,*,a], J. Bernal [1,2], R. Yakupov [1,2], D. Singh [4], M. Dyrba [4], E.I. Incesoy [1,2,3], S. Mukherjee [11,5,6], O. Peters [7,8], E. Ersözlü [7,8], J. Hellmann-Regen [7,8], L. Preis [8], J. Priller [7,8,9,10], E. Spruth [7,8], S. Altenstein [7,8], A. Lohse [8], A. Schneider [11,12], K. Fliessbach [11,12], O. Kimmich [11], J. Wiltfang [14,13,15], N. Hansen [13], B. Schott [13,14,16], A. Rostamzadeh [17], W. Glanz [2,18], M. Butryn [2,18], K. Buerger [19,20], D. Janowitz [20], M. Ewers [19,20], R. Perneczky [19,21,22,23], B. Rauchmann [21,24,25], S. Teipel [4,27], I. Kilimann [4,27], D. Goerss [4,27], C. Laske [26,28,29], S. Sodenkamp [26,29], A. Spottke [11,35], M. Coenjaerts [11], F. Brosseron [11,12], F. Lüsebrink [2], P. Dechent [30], K. Scheffler [31], S. Hetzer [32], L. Kleineidam [11,12], M. Stark [12], F. Jessen [11,33,34], E. Duzel [1,2,a], G. Ziegler [1,2,a]

[1] *Institute of Cognitive Neurology and Dementia Research (IKND), Otto-von-Guericke University (OVGU), Magdeburg, Germany*
[2] *German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany*
[3] *Department for Psychiatry and Psychotherapy, University Clinic Magdeburg, Germany*
[4] *DZNE, Rostock, Germany*
[5] *University of Bonn, Germany*
[6] *University of Cambridge, Cambridge, UK*
[7] *DZNE, Berlin, Germany*
[8] *Charité – Universitätsmedizin Berlin, Institute of Psychiatry and Psychotherapy, Berlin, Germany*
[9] *Technical University of Munich, Department of Psychiatry and Psychotherapy, Munich, Germany*
[10] *University of Edinburgh, UK*
[11] *DZNE, Bonn, Germany*
[12] *Department of Old Age Psychiatry and Cognitive Disorders, University Hospital Bonn and University of Bonn, Germany*
[13] *Department of Psychiatry and Psychotherapy, University Medical Center Goettingen, University of Goettingen, Germany*
[14] *DZNE, Goettingen, Germany*
[15] *Institute of Biomedicine (iBiMED), University of Aveiro, Portugal*
[16] *Leibniz Institute for Neurobiology, Magdeburg, Germany*
[17] *Department of Psychiatry, University of Cologne, Germany*
[18] *Department for Neurology, University Clinic Magdeburg, Germany*
[19] *DZNE, Munich, Germany*
[20] *Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Germany*
[21] *Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, Munich, Germany*
[22] *Munich Cluster for Systems Neurology (SyNergy), Munich, Germany*
[23] *Ageing Epidemiology Research Unit (AGE), Imperial College London, UK*
[24] *Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, UK*
[25] *Department of Neuroradiology, University Hospital, LMU Munich, Germany*
[26] *DZNE, Tübingen, Germany*
[27] *Department of Psychosomatic Medicine, Rostock University Medical Center, Rostock, Germany*
[28] *Section for Dementia Research, Hertie Institute for Clinical Brain Research, Germany*
[29] *Department of Psychiatry and Psychotherapy, University of Tübingen, Germany*
[30] *MR-Research in Neurosciences, Department of Cognitive Neurology, Georg-August-University Goettingen, Germany*
[31] *Department for Biomedical Magnetic Resonance, University of Tübingen, Germany*
[32] *Berlin Center for Advanced Neuroimaging, Charité – Universitätsmedizin Berlin, Germany*
[33] *Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Germany*
[34] *Department of Psychiatry, University of Cologne, Medical Faculty, Germany*

* Corresponding author at: Institute of Cognitive Neurology and Dementia Research (IKND), Otto-von-Guericke University (OVGU), Magdeburg, Germany.
E-mail address: aditya.nemali@dzne.de (A. Nemali).
[a] The two authors contributed equally to this paper.

https://doi.org/10.1016/j.compbiomed.2025.110829
Received 6 December 2024; Received in revised form 16 July 2025; Accepted 20 July 2025
Available online 15 August 2025

[35] Department of Neurology, University of Bonn, Germany

ARTICLE INFO

Keywords:
Brain morphology indices
Bayesian Supervised Variational Autoencoder
Cognitive decline
Alzheimer's disease
Bayesian inference

ABSTRACT

This study introduces the Structural MRI-based Alzheimer's Disease Score (SMAS), a novel index intended to quantify Alzheimer's Disease (AD)-related morphometric patterns using a deep learning Bayesian-supervised Variational Autoencoder (Bayesian-SVAE). The SMAS index was constructed using baseline structural MRI data from the DELCODE study and evaluated longitudinally in two independent cohorts: DELCODE (n=415) and ADNI (n=190). Our findings indicate that SMAS has strong associations with cognitive performance (DELCODE: r=−0.83; ADNI: r=−0.62), age (DELCODE: r=0.50; ADNI: r=0.28), hippocampal volume (DELCODE: r=−0.44; ADNI: r=−0.66), and total gray matter volume (DELCODE: r=−0.42; ADNI: r=−0.47), suggesting its potential as a biomarker for AD-related brain atrophy. Moreover, our longitudinal studies indicated that SMAS may be useful for the early identification and tracking of AD. The model demonstrated significant predictive accuracy in distinguishing cognitively healthy individuals from those with AD (DELCODE: AUC=0.971 at baseline, 0.833 at 36 months; ADNI: AUC=0.817 at baseline, improving to 0.903 at 24 months). Notably, over 36 months, the SMAS index outperformed existing measures such as SPARE-AD and hippocampal volume. The relevance map analysis revealed significant morphological changes in key AD-related brain regions, including the hippocampus, posterior cingulate cortex, precuneus, and lateral parietal cortex, highlighting that SMAS is a sensitive and interpretable biomarker of brain atrophy, suitable for early AD detection and longitudinal monitoring of disease progression.

1. Introduction

Alzheimer's Disease (AD) is a major global health challenge, largely due to its increasing prevalence among the ageing population and its substantial socioeconomic implications. With demographic changes leading to increased prevalence, AD and other age-related dementias are projected to significantly strain healthcare systems and economies worldwide [1–4]. Therefore, it is crucial to prioritise the development of accurate and early diagnostic tools for AD. A promising approach is the development of a risk assessment score to quantify the disease stage in its early phases. Such scores could facilitate early detection, enabling timely interventions at stages when treatments are most effective. In addition, these scores could be extremely valuable in identifying patients and stratifying them for clinical trials, ensuring that research is effectively targeted and adapted to the specific stages and needs of patients, leading to more efficient and personalised care for those affected by AD [5–7].

A key challenge is developing a risk assessment score that characterises AD-related brain atrophy patterns using sMRI. Traditional approaches to quantify brain atrophy related to AD often rely on affected regions, such as hippocampal atrophy (a prominent biomarker of AD) and entorhinal cortex measurements [8–10]. However, they might lack specificity for AD and struggle with sensitivity and specificity at the mild cognitive impairment (MCI) stage, which can also manifest in non-AD dementias [11–13]. Therefore, solely relying on these regions for a risk assessment score may overlook the comprehensive spatio-temporal patterns of brain atrophy associated with AD.

Data-driven machine learning approaches involve extraction of morphometric features from sMRI scans, such as tissue segments (gray matter (GM), white matter (WM) & CSF), which can reveal patterns of brain changes associated with ageing and AD, potentially predicting disease progression and severity [14,15]. However, the high dimensionality and heterogeneity of these imaging features present significant challenges in the analysis, often limiting the effectiveness (in terms of performance) of traditional approaches [16,17]. One promising approach involves identifying low-dimensional latent representations of large, high-dimensional datasets. Numerous studies have employed multivariate approaches such as the support vector machine (SVM) classifier to estimate an index of AD anatomical risk called the spatial pattern of brain abnormality for recognition of early AD (SPARE-AD) [17], partial least squares (PLS), canonical correlation analysis (CCA), and their sparse variants (sPLS and sCCA). These techniques have been applied to associate cognitive scores or symptoms with imaging data [18,19] and for multimodal analysis [20,21]. However, these linear models might fall short in capturing the intricate interplay and non-linear dynamics that characterise AD progression [22–24].

Advancements in multimodal deep learning have shifted the focus towards models that can learn these non-linear relationships and extract meaningful features from complex data [25–27]. Variational Autoencoders (VAEs) [28,29], a class of generative models, have shown promise in learning high-level probabilistic latent embeddings from data, facilitating the identification of common patterns in ageing and AD progression [30,31]. However, the complexity and heterogeneity inherent in sMRI scans used in AD research pose challenges for VAEs. Issues such as posterior collapse, i.e., when the learnt latent variables become too similar to the prior, leading to a failure in capturing meaningful variability such as brain atrophy, and lack of supervision can impede the VAE's effectiveness in extracting meaningful latent representations from sMRI data [32].

To address the challenges, we propose a novel Bayesian supervised VAE (Bayesian-sVAE) [29] to develop a risk assessment index, namely, Structural MRI-based AD Score (SMAS). By introducing a supervised prediction node, conditioned on cognitive performance, our approach enables the model to learn latent representations of sMRI that are relevant to AD-related cognitive decline. Integration of cognitive performance into the model's architecture can potentially improve the specificity of learnt patterns related to AD, allowing for more accurate identification and staging of AD progression.

In medical image analysis, the reliability of ML models is critical for supporting quantification of relevant uncertainties and transparency. Our approach focuses on three key aspects: uncertainty quantification, explainability, and generalisability to improve model robustness and interpretability. To address uncertainty arising from inter-subject variability, scanner heterogeneity, and model complexity, [33,34], we employ a Bayesian framework within the supervised VAE, treating model parameters as random variables with prior distributions. For explainability, we use gradient activation maps to identify the image regions most influential to the model's predictions, facilitating clinicians' understanding and trust in the rationale behind its predictions. Generalisability is evaluated through validation on independent cohorts, including datasets with varying scanner resolution, thus supporting its broader clinical applicability.

The novel SMAS index derived from a Bayesian-sVAE, trained on sMRI images conditioned on cognitive scores, captures unique patterns of brain atrophy associated with AD-related cognitive impairments. The study utilised longitudinal sMRI data from two large multi-centre

cohorts: the DELCODE (DZNE Longitudinal Cognitive Impairment and Dementia) study, which is part of the German Centre for Neurodegenerative Diseases (DZNE) [35], and the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (). The analysis follows three main stages: first, training the Bayesian-sVAE model on baseline DELCODE data, which included subjects from various diagnostic groups (cognitively normal (CN), subjective cognitive decline (SCD), MCI, AD, and AD-relatives (ADR)), allowing the estimation of the SMAS index; second, validating the trained model using the two datasets: the unseen DELCODE follow-up data and the independent ADNI dataset. Finally, the resulting SMAS index underwent rigorous evaluation through several analyses to assess its clinical relevance, longitudinal progression, and disease staging.

2. Materials & methods

2.1. Bayesian supervised variational autoencoder

Variational Autoencoders (VAEs) provide a powerful probabilistic framework for analysing sMRI features by combining deep learning with variational inference. VAEs are particularly suitable for capturing the non-linear dynamics associated with AD [36]. In this context, let x represent MRI-based (potentially preprocessed) structural images (or input features) that capture neuroanatomical changes, and let z denote a latent (or hidden) variable. The VAE is designed to learn the joint distribution $p(x, z)$, which can be decomposed as $p(x, z) = p(z)p(x|z)$, where $p(z)$ is the prior distribution over latent variables and $p(x|z)$ is the likelihood of the sMRI imaging features given the latent representation. The latent variable z is assumed to capture relevant aspects of brain atrophy patterns present in the sMRI images.

In this study, we propose a Bayesian Supervised Variational Autoencoder (Bayesian-sVAE) generative model to learn latent representations of brain structure from MRI features while simultaneously predicting a dependent variable, such as cognitive scores or clinical diagnoses. The Bayesian sVAE extends the traditional sVAE framework by treating the model parameters as random variables and employing variational inference for Bayesian model optimisation.

The Bayesian sVAE model consists of three main components: an encoder $\mathbf{q}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$, a decoder $\mathbf{p}_{\theta}(\mathbf{x}|\mathbf{z})$, and a regressor $\mathbf{q}_{\boldsymbol{\psi}}(\mathbf{y}|\mathbf{z})$, where $\mathbf{y}$ represents the dependent variable. The encoder network transforms the input MRI samples $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ and dependent variables $\mathbf{y} = [y_1, y_2, \ldots, y_n]$ into a lower-dimensional latent space, where each x_i represents MRI features of the ith subject. We assume that each MRI feature x_i is linked to a latent representation z_i, and $\mathbf{z} = [z_1, z_2, \ldots, z_n]$ is a latent vector influenced by the dependent variable $\mathbf{y}$.

The **encoder network** $\mathbf{q}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ comprises a 3D convolutional block, several Residual Network (ResNet) blocks, and a fully connected layer. The ResNet blocks include two 3D convolutional layers with batch normalisation and ReLU activation, along with a skip connection. As the network deepens, the number of filters in these blocks increases progressively (for more details, see [37]). The fully connected layer produces a latent vector z_i of m dimensions that follows a Gaussian distribution: $z_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, where μ_i and σ_i are the mean and standard deviation of the Gaussian distribution, respectively, and are learnt by the encoder network.

The **decoder network** $\mathbf{p}_{\theta}(\mathbf{x}|\mathbf{z})$ reconstructs the input features from the latent representations. It includes a fully connected layer, followed by ResNet blocks, and a final convolutional block. The output of this convolutional block is a reconstructed image $\hat{x}_i$ matching the size of the input image.

The **regressor network** $\mathbf{q}_{\boldsymbol{\psi}}(\mathbf{y}|\mathbf{z})$ is designed to predict the dependent variable from the latent representations. It is structured as a simple linear layer that takes in a latent vector z_i and outputs a scalar value $\hat{y}_i$ representing the predicted dependent variable: $\hat{y}_i = W z_i + b$, where W and b are the learnable weights and bias of the linear layer, respectively.

In the Bayesian framework, the parameters of the encoder (ϕ), decoder (θ), and regressor (ψ) networks are treated as random variables with prior distributions $p(\phi)$, $p(\theta)$, and $p(\psi)$, respectively. Variational inference is employed to approximate the posterior distributions of these parameters given the observed data by introducing variational distributions $q(\phi)$, $q(\theta)$, and $q(\psi)$. Finally, due to the comparably small training data, instead of learning features from MRI scans directly, in this study we focused on using preprocessed morphometric features as inputs $\mathbf{x}$ to the VAE (see [38]) for the processing pipeline. Specifically, we utilised normalised, unmodulated GM, WM, and CSF tissue probability maps. The VAE model learns an m-dimensional latent representation of the sMRI images, which we refer to as the SMAS indices. Given the model construction and the training sample (Fig. C.7), we expect the SMAS indices to learn a condensed representation of cognition-related atrophy patterns.

2.2. Model optimisation basics

In the Bayesian-sVAE framework, the primary goal is to optimise a composite loss function that incorporates Bayesian inference, including the reconstruction and KL divergence terms from the ELBO, along with a supervised learning objective. Optimisation is conducted using backpropagation. The loss function is as follows:

$$\begin{aligned}\mathcal{L}(\theta, \phi, \psi; x, y) = &- \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] + \beta \cdot D_{KL}(q_\phi(z|x) \parallel p(z)) \\ &+ \gamma \cdot \mathbb{E}_{q_\phi(z|x)}\left[-\log \mathbb{E}_{q(\psi)}[p(y|z, \psi)]\right] \\ &+ D_{KL}(q(\theta, \phi, \psi) \parallel p(\theta, \phi, \psi))\end{aligned}$$

where θ, ϕ, and ψ denote the parameters of the decoder, encoder, and regressor, respectively. The loss consists of four components: the reconstruction loss, KL divergence terms, and the supervised regression loss. The reconstruction loss, $\mathbb{E}q_\phi(z|x)[\log p_\theta(x|z)]$, evaluates the sVAE's ability to accurately reconstruct input data x from latent variables z, thereby ensuring that the decoded samples bear a close resemblance to the original input data. The first KL term, $D_{KL}(q_\phi(z|x) \parallel p(z))$, encourages the latent variable distribution to match the prior. The supervised loss, $\mathbb{E}q_\phi(z|x)\left[-\log \mathbb{E}q_\psi[p(y|z, \psi)]\right]$, represents the expected negative log-likelihood of predicting the true target y from the latent variable z marginalised over the regressor's variational posterior $q(\psi)$. The final KL term, $D_{KL}(q(\theta, \phi, \psi) \parallel p(\theta, \phi, \psi))$, assesses the divergence between the variational distribution of the parameters and their prior distribution. The hyperparameters β and γ are employed to balance the relative importance of each term in the loss function, with their optimal values generally determined via grid search on out-of-sample data. The selection of appropriate weight priors for the Bayesian sVAE is a challenging task [39]. In our approach, the MOdel Priors with Empirical Bayes using DNN (MOPED) method, as proposed by [39], is applied to establish well-informed weight priors in Bayesian neural networks. To minimise the loss function, the Adam optimisation algorithm is utilised [40].

2.3. Application to real MRI sample

The DELCODE cohort is a multi-centric observational study conducted at 10 sites of the DZNE. At baseline, the study included 1079 participants representing a broad spectrum from healthy individuals to those clinically diagnosed with dementia. Specifically, the cohort consisted of 236 CN without any cognitive impairment, 444 subjects with SCD, 191 cases with MCI, 126 AD patients, and 82 ADR (for more details refer, [35]). Of the 1079 participants, 973 subjects (aged 60–89) had T1-weighted MRI scans. Thirty subjects were excluded due to MR artifacts and poor processing quality (see also section A), leaving 943 subjects available for analysis. Longitudinal data included 705 scans at the first annual follow-up (M12), 550 scans at the second annual follow-up (M24), and scans of subjects at the third annual follow-up (M36) (see [14] for more details).

The primary objective of the validation analysis was to evaluate how well the novel SMAS index tracks disease progression and staging and to validate this index against simpler or alternative markers such as the SPARE-AD index [17]. In addition to testing the model on longitudinal DELCODE data, we further assessed its generalisability using an independent subsample from the ADNI cohort. In the current work, a total of 200 subjects with T1-weighted scans acquired on 1.5 T MRI scanners were retrieved from the ADNI database (as specified by [41]). The sample consisted of 50 subjects with a stable diagnosis of CN state over the 24 months of follow-up, 50 subjects with a stable diagnosis of mild cognitive impairment (sMCI), 50 subjects with a stable diagnosis of AD, and 50 subjects with an initial diagnosis of MCI who showed progression to AD (pMCI) during the follow-up period.

The Preclinical Alzheimer's Cognitive Composite (PACC5) [42] in DELCODE is calculated by z-standardising and averaging five cognitive tests: free and total recall from the Free and Cued Selective Reminding Test (FCSRT), the Symbol Digit Modalities Test, Logical Memory Delayed Recall, Semantic Fluency (Animals), and the Mini-Mental State Examination (MMSE). In the ADNI dataset, we used the PACC score obtained from the ADNI portal (see [43] for further details on its estimation). In both cohorts, raw scores are z-standardised using the baseline mean and standard deviation of cognitively normal participants.[1]

2.4. Model training & hyperparameter optimisation

We first trained a deterministic sVAE on 75% of the DELCODE baseline sample, conditioned on PACC5. This step derived initial weights for the Bayesian-sVAE, using the MOPED methodology (see above). These pre-trained weights were then used to initialise the Bayesian-sVAE parameters. This approach enhances model convergence and performance by probabilistically modelling uncertainties in both the data and model parameters. The latent representations from the model correspond to SMAS indices. Choosing the dimensionality m of the latent space for VAEs is a challenging task, as it affects all terms in the loss function, e.g., reconstruction error, but particularly interpretability. Our approach focused on optimising the latent space for tracking subjects on their progression towards AD. Therefore, we decided to identify the number of latent dimensions (SMAS) using a predictive task discriminating between CN and AD subjects, assessed by the area under the receiver operating characteristic curve (AUC-ROC). The performance of the predictive task was evaluated on unseen DELCODE scans at the 12-month follow-up. While a marginal improvement in performance was observed with increased latent dimensions, this difference was not statistically significant, favouring the selection of a single latent dimension for simplicity and interpretability (see supplementary Fig. C.10 for more details). Finally, a grid search was employed to optimise hyperparameters β and γ, yielding values of 0.1 and 1, respectively, on out-of-sample data. These values were chosen for their ability to maximise predictive performance while maintaining model stability during training. We used the optimal hyperparameters in our further analysis (see Fig. 1).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

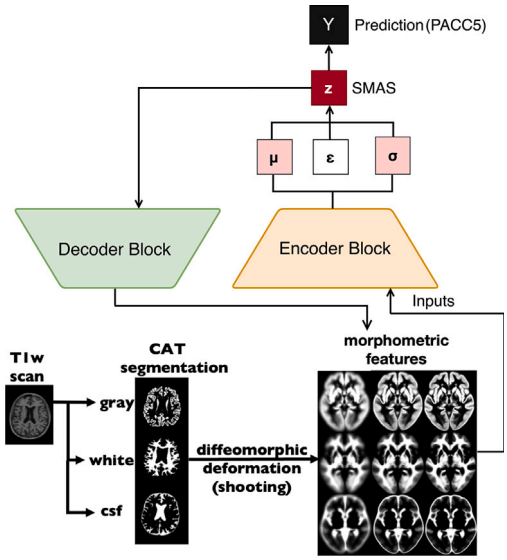

Fig. 1. Bayesian-sVAE model for AD-related atrophy patterns: The T1-weighted scans are segmented into different brain tissue types (GM, WM & CSF) using the CAT12 segmentation algorithm [38]. The model consists of three main components: an encoder, a prediction block, and a decoder. The encoder encodes the segmented brain tissue features into a latent space to derive the SMAS index. The prediction block uses this latent vector, conditioned on the inputs, to predict cognitive performance differences. Finally, the decoder reconstructs the input space from the latent vector, ensuring the latent representation also captures the original features. See supplementary Fig. C.7 for more details on architecture.

2.5. Model evaluation

2.5.1. Validation of SMAS index: Correlation with clinical assessments

To evaluate the effectiveness of the SMAS index in quantifying brain atrophy linked to AD progression, we derive the SMAS index across various validation datasets utilising the optimal hyperparameters established during the training phase (see Section 2.4 for more details). The SMAS index was normalised with min–max scaling, with the scaling factors generated from the training sample used for the validation datasets for ensuring consistency. We hypothesise that the SMAS index, intended to measure the degree and spatial distribution of brain atrophy, ought to correlate with clinical indicators of disease progression, including the PACC5, hippocampal volume, and total gray matter volume. Additionally, we anticipate a higher association between the SMAS index and age in the DELCODE dataset, which is age-specific, compared to the more heterogeneous ADNI dataset.

2.5.2. Tracking longitudinal progression of SMAS index

To investigate the usefulness of the SMAS index in tracking AD progression, we conducted a longitudinal analysis utilising data from the DELCODE and ADNI cohorts. To do this, we applied a linear mixed-effects model to assess temporal changes in the SMAS index across various diagnostic groups (CN, SCD, MCI, & AD).

$$\mathrm{SMAS}_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 \mathrm{Group}_i + \beta_3 (t_{ij} \times \mathrm{Group}_i) + b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij}$$

where SMAS_{ij} is the SMAS index for subject i at timepoint j, t_{ij} is the time, and Group_i is the diagnostic group. Here, the fixed effects (β_0, β_1, β_2, β_3) estimate the average population-level effects, capturing overall trends related to time t_{ij}, diagnostic group $Group_i$, and their interaction $t_{ij} \times \mathrm{Group}_i$ respectively. The random effects (b_{0i}, b_{1i}) allow each subject to have their own baseline (random intercept) and rate of change over time (random slope), thus modelling individual differences in both intercept and slope. The residual term ε_{ij} captures unexplained within-subject variability at each time point.

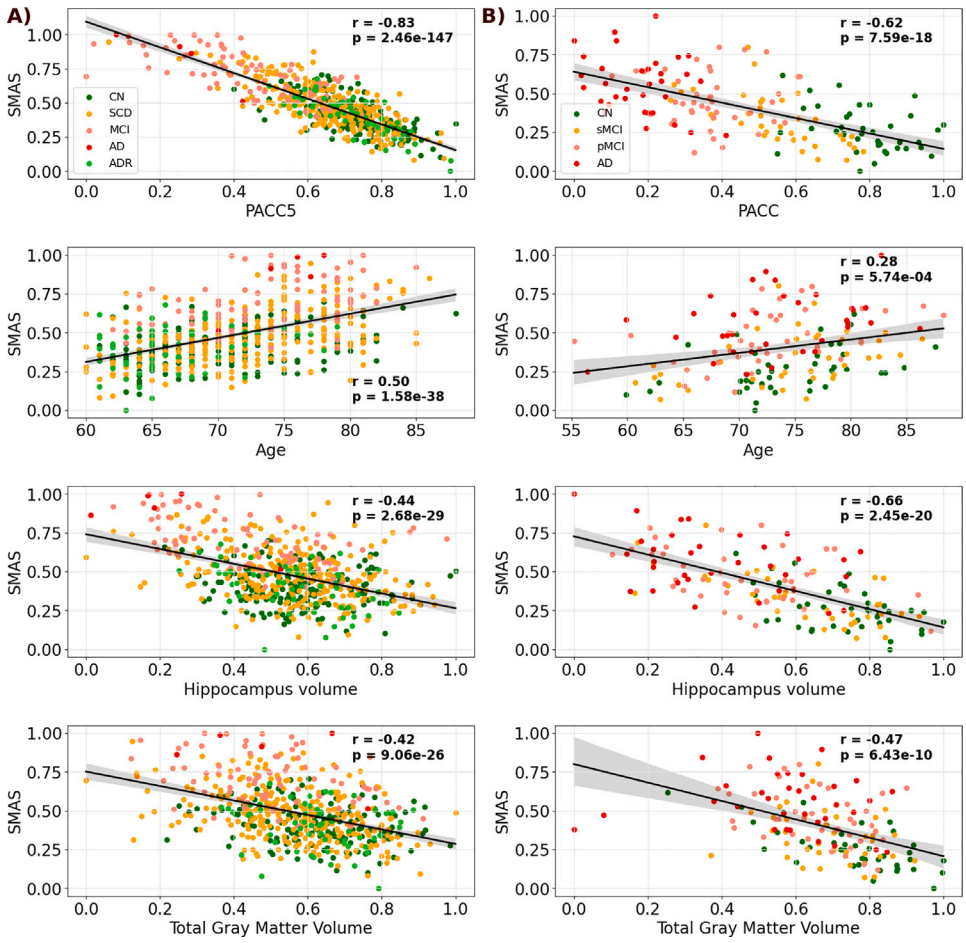

Fig. 2. Correlation analysis between SMAS index and clinical measures in (A) DELCODE and (B) ADNI; The SMAS index is computed using the posterior mean $\mu(x)$ from the encoder of the Bayesian VAE, representing a deterministic point estimate of the latent variable. SMAS shows significant negative correlations with PACC, hippocampus volume, and total gray matter volume, and a positive correlation with age, validating its association with cognitive impairment and brain atrophy.

This aimed to demonstrate whether the SMAS index effectively distinguishes disease stages and captures AD progression dynamics. Additionally, we assessed individual rates of change in the SMAS index and compared them with the rates of atrophy in brain regions typically affected by AD, such as the hippocampus, amygdala, thalamus, ventricles, and total gray matter volume. By comparing these rates, we attempt to demonstrate the association between the SMAS index and region-specific atrophy rates, highlighting the index's capacity to identify brain atrophy patterns associated with AD progression.

2.5.3. Comparative analysis: SMAS vs. SPARE-AD and hippocampal volume

To further validate the SMAS index as a marker of brain atrophy, we examined its ability to distinguish between CN, MCI, and AD individuals. We also compared the performance of the SMAS index against the SPARE-AD index and hippocampal volume in differentiating CN, MCI, and AD in the DELCODE dataset and CN, stable MCI (sMCI), and progressive MCI (pMCI) in the ADNI dataset. SPARE-AD indices were computed using available code from https://github.com/CBICA/spare_score, while hippocampal volume was obtained using Freesurfer (http://surfer.nmr.mgh.harvard.edu/). To ensure methodological consistency with the SPARE-AD index, which was proposed using a Support Vector Machine (SVM) model, we employed the SVM model for the classification task of distinguishing among CN, MCI, and AD individuals.

2.5.4. Model transparency and explainability

To enhance clinical acceptance and support informed decision-making, we provide explainability by generating gradient-based relevance maps (see [44] for more details) highlighting key sMRI regions contributing to estimating the SMAS index. Formally, for a given input sMRI feature $\mathbf{x} \in \mathbb{R}^{H\times W\times D}$ and the associated SMAS index, we estimated the gradient of the output concerning each voxel:

$$\mathbf{G} = \left| \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right|$$

Here, $\mathbf{G} \in \mathbb{R}^{H\times W\times D}$ represents the relevance map, where each element G_{ijk} reflects the absolute value of the derivative of the model's output concerning voxel x_{ijk}. Higher gradient magnitudes imply higher influential voxels. Next, we computed individual gradient maps for all subjects and then calculated the voxel-wise mean gradient across subjects:

$$\bar{\mathbf{G}} = \frac{1}{N}\sum_{n=1}^{N} \left| \frac{\partial f(\mathbf{x}^{(n)})}{\partial \mathbf{x}^{(n)}} \right|$$

where N is the number of subjects, and $\mathbf{x}^{(n)}$ is the input sMRI feature for subject n.

3. Results

3.1. Validation of SMAS index: Correlating with clinical assessments

We observed significant correlations between the SMAS index and various clinical assessments. Higher SMAS scores were associated with

Table 1
Longitudinal validation of the SMAS index across different clinical groups, highlighting the progression trajectories. A) DELCODE dataset B) ADNI dataset. The findings indicate a more rapid progression in AD, succeeded by MCI and SCD individuals.

A)

	No. of observations		1474	Log-Likelihood		−1123.5
	No. Groups		415	Converged		yes
		Coef.	Std.Err	z	P>|z|	[0.025 0.975]
	Intercept	−0.593	0.077	−7.696	<0.001	−0.744 −0.442
	SCD	0.333	0.109	3.060	0.002	0.120 0.547
	MCI	1.887	0.153	12.301	<0.001	1.586 2.188
	AD	3.987	0.494	8.079	<0.001	3.020 4.954
	ADR	−0.053	0.178	−0.300	0.764	−0.402 0.295
	time	0.010	0.001	10.249	<0.001	0.008 0.012
	time * SCD	0.007	0.001	5.047	<0.001	0.004 0.010
	time * MCI	0.014	0.002	7.025	<0.001	0.010 0.018
	time * AD	0.032	0.006	5.125	<0.001	0.020 0.044
	time * ADR	0.001	0.002	0.554	0.580	−0.03 0.006

B)

	No. of observations		759	Log-Likelihood		−115.9
	No. Groups		190	Converged		yes
		Coef.	Std.Err	z	P>|z|	[0.025 0.975]
	Intercept	−0.058	0.090	−0.645	0.519	−0.235 0.119
	sMCI	0.399	0.127	3.147	0.002	0.151 0.648
	pMCI	0.953	0.125	7.631	<0.001	0.709 1.198
	AD	1.363	0.127	10.749	<0.001	1.115 1.612
	time	0.004	0.001	2.676	0.007	0.001 0.006
	time * sMCI	0.006	0.002	3.190	<0.001	0.002 0.010
	time * pMCI	0.016	0.002	8.387	<0.001	0.012 0.019
	time * AD	0.020	0.002	10.290	<0.001	0.016 0.023

cognitive impairment, while lower scores indicated cognitively unimpaired subjects. In the DELCODE dataset (Fig. 2 A), SMAS negatively correlated with PACC5 (r = −0.83), suggesting higher SMAS scores (indicating greater brain atrophy) are associated with lower cognitive performance.

Age correlated positively with SMAS (r = 0.50), and correlations with hippocampus volume and total gray matter volume were r = −0.44 and r = −0.42, respectively, aligning with the expected trends. To assess generalisability, we analysed the SMAS index in the ADNI dataset (Fig. 2 B), finding similar trends. The PACC-SMAS correlation was r = −0.62, reaffirming the negative relationship between cognitive performance and SMAS scores. Age showed a weaker correlation (r = 0.28), while correlations with hippocampus volume (r = −0.66) and total gray matter volume (r = −0.47) were consistent.

Following the validation of the SMAS index, we applied the trained model to analyse the DELCODE and ADNI datasets longitudinally to track disease progression. In the DELCODE dataset (Table 1 A), which includes 1474 observations across 415 groups, significant differences in disease progression trajectories were observed relative to the CN group. AD subjects exhibited the steepest progression (coefficient = 3.987, $p < 0.001$), followed by MCI (coefficient = 1.887, $p < 0.001$) and SCD (coefficient = 0.333,p = 0.002). The progression rates also varied, with AD progressing the fastest (coefficient for time * AD = 0.032, $p < 0.001$), followed by MCI (0.014, $p < 0.001$) and SCD (0.007, $p < 0.001$). Likewise, the ADNI dataset (Table 1 B), with 759 observations across 190 groups, showed similar patterns. Significant differences in progression were noted compared to the CN reference group, with the AD group showing the highest progression (coefficient = 1.363, $p < 0.001$), followed by pMCI (0.953, $p < 0.001$) and sMCI (0.399,p = 0.002). The fastest progression rates were in AD (0.020, $p < 0.001$), followed by pMCI (0.016, $p < 0.001$) and sMCI (0.006, $p < 0.001$).

We also explored the relationship between changes in the SMAS index and brain volume changes. In the DELCODE cohort, the rate of change in the SMAS index was significantly and negatively correlated with hippocampal volume change (r = −0.55, $p < 0.001$), indicating that a higher SMAS index corresponds to reduced hippocampal volume. Similar negative & significant correlations were observed for the thalamus (r = −0.32, $p < 0.001$, $p < 0.001$), amygdala (r = −0.50, $p < 0.001$), and total gray matter volume (r = −0.39, $p < 0.001$), suggesting higher rates of change in the SMAS index are linked with higher volume declines in these regions (see supplementary Fig. C.15). In the ADNI dataset, a negative correlation between SMAS index change and hippocampal volume change (r = −0.32, $p < 0.001$) was consistent with the findings from the DELCODE dataset. Additionally, a positive correlation was found with ventricular volume change (r = 0.51, $p < 0.001$), aligning with typical AD patterns where increased ventricular volume accompanies brain tissue loss (see supplementary Fig. C.16). Furthermore, we compared the rate of change in SMAS indices in relation to amyloid beta (Aβ42/40) and phosphorylated tau (p-tau) and observed significant differences between positive and negative groups (see supplementary Fig. C.17).

We further examined the longitudinal progression of SMAS in the DELCODE sample across clinical groups and age ranges. The analysis showed that SCD and MCI groups exhibited faster atrophy compared to CN and ADR. When comparing SMAS changes between cognitively unimpaired (CU) and cognitively impaired (CI) groups longitudinally, we observed higher progression rates in the older age group (75–85 years) and in CI individuals across all ages (see supplementary Fig. C.14).

3.2. Comparative analysis: SMAS vs. SPARE-AD indices and hippocampus volume

3.2.1. DELCODE

We compared the SMAS index with SPARE-AD and hippocampal volume in distinguishing AD, MCI, and CN individuals over time. The SMAS index consistently outperformed across all timepoints.

For CN vs AD (Fig. 3 A), SMAS AUC values remained high: 0.971 at baseline, 0.972 at 12 months, 0.875 at 24 months, and 0.833 at 36 months. SPARE-AD declined from 0.864 at baseline to 0.5 at 24 and 36 months. Hippocampal volume showed a similar drop, from 0.82 to

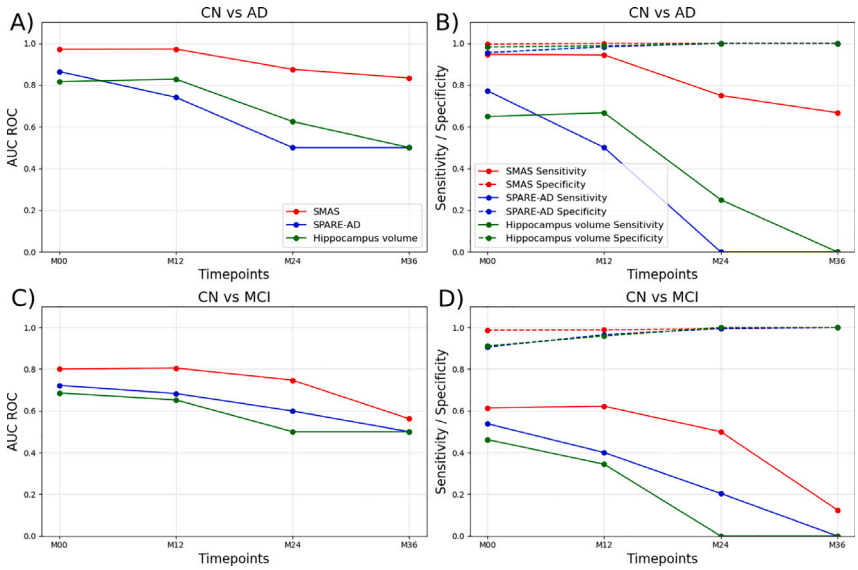

Fig. 3. DELCODE: Performance of hippocampus volume, SPARE-AD, and SMAS index in predicting cognitively normal (CN) vs. Alzheimer's disease (AD) and CN vs. mild cognitive impairment (MCI) across different timepoints. (A) Area under the ROC curve (AUC) for CN vs AD prediction. (B) Sensitivity and specificity for CN vs AD prediction. (C) AUC for CN vs. MCI prediction. (D) Sensitivity and specificity for CN vs. MCI prediction. The SMAS index (red) demonstrates higher AUC, sensitivity, and specificity compared to SPARE-AD (blue) and hippocampus volume (green) for both prediction tasks over the 36-month period.

0.5 by 36 months. In CN vs MCI (Fig. 3 C), SMAS exhibited superior AUC values, from 0.849 at baseline to 0.706 at 36 months. SPARE-AD dropped from 0.751 to 0.605, while hippocampal volume declined from 0.686 to 0.5.

Sensitivity was higher for SMAS in both tasks (Fig. 3). For CN vs AD, it decreased from 0.947 at baseline to 0.667 at 36 months, while SPARE-AD and hippocampal volume dropped to 0 at 24 months. In CN vs MCI, SMAS sensitivity fell from 0.788 to 0.438, outperforming SPARE-AD (0.575 to 0.229) and hippocampal volume (0.462 to 0). Specificity remained high for SMAS, reaching 1.0 for CN vs AD from 12 months onwards and improving in CN vs MCI from 0.911 to 0.974 by 36 months. SPARE-AD and hippocampal volume achieved similar specificity after 24 months (Fig. 3 D).

3.2.2. ADNI

In the ADNI dataset, SMAS continued to outperform SPARE-AD and hippocampal volume.

For CN vs AD, SMAS AUC values increased from 0.82 at baseline to 0.90 at 24 months, with sensitivity rising from 0.81 to 0.92 and specificity consistently above 0.83. SPARE-AD showed lower AUC (0.66 to 0.75), while hippocampal volume had intermediate values (0.752 to 0.81) (Fig. 4 A, B). For CN vs pMCI, SMAS showed higher AUC (0.76 to 0.812) compared to SPARE-AD (0.69 to 0.68) and hippocampal volume (0.73 to 0.79) (Fig. 4 C, D). In CN vs sMCI, SMAS AUC ranged from 0.625 to 0.62, performing similarly to SPARE-AD and better than hippocampal volume (Fig. 4 E, F).

3.3. Comparison of SMAS and CSF-biomarkers

We investigated the relationship between SMAS and CSF biomarkers (Aβ42/40 & pTau) on an independent DELCODE dataset for the available sample size (n=415). We observed a significant moderate negative correlation between baseline SMAS and the Aβ42/40 ratio (r=-0.50, $p < 0.001$). This suggests that lower Aβ42/40 levels, reflecting greater amyloid pathology, are associated with a higher risk of brain atrophy (in cognition-related areas), as indicated by higher SMAS values. As expected, SMAS demonstrated a significant moderate positive correlation with pTau (r=0.38, $p < 0.001$), suggesting that higher pTau concentrations are also linked to more emphasised brain atrophy (see Fig. 5).

Table 2
Classification performance of clinical diagnostic groups (CN vs. AD) using SMAS, CSF biomarkers, and their combinations in n=82 subjects of the DELCODE cohort. Performance was evaluated using leave-one-out cross-validation.

Feature set	AUC score	F1-score	Sensitivity/Specificity
SMAS	0.985	0.992	0.971/1.0
Aβ42/40	0.883	0.917	0.829/0.954
pTau	0.824	0.781	0.800/0.759
SMAS + Aβ42/40	0.986	0.976	0.971/0.977
SMAS + pTau	0.975	0.983	0.943/1.0
SMAS + Aβ42/40 + pTau	0.973	0.983	0.943/1.0

Additionally, we assessed the diagnostic performance of SMAS and CSF biomarkers and their combination using SVM within the DELCODE cohort with the available CSF biomarker data. Performance was evaluated using leave-one-out cross-validation. More specifically, we focused on metrics such as the area under the receiver operating characteristic curve (AUC-ROC) and the weighted F1-score, as well as sensitivity and specificity.

In differentiating CN individuals from those with AD, SMAS alone demonstrated near-perfect classification performance (AUC = 0.985, F1 = 0.992, sensitivity/specificity = 0.971/1.0), outperforming both Aβ42/40 (AUC = 0.883) and pTau (AUC = 0.824) when used individually. Integrating SMAS with either or both CSF biomarkers resulted in marginal improvements or maintained similarly high accuracy (SMAS + Aβ42/40: AUC = 0.986; SMAS + pTau: AUC = 0.975; SMAS + Aβ42/40 + pTau: AUC = 0.973). These findings indicate that pattern-based SMAS scores capture a highly discriminative index for AD classification that is complementary in its definition and calculation based on MR images compared to traditional CSF biomarkers (see Table 2).

A similar pattern was observed for the earlier disease stage classification when classifying CN versus MCI. SMAS outperformed both Aβ42/40 (AUC = 0.702, F1 = 0.706, sensitivity/specificity = 0.560/0.862) and pTau (AUC = 0.650, F1 = 0.652, sensitivity/specificity = 0.5/0.816) individually (SMAS AUC = 0.939, F1 = 0.883, sensitivity/specificity = 0.893/0.84). Combining SMAS with either biomarker led to slight improvements in overall model performance (SMAS + Aβ42/40: AUC = 0.940; SMAS + pTau: AUC = 0.942). This suggests that SMAS might provide a useful index with early-stage discriminative power and might add incremental performance gains when integrated

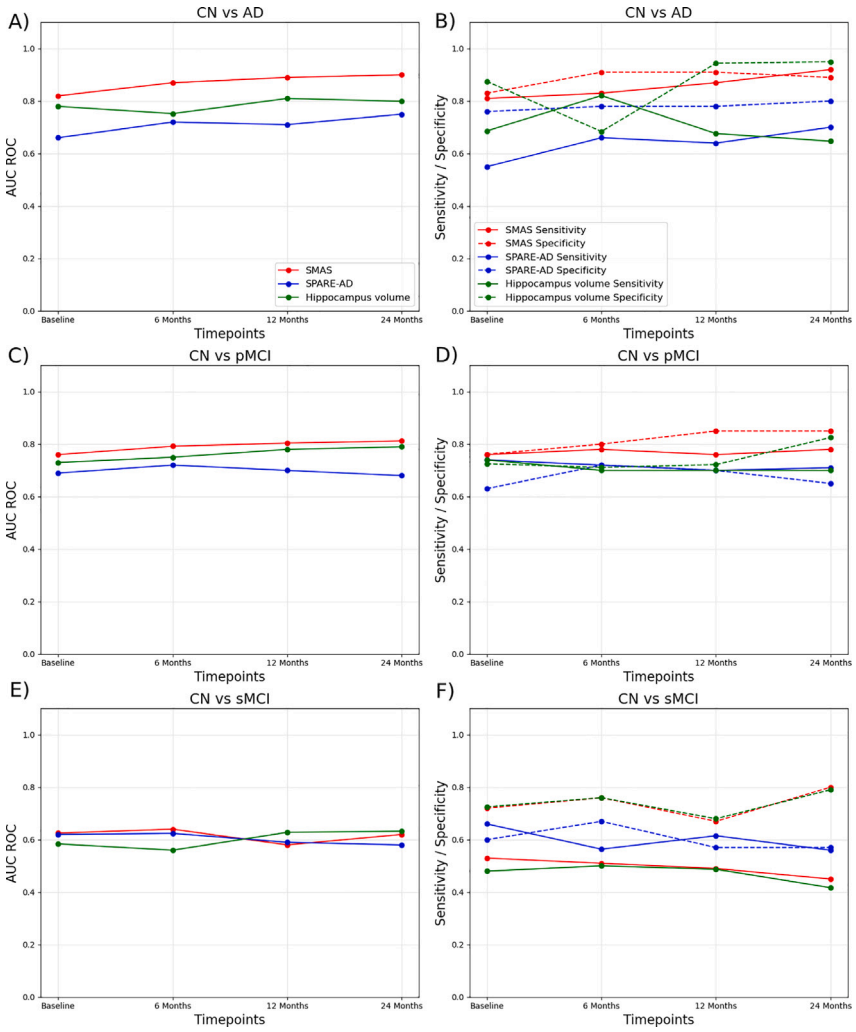

Fig. 4. ADNI: Performance of hippocampal volume, SPARE-AD, and SMAS index in predicting CN vs. AD, CN vs. pMCI, and CN vs. sMCI at various time points using an independent ADNI dataset. (A) For CN vs. AD prediction, the SMAS index (red) consistently showed slightly higher performance compared to SPARE-AD (blue) and hippocampal volume (green) over 24 months. (B) Sensitivity and specificity for CN vs. AD prediction further support the favourable performance of the SMAS index. (C) For CN vs. pMCI prediction, the SMAS index exhibited slightly higher AUC values than SPARE-AD and hippocampal volume. (D) Sensitivity and specificity for CN vs. pMCI prediction also suggest a slight advantage for the SMAS index. (E) For CN vs. sMCI prediction, the SMAS index maintained moderately higher ROC AUC values. (F) Sensitivity and specificity for CN vs. sMCI prediction suggest a tendency for higher sensitivity for the SMAS index, while all models demonstrated reasonable specificity over time.

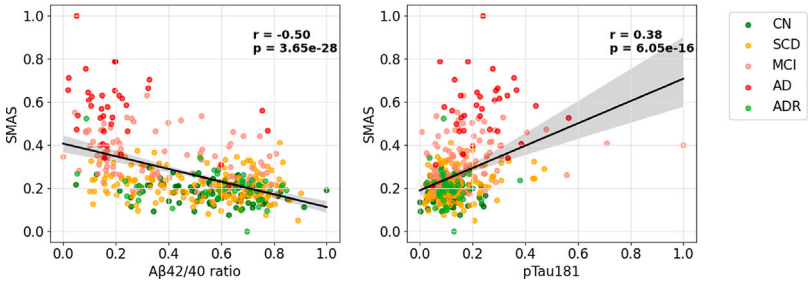

Fig. 5. Illustrates the relationship between SMAS index and CSF biomarkers: Aβ42/40 (left) and pTau (right). Aβ42/40 shows a significant negative correlation with SMAS (r=-0.50,p = 3.65e−28), while pTau shows a significant positive correlation (r=0.38,p = 6.05e−16).

with conventional molecular CSF biomarkers, raising the question of cost-efficiency trade-offs for clinical decision support (see Table 3).

3.4. Model transparency

In our study, we examined the relevant maps to derive the SMAS index, focusing on the DELCODE M12 validation data. These maps were created by averaging the weight contributions from all subjects in the DELCODE M12 validation dataset, representing the average weight distribution and highlighting regions with varying contributions to the model's predictions. Our findings reveal the hippocampus, posterior cingulate cortex, precuneus, and lateral parietal cortex as regions with high relevance contributions in estimating the SMAS index (Fig. 6). Nevertheless, while these regions demonstrate significant model-based relevance, it is crucial to exercise caution when interpreting these

Table 3
Classification performance (CN vs. MCI) using SMAS, CSF biomarkers, and their combinations in the DELCODE cohort.

Feature set	AUC score	F1-score	Sensitivity/Specificity
SMAS	0.939	0.883	0.893/0.894
Aβ42/40	0.702	0.706	0.560/0.862
pTau	0.650	0.652	0.5/0.816
SMAS + Aβ42/40	0.940	0.889	0.893/0.885
SMAS + pTau	0.942	0.883	0.869/0.897
SMAS + Aβ42/40 + pTau	0.943	0.889	0.881/0.897

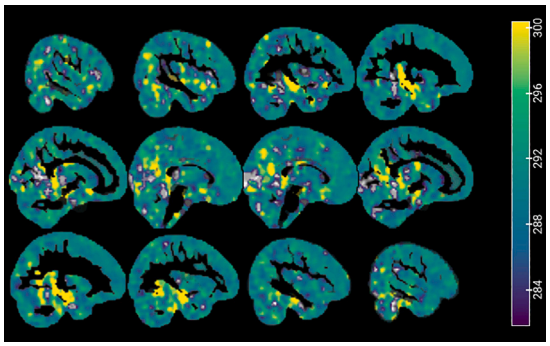

Fig. 6. Relevance map of brain regions (hippocampus, posterior cingulate cortex, precuneus, and lateral parietal cortex) contribution in estimating SMAS index.

findings. Clinical applicability must be rigorously validated through further empirical research and thorough clinical evaluation.

4. Discussion

In this study, we present the SMAS index, a novel measure of AD-related morphometric patterns using a deep learning Bayesian-sVAE, trained on structural brain imaging features of the large-scale DELCODE cohort. Our results, including a validation of the score across unseen data from two datasets (DELCODE and ADNI), do suggest the robustness and potential of SMAS to characterise at-risk individuals along the progression towards AD.

4.1. Choice of latent dimension

The selection of the latent dimensionality, especially determining the latent vector $z_i \in \mathbb{R}^1$, was a key design decision in our Bayesian-sVAE framework. While higher-dimensional latent spaces have the potential to capture more complex patterns in data, our empirical study over various dimensions (m = 1, 2, 4, …, 256) indicated modest improvement in performance (see Fig. C.10 for more details). These minor improvements were insufficient to compensate for the increased complexity and loss of interpretability associated with higher-dimensional latent representation models. Therefore, the decision to utilise a one-dimensional latent space (m = 1) was motivated by two main factors: simplicity and interpretability, particularly in terms of clinical utility. A single latent variable offers a straightforward and easily communicable metric for clinicians, facilitating longitudinal tracking of disease progression and enhancing transparency for risk assessment. Unlike multidimensional representations, which can be difficult to interpret and integrate, a one-dimensional score is intrinsically easier to compare and associate with external clinical factors such as cognitive test scores, genetic markers, and CSF biomarkers. Furthermore, this simplicity might facilitate better adaptation to decision-making in healthcare contexts, where model transparency and trust in AI-assisted approaches are important.

4.2. SMAS index: A reliable indicator of brain atrophy

This study suggests that the SMAS index, derived from a Bayesian-sVAE trained on DELCODE baseline sMRI data, may serve as a useful indicator for assessing brain atrophy. Analysis across independent DELCODE and ADNI datasets shows that the SMAS index is associated with various clinical and structural factors, such as PACC5, age, hippocampal volume, and total gray matter volume. A strong negative correlation between SMAS and PACC5 scores suggests that higher SMAS values are associated with lower cognitive performance, indicating potential relevance of SMAS for tracking cognitive decline in AD. Additionally, positive correlations between SMAS and age support its sensitivity to age-related atrophic patterns (see Fig. 2). The SMAS index also demonstrates negative correlations with both hippocampal and total gray matter volumes regions commonly affected in AD, indicating its capability to reflect atrophy in these areas [45–47]. Longitudinal analyses further support these findings, showing that a higher rate of change in the SMAS index corresponds to reduced volumes in the hippocampus, thalamus, amygdala, and total gray matter, as well as ventricular enlargement. In addition to structural and cognitive correlates, SMAS also demonstrated significant relationships with core cerebrospinal fluid (CSF) biomarkers of AD pathology. We observed a moderate negative correlation between SMAS and the Aβ42/40 ratio, indicating that greater amyloid burden is associated with more severe structural atrophy. Similarly, SMAS showed a moderate positive correlation with pTau levels, further suggesting its relevance as a marker sensitive to neurofibrillary degeneration (see Fig. 5). These findings establish SMAS as a bridge between structural atrophy and underlying molecular pathology. Importantly, the SMAS index reflects underlying pathological processes, as indicated by its correlation with amyloid and tau biomarkers. This suggests that the SMAS index might detect molecular aspects of AD in addition to structural abnormalities. From a clinical staging viewpoint, the SMAS index outperforms established CSF biomarkers in distinguishing individuals with AD and those in the prodromal stage (MCI) from CN. Notably, it demonstrates an increased sensitivity in detecting early-stage disease, making it useful for identifying individuals at risk before significant cognitive decline develops. SMAS, as a non-invasive, imaging-based marker, is a viable alternative or complement to fluid biomarkers, especially in situations when lumbar puncture is not possible. Furthermore, when combined with CSF biomarkers, the SMAS index offers additional predictive value. Aside from clinical and biomarker relationships, we investigated the potential impact of variations in MRI acquisition parameters across cohorts and scanner sites. The results suggest site effects were essentially negligible in the DELCODE cohort and limited to a few sites in ADNI, indicating overall robustness of the SMAS index across diverse imaging conditions.

Together, these findings underscore the SMAS index as a reliable, robust, and data-driven indicator of brain atrophy that integrates well with both structural and molecular aspects of AD. Its strong correlation with cognitive, anatomical, and biomarker-based markers, along with its higher diagnostic utility, supports its potential role in both research and clinical contexts for tracking disease progression and aiding diagnosis.

4.3. SMAS index: A promising tool for early detection and monitoring of AD

4.3.1. Monitoring disease progression

The longitudinal analysis of the SMAS index in both the DELCODE and ADNI cohorts suggests that it may offer value in monitoring the progression of AD across different clinical stages. Notably, AD individuals exhibited the most pronounced increases in SMAS over time, followed by those with MCI. This is consistent with the hypothesis that the SMAS index is associated with disease severity and progression. Importantly, the detection of a small but significant increase in SMAS among SCD individuals raises the possibility that this index may be

sensitive to subtle, early alterations in neurodegenerative processes. This finding is particularly interesting, as SCD is increasingly considered as a potential preclinical stage of AD, and the sensitivity of SMAS to detect these subtle early changes suggests it may serve as an early biomarker for neurodegeneration before clinical symptoms become apparent. Furthermore, the differences observed in the rates of SMAS change between the DELCODE and ADNI cohorts may reflect underlying differences in cohort composition, such as demographic variability, recruitment criteria, or other methodological factors. Collectively, these results suggest that the SMAS index has potential utility in tracking AD progression (see Table 1).

4.3.2. Age-related variations

The analysis of the DELCODE cohort reveals varied SMAS trajectories across age and clinical groups, suggesting that SMAS may reflect differences in brain atrophy associated with these groups. In CN individuals, SMAS increases gradually with age, consistent with expected age-related atrophy. In SCD individuals, a steeper SMAS increase is observed, particularly in older age groups. This pattern may suggest that brain atrophy begins to accelerate earlier in SCD, possibly indicating initial neurodegenerative changes before clinical symptoms become apparent [48]. The MCI group shows an even higher SMAS rate from earlier ages, with a steady increase over time, which could suggest that brain atrophy progresses more rapidly in MCI, reflecting a more advanced stage of atrophy. Additionally, CI individuals exhibit a notably steeper SMAS trajectory compared to CU individuals, suggesting an accelerated rate of brain atrophy in CI. Overall, these findings imply that SMAS could be a useful indicator for assessing brain atrophy, with potential utility in identifying and monitoring early stages of AD.

4.3.3. Disease staging

Our analysis of the DELCODE dataset showed that the SMAS index achieved high predictive accuracy in distinguishing CN and AD subjects, demonstrating strong performance with a slight decline over time (refer Fig. 3 A). For CN versus MCI, the AUC showed a moderate decrease over time 3 C). The SMAS index outperformed the 3D CNN model by [49], which achieved an AUC of 0.953 for CN vs. AD and 0.775 for CN vs. MCI at baseline.

For the independent ADNI dataset, the SMAS index showed strong performance in distinguishing CN from AD and in predicting pMCI to AD conversion, with improving AUC as the time to stable diagnosis decreased. However, the performance for CN vs. sMCI was lower, indicating challenges in differentiating sMCI cases from CN (refer Fig. 3). When compared to other methods, our approach shows competitive performance. For instance, [49] achieved an AUC of 0.949 for AD vs. CN and 0.785 for CN vs. MCI using a 3D CNN. [41,50], which used a support vector machine with PCA+FDR feature extraction on a similar dataset, reported lower performance metrics, particularly for (CN + sMCI) vs. (pMCI + AD) classification at various time points before stable diagnosis. Additionally, the multi-modal feature selection algorithm (FC2FS) by [51] achieved an AUC of 92.84% for NC vs. AD, while the multi-modal multi-task learning (M3T) method by [52] reported an AUC of 0.933 for AD vs. CN and 0.832 for MCI vs. CN. These comparisons indicate the potential of the SMAS index in early disease staging, particularly in distinguishing CN from MCI and pMCI.

4.4. Comparative analysis

To further validate the robustness of the SMAS index, we compared them with the well-established SPARE-AD and hippocampus volume. We evaluated their ability to differentiate between CN individuals and those with AD or MCI across both the ADNI and DELCODE datasets. Our results consistently showed that the SMAS index demonstrated higher performance compared to SPARE-AD and hippocampus volume over a 36-month observation period. In both datasets, the SMAS model achieved higher AUC scores for CN vs. AD and CN vs. MCI, reaffirming its robustness in early detection (refer Fig. 3, 4).

We also confirmed the effectiveness of the SMAS index against those derived from traditional principal component analysis (PCA). We carried out classification analyses over latent dimensions using an independent subset of the DELCODE dataset (M12). The results echoed our previous findings, showing that the SMAS index consistently outperforms traditional methods in distinguishing CN from MCI. This suggests that these indexes might be useful for the early detection of AD stages.

4.5. Transparency and explainability

The quantitative analysis of relevance contributions to the SMAS index derived through Bayesian-sVAE demonstrates the model's capacity to identify and quantify morphological alterations in key brain regions associated with AD. Significant relevance contributions were observed in the hippocampus, posterior cingulate cortex, precuneus, and lateral parietal cortex, aligning with established patterns of AD-related neurodegeneration [53–57]. This correlation emphasises the model's sensitivity to disease-specific structural changes.

Focusing on these critical regions suggests potential for early detection of AD-related atrophy, possibly at the MCI stage; secondly, it offers a means for longitudinal monitoring of disease progression; and thirdly, it presents a promising metric for evaluating the efficacy of potential disease-modifying therapies in clinical trials. This approach offers a transparent and interpretable method for quantifying AD-related brain atrophy.

4.6. SMAS: Clinical utility

To further support the clinical applicability and aid real-world implementation, the SMAS index could potentially be used for stratifying individuals across the AD continuum (CN, MCI, and AD stages). In this context, we identified optimal cut-off thresholds using Gaussian mixture modelling, with values of 0.386 and 0.629 emerging as possible actionable reference points. These thresholds may allow clinicians to identify individuals at increased risk for early neurodegenerative changes (i.e., MCI) or those exhibiting structural atrophy patterns that are consistent with AD. Specifically, SMAS score above 0.386 might indicate early-stage neurodegeneration, suggesting that closer monitoring could be warranted, while a score above 0.629 may be indicative of more advanced atrophy typically associated with AD. Importantly, while the SMAS index may serve as a helpful standalone measure, its integration with other clinical, genetic, and biomarker data could be particularly valuable in identifying individuals at risk for progression from CN to MCI, where early intervention can have the greatest impact. Additionally, the broader range of SMAS values, between 0.629 and 1.0, may reflect underlying heterogeneity among AD patients, potentially indicating variability in atrophy distribution, comorbid pathologies, and further disease progression.

4.7. Limitations and future directions

While our study shows promise, we acknowledge certain limitations and suggest key directions for future work. The current model is primarily based on structural MRI, limiting multimodality to different tissue classes. This choice was based on the increased availability of T1-weighted data within the DELCODE cohort. In future work, we aim to incorporate additional imaging modalities such as fMRI, PET, or DTI, as well as extended demographic and clinical test scores, which could offer a more comprehensive representation of at-risk individuals and capture heterogeneity. This also highlights an important future challenge: addressing the impact of missing data for specific inputs and quantifying uncertainty in latent scores. Our analysis suggests that there is no systematic bias across scanner sites in the estimation of the SMAS index.

However, this does not imply that the SMAS index is entirely free from site-related variability, as differences in MRI acquisition parameters may still influence the estimates. In future work, we plan to further evaluate the performance of SMAS in additional cohorts. Specifically, we aim to investigate the extent to which MRI acquisition factors, such as field strength, scanner manufacturer, and software version, influence SMAS predictions. Additionally, we will explore model-based correction strategies (e.g., calibration approaches), including the use of scanner-related features as nuisance regressors or integrating them directly into the learning framework to account for potential sources of variation. Future studies might incorporate scan quality indicators and scan-related parameters to enhance the robustness of the score for single subjects. Finally, the potential influence of reserve variables should not be overlooked. Future research may aim to study a wider range of subject-specific factors that could mitigate the effects of brain pathology on cognitive outcomes, especially in assessing an individual's risk of disease progression.

Ethics declarations

The study protocol was approved by the ethical committees of the medical faculties of all participating sites: the ethical committees of Berlin (Charité, University Medicine), Bonn, Cologne, Göttingen, Magdeburg, Munich (Ludwig-Maximilians-University), Rostock, and Tübingen. The process was led and coordinated by the ethical committee of the medical faculty of the University of Bonn. The registration number of the trial at the ethical committee in Bonn is 117/13.

Trial registration German Clinical Trials Register DRKS00007966. Registered 4 May 2015.

CRediT authorship contribution statement

A. Nemali: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **J. Bernal:** Writing – review & editing. **R. Yakupov:** Data curation. **D. Singh:** Writing – review & editing. **M. Dyrba:** Writing – review & editing. **E.I. Incesoy:** Writing – review & editing, Data curation. **S. Mukherjee:** Writing – review & editing. **O. Peters:** Data curation. **E. Ersözlü:** Data curation. **J. Hellmann-Regen:** Data curation. **L. Preis:** Data curation. **J. Priller:** Data curation. **E. Spruth:** Data curation. **S. Altenstein:** Data curation. **A. Lohse:** Data curation. **A. Schneider:** Data curation. **K. Fliessbach:** Data curation. **O. Kimmich:** Data curation. **J. Wiltfang:** Data curation. **N. Hansen:** Data curation. **B. Schott:** Data curation. **A. Rostamzadeh:** Data curation. **W. Glanz:** Data curation. **M. Butryn:** Data curation. **K. Buerger:** Data curation. **D. Janowitz:** Data curation. **M. Ewers:** Data curation. **R. Perneczky:** Data curation. **B. Rauchmann:** Data curation. **S. Teipel:** Data curation. **I. Kilimann:** Data curation. **D. Goerss:** Data curation. **C. Laske:** Data curation. **S. Sodenkamp:** Data curation. **A. Spottke:** Data curation. **M. Coenjaerts:** Data curation. **F. Brosseron:** Data curation. **F. Lüsebrink:** Data curation. **P. Dechent:** Data curation. **K. Scheffler:** Data curation. **S. Hetzer:** Data curation. **L. Kleineidam:** Data curation. **M. Stark:** Data curation. **F. Jessen:** Data curation. **E. Duzel:** Supervision, Resources. **G. Ziegler:** Writing – review & editing, Supervision, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT-4 to improve readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the published article.

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.

Appendix A. DELCODE MRI acquisition

MRI scans were acquired in 9 out of 10 involved DZNE sites (3T Siemens scanners: 3 TIM Trio systems, 4 Verio systems, 1 Skyra, and 1 Prisma system). Our main analyses were based on whole-brain T1-weighted MPRAGE (3D GRAPPA PAT 2, 1 mm^3 isotropic, 256 X 256 px, 192 slices, sagittal, 5 min, TR 2500 ms, TE 4.33 ms, TI 110 ms, FA 7°). Further ROI and covariate processing was based on the additionally available FLAIR protocol (for details see Jessen, F., et al. 2018).

Appendix B. Demographics information

See Tables B.4 and B.5.

Appendix C. Supplementary results

C.1. Supervised VAE framework

See Fig. C.11.

C.2. VAE vs. Supervised VAE

To study the influence of integrating the cognitive variable PACC5 as a constraint within the VAE framework, we examined both the reconstruction maps and latent space for constrained and unconstrained models. We first assessed the ability of the latent variable to reconstruct the original MRI morphological features. Figs. C.9A and C.9B illustrate reconstruction maps for an individual subject, representing the outputs of the unconstrained and constrained VAE models, respectively. Visually, both models reproduce the input GM feature map with good fidelity, demonstrating that global anatomical structure is retained regardless of the constraint. To further quantify the differences, we present a voxel-wise difference map in Fig. C.9C, highlighting the absolute intensity differences between the two reconstructions. The results demonstrate that most brain areas exhibit small intensity variations, suggesting that the integration of the PACC5 constraint does not significantly impact reconstruction quality. Notably, localised regions reveal subtle differences, which might be interpreted as reflecting cognitively significant aspects captured by the PACC5 constraint. In addition to reconstruction, we assessed how the presence of the PACC5 constraint impacts the latent representation of the input data. As shown in Fig. C.8, we examined the relationship between the SMAS index, a derived latent measure, and PACC5 scores within the DELCODE cohort (independent - timepoint: M12). The unconstrained model (Fig. C.8, left) showed a minimal correlation ($r = 0.05, p = 0.191$), demonstrating that the latent space and cognitive performance are not well aligned. In contrast, the constrained model (Fig. C.8, right) revealed a strong negative correlation ($r = -0.82, p = 2.83e{-}142$), demonstrating that the latent space is effectively shaped by the PACC5 variable. This alignment enables the constrained model to better capture variations in brain morphology that are predictive of cognitive decline. Moreover, the constrained latent structure more clearly distinguishes between clinical groups (CN, SCD, MCI, AD, ADR), supporting its potential utility in early disease characterisation.

C.3. Influence of site

To evaluate whether heterogeneity in MRI acquisition parameters across different scanner sites influenced the SMAS index, we conducted mixed linear model regression analyses for the DELCODE and ADNI datasets, treating scanner site as a fixed effect. In the DELCODE cohort ($n = 1474$ across 9 sites), none of the sites reached statistical significance after FDR correction (all $p > 0.51$), suggesting that site-specific factors did not substantially contribute to the differences of the SMAS index in our DELCODE cohort (see Table C.6 A for more

Table B.4
Baseline demographic information for the participants from the DELCODE cohort used in this modelling study. The PACC5 score is transformed using min–max normalisation to the unit interval. Age is indicated in years.

Variable	CN	SCD	MCI	AD	ADR
No. of subjects	229	388	158	109	75
Males/females	129/94	173/200	69/82	63/44	44/31
Age (Mean ±SD)	69.46 ±5.42	71.27 ±6.06	72.91 ±5.72	75.19 ±6.25	66.27 ±4.61
PACC5 (Mean ±SD)	0.77 ±0.09	0.72 ±0.11	0.49 ±0.14	0.25 ±0.10	0.76 ±0.11

Table B.5
Baseline demographic information for the participants from the ADNI cohort used in this modelling study. The PACC score is transformed using min–max normalisation to the unit interval. Age is indicated in years.

Variable	CN	sMCI	pMCI	AD
No. of subjects	50	50	50	50
Age (Mean ±SD)	74.84 ±6.10	74.66 ±6.83	74.49 ±7.08	75.14 ±6.75
PACC (Mean ±SD)	0.77 ±0.11	0.52 ±0.12	0.39 ±0.12	0.21 ±0.10

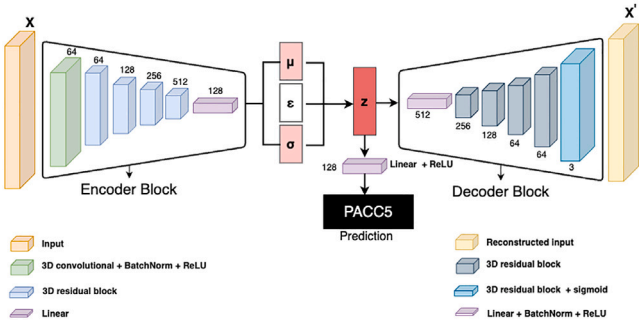

Fig. C.7. Overview of the proposed 3D supervised ResNet-based VAE architecture. The encoder block consists of a 3D convolutional layer followed by a sequence of residual blocks with increasing channel dimensions (64 → 128 → 256 → 512), ending with a linear transformation that produces the latent mean (μ) and log-variance (σ). The latent vector z is sampled via the reparameterisation trick and used for both decoding and prediction. The decoder reconstructs the input volume through symmetric residual blocks and upsampling layers, ending in a 3D residual block with sigmoid activation. A prediction head (pacc5) operates directly on the latent representation. Colour-coded modules indicate different operations, as defined in the legend.

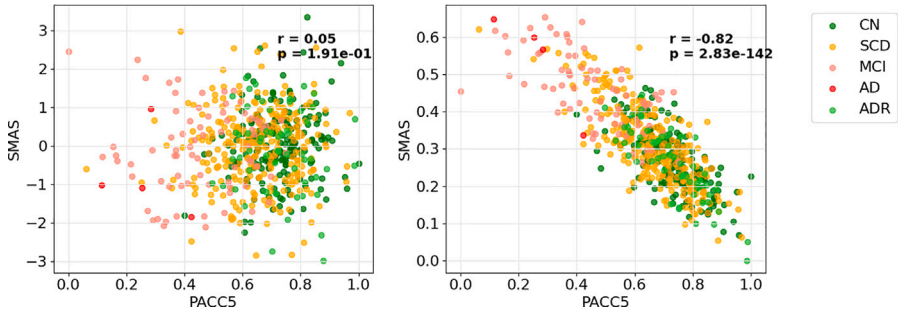

Fig. C.8. Correlation analysis between the SMAS index and PACC5 in the DELCODE cohort. Unconstraint VAE (left), Constraint VAE using additional regression of latents on clinical scores (right).

details). In the ADNI cohort (n = 759 across 38 sites), testing for each site effect on SMAS values, we also found no significant effect of individual sites, while a trend towards a compound hypothesis effect was observed (all $p > 0.06$ FDR) (see Table C.6 B for more details). These results do suggest that, within our datasets, heterogeneity in scanners or acquisition parameters across sites did not significantly introduce considerable bias into our analyses, but differences in harmonisation strategies across cohorts might contribute in larger sample analyses with increased sensitivity. Additionally, these analyses of mean effects do not imply that SMAS might be entirely free from variance differences across sites. In the future, we plan to continue using SMAS in other cohorts and testing whether the extent to which MRI acquisition parameters influence its predictions as well as testing whether model-based corrections (or calibration approaches), such as including scanner-related features (e.g., field strength, manufacturer, software version) as nuisance regressors or embedding them directly into the learning framework. This approach would enable the SMAS index to inherently adjust for site-related variability, thereby enhancing its generalisability and suitability for clinical settings.

Table C.6
Scanner site effects on SMAS scores across cohorts. A) DELCODE dataset B) ADNI dataset. In DELCODE & ADNI, no significant site effects were observed, suggesting stable performance across scanners.

A)

No. of observations		1474	Log-Likelihood			−1123.5
No. Groups		415	Converged			yes
	Coef.	Std.Err	z	P (FDR)	[0.025 0.975]	
Intercept	−0.593	0.077	−7.696	<0.001	−0.744	−0.442
SCD	0.333	0.109	3.060	0.002	0.120	0.547
MCI	1.887	0.153	12.301	<0.001	1.586	2.188
AD	3.987	0.494	8.079	<0.001	3.020	4.954
ADR	−0.053	0.178	−0.300	0.910	−0.402	0.295
site-5	−0.113	0.096	−1.172	0.515	−0.185	0.130
site-8	0.030	0.096	0.308	0.910	−0.301	0.076
site-10	−0.061	0.070	−0.867	0.579	−0.159	0.218
site-11	0.010	0.073	0.133	0.922	−0.199	0.077
site-13	−0.074	0.082	−0.902	0.579	−0.134	0.154
site-14	−0.135	0.120	−1.127	0.515	−0.236	0.087
site-17	−0.016	0.070	−0.222	0.922	−0.371	0.100
site-18	−0.058	0.148	−0.391	0.910	−0.153	0.122
time	0.010	0.001	10.249	<0.001	0.008	0.012
time * SCD	0.007	0.001	5.047	<0.001	0.004	0.010
time * MCI	0.014	0.002	7.025	<0.001	0.010	0.018
time * AD	0.032	0.006	5.125	<0.001	0.020	0.044
time * ADR	0.001	0.002	0.554	0.922	−0.03	0.006

B)

No. of observations		759	Log-Likelihood			−115.9
No. Groups		190	Converged			yes
	Coef.	Std.Err	z	P (FDR)	[0.025 0.975]	
Intercept	−0.058	0.090	−0.645	0.667	−0.235	0.119
sMCI	0.399	0.127	3.147	0.003	0.151	0.648
pMCI	0.953	0.125	7.631	<0.001	0.709	1.198
AD	1.363	0.127	10.749	<0.001	1.115	1.612
time	0.004	0.001	2.676	0.070	0.001	0.070
site-3	−0.587	0.636	−0.923	0.589	−1.834	0.660
site-5	0.721	0.367	1.967	0.181	0.003	1.440
site-6	0.011	0.529	0.020	0.984	−1.026	1.047
site-7	0.517	0.415	1.245	0.432	−0.297	1.330
site-11	0.454	0.367	1.237	0.432	−0.265	1.173
site-13	0.441	0.531	0.831	0.627	−0.599	1.482
site-14	−0.602	0.480	−1.256	0.432	−1.543	0.338
site-16	0.365	0.629	0.580	0.705	−0.868	1.598
site-20	−0.704	0.627	−1.123	0.494	−1.934	0.525
site-22	0.531	0.377	1.408	0.382	−0.208	1.271
site-23	0.461	0.323	1.426	0.382	−0.173	1.094
site-24	−0.214	0.855	−0.251	0.837	−1.891	1.462
site-27	0.182	0.349	0.523	0.709	−0.501	0.866
site-29	0.891	0.478	1.865	0.200	−0.045	1.827
site-33	0.365	0.329	1.109	0.493	−0.280	1.010
site-35	1.119	0.454	2.463	0.110	0.228	2.009
site-36	0.765	0.432	1.771	0.231	−0.082	1.611
site-41	0.693	0.482	1.437	0.382	−0.252	1.638
site-51	1.062	0.537	1.977	0.182	0.009	2.115
site-52	−0.242	0.529	−0.457	0.720	−1.279	0.795
site-53	−0.316	0.550	−0.575	0.705	−1.393	0.762
site-57	−0.341	0.421	−0.810	0.627	−1.167	0.485
site-62	0.178	0.382	0.466	0.720	−0.571	0.927
site-67	0.658	0.631	1.042	0.523	−0.580	1.895
site-73	0.415	0.629	0.660	0.698	−0.818	1.648
site-94	0.273	0.484	0.564	0.705	−0.675	1.221
site-98	−0.534	0.659	−0.809	0.627	−1.825	0.758
site-99	0.183	0.417	0.440	0.720	−0.634	1.001
site-109	1.520	0.544	2.794	0.063	0.453	2.586
site-114	1.392	0.636	2.190	0.152	0.147	2.638
site-126	0.716	0.541	1.322	0.425	−0.345	1.777
site-127	0.263	0.423	0.622	0.705	−0.567	1.093
site-130	1.517	0.637	2.381	0.111	0.268	2.766
site-131	1.095	0.549	1.996	0.182	0.020	2.170
site-133	0.184	0.451	0.408	0.729	−0.699	1.067
site-136	0.037	0.428	0.087	0.951	−0.802	0.877
site-941	0.865	0.854	1.012	0.534	−0.810	2.539
time * sMCI	0.006	0.002	3.190	<0.001	0.002	0.010
time * pMCI	0.016	0.002	8.387	<0.001	0.012	0.019
time * AD	0.020	0.002	10.290	<0.001	0.016	0.023

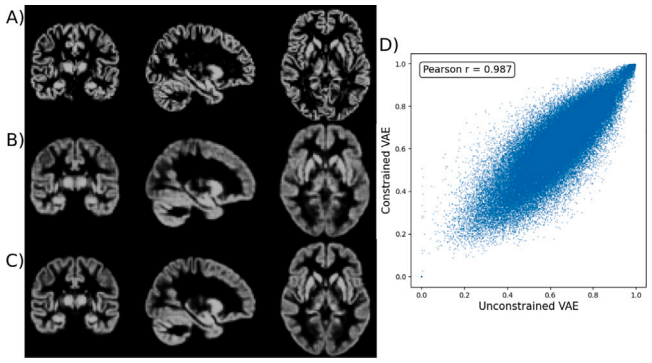

Fig. C.9. Gray matter morphological tissue map of a subject with mild cognitive impairment (MCI). (A) Original tissue map shown in coronal, sagittal, and axial views. (B) Reconstructed map for the same subject obtained using an unconstrained variational autoencoder (VAE). (C) Reconstructed tissue map generated by a constrained VAE for comparison. (D) Scatter plot illustrating the correlation between reconstructed voxel intensities from the unconstrained and constrained VAEs, indicating high similarity (Pearson correlation coefficient r = 0.987) in terms of reconstruction.

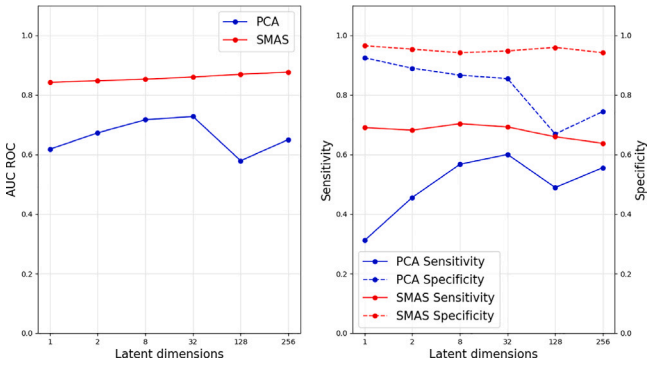

Fig. C.10. Comparative evaluation of CN vs. MCI classification performance using indices derived from PCA and Bayesian-VAE for the DELCODE M12 sample trained on the DELCODE baseline sample: SMAS indices exhibit superior AUC performance over PCA indices across latent dimensions, demonstrating high specificity and sensitivity. Despite a slight performance enhancement with increasing latent dimensions, the difference was not statistically significant, advocating for the choice of a single latent dimension for ease of interpretation and simplicity.

C.4. Longitudinal trajectories: SMAS index

To enhance the interpretability and clinical relevance of our findings, we provide illustrative cases illustrating longitudinal trajectories of the SMAS index alongside established PACC5 and hippocampal volume. The examples provided represent individuals across four diagnostic categories: CN, SCD, MCI, and AD. Fig. C.12 illustrates time-scaled trajectories for three individuals per diagnostic group (selected randomly), demonstrating the progression of SMAS (blue), PACC5 (green), and hippocampus volume (orange) over 36 months. The SMAS index in CN individuals (Fig. C.12 row 1) remains mostly constant throughout time, with trajectories closely matching those of PACC5 and hippocampus volume, indicating no major cognitive or structural degeneration. SCD individuals (Fig. C.12 row 2) demonstrate increased variability, with some exhibiting an early increase in SMAS that is not yet reflected in corresponding cognitive or volumetric changes. Individuals with MCI (Fig. C.12 row 3) exhibit a greater degree of longitudinal SMAS index, which is frequently associated with declining PACC5 scores and hippocampal shrinkage, indicating increasing neurodegeneration. Finally,

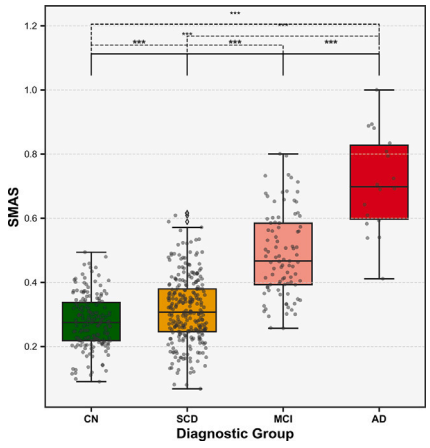

Fig. C.11. Distribution of the SMAS index across an independent DELCODE cohort at 12 months. The clinical groups are: CN (Cognitively Normal, n=172, green), SCD (Subjective Cognitive Decline, n=271, yellow), MCI (Mild Cognitive Impairment, n=91, salmon), and AD (Alzheimer's Disease, n=18, dark red). Box plots display the median (central line), interquartile range (box), and outliers (individual points). SMAS values are normalised between 0 and 1, with higher values indicating greater brain atrophy. Statistical significance was assessed using ANOVA followed by post-hoc pairwise comparisons (***p < 0.001). The progressive increase in SMAS values from CN through SCD and MCI to AD demonstrates the metric's ability to capture the continuum of neurodegeneration across the AD spectrum.

AD subjects exhibit the steepest increases in SMAS, accompanied by pronounced cognitive decline and hippocampus volume loss (Fig. C.12 row 4).

To supplement these trajectory plots, we illustrate voxel-wise relevance maps for selective individuals (last column from Fig. C.12) from each diagnostic group. Fig. C.13 shows maps that highlight the significant influence of the SMAS index for a test subject (subject 3) with warmer colours (red) indicating more voxel importance. In the CN (SMAS = 0.25) and SCD (SMAS = 0.23) subjects, relevance activation was minimal and, with slight activations in the temporal region and thalamus, suggesting limited structural atrophy. However, in the SCD subject, emerging focal activations, particularly in the thalamus region, indicate subtle changes potentially linked to preclinical pathology. In the MCI subject, relevance activations become stronger, especially in the medial temporal region, aligning with known progression patterns in prodromal AD. The AD subject exhibits more intense relevance activations involving regions typically affected by AD, such as the medial temporal lobe, temporal lobe, and thalamus.

Appendix D. SMAS thresholds

To establish clinically meaningful SMAS index cut-off thresholds for differentiating CN, MCI, and AD individuals, we applied Gaussian mixture modelling (GMM). GMM identified two optimal cut-off values: 0.386, which best separated CN from MCI individuals, and 0.629, which distinguished MCI from AD individuals. These thresholds reflect progressive neurodegenerative burden, with scores above 0.386 indicating increased risk of early-stage atrophy (MCI), and scores above 0.629 reflecting advanced atrophy typical of AD. We validated these thresholds using AUC-ROC analysis to assess diagnostic accuracy. For the CN vs. MCI comparison at the SMAS threshold of 0.386, the AUC was 0.945, with 92.3% sensitivity and 82.6% specificity. In distinguishing MCI from AD at the SMAS threshold of 0.629, the AUC was 0.875, yielding 83.3% sensitivity and 81.3% specificity (see Fig. C.18). See Figs. D.19 and D.20.

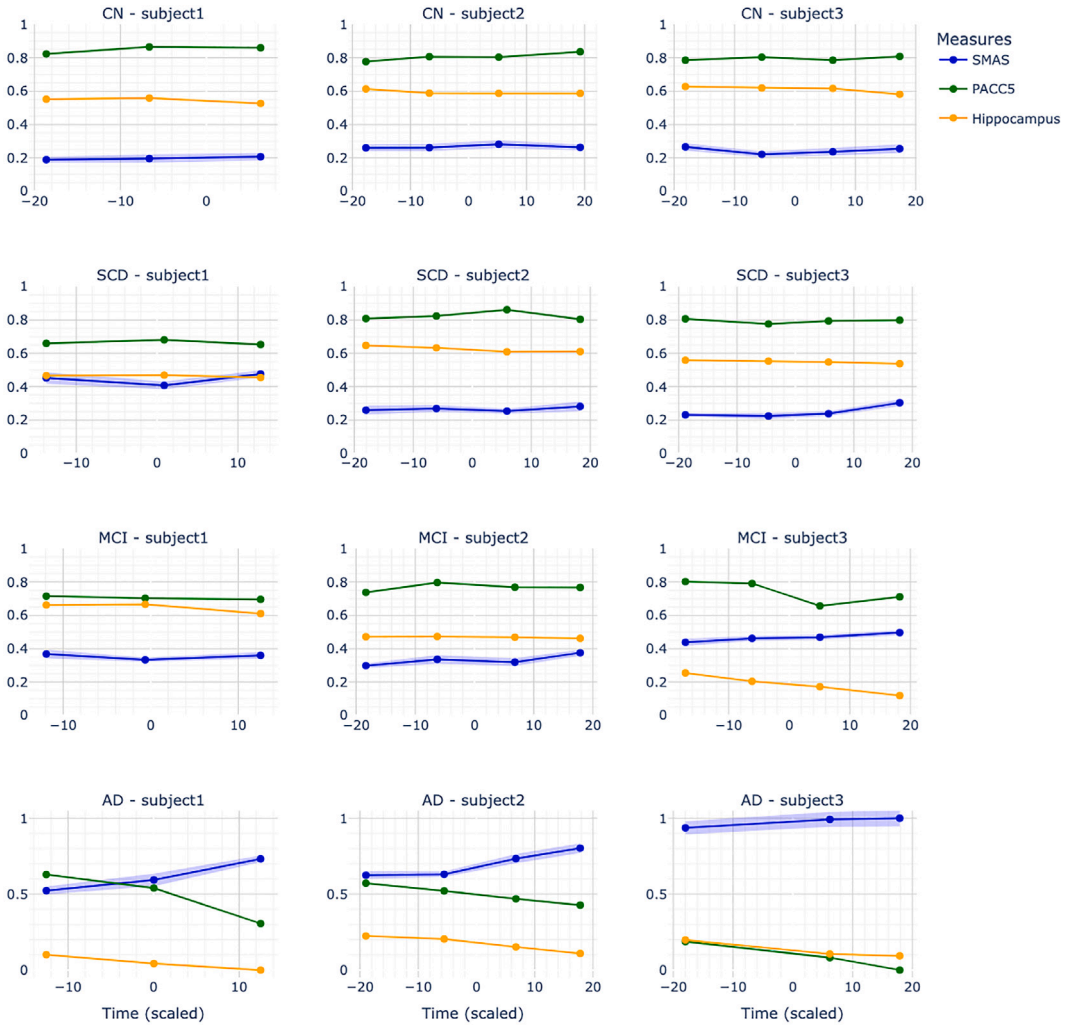

Fig. C.12. Individual longitudinal trajectories of SMAS index, PACC5, and hippocampus volumes across DELCODE diagnostic groups. Each panel displays time-scaled trajectories of SMAS (blue), PACC5 (green), and hippocampal volume (orange). All measures are normalised to a 0–1 scale to enable direct comparisons.

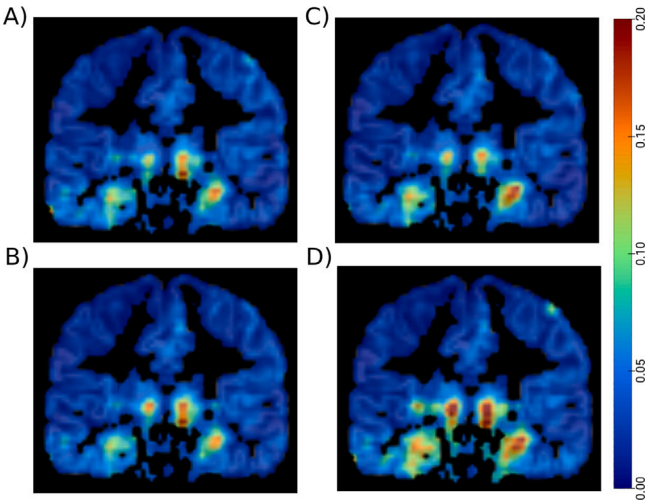

Fig. C.13. Gray matter relevance maps illustrating regions contributing to the estimation of the SMAS index. Representative subject 3 from each clinical group are shown: A) Cognitively Normal (CN; SMAS = 0.25), B) Subjective Cognitive Decline (SCD; SMAS = 0.23), C) Mild Cognitive Impairment (MCI; SMAS = 0.41), and D) Alzheimer's Disease (AD; SMAS = 0.87). The colour bar indicates the magnitude of relevance activation, with warmer colours (yellow-red) indicating higher relevance and cooler colours (blue–green) representing lower relevance. Notably, medial temporal regions demonstrate increased activation in MCI and exhibit the highest activation levels in AD, reflecting structural neurodegeneration along the AD continuum.

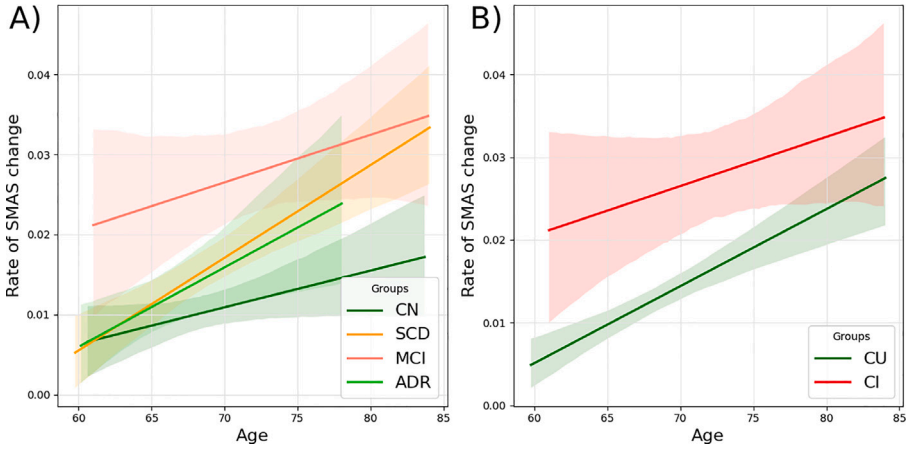

Fig. C.14. Rate of SMAS change over age for different clinical groups A) Comparison of the rate of SMAS change across age for CN, SCD, MCI, AD, and ADR groups. B) Rate of SMAS change over age for cognitively unimpaired (CU) and cognitively impaired (CI) groups.

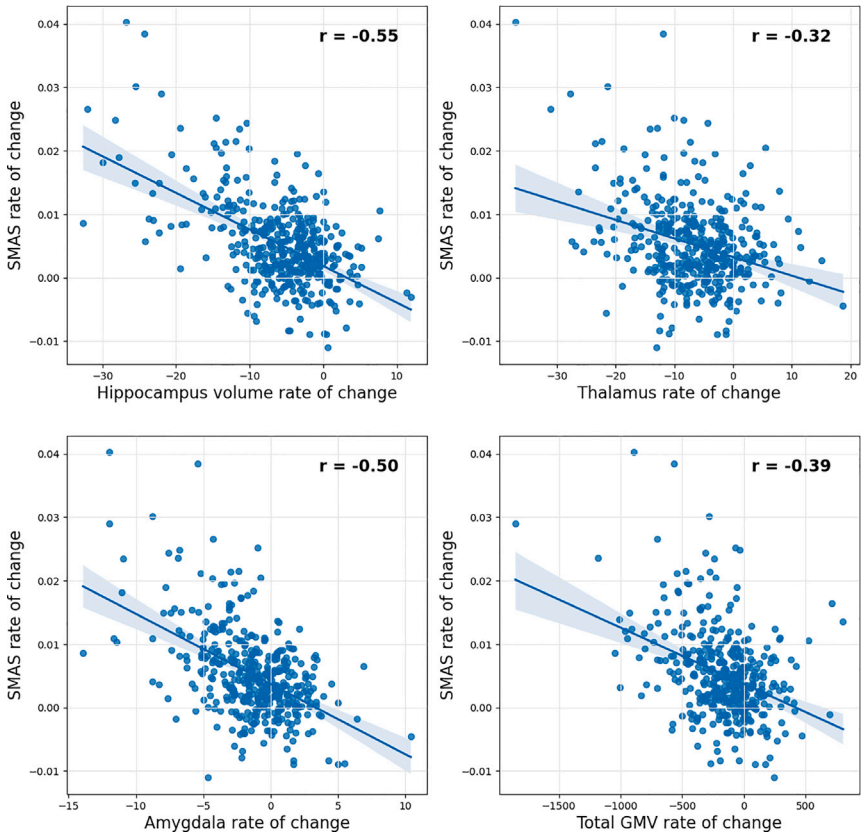

Fig. C.15. Rate of change of SMAS indices and brain volume changes in the DELCODE sample. The scatter plots depict the correlation between the rate of change of SMAS indices and the rate of change in the volume of the hippocampus (r = −0.55), thalamus (r = −0.32), amygdala (r = −0.50), and total gray matter volume (r = −0.39).

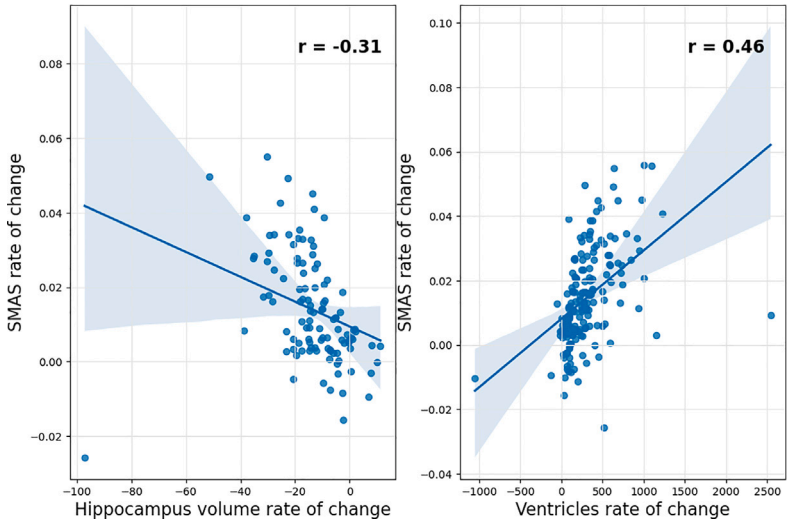

Fig. C.16. Rate of change of SMAS indices and brain volume changes in the ADNI sample. The scatter plots depict the correlation between the rate of change of SMAS indices and the rate of change in the volume of the hippocampus (r = −0.32) and ventricles (r = 0.51).

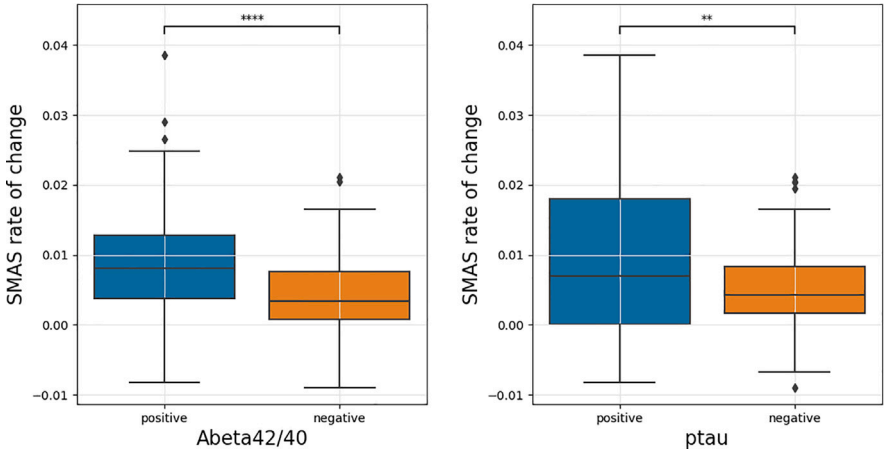

Fig. C.17. Comparison of SMAS indices rate of change in relation to CSF biomarkers (Aβ 42/40) and ptau. The left side represents the significant rate of change in SMAS indices for both Aβ42/40 positive and negative groups. On the right side, for both the ptau-positive and negative groups, we observe a significant rate of change.

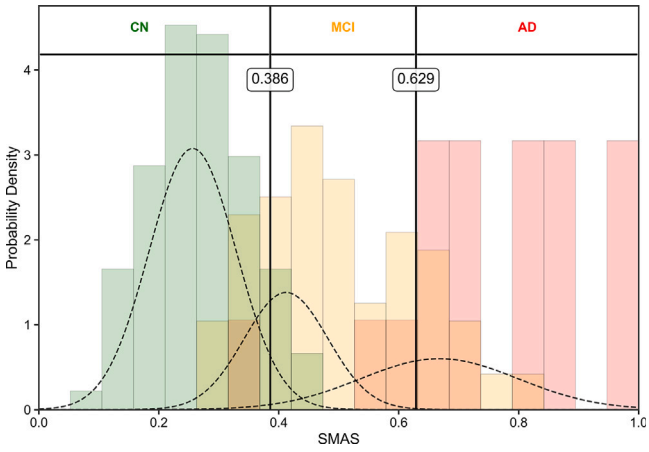

Fig. C.18. Illustrates the distribution of SMAS values for three clinical groups: Cognitively Normal (CN, green), Mild Cognitive Impairment (MCI, orange), and Alzheimer's Disease (AD, red). Vertical black lines indicate threshold values (SMAS = 0.386 and SMAS = 0.629) determined using Gaussian mixture modelling (GMM) to optimise separation between the diagnostic categories. SMAS > 0.386 denotes increased risk for early neurodegenerative changes (MCI), while SMAS > 0.629 indicates significant structural atrophy consistent with an AD-like pattern. Dashed lines represent GMM-derived probability density estimates for each group, illustrating the progression of neurodegeneration risk with increasing SMAS.

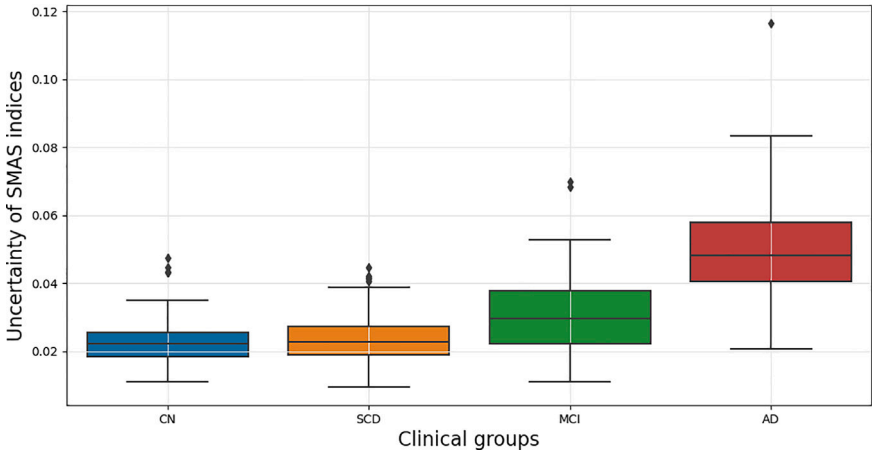

Fig. D.19. Uncertainty of SMAS indices for the DELCODE cohort across clinical groups. The CN and SCD groups exhibit relatively low uncertainty, indicating more consistent SMAS index measurements in these groups. The MCI group shows moderate uncertainty with a broader distribution, reflecting increased variability. The AD group presents the highest level of uncertainty, suggesting significant variability in SMAS indices within this population.

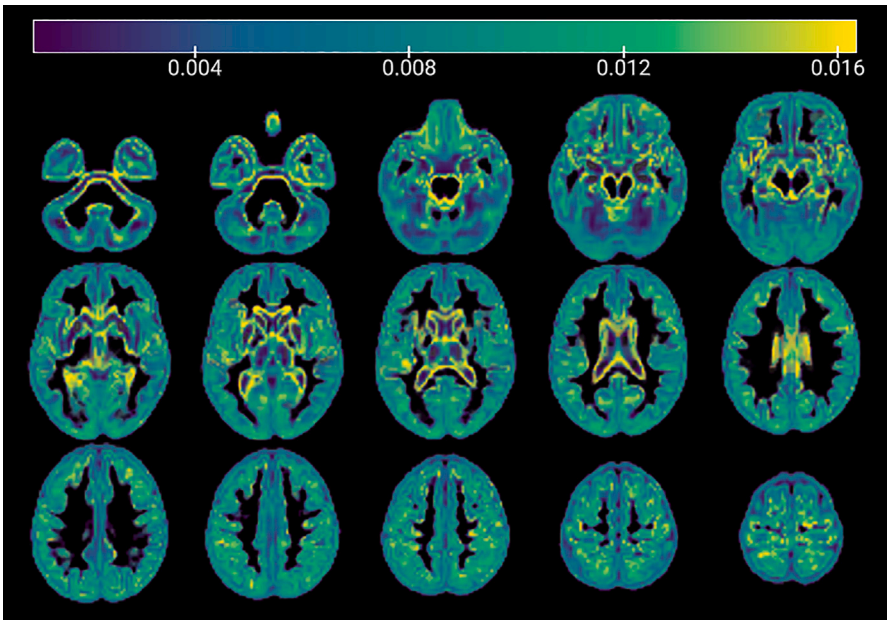

Fig. D.20. Reconstruction uncertainty from the Bayesian-sVAE model in estimating the SMAS index. Higher uncertainty areas are observed primarily in the central regions around the ventricles, the boundaries and edges of brain structures, and some cortical regions in the lower slices. These areas reflect the regions where the Bayesian-sVAE model has more difficulty in accurate reconstruction, potentially due to the variability in the sMRI.

References

[1] E. Nichols, et al., Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the global burden of disease study 2019, Lancet Public Heal. 7 (2) (2022) e105–e125.

[2] J.-H. Shin, Dementia epidemiology fact sheet 2022, Ann. Rehabil. Med. 46 (2) (2022) 53–59.

[3] M.G. Ulep, S.K. Saraon, S. McLea, Alzheimer disease, J. Nurse Pr. 14 (3) (2018) 129–135.

[4] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, C.P. Ferri, The global prevalence of dementia: A systematic review and metaanalysis, Alzheimer' s Dement. 9 (1) (2013) 63–75.

[5] S. Qiu, et al., Multimodal deep learning for Alzheimer's disease dementia assessment, Nat. Commun. 13 (1) (2022) 3404.

[6] K. Blennow, B. Dubois, A.M. Fagan, P. Lewczuk, M.J. De Leon, H. Hampel, Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early Alzheimer's disease, Alzheimer' s Dement. 11 (1) (2015) 58–69.

[7] J.L. Cummings, T. Morstorf, K. Zhong, Alzheimer's disease drug-development pipeline: Few candidates, frequent failures, Alzheimer's Res. Ther. 6 (4) (2014) 37.

[8] R. Casanova, et al., Using high-dimensional machine learning methods to estimate an anatomical risk factor for Alzheimer's disease across imaging databases, Neuroimage 183 (2018) 401–411.

[9] L. Pini, M. Pievani, M. Bocchetta, D. Altomare, P. Bosco, E. Cavedo, S. Galluzzi, M. Marizzoni, G.B. Frisoni, Brain atrophy in Alzheimer's disease and aging, Ageing Res. Rev. 30 (2016) 25–48.

[10] M. Zhou, F. Zhang, L. Zhao, J. Qian, C. Dong, Entorhinal cortex: A good biomarker of mild cognitive impairment and mild Alzheimer's disease, Rev. Neurosci. 27 (2) (2016) 185–195.

[11] A.J. Bastos-Leite, W.M. Van Der Flier, E.C. Van Straaten, S.S. Staekenborg, P. Scheltens, F. Barkhof, The contribution of medial temporal lobe atrophy and vascular pathology to cognitive impairment in vascular dementia, Stroke 38 (12) (2007) 3182–3185.

[12] L.A. Van De Pol, et al., Hippocampal atrophy on MRI in frontotemporal lobar degeneration and Alzheimer's disease, J. Neurol. Neurosurg. Psychiatry 77 (4) (2006) 439–442.

[13] D. Chan, et al., Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease, Ann. Neurol. 49 (4) (2001) 433–442.

[14] A. Nemali, et al., Gaussian process-based prediction of memory performance and biomarker status in ageing and Alzheimer's disease—A systematic model evaluation, Med. Image Anal. 90 (2023) 102913.

[15] J.M. Mateos-Pérez, M. Dadar, M. Lacalle-Aurioles, Y. Iturria-Medina, Y. Zeighami, A.C. Evans, Structural neuroimaging as clinical predictor: A review of machine learning applications, NeuroImage: Clin. 20 (2018) 506–522.

[16] M. Habes, M.J. Grothe, B. Tunc, C. McMillan, D.A. Wolk, C. Davatzikos, Disentangling heterogeneity in Alzheimer's disease and related dementias using data-driven methods, Biol. Psychiatry 88 (1) (2020) 70–82.

[17] C. Davatzikos, F. Xu, Y. An, Y. Fan, S.M. Resnick, Longitudinal progression of alzheimer's-like patterns of atrophy in normal older adults: The SPARE-AD index, Brain 132 (8) (2009) 2026–2035.

[18] A. Mihalik, J. Chapman, R.A. Adams, N.R. Winter, F.S. Ferreira, J. Shawe-Taylor, J. Mourão-Miranda, A.D.N. Initiative, Canonical correlation analysis and partial least squares for identifying brain–behavior associations: A tutorial and a comparative study, Biol. Psychiatry: Cogn. Neurosci. Neuroimaging 7 (11) (2022) 1055–1067.

[19] G. Ziegler, R. Dahnke, A.D. Winkler, C. Gaser, Partial least squares correlation of multivariate cognitive abilities and local brain structure in children and adolescents, NeuroImage 82 (2013) 284–294.

[20] M. Lorenzi, et al., Partial least squares modelling for imaging-genetics in Alzheimer's disease: Plausibility and generalization, in: 2016 IEEE 13th International Symposium on Biomedical Imaging, ISBI, IEEE, 2016, pp. 838–841.

[21] A.Z. Burzynska, et al., A scaffold for efficiency in the human brain, J. Neurosci. 33 (43) (2013) 17150–17159.

[22] Y. Wang, et al., Predicting long-term progression of Alzheimer's disease using a multimodal deep learning model incorporating interaction effects, J. Transl. Med. 22 (1) (2024) 265.

[23] H. Nakua, et al., Comparing the stability and reproducibility of brain-behavior relationships found using canonical correlation analysis and partial least squares within the ABCD sample, Netw. Neurosci. (2024) 1–21.

[24] V. Diogo, H. Ferreira, D. Prata, Early diagnosis of alzheimer's disease using machine learning: A multi-diagnostic, generalizable approach, IBRO Neurosci. Rep. 15 (2023) S391–S392.

[25] J. Venugopalan, L. Tong, H.R. Hassanzadeh, M.D. Wang, Multimodal deep learning models for early detection of Alzheimer's disease stage, Sci. Rep. 11 (1) (2021) 3254.

[26] M.A. Ebrahimighahnavieh, S. Luo, R. Chiong, Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review, Comput. Methods Programs Biomed. 187 (2020) 105242.

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

[28] D.P. Kingma, M. Welling, An introduction to variational autoencoders, Found. Trends® Mach. Learn. 12 (4) (2019) 307–392.

[29] D.P. Kingma, S. Mohamed, D. Jimenez Rezende, M. Welling, Semi-supervised learning with deep generative models, Adv. Neural Inf. Process. Syst. 27 (2014).

[30] B. Sauty, S. Durrleman, Progression models for imaging data with longitudinal variational auto encoders, in: L. Wang, Q. Dou, P.T. Fletcher, S. Speidel, S. Li (Eds.), in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, vol. 13431, Springer Nature Switzerland, Cham, 2022, pp. 3–13.

[31] S. Basu, K. Wagstyl, A. Zandifar, L. Collins, A. Romero, D. Precup, Early prediction of alzheimer's disease progression using variational autoencoders, in: D. Shen, T. Liu, T.M. Peters, L.H. Staib, C. Essert, S. Zhou, P.-T. Yap, A. Khan (Eds.), in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, vol. 11767, Springer International Publishing, Cham, 2019, pp. 205–213.

[32] Q. Zhao, E. Adeli, N. Honnorat, T. Leng, K.M. Pohl, Variational AutoEncoder for regression: application to brain aging analysis, in: D. Shen, T. Liu, T.M. Peters, L.H. Staib, C. Essert, S. Zhou, P.-T. Yap, A. Khan (Eds.), in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, vol. 11765, Springer International Publishing, Cham, 2019, pp. 823–831.
[33] L. Huang, S. Ruan, Y. Xing, M. Feng, A review of uncertainty quantification in medical image analysis: Probabilistic and non-probabilistic methods, Med. Image Anal. (2024) 103223.
[34] B. Lambert, F. Forbes, S. Doyle, H. Dehaene, M. Dojat, Trustworthy clinical AI solutions: A unified review of uncertainty quantification in deep learning models for medical image analysis, Artif. Intell. Med. (2024) 102830.
[35] F. Jessen, et al., Design and first baseline data of the DZNE multicenter observational study on predementia alzheimer's disease (DELCODE), Alzheimer's Res. Ther. 10 (1) (2018) 15.
[36] C.R. Jack Jr., et al., Update on hypothetical model of Alzheimer's disease biomarkers, Lancet Neurol. 12 (2) (2013) 207.
[37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[38] C. Gaser, R. Dahnke, P.M. Thompson, F. Kurth, E. Luders, A.D.N. Initiative, CAT: A computational anatomy toolbox for the analysis of structural MRI data, GigaScience 13 (2024) giae049.
[39] R. Krishnan, M. Subedar, O. Tickoo, Specifying weight priors in Bayesian deep neural networks with empirical Bayes, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 4477–4484.
[40] I.K.M. Jais, A.R. Ismail, S.Q. Nisa, Adam optimization algorithm for wide and deep neural network, Knowl. Eng. Data Sci. 2 (1) (2019) 41–46.
[41] C. Salvatore, A. Cerasa, I. Castiglioni, MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia 24 months before probable diagnosis, Front. Aging Neurosci. 10 (2018) 135.
[42] K.V. Papp, D.M. Rentz, I. Orlovsky, R.A. Sperling, E.C. Mormino, Optimizing the preclinical alzheimer's cognitive composite with semantic processing: the PACC5, Alzheimer's Dement.: Transl. Res. Clin. Interv. 3 (4) (2017) 668–677.
[43] M.C. Donohue, R.A. Sperling, D.P. Salmon, D.M. Rentz, R. Raman, R.G. Thomas, M. Weiner, P.S. Aisen, et al., The preclinical alzheimer cognitive composite: measuring amyloid-related decline, JAMA Neurol. 71 (8) (2014) 961–970.
[44] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.
[45] Y.L. Rao, B. Ganaraja, B.V. Murlimanju, T. Joy, A. Krishnamurthy, A. Agrawal, Hippocampus and its involvement in Alzheimer's disease: A review, 3 Biotech 12 (2) (2022) 55.
[46] S.G. Mueller, N. Schuff, K. Yaffe, C. Madison, B. Miller, M.W. Weiner, Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's disease, Hum. Brain Mapp. 31 (9) (2010) 1339–1347.
[47] D.D. Cvetković, G.M. Skender, S. Dožić, Neuropathological hallmarks of Alzheimer's disease, Arch. Oncol. 9 (3) (2001) 195–199.
[48] O. Lerch, D. Ferreira, E. Stomrud, D. Van Westen, P. Tideman, S. Palmqvist, N. Mattsson-Carlgren, J. Hort, O. Hansson, E. Westman, Predicting progression from subjective cognitive decline to mild cognitive impairment or dementia based on brain atrophy patterns, Alzheimer's Res. Ther. 16 (1) (2024) 153.
[49] M. Dyrba, et al., Improving 3D convolutional neural network comprehensibility via interactive visualization of relevance maps: Evaluation in Alzheimer's disease, Alzheimer's Res. Ther. 13 (1) (2021) 191.
[50] C. Salvatore, A. Cerasa, P. Battista, M.C. Gilardi, A. Quattrone, I. Castiglioni, A.D.N. Initiative, Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach, Front. Neurosci. 9 (2015) 307.
[51] Z. Jiao, S. Chen, H. Shi, J. Xu, Multi-modal feature selection with feature correlation and feature structure fusion for MCI and AD classification, Brain Sci. 12 (1) (2022) 80.
[52] D. Zhang, D. Shen, A.D.N. Initiative, Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease, NeuroImage 59 (2) (2012) 895–907.
[53] K. Poulakis, J.B. Pereira, J.-S. Muehlboeck, L.-O. Wahlund, Ö. Smedby, G. Volpe, C.L. Masters, D. Ames, Y. Niimi, T. Iwatsubo, Multi-cohort and longitudinal Bayesian clustering study of stage and subtype in Alzheimer's disease, Nat. Commun. 13 (1) (2022) 4566.
[54] H. Yang, et al., Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls, Gen. Psychiatry 32 (2) (2019).
[55] F. Levin, et al., Data-driven FDG-PET subtypes of Alzheimer's disease-related neurodegeneration, Alzheimer's Res. Ther. 13 (1) (2021) 49.
[56] P.-L. Lee, K.-H. Chou, C.-P. Chung, T.-H. Lai, J.H. Zhou, P.-N. Wang, C.-P. Lin, Posterior cingulate cortex network predicts Alzheimer's disease progression, Front. Aging Neurosci. 12 (2020) 608667.
[57] L. Rami, et al., Distinct functional activity of the precuneus and posterior cingulate cortex during encoding in the preclinical stage of Alzheimer's disease, J. Alzheimer's Dis. 31 (3) (2012) 517–526.