AN INTERPRETABLE GENERATIVE MULTIMODAL NEUROIMAGING-GENOMICS FRAMEWORK FOR DECODING ALZHEIMER'S DISEASE

Giorgio Dolci[1,2], Federica Cruciani[1], Md Abdur Rahaman[2], Anees Abrol[2], Jiayu Chen[2], Zening Fu[2],
Ilaria Boscolo Galazzo[1], Gloria Menegaz[1,+], and Vince D. Calhoun[2,+],
for the Alzheimer's Disease Neuroimaging Initiative[*]

[1]Department of Engineering for Innovation Medicine, University of Verona, Verona, Italy
[2]Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS),
Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA

[+]V.D. Calhoun and G. Menegaz equally contributed as last authors to this work.
[*]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

November 15, 2024

ABSTRACT

Alzheimer's disease (AD) is the most prevalent form of dementia, affecting millions worldwide with a progressive decline in cognitive abilities. The AD continuum encompasses a prodromal stage known as Mild Cognitive Impairment (MCI), where patients may either progress to AD (MCIc) or remain stable (MCInc).

Understanding the underlying mechanisms of AD requires complementary analyses relying on different data sources, leading to the development of multimodal deep learning models. In this study, we leveraged structural and functional Magnetic Resonance Imaging (sMRI/fMRI) to investigate the disease-induced grey matter and functional network connectivity changes. Moreover, considering AD's strong genetic component, we introduced Single Nucleotide Polymorphisms (SNPs) as a third channel. Given such diverse inputs, missing one or more modalities is a typical concern of multimodal methods. We hence propose a novel deep learning-based classification framework where a generative module employing Cycle Generative Adversarial Networks (cGAN) was adopted for imputing missing data within the latent space. Additionally, we adopted an Explainable Artificial Intelligence (XAI) method, Integrated Gradients (IG), to extract input features' relevance, enhancing our understanding of the learned representations.

Two critical tasks were addressed: AD detection and MCI conversion prediction. Experimental results showed that our framework was able to reach the state-of-the-art in the classification of CN vs AD with an average test accuracy of 0.926 ± 0.02. For the MCInc vs MCIc task, we achieved an average prediction accuracy of 0.711 ± 0.01 using the pre-trained model for CN and AD. The interpretability analysis revealed that the classification performance was led by significant grey matter modulations in cortical and subcortical brain areas well known for their association with AD. Moreover, impairments in sensory-motor and visual resting state network connectivity along the disease continuum, as well as mutations in SNPs defining biological processes linked to endocytosis, amyloid-beta, and cholesterol, were identified as contributors to the achieved performance. Overall, our integrative deep learning approach shows promise for AD detection and MCI prediction, while shading light on important biological insights.

Keywords Alzheimer's disease continuum · Generative model · Imaging genetics · Integrated Gradients

1 Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disorder that affects millions of people worldwide (approximately 30 million in 2015 ([1]). It is the most common cause of cognitive impairment, gradually impacting the activities of a patient's daily life. It is characterized by the progressive loss of cognitive and functional abilities, including memory, language, and executive functions ([2]), with a temporal progression. Amyloid accumulation represents the first event, followed by tau accumulation, hypometabolism (assessed with Positron Emission Tomography (PET)), atrophy, and cognitive decline ([3, 4]). It is hence evident that the pathology changes of AD actually begin several years before the first clinical symptoms. Therefore, a timely AD diagnosis is highly beneficial for optimizing patient care and enabling appropriate therapeutic interventions. Mild cognitive impairment (MCI) is the intermediate stage from normal cognitive function to AD, hence representing an opportunity for an early targeting of the disease. However, it includes a very heterogeneous class of patients, including subjects that will likely convert to AD, known as MCI converters (MCIc), with an estimated annual conversion rate around the 16.5% ([5]), and subjects that remain stable after several years, being part of the MCI non-converters (MCInc) group ([6]).

Among the available neuroimaging technologies, structural Magnetic Resonance Imaging (sMRI) and resting-state functional MRI (rs-fMRI) have provided unprecedented opportunities for deriving biomarkers allowing the early diagnosis of AD. For instance, sMRI is currently a key part of the diagnostic criteria for the differential diagnosis and longitudinal monitoring of patients with dementia, enabling the estimation of brain atrophy. Several studies have consistently observed both global and local atrophic changes in AD, lying along the hippocampal pathway (entorhinal cortex, hippocampus, parahippocampal gyrus, and posterior cingulate cortex) in the early stages of the disease, while atrophy in temporal, parietal and frontal neocortices emerges at later stages being associated with neuronal loss as well as with language, visuospatial and behavioral impairments ([7, 8]). Rs-fMRI, in turn, indirectly measures neural activity by detecting changes in the Blood Oxygenation Level Dependent (BOLD) signals, which depend on the neurovascular coupling ([9]). In particular, investigating functional connectivity (FC; inter-regional coupling), functional network connectivity (FNC; inter-network coupling), and functional networks from BOLD rs-fMRI provides a means for understanding the mechanisms and relevance of the functional relationships across brain regions. A growing body of rs-fMRI studies suggests that failure of specific resting-state networks (RSNs) is closely related to AD, with the default-mode network (DM) and the salience network (SN) playing a pivotal role and being disrupted before clinically evident symptoms ([10]). Specific alterations in these functional networks have been reported in AD patients, with prominent FC decreased within the DM and increased FC in the SN ([11, 12]). Moreover, disconnections within and between the different RSNs have been consistently demonstrated, particularly over long connection distances ([13]). All of these factors have contributed to the widespread view of AD as a disconnection syndrome being characterized by a cascading network failure, beginning in the posterior DM and then shifting to other systems containing prominent connectivity hubs, possibly associated with amyloid accumulation ([14]). Moreover, increased evidence supports the view that tau depositions are also related to functional brain architecture and FC changes, supporting the view of transneuronal tau propagation in AD ([15]).

AD also features a strong genetic component. Strategies to extract linked genotype traits are commonly based on Single Nucleotide Polymorphism (SNPs) analysis ([16]), representing variations of single nucleotides in DNA sequences that vary from person to person ([16]) and are present in at least 1% of the population. Several Genome-Wide Association Studies (GWAS) have identified more than 40 AD-associated genes/loci, which are likely to increase the risk of developing the disease ([17, 18, 19]). Among them, the apolipoprotein E (APOE), in particular the $\epsilon 4$ allele, PICALM, CLU, ABCA7, and CR1 are the most important genes being associated with AD risk factor or the progression from MCI to AD ([20], [21], [22], [23], [24], [25], [26, 27]).

This great heterogeneity of available biomarkers offers a unique opportunity to explore various aspects of the disease continuum. Each biomarker provides valuable information about specific characteristics, enabling a multifaceted investigation of AD which calls for methods that can effectively integrate and leverage complementary information. In this regard, artificial intelligence, Machine Learning (ML) and particularly Deep Learning (DL), emerge as promising technologies to tackle this complex task. Different ML algorithms have been developed for integrating heterogeneous data, such as Canonical Correlation Analysis (CCA), Partial Least Squares (PLS), and ensemble methods, e.g. consisting of different Support Vector Machines (SVM). Even if these methods can perform well in studying neurodegenerative diseases ([28, 29]), they have some limitations, such as requiring vectorized input data, hence denaturing the spatial information of 3D volumes (e.g., sMRI or PET 3D images), and potentially being too simple for solving complex tasks. On the other hand, by employing multiple layers of processing, DL models allow extracting progressively higher-level and more informative features from input data, with the option of preserving spatial information, without the need of vectorizing the input features. Moreover, the inclusion of multiple data sources into a single model can uncover

complex and deep non-linear associations across the input features from a multimodal perspective ([30]). As a result, multimodality is gaining significant popularity representing the key approach to derive valuable insights into complex and multifaceted neurodegenerative diseases such as AD ([31], [32], [33]). However, despite the promising foreseen of such an approach, multiple drawbacks are present and represent the focus of the current research. The main concern resides in missing data management. Particularly in the biomedical domain, it is very common to incur in missing values or acquisitions for certain subjects or entire study cohorts due to different reasons, such as missing acquisition, corrupted data, patients dropout from a study ([34]), and the requirements of privacy and expensive tests. Additionally, the interpretability of a model's predictions is a central characteristic of decision-aiding models, which is still not pervasively addressed. In the past years, many high-performing prediction models have been proposed lacking a clear rationale to be effectively considered for practical use. Strong and validated explanations associated with a given prediction are fundamental for increasing trust in the results as well as for their applicability in a real-world scenario. Innovative and viable missing modality management as well as interpretability are the key attributes that a multimodal model for diagnosis detection should have.

Diving into missing data management, the simplest and most common solutions consist either in discarding samples with missing modalities, or in filling the missing values with zeros [35], and computing imputations based on data interpolation [36]. Such solutions are evidently suboptimal since they could significantly reduce the number of training samples, already small when addressing biomedical-related tasks, or introduce important biases in the data and the model due to the interpolation or the zero filling. Alternative methods to exploit the information of all the available subjects were proposed. A possible approach is the complementation of incomplete data representation, which consists in extracting a latent representation from both the complete and incomplete modalities, avoiding the need for data imputation ([37, 38]). Most recent approaches aim at generating missing data either in the input space ([39]) or in an intermediate latent representation ([40, 41]) relying on generative models such as Generative Adversarial Networks (GANs) ([42]) and their variants, or Variational Autoencoders ([43]). This last approach has been successfully applied to the AD continuum investigation. It allows the exploitation of the availability of multimodal data for capturing the relationship among different data sources in the latent space, enabling the generation of one modality from another. Generative models, hence, appear as the route to be followed when dealing with missing data. Interestingly, the recently proposed Cycle-consistent GANs (Cycle-GANs) [44] have shown impressive results in various knowledge translation tasks since they offer a flexible and effective approach for learning mappings across different domains without relying on paired data.

Moving to model interpretability, it is well established that with increasing model complexity, interpretability decreases drastically. However, in order to better understand the mechanisms that underlie the AD continuum and allow the models to be applied in clinical and real-life applications, it is vital to understand the reasons behind a certain output. In this case, Explainable Artificial Intelligence (XAI), becomes essential to understand why a given model made a certain prediction. XAI encompasses *post-hoc* methods that allow the assignment of an importance score to each feature, reflecting its role in the classification task. This means finally opening the 'black box' of complex models hopefully increasing their exploitability in clinical practice. XAI is starting to be applied in multimodal frameworks to study neurodegenerative and psychiatric diseases. Few works could be found adopting simple gradient-based or feature perturbation methods ([45, 46, 47]). On top of this, another overlooked yet stringent aspect is the strong validation of XAI methods and the obtained explanations. The evaluation of explanation methods is still under-investigated, however, since explainability is meant to increase confidence in AI, it is vital to systematically analyze the obtained results referring also to the prior knowledge derived from the state-of-the-art.

In this rapidly evolving landscape, where multimodality holds a central role, we aim at shading light on the importance of the complementary information that could be derived from advanced biomarkers such as brain FC and SNPs mutations, together with the well-established brain atrophy measures derived from sMRI, in detecting AD-related modulations. With this purpose, we will present a novel multimodal framework for AD detection and MCI conversion prediction, which addresses (i) heterogeneous data integration, (ii) missing data management, and (iii) interpretability. Regarding the former, our framework allows the integration of heterogeneous data featuring different dimensionality, e.g., 3D volumes and vectorized data, and nature, meaning from multi-domain, e.g., imaging and genetics. Concerning the second, we propose an approach that allows generating missing modalities in the latent space obtained after input feature reduction, ensuring high accuracy in reconstructing latent features while minimizing computational demands. The goal is to define a framework that can be generalized to any missing modality, without requiring to have at least one specific modality shared by all the subjects in the considered population. Finally, we aimed at emphasizing the interpretability of the proposed framework in order to reinforce its transparency and reliability by conducting a post-hoc interpretability analysis, supplemented by a robust validation step, which enables to precisely discern the contribution of each input feature to the classification of subjects in the AD continuum.

Table 1: Demographic information of the AD and CN study cohort. MMSE = Mini-Mental State Examination; wild (W) and mutated (M) alleles in APOE e4: 0 = homozygous (W/W), 1 = heterozygous (W/M), 2 = homozygous (M/M).

Cohorts	CN			AD		
	All subjects	sMRI and fMRI	sMRI and SNPs	All subjects	sMRI and fMRI	sMRI and SNPs
Count	644	321	253	332	66	152
Age (y)	73.6 ± 6.6	72.4 ± 7.1	76.4 ± 5.4	75.1 ± 7.9	74.7 ± 8.1	75.8 ± 7.7
Gender (M/F)	284/360	117/204	137/116	181/151	35/31	84/68
MMSE	29.1 ± 1.1	29.1 ± 1.2	29.1 ± 1.1	23.1 ± 2.2	22.6 ± 2.6	23.3 ± 2.0
APOE e4 (0/1/2)	455/168/21	224/87/10	189/57/7	109/156/67	19/34/13	49/76/27

Table 2: Demographic information and missing data percentage of the MCI study cohort. MMSE = Mini-Mental State Examination; wild (W) and mutated (M) alleles in APOE4: 0 = homozygous (W/W), 1 = heterozygous (W/M), 2 = homozygous (M/M).

	MCInc	MCIc
Count	646	289
Age (y)	73 ± 8.5	74 ± 8.5
Gender (M/F)	374/272	172/117
MMSE	27.9 ± 1.8	26.9 ± 1.8
APOE4 (0/1/2)	377/205/64	103/143/43
Missing SNPs (%)	59.6%	30.8%
Missing fMRI (%)	60.1%	92.0%

2 Material and Methods

2.1 Dataset

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`https://adni.loni.usc.edu/`). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see `www.adni-info.org`.

For each subject, the baseline timepoint was retained for all the selected imaging modalities, that is, sMRI and rs-fMRI. The respective genetic variants (SNPs) were then selected, if available. The final study cohort included 1911 subjects, divided into healthy controls (CN), AD, MCI non-converter (MCInc), and MCI converter (MCIc). In detail, following the classification available on the ADNI website, an MCI subject was considered as MCIc if dementia was diagnosed at any timepoint after MCI diagnosis.

Table 1 shows the AD and CN study cohorts demographic information highlighting the subjects sharing modalities, in particular, sMRI-fMRI and sMRI-SNPs, while Table 2 details the MCI study cohort including the percentage of missing data for each considered modality.

3D T1-w MRI and rs-fMRI acquisitions were considered as imaging input channels and were acquired with the following sequence parameters: 1) sagittal accelerated MPRAGE, TR/TE = shortest, TI = 900 ms, flip angle = $9°$, Field Of View = 256×256 mm^2, spatial resolution = $1 \times 1 \times 1$ mm^3, slices = $176 - 211$), 2) rs-fMRI: TR/TE =3000/30 ms, FA = $90°$, FOV = $220 \times 220 \times 163$ mm^3, 3.4-mm isotropic voxel size. 200 fMRI volumes were acquired in almost all subjects, with minimal variations in a small subset (e.g., 197 or 195 volumes). More details about the data acquisition can be found in [48]. Concerning the genetic data, DNA samples were genotyped using Illumina Human610-Quad or Illumina HumanOmniExpress BeadChip.

2.2 Preprocessing and feature engineering

The sMRI volumes preprocessing included tissue segmentation in Gray Matter (GM), White matter, and Cerebrospinal fluid (CSF) using the modulated normalization algorithm. Only the GM volume was considered for this study, to which a smoothing using a Gaussian kernel (FWHM = 6mm) was applied. Quality control (QC) included discarding images that exhibited low correlation with individual and/or group level masks. The full preprocessed GM volume was used as input for the sMRI channel resulting in an input size of $121 \times 145 \times 121$ for each subject.

The rs-fMRI data was preprocessed using the statistical parametric mapping toolbox (SPM12, `http://www.fil.ion.ucl.ac.uk/spm/`) including rigid body motion correction to correct subject head motion, slice-timing correction, warping to the standard MNI space using the EPI template, resampling to $(3mm)^3$ isotropic voxels, and smoothing using a Gaussian kernel (FWHM = 6mm), following the preprocessing proposed in [49]. QC procedure was the same as for sMRI. Fifty-three independent components (ICs) covering the whole brain were extracted using spatially constrained ICA with the Neuromark_fMRI_1.0 template (available in the GIFT software; `http://trendscenter.org/software/gift`). For each subject, a correlation matrix was then created computing the Pearson correlation between the ICs time courses, resulting in a 53x53 static functional network connectivity (sFNC) matrix. This was divided into 7 RSNs, named: (i) Sub-cortical network (SC), (ii) Auditory network, (iii) Sensorimotor network (SM), (iv) Visual network (VI), (v) Cognitive-control network (CC), (vi) Default-mode network (DM), and (vii) Cerebellar network (CB) (please, refer to Supplementary Table 2 for more information). The upper triangular matrix was then vectorized, resulting in an input vector of 1378 features for each subject.

Moving to the genomics data, pre-imputation QC was performed to remove Single Nucleotide Polymorphisms (SNPs) with minor allele frequency (MAF) < 0.05, call rate < 0.98, and Hardy Weinberg Equilibrium $< 10^{-3}$ (details in `https://www.synapse.org/#!Synapse:syn2290704/wiki/64710`). Imputation was performed with the same reference panel. Only the SNPs with imputation $r^2 > 0.4$ were included. Linkage disequilibrium (LD)-based pruning with $r^2 = 0.8$ in a window of 50kb was applied, yielding 445838 SNPs for further analyses ([50]). A feature selection leveraging on the genome-wide association study (GWAS) on AD conducted by [17] was carried on to include all the relevant SNPs having GWAS p-values less than $1e^{-04}$. A total of 565 polymorphisms was selected and a value between 0 and 2 was assigned based on the presence of mutated alleles. In detail, defining wild and mutated alleles respectively as W and M, a score of 0 means wild homozygous alleles (W/W), a score of 1 indicates heterozygous alleles (W/M), and a score of 2 defines mutated homozygous alleles (M/M).

2.3 Classification tasks

Following the aim of this study, the proposed architecture was devised with a twofold classification aim: (1) AD detection, that is the differentiation of AD and CN subjects, also referred as Task 1 and (2) MCI conversion prediction, that is the stratification of MCIc and MCInc patients, also referred to as Task 2. In detail, the network was trained and tested on Task 1 and subsequently used to solve Task 2 allowing to assess its ability in discriminating different stages of disease and also in capturing and highlighting shared patterns across the different categories.

2.4 Framework architecture

The proposed framework builds on our previous work [32] and is shown in Figure 1. In detail, the architecture for disease detection consists of three modules: i) *Feature reduction module*, which performs a CNN-based feature extraction to derive a lower dimensionality latent space, separately for each input channel; ii) *Generative module* that, in the eventuality of missing modalities, actuates a generative process in the latent space transferring the knowledge from one domain to another; and iii) *Data fusion & classification module* that fuses the latent features obtained for the three modalities and then performs the classification. Post-hoc interpretability analysis was then carried on in order to retrieve features' contributions to the classification task. The three modules will be detailed in the following paragraphs.

Feature reduction module The feature reduction module consists of three different CNNs, one for each input channel, leading to a latent low dimensionality representation consisting of 100 latent features for each channel. The sMRI channel was analyzed through a 3D CNN defined by three successive convolutional blocks, each composed of two consecutive convolutional layers (filter sizes of $3 \times 3 \times 3$ for the first four convolutional layers and $2 \times 2 \times 2$ for the last two, number of filters: 64, 64, 64, 128, 128, and 128, respectively, for the six convolutional layers) and one max-pooling layer, followed by four fully-connected layers (FCLs) (number of nodes: 1536, 768, 384, and 192, respectively).
The rs-fMRI channel was analyzed through a 1D CNN, consisting of four convolutional layers (filter sizes 5, number of filters: 64, 128, 128, and 128, respectively), two max-pooling layers and two FCLs (number of nodes: 384 and 100, respectively).
Finally, for genetic data, a 1D CNN was employed, enclosing three convolutional blocks, each including a sequence of convolutional layers (filter sizes 3, number of filters: 64, 64, and 128, respectively, for the three convolutional layers), batch-normalization and max-pooling, followed by three FCLs (number of nodes: 1024, 512, and 128, respectively). The ReLU activation function ([51]) was used for all the layers of these three CNNs.

Generative module The generative module allows imputing the missing modalities in the latent space given the others, when necessary. This task is possible thanks to the injection in the complete framework of four pre-trained generators derived from two different cGANs ([44]). The first, the sMRI-SNPs-cGAN allows imputing the latent genetics features
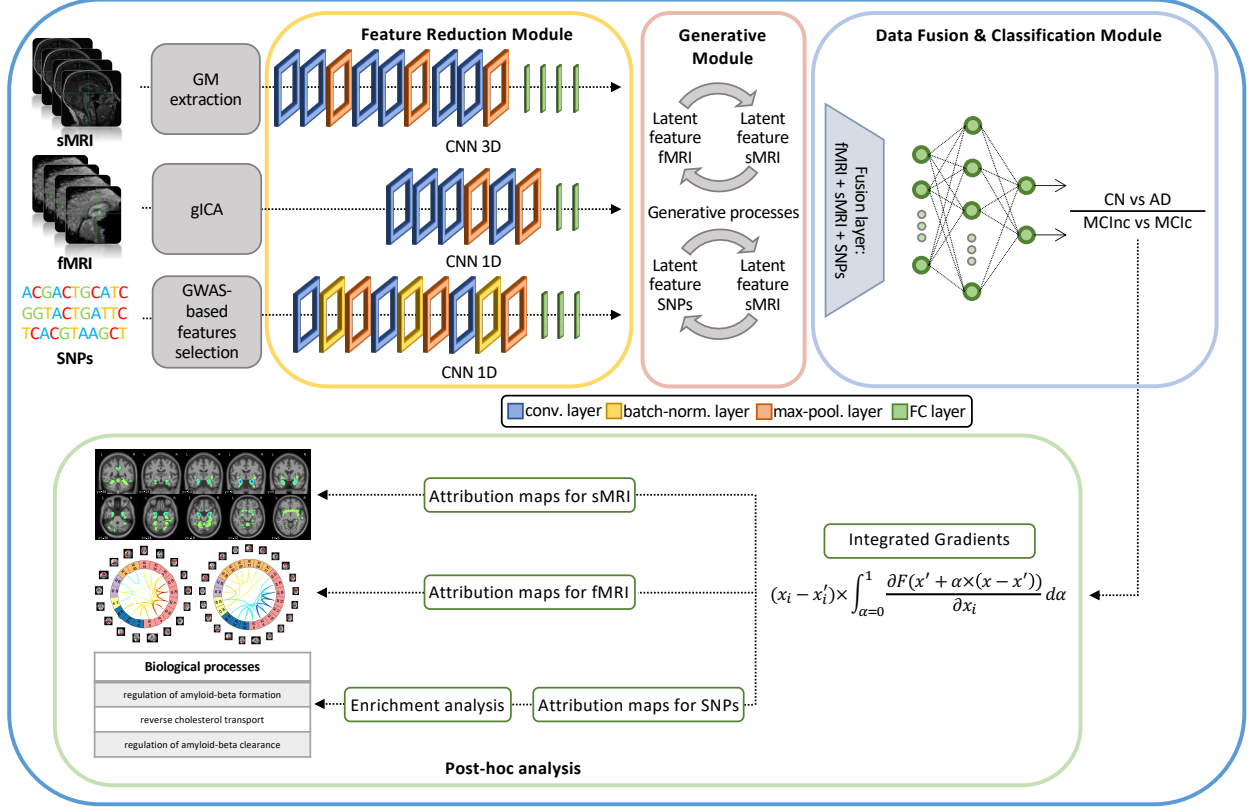
Figure 1: Schematic representation of the proposed framework. Given three input modalities, structural Magnetic Resonance Imaging (sMRI), functional MRI (fMRI) and Single Nucleotide Polymorphisms (SNPs), the framework is articulated in three different modules: i) A feature reduction module where the input channels are transformed in their low dimensionality latent representations; ii) A generative module where, if necessary, the missing modalities are imputed transferring the knowledge from one domain to another; and iii) A data fusion & classification module where the latent features are fused together and then classified using a Multilayer Perceptron. Finally, *post-hoc* interpretability analysis is performed relying on the Integrated Gradients (IG) method for feature attribution derivation.

transferring the knowledge from the latent sMRI one and *vice-versa*, and the second, the sMRI-fMRI-cGAN, transfers the knowledge again from the latent sMRI features to generate the respective rs-fMRI ones, and *vice-versa*. The scarcity of subjects sharing all modalities impeded the development of an fMRI-SNPs-cGAN model.

The cGANs were built and trained prior to the full framework. In detail, to obtain the pretrained generators, the first step required to obtain the channels' latent representations (i.e., the 100 features vector for each modality) needed to subsequently train the cGANs. This was necessary since, as already highlighted, the proposed framework performs the data generation directly in the latent space and not in the original input space. Since the subset of subjects having all three modalities was not numerous enough to train a classification model, two separate bi-modal models were developed, one having sMRI and SNPs (sMRI-SNPs-NN) as input channels and the other having sMRI and rs-fMRI as input channels (sMRI-fMRI-NN). Fig. 2a shows the schematic representation of the sMRI-fMRI-NN, as a reference example. In order to obtain the same latent features as the complete framework, the feature reduction for each channel and the data fusion & classification modules of these bi-modal models were the same as the ones adopted in the complete framework, and described in the previous paragraph. Once the latent vectors were obtained, they were given as input to two cGANs, namely sMRI-fMRI-cGAN and sMRI-SNPs-cGAN, one for each bi-modal model, and hence for each information transfer, from sMRI to rs-fMRI and from sMRI to SNPs, respectively. Fig. 2b shows the sample sMRI-fMRI-cGAN architecture used to generate sMRI from rs-fMRI and *vice-versa*. The generators are composed of six FCLs (number of nodes: $256, 512, 512, 1024, 512,$ and 100, respectively) activated by a ReLU activation function. The discriminators consisted of four FCLs (number of nodes: $256, 128, 64,$ and 1, respectively) activated by the LeakyRelu function ([51]), alternated with three dropout layers (dropout probability: 0.3). The same architecture was considered for the sMRI-SNPs-cGAN.
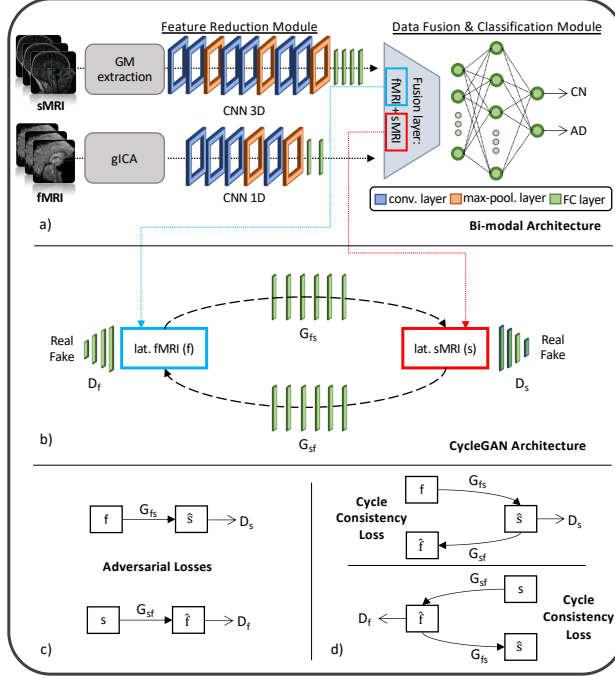
Figure 2: Schematic representation of a sample bi-modal framework (sMRI-fMRI-NN) followed by the respective cGAN (sMRI-fMRI-cGAN). (a) The sMRI-fMRI-NN is composed of two different modules: feature reduction module, equal to the respective module for each channel in the full framework, and data fusion & classification module where the latent features are fused together and then classified. (b) After training, the latent features of both modalities obtained from the feature reduction module were extracted for each subject and given as input for the training of the associated cGAN, whose loss is described in (c) and (d).

Data fusion & classification module   The data fusion & classification module consists of a fusion layer and a classifier. In this module, the latent features obtained either from the data reduction module or the data generation module (for the subjects with missing modalities) are fused together through vector concatenation, resulting in 300 features. A Multilayer Perceptron (MLP) composed of three FCLs (number of nodes: 150, 75, and 2, respectively) activated by the ReLU function was then used for classification. Softmax was used as activation function for the output layer, allowing to obtain the classification probability associated with the input data.

2.5   Training scheme

Figure 3 illustrates the training workflow of the entire framework. As already specified, the training phase was not performed end-to-end but was divided into two steps, necessary for the correct training of the generation module: (i) Step 1: Training of bi-modal models and cGANs, and (ii) Step 2: Training of the full multimodal framework.

Hyperparameter tuning was performed empirically. The number of layers, channels, nodes, and the filters' size were tuned considering the single-modality architectures for the classification (Task 1) before training the multimodal framework. The weights of the CNNs and classifier, used in both bi-modal models and the final pipeline, were initialed as follows: the 3D CNN for sMRI, the 1D CNN for SNPs, and the classifier using the Xavier Uniform distribution, while Xavier Normal distribution was used for the 1D CNN for rs-fMRI. The weights of the generators and discriminators were initialized following a Xavier Normal distribution ([52]) for both the sMRI-fMRI-cGAN and sMRI-SNPs-cGAN. The weights initialization was independent between Step 1 and Step 2.

Training Step 1: Bi-modal models and cGANs   In the training Step 1 the two sMRI-fMRI-NN and sMRI-SNPs-NN bi-modal models were trained for the AD versus CN classification (Step 1a, Fig. 3) in order to subsequently extract the latent features to be fed to the two cGANs, sMRI-fMRI-cGAN and sMRI-SNPs-cGAN (Step 1b, Fig. 3). More in-depth, the bi-model models and subsequently the cGANs were trained relying on the AD and CN subjects sharing respectively the sMRI and rs-fMRI acquisition and the sMRI and SNPs acquisition, presented in Table 1. A 10-folds stratified Cross Validation (CV) procedure for a total of 40 epochs, Adam optimizer ([53]) with a learning rate of
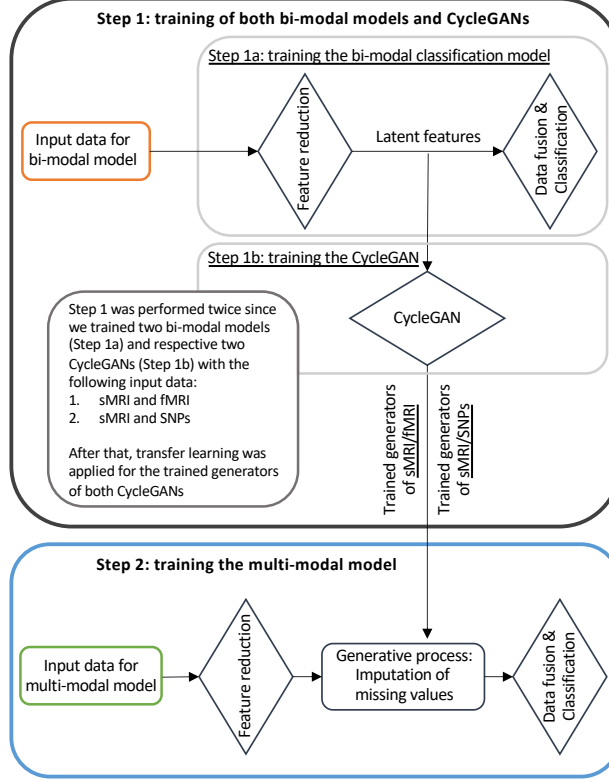
Figure 3: Framework training scheme. The training was performed in two steps: Step 1 involved the training of two bi-modal models (sMRI-fMRI-NN and sMRI-SNPs-NN, Step 1a) and the respective cGANs (sMRI-fMRI-cGAN and sMRI-SNPs-cGAN, Step 1b); Step 2 was needed for the training of the complete multimodal framework.

0.0001, and mini-batch technique considering a batch size of 14 and 9 were selected, respectively, for the training of the sMRI-fMRI-NN and the sMRI-SNPs-NN. Weighted Cross Entropy (CE) was used as the loss function. Differently from the classical CE, it allows computing a weight guaranteeing higher value to the less numerous class, to address the unbalanced classes issue. The sMRI-fMRI-NN and the sMRI-SNPs-NN achieved an average validation accuracy of 0.89 ± 0.03 and 0.864 ± 0.03, respectively. The best model in terms of validation accuracy was retained for both the NNs and used to extract the 100 latent features for each subject and for each modality.

The newly obtained feature set was used to train the two cGANs. A mini-batch technique, keeping the same subjects in each batch as the bi-modal models, was used for both the sMRI-fMRI-cGAN and the sMRI-SNPs-cGAN training. Adam optimizer with a learning rate of 0.0001 was used to train the generators for a total of 600 epochs. As in [44], the *adversarial losses* were considered as optimization targets for each cGAN as in Fig. 2c. In detail, considering the sMRI-fMRI-cGAN as reference, given s and f the latent feature vectors for the sMRI and rs-fMRI, G_{fs} and G_{sf} the generators to obtain s from f and *vice-versa* and D_s and D_f the discriminators to distinguish the s from the generated $\hat{s}$ and the f from the generated $\hat{f}$, respectively, the adversarial loss penalizes the D_s and D_f errors in discriminating between $\hat{s} = G_{fs}(f)$ and s, and $\hat{f} = G_{sf}(s)$ and f, respectively. The major issue arising when considering this loss solely is that with a large enough capacity, the generators G_{fs} and G_{sf} can map the same set of input images to any random permutation of images in the target domain (latent features in our case) ([44]). The cycle consistency losses were hence introduced to limit the space of possible mapping functions by also penalizing the difference between f and $G_{sf}(G_{fs}(f))$, and between s and $G_{fs}(G_{sf}(s))$, respectively, hence completing the cycle (Fig. 2d).

The four pre-trained generators, two for the sMRI-fMRI-cGAN and two for the sMRI-SNPs-cGAN, were then extracted and injected in the latent space of the proposed multimodal framework for the on-line generation of the missing modalities during the full framework training.

Of note, no data leakage was present between the training Step 1 (bi-modal models and cGANs) and the training Step 2 (full multimodal framework) since Step 2 involved end-to-end training using subjects with no missing modalities (i.e., using input data for these subjects only) and generative models trained in Step 1b for deriving latent features for

subjects with missing modalities only. Of note, the full cohort of subjects was not used to train the bi-modal models and corresponding cGANs (Step 1a and 1b), only relying on those subjects sharing the modalities under analysis (sMRI and rs-fMRI for sMRI-fMRI-NN/cGAN, and sMRI and SNPs for sMRI-SNPs-NN/cGAN).

The cycle consistency loss was used in both sMRI-fMRI-cGAN and sMRI-SNPs-cGAN in order to assess the performance of the generators in imputing missing modalities using the Mean Absolute Error (MAE).

Training Step 2: Full multimodal framework   The full multimodal framework including the three input channels, sMRI, rs-fMRI, and SNPs, was trained for classification Task 1, AD detection, relying on the complete AD and CN data cohort.

Subjects were split into training and testing sets, respectively, including the 80%-20% of the study cohort. The testing set was kept unseen until the last testing phase. A stratified 10-fold CV procedure was applied to the training set. A mini-batch strategy (12 subjects per batch) was adopted during training. Adam optimizer with a learning rate of 0.0001 was chosen and the model was trained for a total of 70 epochs. Weighted CE was considered as the loss function. During this training phase, the four pre-trained cGAN generators' weights were kept frozen. Indeed, during the backpropagation phase, only the weights of the feature reduction module and the data fusion and classification module were updated.

2.5.1   Testing and performance evaluation

The best model, in terms of validation accuracy, obtained from the CV procedure of training Step 2 was considered for the testing phase for the two classification tasks presented in Section 2.3. The testing set retained from the AD versus CN study cohort splitting was used for classification Task 1, AD detection. The full MCI cohort, never considered during the training phase, hence kept completely unseen by the model, was considered as a testing set for the classification Task 2, MCI conversion prediction. The framework was finally evaluated in terms of accuracy (ACC), precision (PRE), recall (REC), F1 score, Area Under Precision Recall Curve (AUPRC), and Area Under the Curve (AUC) on both testing sets for the two classification tasks.

Five independent runs reshuffling the train and test sets were performed for probing the generalization capabilities of the model, and the average test performance over the different runs was retained. All the training and testing described were performed in Python, specifically relying on the Pytorch library ([54]).

3   Post-hoc interpretability analysis

The post-hoc interpretability analysis was performed relying on Integrated Gradients (IG) ([55]), aiming at obtaining attribution maps describing the relevance score associated with each input feature on the different input channels. IG maps were derived for the subjects in the testing set of the best model over the five generalization runs for both the classification tasks (Task 1: AD detection and Task 2: MCI prediction conversion, which coincides with the full MCI cohort). A brief introduction of the IG method and its properties followed by the attribution analysis performed for each input channel will be given in the following sections.

3.1   Integrated Gradients (IG)

Integrated Gradients (IG) ([55]) is an interpretability method allowing to find the relevance of the input features with respect to the outcome of the model without requiring any modification to the network under analysis. It is a Baseline Attribution Method (BAM) ([56]). To obtain feature attributions, IG computes the path integral of the gradients along the straightline path from a given baseline x' to the input x ([55]) following the equation

$$IG_i(x) ::= (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial F(x_i' + \alpha \times (x_i - x_i'))}{\partial x_i} d\alpha \tag{1}$$

where F is the model and α is a factor ranging in $[0, 1]$, determining the step size along the straightline path from x to x'.

This method is gaining high popularity among the *post-hoc* interpretability methods since it satisfies three important axioms: (i) *Sensitivity*, for which a null score is given to the input features which do not contribute to the prediction, (ii) *Implementation invariance*, hence given a certain input, the derived attributions for two functionally equivalent networks are as well equivalent, and (iii) *Completeness*, which ensures that the attributions approximately add up to the difference between the input and the baseline prediction scores ([55]).

Following the definition, IG attribution scores can assume both positive and negative values. The larger the absolute value of the score of a given feature is, the higher its contribution to the prediction outcome is, either with a positive or negative effect.

It is evident how the baseline choice impacts on the obtained attribution maps. Some recent studies have investigated this issue in order to raise awareness among the users and show the impact of different baselines on the IG scores obtained for the same input and model ([57, 58, 59]). Of note, the main requirement for the baseline is to represent a neutral input for the model under investigation. In other words, it should produce a zero score or, equivalently, around 0.5 prediction probability for both classes resulting from the activation function of the final classification layer. For the proposed framework, a triplet of neutral baselines, one for each input modality, was selected after exhaustive empirical research, defined as follows: (i) *3D matrix of zeros* for sMRI, (ii) *1D vector of zeros* for rs-fMRI, and (iii) *1D vector of Gaussian noise* for SNPs, which allowed on average to obtain random classification scores.

3.2   IG feature attribution analysis

IG allows obtaining attribution maps for each input and channel, that lies in the same input space. In our study IG was applied on the testing set of each classification task, resulting in a triplet of attribution maps representing sMRI, rs-fMRI, and SNPs feature attributions lying in the same space as the input data, for each subject and task. The different maps will be referred to as sMRI-IG, fMRI-IG, and SNPs-IG in what follows.

3.2.1   Neuroimaging

In order to retain the most relevant features, an initial sMRI-IG map thresholding was applied, selecting only the attribution values exceeding in absolute value the 99.5^{th} percentile of the respective IG values distribution voxelwise. The average sMRI-IG values for 55 cortical and subcortical brain regions based on the Harvard-Oxford atlas ([60]) were then extracted for all the subjects in the testing sets of both classification tasks. Only the regions where at least 99% of the subjects had average values exceeding the percentile threshold were kept for subsequent analysis.

For fMRI-IG, only the connections whose attribution scores were in absolute value above the 98th percentile of the respective IG value distribution connection-wise were considered for further analysis. Different thresholds were chosen for the IG maps due to the different input sizes.

A statistical group-based analysis was then performed considering the attribution maps derived for each class, namely CN, MCInc, MCIc, and AD. Shapiro-Wilk test of normality ([61]) was initially performed on the sMRI-IG and fMRI-IG derived features, revealing non-normally distributed values. The non-parametric Kruskal-Wallis test ([62]) was hence performed to check the statistical difference across the four classes considered in each feature, derived either from rs-fMRI (most relevant connections resulting from fMRI-IG) or sMRI (most relevant brain regions derived from the sMRI-IG). When significance was found, the pairwise Wilcoxon Rank Sum Test ([63]) was performed between all the possible pairs of the four considered groups (AD-CN, AD-MCIc, AD-MCInc, CN-MCIc, CN-MCInc, MCIc-MCInc) and adjusted for multiple comparisons with Bonferroni correction. All the statistical analyses were performed in R.

3.2.2   Genetics

Only the SNPs with attribution greater than the 65th percentile of the SNPs-IG distribution, hence those having positive attribution, were kept. The Ensemble Variant Effect Predictor (VEP) tool ([64]) was used to annotate the selected SNPs in their corresponding genes. Web interface of VEP (`https://useast.ensembl.org/Tools/VEP`) and default settings were used to run the analysis. Selected settings included finding the co-located known variants and 1000 Genomes global minor allele frequency as frequency data for co-located variants parameters. No filtering was applied to the analysis to not exclude relevant biological data. Enrichment analysis was then performed using the Gene Ontology enrichment analysis tool ([65, 66, 67]) to derive the most important biological processes in which the annotated genes were involved. A background of 334 genes was generated by annotating the 565 SNPs-set used as input data for the model. Bonferroni correction for multiple testing was performed, and both corrected and no corrected p-values were reported. The SNPs annotation, as well as the enrichment analysis, was computed only on the classes of patients, namely the AD, MCInc, and MCIc individuals.

Classification performance and IG attribution analysis results will be firstly presented separately for classification Task 1 and Task 2. An overview of the overall group-based analysis will then follow.

3.3   Task 1: AD detection

The generators for both sMRI-fMRI-cGAN and sMRI-SNPs-cGAN models were tested by computing the MAE across the validation sets to compare the true latent space and the reconstructed one. For the sMRI-fMRI-cGAN, the generators achieved a MAE of 0.065 ± 0.01 and 0.074 ± 0.01 for the fMRI and the sMRI modalities, respectively. Meanwhile, the sMRI-SNPs-cGAN generators reached a MAE of 0.506 ± 0.04 and 0.333 ± 0.10 for SNPs and sMRI data, respectively.

The proposed framework reached 0.926 ± 0.02 in ACC, 0.910 ± 0.05 in PRE, 0.876 ± 0.03 in REC, 0.891 ± 0.03 in F1 score, 0.829 ± 0.03 in AUPRC, and 0.960 ± 0.01 in AUC for the differentiation between AD and CN subjects. Of note, the best model over the five generalization runs reached 0.964 in ACC, 0.984 in PRE, 0.909 in REC, 0.945 in F1, 0.876 in AUPRC, and 0.970 in AUC.
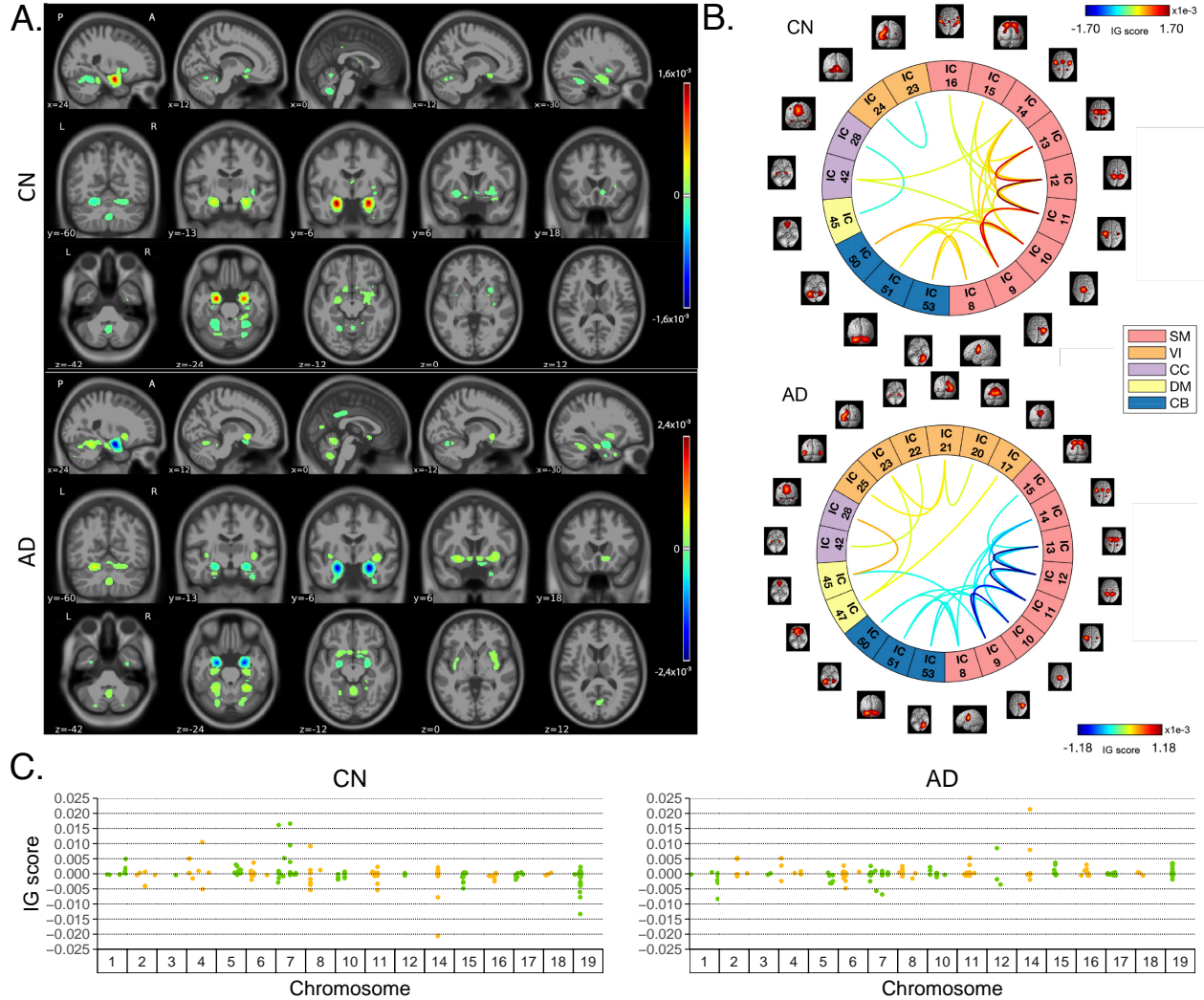
3.3.1 Feature relevance

Figure 4: Overview of IG attribution maps for the classification Task 1, AD detection. All the IG maps are presented for the correctly classified CN and AD subjects in the testing set. A. Average sMRI-IG maps, thresholded to retain IG scores exceeding the 99.5^{th} percentile and overlaid to MNI152 template; B. Average fMRI-IG derived connectograms, thresholded to retain the connections with an associated IG score over the 98th percentile; C. Average SNPs-IG scores highlighting the SNPs with an associated positive IG score exceeding the 65th percentile, SNPs are grouped by chromosome.

Fig. 4 shows the average IG maps obtained for the AD detection task and for each input channel, averaged over the correctly classified subjects per class (test set). The *jet* colorbar was used to highlight the attribution scores, where red and blue correspond to positive and negative attribution values, respectively. In detail, Fig. 4A shows the average sMRI-IG maps for the correctly classified CN and AD patients, overlaid onto the MNI152 template (1.5 mm). A thresholding was applied in order to visualize only the relevance scores above the 99.5^{th} percentile of the relevance distribution, followed by a Gaussian smoothing with FWHM = 3 mm for visualization purposes. It is evident that the subcortical regions, in particular the hippocampus, are associated with high IG scores in absolute value, with the highest

positive (negative) values for CN (AD). Cortical regions generally showed almost null relevance for both the CN and the AD-derived sMRI-IG maps. Of note, higher absolute values were found for the AD maps compared with the CN ones.

Moving to the fMRI-IG qualitative analysis, Fig. 4B shows the average connectograms for CN and AD subjects, including only the connections above the 98th percentile of the relevance distribution. The ICs-brain region correspondences are detailed in Supplementary Table 2.

The most relevant RSNs were the primary information processing-related networks (SM and VI) followed by multi-sensory integration networks (CC and DM), and cerebellum (CB), showing high relevance for both the CN and the AD-derived fMRI-IG. In particular, the 53% of the most relevant connections for the CN group belong to the SM network, while a percentage of 38% was found for the AD-derived fMRI-IG maps. On the other hand, the VI was highly involved for AD subjects, with the 28% of relevant connections belonging to this RSN, differently from the CN where only 2 VI ICs resulted as relevant. Of note, a total of 15 ICs were marked as relevant for both the CN and the AD-derived fMRI-IG, but with an opposite sign.

More in detail of the relevant connections between the different ICs, the CN subjects showed positive relevant intra-network connections in the CB and SM RSNs, with particularly high scores for the connections involving the post/paracentral and parietal gyri (ICs 9, 10, 11, 12, 13). Inter-network positively relevant connections were also found between the SM and the CB, and the SM and the CC RSNs. Only two negative IG scores were instead retrieved, related to an intra-network connection in the VI network and to an inter-network connection between CC and DM.

A similar pattern was found for the AD fMRI-IG relevant connections but with generally opposite IG-associated scores. In detail, negative relevance was mainly found for both the intra- and inter-network connections encompassing the SM and CB RSNs, with an intra-connection in SM (ICs $11 - 12$) showing the highest negative relevance between the same ICs highlighted with the opposite sign in the CN. On the contrary, positive relevance was recorded for the connections between multiple VI ICs, in particular involving right middle occipital gyrus (IC 21) with cuneus (IC 20), inferior occipital gyrus (IC 23), and middle temporal gyrus (IC 25) as well as the inter-network connections VI-DM and CC-DM, with the latter being the most relevant one.

Finally, concerning the SNPs-IG qualitative analysis, Fig. 4C presents a Manhattan plot including only the SNPs with an attributed positive IG score higher than the 65th percentile, grouped by chromosome. The y-axis reports the associated IG score. A complementary trend between the IG values associated with the SNPs was found between the two classes. This was particularly evident for Chr 1, 7, 11, 14, 15 and 19. A generally high involvement of the Chr 7, 8, 11, 14, and 19 was present for both the AD and CN, with the majority of SNPs selected in these Chr showing high IG scores, either positively or negatively. Of interest, the most relevant SNPs, with positive IG were found in Chr 4, 7, and 8 for the CN and in Chr 12 and 14 for the AD. On the contrary, the most negatively relevant SNPs for CN were found in Chr 14 and 19, as opposed to AD where they were mainly in Chr 1, 6, and 7.

3.4 Task 2: MCI conversion prediction

The MCI prediction task was performed by testing the full MCI cohort on the best model obtained for the AD detection task, for each one of the five independent runs. This allowed to assess its viability in predicting MCI conversion to AD, by correctly stratifying MCIc and MCInc subjects, without being trained for the specific task. This procedure allowed to reach an ACC of 0.711 ± 0.01, a PRE of 0.558 ± 0.03, a REC of 0.610 ± 0.03, 0.581 ± 0.01 in F1 score, 0.470 ± 0.02 in AUPRC, and finally an AUC of 0.755 ± 0.01 for the differentiation between MCIc and MCInc. Regarding the performance obtained using the best model trained on the classification Task 1, we achieved 0.711 in ACC, 0.612 in PRE, 0.550 in REC, 0.579 in F1 score, 0.505 in AUPRC, and 0.766 in AUC.

3.4.1 Feature relevance

In parallel with the results shown for Task 1, Fig. 5 shows a summary of the IG results obtained for the MCI conversion prediction task. The same visualization techniques presented for Task 1 were applied also for this figure. The maps were averaged over the correctly classified subjects for each class, the *jet* colorbar was used to highlight the attribution scores, and thresholding was applied to retain only the values above the 99.5^{th}, 98th and 65th percentile of the relevance distribution, respectively for the sMRI-IG, fMRI-IG, and the SNPs-IG. The sMRI-IG maps were smoothed and overlaid onto the MNI152 template. In detail, Fig. 5A shows the average sMRI-IG maps for the correctly classified MCIc and MCInc patients. Similarly to what was shown for Task 1, the subcortical regions held the highest relevance, with the hippocampus being clearly highlighted as the most important region with positive attributions for MCInc and negative values for MCIc. Cortical regions did not show particular relevance, except for some temporal and occipital areas, which however showed low IG scores.
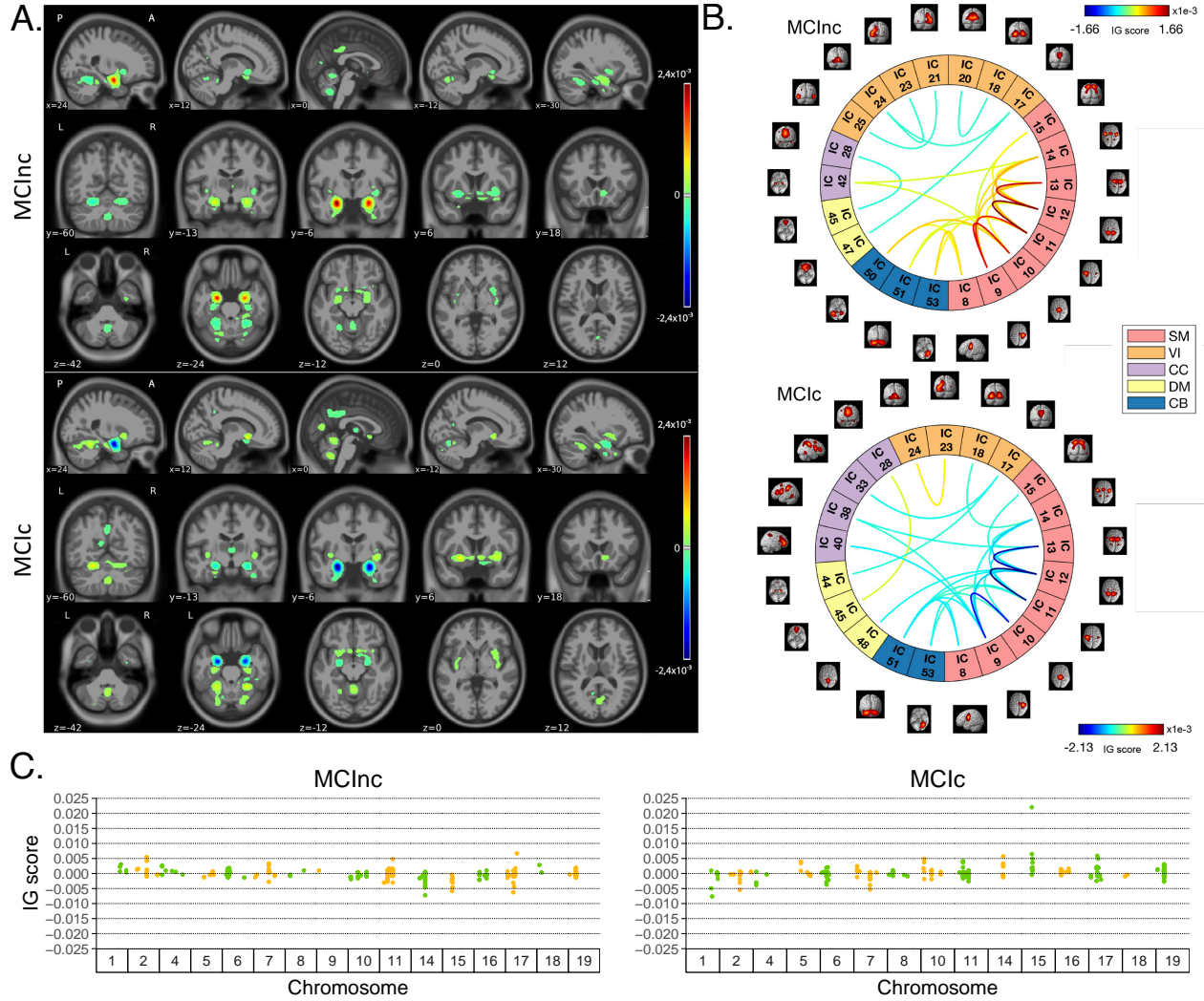
Figure 5: Overview of IG attribution maps for the classification Task 2, MCI conversion prediction. All the IG maps are presented for the correctly classified MCIc and MCInc subjects in the testing set. A. Average sMRI-IG maps, thresholded to retain IG scores exceeding the 99.5^{th} percentile overlaid to MNI152 template; B. Average fMRI-IG derived connectograms, thresholded to retain the connection with an associated IG score over the 98th percentile; C. Average SNPs-IG scores highlighting the SNPs with an associated positive IG score exceeding the 65th percentile, grouped by chromosome.

Moving to the fMRI-IG, Fig. 5B shows the averaged connectograms. In accordance with Task 1, the same five RSNs resulted as the most relevant also for the MCI prediction conversion task, namely the SM, VI, CC, DM and CB. The MCInc fMRI-IG showed more involvement of VI (25% of the relevant ICs), compared with the MCIc one (14% of the relevant ICs). On the contrary, the MCIc IG connectome showed a higher number of ICs belonging to the CC network (4 ICs) compared to the two found for the MCInc one. Of note, the two connectograms had 11 common connections with opposite trends.

More in details of the relevant connections between the different ICs, of interest, the most relevant connections between the ICs in the SM were exactly the same for MCIc and MCInc, with opposite signs, and mainly involved the post/paracentral and parietal gyri (ICs $9, 10, 11, 12, 13$), as for Task 1. Moreover, the fMRI-IG for the MCInc subjects showed positive relevance also for the intra-network connections in the CB, and for the inter-network connections involving CB-SM and CC-SM. On the other hand, negative relevance was found for the intra-network connections in the VI, mainly involving occipital, lingual, calcarine, middle temporal gyri, and cuneus areas (ICs $17-24, 18-20, 21-25, 23-24$), and for the inter-network connections between VI and DM, and CC and DM. A slightly different

13

pattern was found for the MCIc fMRI-IG, whose relevant connections generally had a negative IG-associated score. In detail, intra-network connections with negative scores were found for SM and CB, as already presented, while inter-network negative connections were depicted for VI-CC, VI-DM, and SM-DM. Of interest, negative connections between CC and SM were also retrieved, never reported for the other classes in the study. Positive attribution was found only for an intra-network connection in the VI (ICs $23 - 24$) and a connection between CC and DM (ICs $28 - 45$).

Finally, concerning the SNPs-IG, Fig. 5C presents a Manhattan plot highlighting the SNPs with an attributed positive IG score higher than the 65th percentile. As expected and as already seen for the other modalities, a complementary trend between the IG values associated with the SNPs was found for the two classes, with positive weights being associated with the MCInc and negative weights associated with the MCIc for the same SNPs and *vice-versa*. This was particularly evident for Chr 1, 2, 7, 14, 15 and 17. A generally high involvement of the Chr 11 and 17 was present for both the MCIc and MCInc, with the majority of SNPs showing high IG scores. Of interest, the most relevant SNPs, with positive IG were found in Chr 17 for the MCInc and in Chr 15 for the MCIc. On the contrary, the most negatively relevant SNPs for MCInc were found in Chr 14, 15, and 17, as opposed to the MCIc where they were mainly in Chr 1, 2, and 7.

3.5 Group-based statistical IG analysis

The IG scores obtained for the four groups of subjects in the study were analyzed in order to quantify the differences and similarities outlined in the qualitative analysis. In what follows, the statistical analyses will be presented separately for the neuroimaging (sMRI and rs-fMRI) and the genetic channels.

Neuroimaging channels Fig. 6A shows the boxplots representing the distribution of the average IG values in the most relevant cortical and subcortical brain regions (ROIs). Of note, only those regions where at least the 99% of the subjects had an IG value were retained. The complete acronym list is reported in Supplementary Table 1. A general agreement between the CN and the MCInc subjects, as well as between the AD and the MCIc patients is evident in all the considered ROIs, with the AD/MCIc subjects showing a generally higher variance. As expected from the qualitative analysis, the Hipp and the Amy resulted as the highest relevant ROIs, as well as the ones showing the highest group distance, with high positive attributions for the CN and the MCInc, and strong negative attributions for the AD and MCIc. Among the other subcortical ROIs, the Acc showed a notable relevance for all the groups, with the MCInc and CN having negative scores and the MCIc and AD positive ones. Cau and Put had an associated high positive relevance score for AD and MCIc, while almost null scores for the other two groups. Interestingly, among the cortical regions, the TOF was the most relevant for both the MCIc and AD (positive IG scores) and the MCInc and CN (negative IG scores). The largest distance between the group IG score distributions was found for PhGa, followed by the TFCp with the AD and MCIc having high negative values. On the contrary, the OFG, the POpC, and the PaGp showed high positive relevance for AD and MCIc patients associated with negative scores for the CN and MCInc.

In order to establish whether these differences were significant, a Kruskal-Wallis test was performed, separately for each ROI, considering the four groups together. The significant ROIs (all except the Ins and the TP) were further analyzed to investigate the group-related differences through the Wilcoxon Rank Sum test. Results are shown in Fig. 6B. The heatmap reports the Bonferroni corrected p-values (6 pairwise comparisons) in logarithmic scale and reverse hot colorbar where yellow/red correspond to higher/lower p-value, and white indicates non-significant outcomes. The most significant differences were found between MCIc and MCInc in the subcortical ROIs, resulting as generally significant (except for Put and Pall) with the lowest p-values being recorded for Hipp and Amy. Moreover, significant differences were also recorded for the temporal (PcC, PhGa, PaGp, TFCp) and occipital (TOF, OFG, POpC) cortical regions, with the most significant difference being recorded for PcC and PhGa. A coherent significance pattern was found between the contrasts MCIc-CN, MCInc-AD, and the CN-AD with generally higher p-values. Of interest, no statistically significant differences were found in the sMRI-IG score distributions between AD and MCIc patients, except for PcC, TFCp, and POpC, while some significant differences were found between CN and MCInc but with relatively high p-values.

The same statistical analysis was carried out for fMRI-IG, considering all the most relevant connections resulting from the fMRI-IG scores for all the considered classes. The connections between IC18 (VI) – IC14 (SM) and IC44 (DM) – IC17 (VI) did not reach significance at the Kruskal-Wallis test, therefore they were excluded from the post-hoc analysis. The Bonferroni-corrected pairwise tests (Wilcoxon rank sum) are reported in Fig. 6C, separately for intra- and inter-network connections, with the p-values being represented in logarithmic scale. The related boxplots are reported in Supplementary Fig. 1. Differently from the sMRI-IG significance pattern, the most significant differences were recorded between the CN-AD and MCInc-AD contrasts, which showed significant differences for almost all the considered connections. In particular, for the MCInc-AD contrast, significant intra-network connections were part of the SM, involving connections in the parietal lobe, and in the VI RSNs. Concerning the inter-network connection, significance was found between ICs in the DM, part of the cingulate gyrus (IC47, IC45), and ICs in the CC or VI
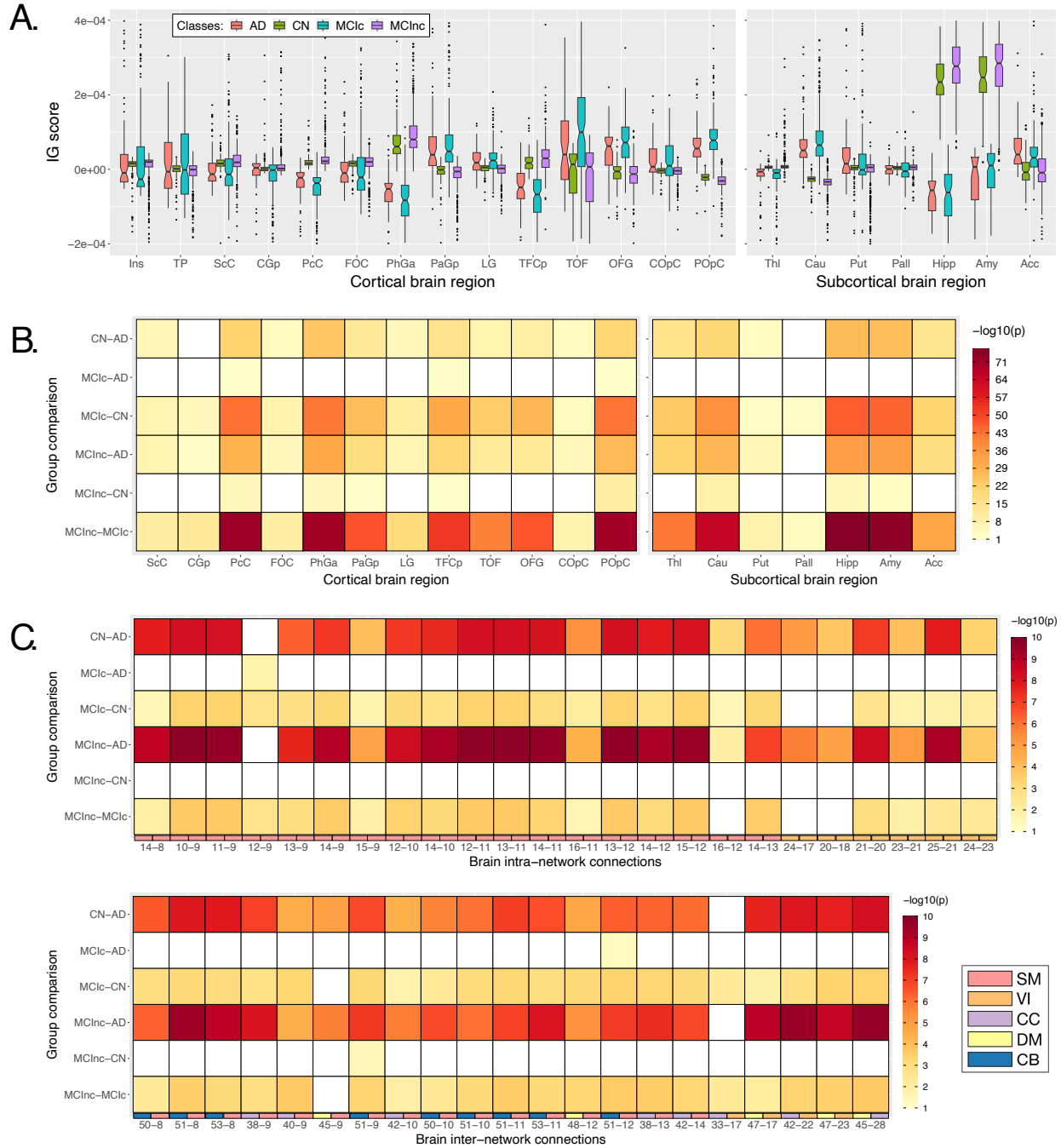
14

Figure 6: Overview of the neuroimaging statistical analysis. A. Boxplots showing the CN, AD, MCInc, and MCIc correctly classified subjects distribution in the most relevant cortical and subcortical brain regions, resulting from the sMRI-IG; B. sMRI-IG Wilcoxon test results for each group comparison and each brain region, divided into cortical and subcortical regions; C. fMRI-IG Wilcoxon test results for each group comparison for each relevant brain connection divided into extra- (upper) and intra-network (lower) connectivity. The *p*-values for B. and C. are reported in negative logarithmic scale, dark red the most significant. White cells represent no statistical significance.

RSNs, as well as between ICs in the SM, located in the parietal lobe and ICs in the CC or CB RSNs. MCIc-CN and MCInc-MCIc showed the same significant pattern as the CN-AD contrast, showing generally higher *p*-values. Of interest, as for the sMRI-IG statistical analysis, no significant differences were detected for the contrasts MCIc-AD
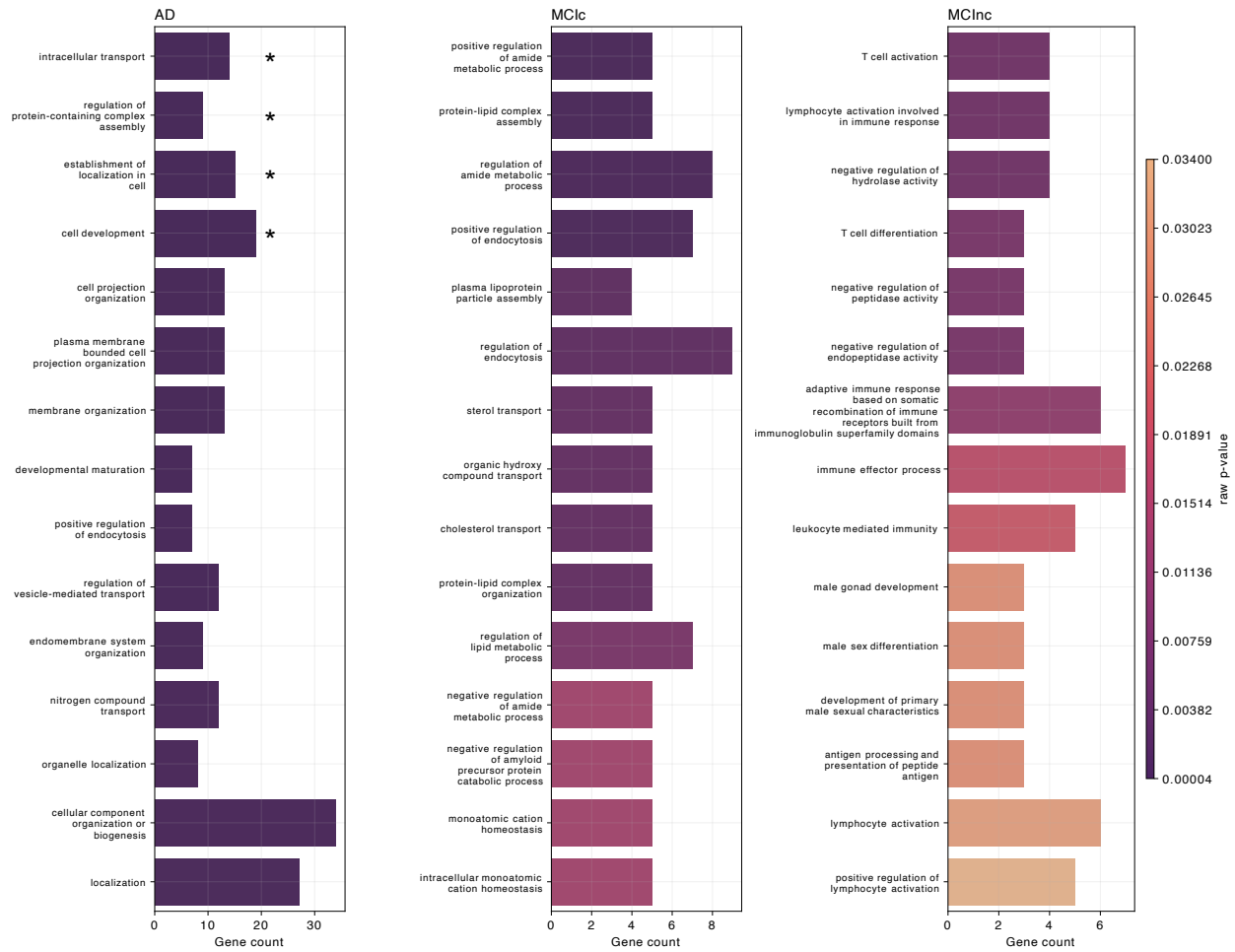
Figure 7: Enrichment analysis results. The top 15 biological processes are reported for the three classes of patients, namely AD, MCIc, and MCInc (columns). The enrichment analysis was performed over the genes annotated from the SNPs-IG. The colorbar represents the associated p-value, dark violet the lowest, while the number of genes involved is reported as the bar height. The biological processes surviving the Bonferroni correction are marked with $*$.

and MCInc-CN, except for two connections between CB and SM (IC51-IC9 and the IC51-IC12 respectively for the MCInc-CN and the MCIc-AD).

Genetic channel Figure 7 shows the biological processes derived from enrichment analysis for AD, MCIc, and MCInc classes as well as their corresponding raw p-values. Only the SNPs having positive attributions higher than the 65th percentile were retained for gene annotation. The left plot of Figure 7, regarding AD patients, highlights the four AD-related biological processes that were statistically significant after p-value correction, that is: *intracellular transport* ($p_{\mathrm{bonf}} = 0.0225$), regulation of protein-containing complex assembly ($p_{\mathrm{bonf}} = 0.0413$), establishment of localization in cell ($p_{\mathrm{bonf}} = 0.0446$), and cell development ($p_{\mathrm{bonf}} = 0.0470$). The other AD-related biological processes that emerged involved the endocytosis mechanism, hence processes related to the transport, such as *positive regulation of endocytosis*, *regulation of vesicle-mediated transport*, and *nitrogen compound transport*. Additionally, biological processes related to the membrane and cell organization were detected, such as *cell organization*, *cell projection organization*, *cellular component organization or biogenesis*, *organelle organization*, *membrane organization*, *plasma membrane bounded cell projection organization*, and *endomembrane system organization*. Similarly, the central plot, concerning the MCIc-related biological processes, also identifies the endocytosis mechanism, such as *regulation of endocytosis* and *positive regulation of endocytosis*. Of note, one process was related to the amyloid-β (Aβ), that was *negative regulation of amyloid precursor protein catabolic process*. Furthermore, important biological processes related to cholesterol, like *sterol transport* and *cholesterol transport*, and processes involving lipids and proteins, such as *protein-lipid complex assembly*, *plasma lipoprotein particle assembly*, *protein-lipid complex organization*, and

16

regulation of lipid metabolic process were also highlighted. Different processes involved the amide compound, e.i., *positive regulation of amide metabolic process*, *regulation of amide metabolic process*, and *negative regulation of amide metabolic process*. The ten most frequent genes in terms of occurrences in the enriched biological processes for AD and MCIc patients were CLU, APOE, PICALM, APOA2, ABCA7, TREM106B, NECTIN2, TREM2, BIN1, and BLOC1S3. Regarding the MCInc, right plot, two biological processes were related to the T cell, the *T cell activation* and the *T cell differentiation*, while different processes involved the white blood cells and immunity, such as *lymphocyte activation involved in immune response*, *lymphocyte activation*, *positive regulation of lymphocyte activation*, *adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains*, *immune effector process*, and *leukocyte mediated immunity*. For this group, the five most frequent genes were FCER1G, ERCC1, CR1, RELB, and ACE. Supplementary Tables 3, 4, and 5 provide additional information, such as the most frequent genes, the GO index, and the corresponding *p*-value for each highlighted biological process for AD, MCIc, and MCInc, respectively.

4 Discussion

In this work, we proposed a novel multimodal generative and interpretable method, which holds the potential of addressing missing data management while focusing on model interpretability. We specifically applied this method to the crucial tasks of segregating AD patients from CN, and assessed its viability in detecting MCI conversion to AD. By integrating neuroimaging (sMRI and rs-fMRI) and genetics (SNPs) data through DL-based feature extraction, we introduced a novel multimodal data fusion framework, relying on input data not yet simultaneously investigated in the AD continuum in the current literature. Incomplete data is handled by generating missing modalities in the latent space, obtained after feature reduction. This ensures high accuracy in reconstructing latent features while minimizing computational demands. To accomplish this, we relied on pre-trained generators from two cGANs, one trained to generate sMRI and SNPs data, and the other to generate sMRI and rs-fMRI patterns. However, the approach is generalizable to any missing modality, providing a viable and effective solution for overcoming the missing data bottleneck. The framework is trained to detect AD in a cohort including CN and AD subjects, while its generalization capabilities were then tested by predicting MCI conversion to AD in an independent study cohort. Furthermore, we meticulously conducted a post-hoc interpretability analysis, supplemented by a robust validation step, consolidating the findings' impact in the field. The next Subsections are dedicated, respectively, to the assessment of the classification performance and missing data processing (Subsection 4.1), and to interpretability analysis in the light of neurosciences (Subsection 4.2).

4.1 Classification performance and missing data management

Focusing on the Task 1, AD detection, our model reached the state-of-the-art performance, obtaining an average accuracy of 0.926 ± 0.02 on the testing set, with the best model reaching an accuracy score of 0.964. Noteworthy, this result was achieved on a cohort where only the 6.5% of the subjects had complete data across all three modalities. Of note, our dataset was constructed considering the subjects having at least the sMRI modality, however, the proposed framework showcases the capability to potentially impute all three modalities by transferring knowledge across domains, not being limited to having at least one acquisition for all the subjects. Indeed, the injection of the pre-trained generators, that is the two cGANs, allows both to impute the missing rs-fMRI or SNPs from the sMRI and the *vice-versa*, meaning that it would allow deriving the missing sMRI from the other two modalities, if needed. This allows relaxing the requirement of having at least one acquisition for all the subjects which is one of the strengths of the proposed model compared with the state-of-the-art.

Table 3a shows the performance of the proposed model compared with the state-of-the-art competitors allowing multimodality and missing data management for Task 1. The simplest model([35]) proposed a DL-based multimodal method based on sMRI, clinical, and genetics data, excluding features if missing in more than 70% of subjects, and filling the remaining missing data with zeros. Their classification task was different from ours since they included the MCI subjects in a three-class classification problem. While they achieved an accuracy of 0.780 for the classification between CN, MCI, and AD, the accuracy dropped to 0.630 when classifying CN from the full patient cohort using only sMRI and genetics. Moreover, the proposed technique had some drawbacks, mainly linked to the biases introduced in the network, as already discussed.

Alternatively, methods to exploit the information of all the available subjects, without explicitly generating the missing modalities, were proposed relying on latent representation learning. [37] developed an Auto-Encoder-based multi-view missing data completion framework using sMRI and FDG-PET ROI-based features. Their method achieved a classification accuracy of 0.836 for classifying CN versus AD, even with 50% missing PET. Similarly, [68] proposed a GAN with an attention layer to generate missing FDG-PET from available MRI features, resulting in a classification

Table 3: Performance comparison of the proposed model with other state-of-the-art methods dealing with missing modalities for AD detection and MCI conversion tasks.

(a) AD detection task. Accuracy (ACC), precision (PRE), and recall (REC) metrics are reported on the test set or averaged during the cross-validation phase ($mean \pm std$).

Authors	Modalities	Study cohort	Input data	Missing data rate	Missing data management	ACC	REC	PRE	XAI
[35]	sMRI, Genetics, Clinics	1406 PAT (266 AD, 699 MCI), 598 CN	Full MRI volume WGS data, Set of clinical scores	$\sim 67\%$ missing sMRI $\sim 61\%$ missing SNP	Feature exclusion if the 70% had missing value Zero filling for the others	0.630^* $0.780^\dagger$	0.570^* $0.780^\dagger$	0.620^* $0.770^\dagger$	Yes
[37]	sMRI, FDG-PET	85 AD, 90 CN	ROI based features	50% missing PET	Auto-Encoder for the complementation of complete and incomplete latent representation	0.836 ± 0.08	n.d.	n.d.	No
[40]	sMRI, FDG-PET	352 AD, 427 CN	sMRI and PET volumes	30% missing PET	Task-induced pyramid and attention generative adversarial network (TPA-GAN)	0.927	0.917	n.d.	No
[68]	sMRI, FDG-PET	160 AD, 210 CN	ROI based features for VBM, FDG	50% missing PET	GAN with attention layer in latent space	0.914 ± 0.19	0.881 ± 0.11	n.d.	Yes
Proposed framework	sMRI, fMRI, Genetics	332 AD, 664 CN	GM volume sFNC matrix, SNPs	$\sim 59\%$ missing fMRI $\sim 57\%$ missing SNPs	Multiple cGANs to generate missing latent representation	0.926 ± 0.02	0.876 ± 0.03	0.910 ± 0.05	Yes

(b) MCI conversion prediction. Accuracy (ACC), precision (PRE), and recall (REC) metrics are reported on the test set.

Authors	Modalities	Study cohort	Input data	Missing data rate	Missing data management	Trained on the specific task	ACC	REC	PRE	XAI
[36]	10 modalities including imaging and clinics	86 MCIc 237 MCInc	Tabular features	$\sim 8\%$ missing features	Mean imputation and imputation by the EM	Yes	0.670	n.d.	n.d.	No
[39]	sMRI, FDG-PET	76 MCIc 128 MCInc	GM volume PET volumes	50% missing PET	3D encoder-decoder network to generate the full PET volume	Yes	0.657	n.d.	n.d.	No
[41]	sMRI, FDG-PET Genetics	157 MCIc 205 MCInc	ROI based features for PET and MRI, SNPs	51% missing PET	Latent representation learning without missing data imputation	Yes	0.743	n.d.	n.d.	Yes
[40]	sMRI, FDG-PET	234 MCIc 342 MCInc	sMRI and PET volumes	30% missing PET	Task-induced pyramid and attention generative adversarial network (TPA-GAN)	Yes	0.753	0.708	n.d.	No
Proposed framework	sMRI, fMRI, Genetics	289 MCIc 646 MCInc	GM volume sFNC matrix, SNPs	$\sim 76\%$ missing fMRI $\sim 45\%$ missing SNPs	Multiple cGANs to generate missing latent representation	No	0.711 ± 0.01	0.610 ± 0.03	0.558 ± 0.03	Yes

* CN vs PAT (AD+MCI) classification relying only on sMRI and SNPs, † Three classes classification, CN vs MCI vs AD relying on the three modalities

accuracy of 0.914 ± 0.19. While these approaches demonstrated excellent results, their framework was limited in handling arbitrary missing modalities, allowing only the completion of missing PET from available sMRI and FDG-PET features.

Finally, [40] presented a task-induced pyramid and attention generative adversarial network (TPA-GAN) to generate missing FDG-PET data from sMRI, achieving an accuracy of 0.927. However, the model was trained and tested on two independent databases with different acquisition protocols, which on one side allowed to evaluate the generalizability of the model, but on the other could bias the results due to the different data source. Moreover, it still lacked the ability to reconstruct sMRI from FDG-PET data. Despite the aforementioned limitations, they are currently the state-of-the-art performance for multimodal MCI conversion prediction, accounting for the 30% of missing PET data.

As mentioned before, the trained model performance was subsequently assessed on Task 2. Only the 7% of the dataset shared all the modalities, making it a challenging scenario. Nevertheless, our framework achieved an average accuracy of 0.711 ± 0.01 for the independent test sets. Although it did not outperform methods specifically trained on this particular task, it demonstrated competitive results. Other approaches have focused on different input data and missing data management to solve this task. For instance, [36] proposed simple machine learning methods with tabular features and limited imputation techniques, achieving an accuracy of 0.670 to stratify MCI subjects. [39] used a GAN to impute missing PET images from sMRI scans, achieving an accuracy of 0.657 for discriminating MCInc vs MCIc patients. Furthermore, [41] proposed a framework for projecting original features into a latent representation, resulting in an accuracy of 0.743 in case of missing the 51% of PET data. Finally, [40] applied the approach already discussed also to address the classification Task 2, reaching the best accuracy score of 0.753.

Despite the promising results obtained by the state-of-the-art generative models for missing data imputation, it is essential to acknowledge their shared limitation: they all relied on having sMRI as a prerequisite to impute PET data, and very few included genetics or rs-fMRI information in their analyses, which have instead been demonstrated as relevant biomarkers for AD ([10, 15]). In contrast, our framework does not necessitate a shared modality across all subjects. The exploitation of two cGANs during the generation phase enables to produce the missing latent rs-fMRI and/or SNPs from sMRI. In addition, this process is applicable bidirectionally, allowing the generation of sMRI latent features from either SNPs or rs-fMRI as well. The versatility of this approach opens the way to its generalization to additional modalities through training distinct cGANs and integrating the resulting generators into the complete classification framework. Furthermore, our model's performance confirms that generative models allow to obtain realistic data and learn nonlinear mappings across the different acquisitions, achieving competitive prediction accuracy also with a substantial proportion of missing modalities.

4.2   Interpretability analyses

Interpretability analysis was performed relying on IG, followed by group-based statistical analysis for the sake of validation. More in detail, for each input modality and each task, relevance maps were extracted and analyzed from a qualitative and quantitative point of view. We recall that being IG baseline-based, the feature relevance is to be considered baseline-dependent. The baseline represents a neutral input to the network, that is an input surrogate subject that for the network could equivalently belong to one class or the other. Hence, the baseline chosen should be validated ([57, 58, 59]). In a binary classification, the sign of IG could generally point to an increase in the associated feature, and, viceversa, a negative value could be associated with a decrease in the corresponding feature value.

Interestingly, among the state-of-the-art methods for multimodal AD or MCI conversion detection, only a few works introduced interpretability analysis. [35] proposed an occlusion-based approach consisting of masking one feature at a time and measuring the drop in accuracy, uncovering that when considering the three modalities together (namely clinics, sMRI, and genetics) the most relevant features were AD-related clinical biomarkers, while when considering solely sMRI and genetics the most relevant features were brain regions such as the hippocampus and amygdala, with limited relevance of the SNP features. Concerning DL and generative-based frameworks, [68] relied on an attention layer in their generation module for obtaining the feature importance for both the generator and the discriminator. This allowed to analyze the most relevant sMRI features for the generation of the missing PET, as well as the most relevant PET features for the discriminator. However, the outcomes were not widely discussed in the paper and were limited to the generation step. Finally, [41] exploited the weights learned for their latent representation learning to derive the input features' importance, finding relevant regions highly related to AD and MCI diagnosis such as the hippocampus and amygdala.

In our results, the interpretability analysis revealed similarities between the MCIc and the AD phenotypes, as well as between the MCInc and CN, while highlighting significant differences between the CN and AD, as expected, but also between MCIc and MCInc which are of high interest since they would allow to identify early AD biomarkers.

Regarding the sMRI data, our findings highlighted that sub-cortical regions carried more informative and relevant information for the final classification compared to cortical areas. The hippocampus resulted as the most critical region for all classes under analysis, alongside the amygdala. In AD and MCIc patients, such regions exhibited negative attribution values, suggesting a decrease in GM probability, as opposed to positive values in CN and MCInc subjects. Of interest, the most significant differences in the relevance assigned to such regions were found between MCIc and MCInc. These findings were in line with the well-known literature and established hallmarks of AD. Indeed, the pathological process initially affects the hippocampus and amygdala before extending to other nearby structures ([69, 70]). Focusing on MCI, [71] demonstrated that MCIc patients exhibited higher atrophy in the hippocampus and amygdala compared to MCInc and CN subjects, which aligns well with our findings. Regarding the caudate region, [72] and [73] highlighted a negative correlation between caudate and hippocampus volumes in healthy subjects, and, in addition, the second work also showed that patients with AD and non-specified dementia have a larger caudate volume compared to non-dementia subjects. However, this is still debated in literature with other works suggesting that the caudate is susceptible to atrophy, resulting in a reduction of the GM volume of this region in AD patients ([74], [75], [76]).

The cortical areas had a generally lower relevance to the classification for all the phenotypes. The parahippocampal gyrus (anterior division) and the occipital fusiform gyrus showed the highest attribution scores, with positive values assigned to CN. This was in line with [77] and [71] findings which revealed smaller volumes in the anterior parahippocampal area in AD, MCIc, and MCInc subjects compared to CN individuals. Additionally, [78] highlighted the importance of the occipital fusiform gyrus for AD classification. As for the other relevant brain regions, the precuneus and fusiform cortices showed a reduction of the GM volume in our results, coherent with what is known from the literature ([79], [80], [81]).

19

Overall, our findings on sMRI revealed that the relevance scores were sensible to an increase in neurodegeneration in some focal brain areas in AD and MCIc compared with CN and MCInc, which aligns well with the literature findings.

Based on rs-fMRI attributions, five functional RSNs emerged as highly relevant for the final classification of both tasks, namely SM, DM, CB, VI, and CC. The DM has a central role in information integration and processing, and its involvement in AD is well-known in the current literature, with several studies consistently demonstrating that this is the first RSN to be affected by abnormal protein aggregation ([82, 14]). In our results, we did not retrieve intra-network connections in the DM mode. However, for the MCIc subjects, a negative relevance score, suggesting a decreased FC, was found for inter-network connections between DM and visual/sensorimotor areas (DM-VI, DM-SM), while a positive inter-network connection was highlighted between DM and CC. Conversely, AD subjects revealed a positive relevance for a few connections between the DM and VI and between the DM and the CC, while one negative connection was found between DM and SM. These patterns are not commonly reported in the literature but deserve further investigation since they could represent compensation connectivity patterns.

Among the other RSNs, the SM was the most present and relevant, showing negative relevance scores for AD and MCIc for both the intra-connections, involving post/para central and partial gyri, and inter-connections with the other relevant RSNs. On the contrary, for the CN and, importantly, for the MCInc, the SM showed positive relevance, leading to significant differences for both intra- and inter-network connections with the most severe groups (AD and MCIc). The lowest p-values were found when comparing MCInc vs AD as well as CN vs AD over the different SM intra-network connections. Inter- and intra-network SM connections were indeed demonstrated to be affected by the AD pathology, generally showing an overall decreased connectivity in the later stages of the disease ([83], [84], [82]). Moreover, [85] showed that many pyramidal and extrapyramidal motor impairments affect a substantial portion of AD patients, even at an early stage of the disease, and progressively worsen along with cognitive impairment, reflecting a possible decreased connectivity in the SM network.

Interestingly, the VI RSN was involved in the classification of the different AD stages. In detail, positive relevance was assigned to AD subjects in both intra- and inter-connections (VI-DM and VI-CC), while a different pattern was found for MCInc, which showed negative relevance in almost the same ICs. Of note, the MCIc did not show high involvement of the VI intra-connections, while few negative inter-connections were found between VI and the other relevant RSN (SM, CC, and DM). This revealed another difference between AD and its prodromal stages, indicating a decrease in VI connectivity in the MCI stage, particularly evident for the MCInc, which then converts into hyperconnectivity in the most severe stage, full-blown dementia. The damage in VI due to AD was previously discussed by [82]. In detail, [85] found that subgroups of AD patients have concomitant eye diseases, and some visual functions are impaired, which could be caused by impairments in the VI RSN. Moreover, recent studies demonstrated a hyperconnectivity pattern in the most severe stages of the disease present in the VI network ([86]). Hence, further investigation would be needed to elucidate the involvement of VI RSN in the AD continuum.

Finally, the inter-network connections between the CB and the SM resulted negatively relevant, hence affected by the disease with a decreased connectivity in both AD and MCIc, while the same was not recorded for the MCInc, unraveling a possible impact on this RSN in a later stage compared to the others.

Regarding the genetics impact, the significant biological processes for MCInc, MCIc, and AD were derived from the genes annotated starting from the most significant SNPs resulting from the interpretability analysis on the two tasks. Biological processes related to the regulation of endocytosis were found both in AD and MCIc. Previous studies highlighted how endocytosis is strongly related to AD ([87]). Endocytosis is a pathway that, with all the components, proteins, and membrane organization and modulations related to it, plays an important role in AD pathology since it is involved in the trafficking and clearance of Aβ proteins ([87]). Moreover, particularly in MCIc, we found biological processes related to cholesterol, like *cholesterol transport* and *sterol transport*, which, on a deeper analysis, were shown as possibly being associated with the AD continuum ([87]). Excess cholesterol deposit in the brain was demonstrated to be related to an increase of Aβ plaques and amyloid cascade leading to synaptic plasticity annihilation, promotion of tau phosphorylation, hence contributing to the risk and pathogenesis of AD, possibly in an early phase ([88, 89]). Additionally, for the MCIc individuals, a biological process related to Aβ was highlighted, the *negative regulation of amyloid precursor protein catabolic process* that is related to AD ([90, 91]). In relation to Aβ proteins, some biological pathways related to the regulation of amide were found in MCIc individuals. A previous study found that fatty acid amides were strongly associated with Aβ and the hippocampal volume ([92]). Regarding the particular genes, APOE, CLU, PICALM, APOA2, ABCA7, and BIN1 were among the most frequent in the biological processes. Of note, they all have a relevant impact in the development of AD ([93, 47, 94]). In particular, APOE is notably the major risk factor of AD ([95, 93]) and is mainly expressed in both the brain and the liver. More in-depth, the ApolipoproteinE is a ligand receptor-mediated endocytosis of lipoprotein particles ([95]) and is the major cholesterol transport and other lipids in the brain ([96], [97]). This gene is hence strongly related to one of the most important biomarkers for AD, namely Aβ plaques containing Aβ peptides and the neurofibrillary tangles containing hyperphosphorylated tau proteins ([98]).

TREM2 is another gene-coding protein with a key role in AD progression, and the rare variant of TREM2 R47H has a high-risk factor for AD comparable to the strongest biomarker APOE. TREM2 is a transmembrane receptor on the microglia that, in the condition of Aβ aggregation, rapidly migrates closer to the aggregation, transforming them and promoting the phagocytosis and clearance of Aβ formation. Hence, one of the roles of TREM2 is the regulation of microglia recruitment closer to Aβ formations for the uptake and degradation of them ([99]). ABCA7 is one of the most important risk genes for AD ([27]) that mainly regulates the processes related to cholesterol and the processing of Aβ proteins ([100]). ABCA7-expressed proteins have a relevant role in the control of cholesterol metabolism, and then the cholesterol has a strong influence in the regulation of Aβ synthesis ([100, 101]). Additionally, some studies reported also that different variants of ABCA7 are associated with an increase of Aβ deposition in MCI patients rather than AD ([102]). Similarly, APOA2 is strictly related to cholesterol and HDL since it transports the cholesterol to the liver and is an important component for the formation of HDL ([94]). As for the other relevant genes, PICALM is involved in endocytic-related processes, suited for the production, modulation, and clearance of Aβ complexes ([87, 23]). Also, for this gene, interactions with APOE have been demonstrated ([24]). Along with PICALM, the CLU variant was identified as an important biomarker for AD ([25]) since it is implicated in white matter integrity, membrane recycling, and lipid transportation ([103]). Finally, [104] and [105] found that BIN1 is a gene strictly related to Aβ and τ pathologies, and presumably regulate the APOE ([93]), hence assuming high relevance in AD.

4.3 Main contributions and outcomes

We proposed an interpretable, flexible, and generative framework for multimodal AD detection and MCI conversion prediction. The main contributions of this work are as follows: (i) reaching the state-of-the-art of multimodal and generative models in the classification of CN vs AD; (ii) reaching competitive performance in the segregation of MCInc/MCIc phenotypes using a pretrained framework; (iii) managing missing data introducing a generation module in the latent space allowing to impute missing modalities relaxing the constraint of having at least one modality shared by the whole cohort, and (iv) proposing an interpretability analysis based on IG for extracting biological information.

The complementary analysis of three different input modalities allowed to uncover the disease signatures at multiple levels of analysis, providing complementary yet interdependent information. Besides some results being in agreement with what is known from the literature, providing evidence of the trustability of the outcomes of the proposed analysis, other findings suggest the involvement of additional mechanisms that could contribute to elucidate the mechanisms underlying the pathogenesis and progression of the disease and would deserve further investigation.

The main findings concern the presence of atrophy, particularly involving the hippocampus in later stages (MCIc and AD), that is well known in the literature, FNC modulation, particularly in the information processing-related RSN (SM and VI) as well as SNPs mutations in phenotype-specific genes. Regarding FNC, for SM negative relevance scores were recorded, hence suggesting a decreased inter- and intra-connectivity, in both the MCIc and the AD stages, while for VI a different pattern was observed for MCInc/MCIc/AD subjects. In particular, negative attributions were recorded in the MCInc stage, with reduced relevance in the MCIc, while positive attribution values characterized the AD, suggesting a compensation mechanism leading to a hyperconnectivity in such RSN in the later stages of the disease. Such brain modulations were present along with the mutation of relevant SNPs linked to the involvement of different biological processes related to endocytosis, cell and membrane transportation and organization, amyloid and cholesterol regulation for the AD and MCIc subjects, while mostly related to the T cell processes, white blood cells, and immune response pathways for the MCInc.

4.3.1 Limitations and future works

In what follows, the main open issues will be briefly summarized, paving the way for further research. First, concerning the MCI task, in this work, our aim was to test the model generalizability by straightforwardly testing it on the MCI cohort. Training and fine-tuning on the specific task would allow obtaining better classification results and will be the object of future analysis. Being the model intrinsically multimodal and easily flexible, additional channels could be included, such as the clinical information and PET imaging, which have been demonstrated to be highly discriminative for the disease ([4]), as well as advanced structural connectivity metrics derived from diffusion MRI which could help to shed light on early disease signatures. On top of this, it would be interesting to explore different phenotype stratification tasks, either proposing a multiclass classification or defining biologically homogeneous groups following the A/T/N (amyloid, tau, neurodegeneration) system which has been attracting increasing attention in recent years ([106]). Finally, concerning the interpretability analysis, we strongly point toward deeper exploitation of XAI methods to open the way to not yet studied associations or mechanisms that are captured by the model. However, we acknowledge that the interpretation of results is not always straightforward, especially when relying on baseline-based XAI approaches, where the choice of the baselines strongly impacts on the results. One future aim is, hence, to establish clear rules for the choice of the baseline, enabling the unambiguous interpretation of the attributions. The next step would be the post-hoc

validation of the XAI outcomes pursuing the robustness, reliability, and trustfulness of the interpretability analysis outcomes. Last but not least, the analysis of the associations between the different input features would support the interpretation of the results as well as further analysis for elucidating causal relations across the investigated underlying mechanisms.

5   Conclusions

In this work, we presented a multimodal, flexible, generative, and interpretable DL-based framework for AD detection and MCI conversion classification. Neuroimaging (structural and functional features) and genetics data were used to address these classification tasks. The generation of missing modalities in the latent space using four pre-trained generators of two different cGANs allowed to obtain competitive classification performance in both tasks. The application of an interpretability method yielded our model to be interpretable extracting the relevance of each input feature and revealing the most important ones for each class, highlighting disease structural, functional, and genetic signatures and opening the way to further analyses.

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.

Acknowledgments

References

[1] Theo Vos, Christine Allen, Megha Arora, Ryan M Barber, Zulfiqar A Bhutta, Alexandria Brown, Austin Carter, Daniel C Casey, Fiona J Charlson, Alan Z Chen, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015. *The lancet*, 388(10053):1545–1602, 2016.

[2] Sheria G Robinson-Lane and Xingyu Zhang. Inclusive support: Addressing the needs of black family caregivers of persons with dementia: Dementia care research (research projects; nonpharmacological)/family/lay caregiving. *Alzheimer's & Dementia*, 16:e046712, 2020.

[3] Zeinab Breijyeh and Rafik Karaman. Comprehensive review on alzheimer's disease: Causes and treatment. *Molecules*, 25(24):5789, 2020.

[4] David S Knopman, Helene Amieva, Ronald C Petersen, Gäel Chételat, David M Holtzman, Bradley T Hyman, Ralph A Nixon, and David T Jones. Alzheimer disease. *Nature reviews Disease primers*, 7(1):33, 2021.

[5] Ronald Carl Petersen, Paul S Aisen, Laurel A Beckett, Michael C Donohue, Anthony Collins Gamst, Danielle J Harvey, Clifford R Jack, William J Jagust, Leslie M Shaw, Arthur W Toga, et al. Alzheimer's disease neuroimaging initiative (adni): clinical characterization. *Neurology*, 74(3):201–209, 2010.

[6] Renske Hamel, Sebastian Köhler, Nicole Sistermans, Ted Koene, Yolande Pijnenburg, Wiesje van der Flier, Philip Scheltens, Pauline Aalten, Frans Verhey, Pieter Jelle Visser, et al. The trajectory of cognitive decline in the pre-dementia phase in memory clinic visitors: findings from the 4c-mci study. *Psychological Medicine*, 45(7):1509–1519, 2015.

[7] Giovanni B Frisoni, Nick C Fox, Clifford R Jack Jr, Philip Scheltens, and Paul M Thompson. The clinical use of structural mri in alzheimer disease. *Nature Reviews Neurology*, 6(2):67–77, 2010.

[8] Meredith N Braskie, Arthur W Toga, and Paul M Thompson. Recent advances in imaging alzheimer's disease. *Journal of Alzheimer's Disease*, 33(s1):S313–S327, 2013.

[9] Keith A Johnson, Nick C Fox, Reisa A Sperling, and William E Klunk. Brain imaging in alzheimer disease. *Cold Spring Harbor perspectives in medicine*, 2(4):a006213, 2012.

[10] Lorenzo Pini, Alexandra M Wennberg, Alessandro Salvalaggio, Antonino Vallesi, Michela Pievani, and Maurizio Corbetta. Breakdown of specific functional brain networks in clinical variants of alzheimer's disease. *Ageing Research Reviews*, 72:101482, 2021.

[11] Juan Zhou, Michael D Greicius, Efstathios D Gennatas, Matthew E Growdon, Jung Y Jang, Gil D Rabinovici, Joel H Kramer, Michael Weiner, Bruce L Miller, and William W Seeley. Divergent network connectivity changes in behavioural variant frontotemporal dementia and alzheimer's disease. *Brain*, 133(5):1352–1367, 2010.

[12] Kim A Celone, Vince D Calhoun, Bradford C Dickerson, Alireza Atri, Elizabeth F Chua, Saul L Miller, Kristina DePeau, Doreen M Rentz, Dennis J Selkoe, Deborah Blacker, et al. Alterations in memory networks in mild cognitive impairment and alzheimer's disease: an independent component analysis. *Journal of Neuroscience*, 26(40):10222–10231, 2006.

[13] Yong Liu, Chunshui Yu, Xinqing Zhang, Jieqiong Liu, Yunyun Duan, Aaron F Alexander-Bloch, Bing Liu, Tianzi Jiang, and Ed Bullmore. Impaired long distance functional connectivity and weighted network architecture in alzheimer's disease. *Cerebral Cortex*, 24(6):1422–1435, 2014.

[14] David T Jones, David S Knopman, Jeffrey L Gunter, Jonathan Graff-Radford, Prashanthi Vemuri, Bradley F Boeve, Ronald C Petersen, Michael W Weiner, and Clifford R Jack Jr. Cascading network failure across the alzheimer's disease spectrum. *Brain*, 139(2):547–562, 2016.

[15] Nicolai Franzmeier, Julia Neitzel, Anna Rubinski, Ruben Smith, Olof Strandberg, Rik Ossenkoppele, Oskar Hansson, and Michael Ewers. Functional brain architecture is associated with the rate of tau accumulation in alzheimer's disease. *Nature communications*, 11(1):347, 2020.

[16] Pui-Yan Kwok and Zhijie Gu. Single nucleotide polymorphism libraries: why and how are we building them? *Molecular medicine today*, 5(12):538–543, 1999.

[17] Douglas P Wightman, Iris E Jansen, Jeanne E Savage, Alexey A Shadrin, Shahram Bahrami, Dominic Holland, Arvid Rongve, Sigrid Børte, Bendik S Winsvold, Ole Kristian Drange, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for alzheimer's disease. *Nature genetics*, 53(9):1276–1282, 2021.

[18] Jean-Charles Lambert, Carla A Ibrahim-Verbaas, Denise Harold, Adam C Naj, Rebecca Sims, Céline Bellenguez, Gyungah Jun, Anita L DeStefano, Joshua C Bis, Gary W Beecham, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for alzheimer's disease. *Nature genetics*, 45(12):1452–1458, 2013.

[19] Brian W Kunkle, Benjamin Grenier-Boley, Rebecca Sims, Joshua C Bis, Vincent Damotte, Adam C Naj, Anne Boland, Maria Vronskaya, Sven J Van Der Lee, Alexandre Amlie-Wolf, et al. Genetic meta-analysis of diagnosed alzheimer's disease identifies new risk loci and implicates aβ, tau, immunity and lipid processing. *Nature genetics*, 51(3):414–430, 2019.

[20] Berndt Winblad, Katie Palmer, Miia Kivipelto, Vesna Jelic, Laura Fratiglioni, L-O Wahlund, Agneta Nordberg, Lars Bäckman, Michael Albert, Ove Almkvist, et al. Mild cognitive impairment–beyond controversies, towards a consensus: report of the international working group on mild cognitive impairment. *Journal of internal medicine*, 256(3):240–246, 2004.

[21] Amir Abbas Tahami Monfared, Michael J Byrnes, Leigh Ann White, and Quanwu Zhang. Alzheimer's disease: epidemiology and clinical progression. *Neurology and therapy*, 11(2):553–569, 2022.

[22] Leo Ungar, Andre Altmann, and Michael D Greicius. Apolipoprotein e, gender, and alzheimer's disease: an overlooked, but potent and promising interaction. *Brain imaging and behavior*, 8:262–273, 2014.

[23] Wei Xu, Lan Tan, and Jin-Tai Yu. The role of picalm in alzheimer's disease. *Molecular neurobiology*, 52:399–413, 2015.

[24] Gyungah Jun, Adam C Naj, Gary W Beecham, Li-San Wang, Jacqueline Buros, Paul J Gallins, Joseph D Buxbaum, Nilufer Ertekin-Taner, M Daniele Fallin, Robert Friedland, et al. Meta-analysis confirms cr1, clu, and picalm as alzheimer disease risk loci and reveals interactions with apoe genotypes. *Archives of neurology*, 67(12):1473–1484, 2010.

[25] Denise Harold, Richard Abraham, Paul Hollingworth, Rebecca Sims, Amy Gerrish, Marian L Hamshere, Jaspreet Singh Pahwa, Valentina Moskvina, Kimberley Dowzell, Amy Williams, et al. Genome-wide association study identifies variants at clu and picalm associated with alzheimer's disease. *Nature genetics*, 41(10):1088–1093, 2009.

[26] Jason J Corneveaux, Amanda J Myers, April N Allen, Jeremy J Pruzin, Manuel Ramirez, Anzhelika Engel, Michael A Nalls, Kewei Chen, Wendy Lee, Kendria Chewning, et al. Association of cr1, clu and picalm with alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. *Human molecular genetics*, 19(16):3295–3301, 2010.

[27] Arne De Roeck, Christine Van Broeckhoven, and Kristel Sleegers. The role of abca7 in alzheimer's disease: evidence from genomics, transcriptomics and methylomics. *Acta neuropathologica*, 138(2):201–220, 2019.

[28] Federica Cruciani, Antonino Aparo, Lorenza Brusini, Carlo Combi, Silvia F Storti, Rosalba Giugno, Gloria Menegaz, Ilaria Boscolo Galazzo, Alzheimer's Disease Neuroimaging Initiative, et al. Identifying the joint signature of brain atrophy and gene variant scores in alzheimer's disease. *Journal of Biomedical Informatics*, 149:104569, 2024.

[29] Xiaoke Hao, Chanxiu Li, Lei Du, Xiaohui Yao, Jingwen Yan, Shannon L Risacher, Andrew J Saykin, Li Shen, and Daoqiang Zhang. Mining outcome-relevant brain imaging genetic associations via three-way sparse canonical correlation analysis in alzheimer's disease. *Scientific reports*, 7(1):44272, 2017.

[30] Andrew Saxe, Stephanie Nelli, and Christopher Summerfield. If deep learning is the answer, what is the question? *Nature Reviews Neuroscience*, 22(1):55–67, 2021.

[31] Md Abdur Rahaman, Jiayu Chen, Zening Fu, Noah Lewis, Armin Iraji, Theo GM van Erp, and Vince D Calhoun. Deep multimodal predictome for studying mental disorders. *Human Brain Mapping*, 44(2):509–522, 2023.

[32] Giorgio Dolci, Md Abdur Rahaman, Jiayu Chen, Kuaikuai Duan, Zening Fu, Anees Abrol, Gloria Menegaz, and Vince D Calhoun. A deep generative multimodal imaging genomics framework for alzheimer's disease prediction. In *2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 41–44. IEEE, 2022.

[33] Anees Abrol, Zening Fu, Yuhui Du, and Vince D Calhoun. Multimodal data fusion of deep learning and dynamic functional connectivity features to predict alzheimer's disease progression. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4409–4413. IEEE, 2019.

[34] Johann de Jong, Mohammad Asif Emon, Ping Wu, Reagon Karki, Meemansa Sood, Patrice Godard, Ashar Ahmad, Henri Vrooman, Martin Hofmann-Apitius, and Holger Fröhlich. Deep learning for clustering of multivariate clinical patient trajectories with missing values. *GigaScience*, 8(11):giz134, 2019.

[35] Janani Venugopalan, Li Tong, Hamid Reza Hassanzadeh, and May D Wang. Multimodal deep learning models for early detection of alzheimer's disease stage. *Scientific reports*, 11(1):3254, 2021.

[36] Kerstin Ritter, Julia Schumacher, Martin Weygandt, Ralph Buchert, Carsten Allefeld, John-Dylan Haynes, Alzheimer's Disease Neuroimaging Initiative, et al. Multimodal prediction of conversion to alzheimer's disease based on incomplete biomarkers. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(2):206–215, 2015.

[37] Yanbei Liu, Lianxi Fan, Changqing Zhang, Tao Zhou, Zhitao Xiao, Lei Geng, and Dinggang Shen. Incomplete multi-modal representation learning for alzheimer's disease diagnosis. *Medical Image Analysis*, 69:101953, 2021.

[38] Li Zhong and Xiao-Fen Chen. The emerging roles and therapeutic potential of soluble trem2 in alzheimer's disease. *Frontiers in aging neuroscience*, 11:328, 2019.

[39] Lei Cai, Zhengyang Wang, Hongyang Gao, Dinggang Shen, and Shuiwang Ji. Deep adversarial learning for multi-modality missing data completion. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1158–1166, 2018.

[40] Xingyu Gao, Feng Shi, Dinggang Shen, and Manhua Liu. Task-induced pyramid and attention gan for multimodal brain image imputation and classification in alzheimer's disease. *IEEE journal of biomedical and health informatics*, 26(1):36–43, 2021.

[41] Tao Zhou, Mingxia Liu, Kim-Han Thung, and Dinggang Shen. Latent representation learning for alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. *IEEE transactions on medical imaging*, 38(10):2411–2422, 2019.

[42] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[43] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[44] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[45] Md Abdur Rahaman, Jiayu Chen, Zening Fu, Noah Lewis, Armin Iraji, and Vince D Calhoun. Multi-modal deep learning of functional and structural neuroimaging and genomic data to predict mental illness. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3267–3272. IEEE, 2021.

[46] Shaker El-Sappagh, Jose M Alonso, SM Islam, Ahmad M Sultan, and Kyung Sup Kwak. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for alzheimer's disease. *Scientific reports*, 11(1):1–26, 2021.

[47] Yan-Shi Hu, Juncai Xin, Ying Hu, Lei Zhang, and Ju Wang. Analyzing the genes related to alzheimer's disease via a network and pathway-based approach. *Alzheimer's research & therapy*, 9:1–15, 2017.

[48] Michael W Weiner, Dallas P Veitch, Paul S Aisen, Laurel A Beckett, Nigel J Cairns, Robert C Green, Danielle Harvey, Clifford R Jack Jr, William Jagust, John C Morris, et al. The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & Dementia*, 13(5):561–571, 2017.

[49] Yuhui Du, Zening Fu, Jing Sui, Shuang Gao, Ying Xing, Dongdong Lin, Mustafa Salman, Anees Abrol, Md Abdur Rahaman, Jiayu Chen, et al. Neuromark: An automated and adaptive ica based pipeline to identify reproducible fmri markers of brain disorders. *NeuroImage: Clinical*, 28:102375, 2020.

[50] Iris E Jansen, Jeanne E Savage, Kyoko Watanabe, Julien Bryois, Dylan M Williams, Stacy Steinberg, Julia Sealock, Ida K Karlsson, Sara Hägg, Lavinia Athanasiu, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing alzheimer's disease risk. *Nature genetics*, 51(3):404–413, 2019.

[51] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.

[52] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[53] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[54] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[55] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR, 2017.

[56] Daniel D Lundstrom, Tianjian Huang, and Meisam Razaviyayn. A rigorous study of integrated gradients method and extensions to internal neuron attributions. In *International Conference on Machine Learning*, pages 14485–14508. PMLR, 2022.

[57] Pascal Sturmfels, Scott Lundberg, and Su-In Lee. Visualizing the impact of feature attribution baselines. *Distill*, 5(1):e22, 2020.

[58] Antonios Mamalakis, Elizabeth A Barnes, and Imme Ebert-Uphoff. Carefully choose the baseline: Lessons learned from applying xai attribution methods for regression tasks in geoscience. *Artificial Intelligence for the Earth Systems*, 2(1):e220058, 2023.

[59] Giorgio Dolci, Federica Cruciani, Ilaria Boscolo Galazzo, Vince D Calhoun, and Gloria Menegaz. Objective assessment of the bias introduced by baseline signals in xai attribution methods. In *2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, pages 266–271. IEEE, 2023.

[60] Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

[61] J Patrick Royston. An extension of shapiro and wilk's w test for normality to large samples. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(2):115–124, 1982.

[62] Myles Hollander, Douglas A Wolfe, and Eric Chicken. *Nonparametric statistical methods*. John Wiley & Sons, 2013.

[63] David F Bauer. Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, 67(339):687–690, 1972.

[64] William McLaren, Laurent Gil, Sarah E Hunt, Harpreet Singh Riat, Graham RS Ritchie, Anja Thormann, Paul Flicek, and Fiona Cunningham. The ensembl variant effect predictor. *Genome biology*, 17(1):1–14, 2016.

[65] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000.

[66] Suzi A Aleksander, James Balhoff, Seth Carbon, J Michael Cherry, Harold J Drabkin, Dustin Ebert, Marc Feuermann, Pascale Gaudet, Nomi L Harris, et al. The gene ontology knowledgebase in 2023. *Genetics*, 224(1):iyad031, 2023.

[67] Paul D Thomas, Dustin Ebert, Anushya Muruganujan, Tremayne Mushayahama, Laurent-Philippe Albou, and Huaiyu Mi. Panther: Making genome-scale phylogenetics accessible to all. *Protein Science*, 31(1):8–22, 2022.

[68] Haizhou Ye, Qi Zhu, Yuan Yao, Yichao Jin, and Daoqiang Zhang. Pairwise feature-based generative adversarial network for incomplete multi-modal alzheimer's disease diagnosis. *The Visual Computer*, pages 1–10, 2022.

[69] Stephane Lehericy, Michel Baulac, Jacques Chiras, Laurent Pierot, Nadine Martin, Bernard Pillon, Bernard Deweer, Bruno Dubois, and Claude Marsault. Amygdalohippocampal mr volume measurements in the early stages of alzheimer disease. *American Journal of Neuroradiology*, 15(5):929–937, 1994.

[70] Josephine Barnes, Jonathan W Bartlett, Laura A van de Pol, Clement T Loy, Rachael I Scahill, Chris Frost, Paul Thompson, and Nick C Fox. A meta-analysis of hippocampal atrophy rates in alzheimer's disease. *Neurobiology of aging*, 30(11):1711–1723, 2009.

[71] Yawu Liu, Teemu Paajanen, Yi Zhang, Eric Westman, Lars-Olof Wahlund, Andrew Simmons, Catherine Tunnard, Tomasz Sobow, Patrizia Mecocci, Magda Tsolaki, et al. Analysis of regional mri volumes and thicknesses as predictors of conversion from mild cognitive impairment to alzheimer's disease. *Neurobiology of aging*, 31(8):1375–1385, 2010.

[72] Devin J Sodums and Véronique D Bohbot. Negative correlation between grey matter in the hippocampus and caudate nucleus in healthy aging. *Hippocampus*, 30(8):892–908, 2020.

[73] K Persson, VD Bohbot, Nenad Bogdanovic, Geir Selbæk, Anne Brækhus, and Knut Engedal. Finding of increased caudate nucleus in patients with alzheimer's disease. *Acta Neurologica Scandinavica*, 137(2):224–232, 2018.

[74] Serge ARB Rombouts, Frederik Barkhof, Menno P Witter, and Philip Scheltens. Unbiased whole-brain analysis of gray matter loss in alzheimer's disease. *Neuroscience letters*, 285(3):231–233, 2000.

[75] Sudevan Jiji, Karavallil Achuthan Smitha, Arun Kumar Gupta, Vellara Pappukutty Mahadevan Pillai, and Ramapurath S Jayasree. Segmentation and volumetric analysis of the caudate nucleus in alzheimer's disease. *European journal of radiology*, 82(9):1525–1530, 2013.

[76] GB Frisoni, C Testa, A Zorzan, F Sabattoli, A Beltramello, H Soininen, and MP Laakso. Detection of grey matter loss in mild alzheimer's disease with voxel based morphometry. *Journal of Neurology, Neurosurgery & Psychiatry*, 73(6):657–664, 2002.

[77] Wen-Ying Wang, Jin-Tai Yu, Yong Liu, Rui-Hua Yin, Hui-Fu Wang, Jun Wang, Lin Tan, Joaquim Radua, and Lan Tan. Voxel-based meta-analysis of grey matter changes in alzheimer's disease. *Translational neurodegeneration*, 4(1):1–9, 2015.

[78] Zhengjia Dai, Chaogan Yan, Zhiqun Wang, Jinhui Wang, Mingrui Xia, Kuncheng Li, and Yong He. Discriminative analysis of early alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (m3). *Neuroimage*, 59(3):2187–2195, 2012.

26

[79] Liana G Apostolova, Calen A Steiner, Gohar G Akopyan, Rebecca A Dutton, Kiralee M Hayashi, Arthur W Toga, Jeffrey L Cummings, and Paul M Thompson. Three-dimensional gray matter atrophy mapping in mild cognitive impairment and mild alzheimer disease. *Archives of neurology*, 64(10):1489–1495, 2007.

[80] Giorgos Karas, Philip Scheltens, Serge Rombouts, Ronald Van Schijndel, Martin Klein, Bethany Jones, Wiesje Van Der Flier, Hugo Vrenken, and Frederik Barkhof. Precuneus atrophy in early-onset alzheimer's disease: a morphometric structural mri study. *Neuroradiology*, 49:967–976, 2007.

[81] Christiane Möller, Hugo Vrenken, Lize Jiskoot, Adriaan Versteeg, Frederik Barkhof, Philip Scheltens, and Wiesje M van der Flier. Different patterns of gray matter atrophy in early-and late-onset alzheimer's disease. *Neurobiology of aging*, 34(8):2014–2022, 2013.

[82] Yafeng Zhan, Hongxiang Yao, Pan Wang, Bo Zhou, Zengqiang Zhang, Ningyu An, Jianhua Ma, Xi Zhang, Yong Liu, et al. Network-based statistic show aberrant functional connectivity in alzheimer's disease. *IEEE Journal of Selected Topics in Signal Processing*, 10(7):1182–1188, 2016.

[83] Pan Wang, Bo Zhou, Hongxiang Yao, Yafeng Zhan, Zengqiang Zhang, Yue Cui, Kaibin Xu, Jianhua Ma, Luning Wang, Ningyu An, et al. Aberrant intra-and inter-network connectivity architectures in alzheimer's disease and mild cognitive impairment. *Scientific reports*, 5(1):1–12, 2015.

[84] Jessica S Damoiseaux, Katherine E Prater, Bruce L Miller, and Michael D Greicius. Functional connectivity tracks clinical deterioration in alzheimer's disease. *Neurobiology of aging*, 33(4):828–e19, 2012.

[85] Mark W Albers, Grover C Gilmore, Jeffrey Kaye, Claire Murphy, Arthur Wingfield, David A Bennett, Adam L Boxer, Aron S Buchman, Karen J Cruickshanks, Davangere P Devanand, et al. At the interface of sensory and motor dysfunctions and alzheimer's disease. *Alzheimer's & Dementia*, 11(1):70–98, 2015.

[86] Lucía Penalba-Sánchez, Patrícia Oliveira-Silva, Alexander Luke Sumich, and Ignacio Cifre. Increased functional connectivity patterns in mild alzheimer's disease: A rsfmri study. *Frontiers in Aging Neuroscience*, 14:1037347, 2023.

[87] ŁLukasz Zadka, Marta Sochocka, Naomi Hachiya, Justyna Chojdak-ŁLukasiewicz, Piotr Dziegiel, Egbert Piasecki, and Jerzy Leszek. Endocytosis and alzheimer's disease. *Geroscience*, 46(1):71–85, 2024.

[88] Leila A Shobab, Ging-Yuek R Hsiung, and Howard H Feldman. Cholesterol in alzheimer's disease. *The Lancet Neurology*, 4(12):841–852, 2005.

[89] Makoto Michikawa. The role of cholesterol in pathogenesis of alzheimer's disease: dual metabolic interaction between amyloid β-protein and cholesterol. *Molecular neurobiology*, 27:1–12, 2003.

[90] Nikolas Dovrolis, Maria Nikou, Alexandra Gkrouzoudi, Nikolaos Dimitriadis, and Ioanna Maroulakou. Unlocking the memory component of alzheimer's disease: Biological processes and pathways across brain regions. *Biomolecules*, 12(2):263, 2022.

[91] Julia Mills and Peter B Reiner. Regulation of amyloid precursor protein cleavage. *Journal of neurochemistry*, 72(2):443–460, 1999.

[92] Min Kim, Stuart Snowden, Tommi Suvitaival, Ashfaq Ali, David J Merkler, Tahmina Ahmad, Sarah Westwood, Alison Baird, Petroula Proitsi, Alejo Nevado-Holgado, et al. Primary fatty amides in plasma associated with brain amyloid burden, hippocampal volume, and memory in the european medical information framework for alzheimer's disease biomarker discovery cohort. *Alzheimer's & Dementia*, 15(6):817–827, 2019.

[93] Philip Scheltens, Bart De Strooper, Miia Kivipelto, Henne Holstege, Gael Chételat, Charlotte E Teunissen, Jeffrey Cummings, and Wiesje M van der Flier. Alzheimer's disease. *The Lancet*, 397(10284):1577–1590, 2021.

[94] Cheng Ma, Jin Li, Zhijun Bao, Qingwei Ruan, Zhuowei Yu, et al. Serum levels of apoa1 and apoa2 are associated with cognitive status in older men. *BioMed research international*, 2015, 2015.

[95] Jungsu Kim, Jacob M Basak, and David M Holtzman. The role of apolipoprotein e in alzheimer's disease. *Neuron*, 63(3):287–303, 2009.

[96] David M Holtzman, Joachim Herz, and Guojun Bu. Apolipoprotein e and apolipoprotein e receptors: normal biology and roles in alzheimer disease. *Cold Spring Harbor perspectives in medicine*, 2(3):a006312, 2012.

[97] Hao Wang, Joshua A Kulas, Chao Wang, David M Holtzman, Heather A Ferris, and Scott B Hansen. Regulation of beta-amyloid production in neurons by astrocyte-derived cholesterol. *Proceedings of the National Academy of Sciences*, 118(33):e2102191118, 2021.

[98] Sonia Sanz Muñoz, Brett Garner, and Lezanne Ooi. Understanding the role of apoe fragments in alzheimer's disease. *Neurochemical research*, 44(6):1297–1305, 2019.

[99] Li Zhong, Ying Xu, Rengong Zhuo, Tingting Wang, Kai Wang, Ruizhi Huang, Daxin Wang, Yue Gao, Yifei Zhu, Xuan Sheng, et al. Soluble trem2 ameliorates pathological phenotypes by modulating microglial functions in an alzheimer's disease model. *Nature communications*, 10(1):1365, 2019.

[100] Shiraz Dib, Jens Pahnke, and Fabien Gosselet. Role of abca7 in human health and in alzheimer's disease. *International Journal of Molecular Sciences*, 22(9):4603, 2021.

[101] Ian J Martins, Tamar Berger, Matthew J Sharman, Giuseppe Verdile, Stephanie J Fuller, and Ralph N Martins. Cholesterol metabolism and transport in the pathogenesis of alzheimer's disease. *Journal of neurochemistry*, 111(6):1275–1308, 2009.

[102] Liana G Apostolova, Shannon L Risacher, Tugce Duran, Eddie C Stage, Naira Goukasian, John D West, Triet M Do, Jonathan Grotts, Holly Wilhalme, Kwangsik Nho, et al. Associations of the top 20 alzheimer disease risk variants with brain amyloidosis. *JAMA neurology*, 75(3):328–341, 2018.

[103] Meredith N Braskie, Neda Jahanshad, Jason L Stein, Marina Barysheva, Katie L McMahon, Greig I de Zubicaray, Nicholas G Martin, Margaret J Wright, John M Ringman, Arthur W Toga, et al. Common alzheimer's disease risk variant within the clu gene affects white matter microstructure in young adults. *Journal of Neuroscience*, 31(18):6764–6770, 2011.

[104] Lei Yu, Lori B Chibnik, Gyan P Srivastava, Nathalie Pochet, Jingyun Yang, Jishu Xu, James Kozubek, Nikolaus Obholzer, Sue E Leurgans, Julie A Schneider, et al. Association of brain dna methylation in sorl1, abca7, hla-drb5, slc24a4, and bin1 with pathological diagnosis of alzheimer disease. *JAMA neurology*, 72(1):15–24, 2015.

[105] J Chapuis, F Hansmannel, Marc Gistelinck, A Mounier, C Van Cauwenberghe, KV Kolen, F Geller, Y Sottejeau, D Harold, P Dourlen, et al. Increased expression of bin1 mediates alzheimer genetic risk by modulating tau pathology. *Molecular psychiatry*, 18(11):1225–1234, 2013.

[106] Clifford R Jack, David A Bennett, Kaj Blennow, Maria C Carrillo, Howard H Feldman, Giovanni B Frisoni, Harald Hampel, William J Jagust, Keith A Johnson, David S Knopman, et al. A/t/n: An unbiased descriptive classification scheme for alzheimer disease biomarkers. *Neurology*, 87(5):539–547, 2016.

Supplementary Materials: An interpretable generative multimodal neuroimaging-genomics framework for decoding Alzheimer's disease

Giorgio Dolci[a,b,], Federica Cruciani[a*], Md Abdur Rahaman[b], Anees Abrol[b], Jiayu Chen[b],
Zening Fu[b], Ilaria Boscolo Galazzo[a], Gloria Menegaz[a,**], Vince D. Calhoun[b,**],
for the Alzheimer's Disease Neuroimaging Initiative[1]

[a]*Department of Engineering and Innovation Medicine, University of Verona, Verona, Italy*
[b]*Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA*

Preprocessing quality control

A thorough quality control (QC) was performed to retain scans with good normalization to the standard MNI space, which involved discarding sMRI and fMRI images that exhibited low correlation with individual and/or group-level masks. In this spatial correlation process, we first calculated subject-level masks using the subject MRI scans (only the first volume in case of fMRI) by setting the brain voxels to 1 if the values of these voxels were greater than 80% of the average value across whole-brain voxels, and 0 otherwise. Next, after computing the subject-level masks, we calculated a group mask by setting the voxels to 1 for which at least 70% of the subject-level masks had a value of 1. Lastly, we examined the spatial correlations of the subject and group level masks and retained subjects that showed a correlation value greater than 0.85. Additionally, for fMRI, scans with larger head motion parameters ($> 3°$ rotations and > 3 mm translations) were discarded.

---

*Corresponding author: Department of Engineering and Innovation Medicine, University of Verona, Verona, Italy.
e-mail: federica.cruciani@univr.it
**Equally contributed as last authors to this work.

Results

Table 1: In this Table are reported the correspondences acronym - full name of the sMRI brain regions under analysis.

Acronym	Full name	Acronym	Full name
Ins	Insular Cortex	OFG	Occipital Fusiform Gyrus
TP	Temporal Pole	COpC	Central Opercular Cortex
ScC	Subcallosal Cortex	POpC	Parietal Operculum Cortex
CGp	Cingulate Gyrus, posterior division	Thl	Thalamus
PcC	Precuneous Cortex	Cau	Caudate
FOC	Frontal Orbital Cortex	Put	Putamen
PhGa	Parahippocampal Gyrus, anterior division	Pall	Pallidum
PaGp	Parahippocampal Gyrus, posterior division	Hipp	Hippocampus
LG	Lingual Gyrus	Amy	Amygdala
TFCp	Temporal Fusiform Cortex, posterior division	Acc	Accumbens
TOF	Temporal Occipital Fusiform Cortex		

Table 2: In this Table are reported the 53 ICs present in the sFNC matrices with corresponding brain region name, network at which it belongs to, and spatial location in the brain along X, Y, and Z axis. SC=Sub-cortical; AU=Auditory; SM=sensorimotor; VI=visual; CC=cognitive-control; DM=default-mode; and CB=cerebellar. In *italic* are highlighted the brain regions present in the fMRI connectograms of Results section.

IC ID	Brain region	Network	X	Y	Z	IC ID	Brain region	Network	X	Y	Z
1	Caudate	SC	6.5	10.5	5.5	26	Inferior parietal lobule	CC	45.5	-61.5	43.5
2	Subthalamus/hypothalamys	SC	-2.5	-13.5	-1.5	27	Insula	CC	-30.5	22.5	-3.5
3	Putamen	SC	-26.5	1.5	-0.5	28	*Superior medial frontal gyrus*	CC	-0.5	50.5	29.5
4	Caudate	SC	21.5	10.5	-3.5	29	Inferior frontal gyrus	CC	-48.5	34.5	-0.5
5	Thalamus	SC	-12.5	-18.5	11.5	30	Right inferior frontal gyrus	CC	53.5	22.5	13.5
6	Superior temporal gyrus	AU	62.5	-22.5	7.5	31	Middle frontal gyrus	CC	-41.5	19.5	26.5
7	Middle temporal gyrus	AU	-42.5	-6.5	10.5	32	Inferior parietal lobule	CC	-53.5	-49.5	43.5
8	*Postcentral gyrus*	SM	56.5	-4.5	28.5	33	*Left inferior parietal lobule*	CC	44.5	-34.5	46.5
9	*Left postcentral gyrus*	SM	-38.5	-22.5	56.5	34	Supplementary motor area	CC	-6.5	13.5	64.5
10	*Paracentral lobule*	SM	0.5	-22.5	65.5	35	Superior frontal gyrus	CC	-24.5	26.5	49.5
11	*Right postcentral gyrus*	SM	38.5	-19.5	55.5	36	Middle frontal gyrus	CC	30.5	41.5	28.5
12	*Superior parietal lobule*	SM	-18.5	-43.5	65.5	37	Hippocampus	CC	23.5	-9.5	-16.5
13	*Paracentral lobule*	SM	-18.5	-9.5	56.5	38	*Left inferior parietal lobule*	CC	45.5	-61.5	43.5
14	*Precentral gyrus*	SM	-42.5	-7.5	46.5	39	Middle cingulate cortex	CC	-15.5	20.5	37.5
15	*Superior parietal lobule*	SM	20.5	-63.5	58.5	40	Inferior frontal gyrus	CC	39.5	44.5	-0.5
16	Postcentral gyrus	SM	-47.5	-27.5	43.5	41	Middle frontal gyrus	CC	-26.5	47.5	5.5
17	*Calcarine gyrus*	VI	-12.5	-66.5	8.5	42	*Hippocampus*	CC	-24.5	-36.5	1.5
18	*Middle occipital gyrus*	VI	-23.5	-93.5	-0.5	43	Precuneus	DM	-8.5	-66.5	35.5
19	Middle temporal gyrus	VI	48.5	-60.5	10.5	44	*Precuneus*	DM	-12.5	-54.5	14.5
20	*Cuneus*	VI	15.5	-91.5	22.5	45	*Anterior cingulate cortex*	DM	-2.5	35.5	2.5
21	*Right middle occipital gyrus*	VI	38.5	-73.5	6.5	46	Posterior cingulate cortex	DM	-5.5	-28.5	26.5
22	*Fusiform gyrus*	VI	29.5	-42.5	-12.5	47	*Anterior cingulate cortex*	DM	-9.5	46.5	-10.5
23	*Inferior occipital gyrus*	VI	-36.5	-76.5	-4,5	48	*Precuneus*	DM	-0.5	-48.5	49.5
24	*Lingual gyrus*	VI	-8.5	-81.5	-4.5	49	Posterior cingulate cortex	DM	-2.5	54.5	31.5
25	*Middle temporal gyrus*	VI	-44.5	-57.5	-7.5	50	*Cerebellum*	CB	-30.5	-54.5	-42.5
						51	*Cerebellum*	CB	-32.5	-79.5	-37.5
						52	Cerebellum	CB	20.5	-48.5	-40.5
						53	*Cerebellum*	CB	30.5	-63.5	-40.5

Table 3: Most significant biological processes for AD patients obtained from the analysis of the most relevant SNPs with positive IG attributions. In this Table are shown the biological processes, their raw p-value, and the overlap genes. * indicates statistically significant biological processes after Bonferroni correction.

Biological processes	GO term	p-value	Overlap genes
Intracellular transport*	GO:0046907	0.000044 ($p_{bonf} = 0.0225$)	BLOC1S3; KLC3; TOMM40; PICALM; MME; NUP88; TMEM106B; NSF; NUP43; BIN1; RAB12; APOE; SORL1; CLU
Regulation of protein-containing complex assembly*	GO:0043254	0.000080 ($p_{bonf} = 0.0413$)	PTK2B; MARK4; FNIP2; PLCG2; BIN1; APOE; SORL1; TREM2; CLU
Establishment of localization in cell*	GO:0051649	0.000086 ($p_{bonf} = 0.0446$)	BLOC1S3; KLC3; TOMM40; PICALM; MME; NUP88; TMEM106B; NSF; NUP43; BIN1; RAB12; APOE; SORL1; NECTIN2; CLU
Cell development*	GO:0048468	0.000091 ($p_{bonf} = 0.0470$)	BLOC1S3; KLC3; PICALM; PTK2B; ALDH1A2; TNXB; TMEM106B; FCER1G; OOSP2; PLCG2; IGSF23; BIN1; ERCC2; RELB; APOE; ETV1; NECTIN2; TREM2; CLU
Cell projection organization	GO:0030030	0.000253	BLOC1S3; KLC3; PICALM; PTK2B; TNXB; TMEM106B; MARK4; CDH13; APOE; ETV1; MTSS2; NECTIN2; TREM2
Plasma membrane bounded cell projection organization	GO:0120036	0.000253	BLOC1S3; KLC3; PICALM; PTK2B; TNXB; TMEM106B; MARK4; CDH13; APOE; ETV1; MTSS2; NECTIN2; TREM2
Membrane organization	GO:0061024	0.000253	TOMM40; PICALM; RABEP1; NSF; APOA2; BIN1; CR1; RAB12; APOE; MTSS2; NECTIN2; TREM2; CLU
Developmental maturation	GO:0021700	0.000417	BLOC1S3; PICALM; PTK2B; ALDH1A2; SLC24A4; OOSP2; ERCC2
Positive regulation of endocytosis	GO:0045807	0.000417	FCER1G; PLCG2; APOA2; BIN1; APOE; TREM2; CLU
Regulation of vesicle-mediated transport	GO:0060627	0.000427	PICALM; NSF; FCER1G; PLCG2; APOA2; BIN1; RAB12; CDH13; APOE; SORL1; TREM2; CLU
Endomembrane system organization	GO:0010256	0.000500	BLOC1S3; NSF; USP8; BIN1; CR1; USP6NL; MTSS2; NECTIN2; CLU
Nitrogen compound transport	GO:0071705	0.000741	BLOC1S3; TOMM40; RABEP1; MME; NUP88; NSF; NUP43; CR1; RAB12; APOE; SORL1; CLU
Organelle localization	GO:0051640	0.000771	BLOC1S3; PICALM; NUP88; TMEM106B; CLU; BIN1; RAB12; NECTIN2
Cellular component organization or biogenesis	GO:0071840	0.000792	BLOC1S3; KLC3; TOMM40; PICALM; RABEP1; MME; NUP88; PTK2B; ALDH1A2; PHB1; TNXB; TMEM106B; NSF; MARK4; KCTD1; CCDC6; CLU; MINDY2; LTBP4; APOA2; USP8; BIN1; CR1; ERCC2; USP6NL; RAB12; GEMIN7; CDH13; APOE; ETV1; MTSS2; NECTIN2; TREM2; CLU
Localization	GO:0051179	0.000950	BLOC1S3; KLC3; TOMM40; PICALM; RABEP1; MME; NUP88; TMEM106B; NSF; FCER1G; SLC24A4; CLU; PLCG2; LTBP4; APOA2; NUP43; BIN1; CR1; USP6NL; RAB12; NBEAL1; CDH13; APOE; SORL1; NECTIN2; TREM2; CLU

Table 4: Most significant biological processes for MCIc patients obtained from the analysis of the most relevant SNPs with positive IG attributions. In this Table are shown the biological processes, their raw p-value, and the overlap genes.

Biological processes	GO term	p-value	Overlap genes
Positive regulation of amide metabolic process	GO:0034250	0.00109	PICALM; RPS27L; PTK2B; APOE; CLU
Protein-lipid complex assembly	GO:0065005	0.00109	APOC1; ABCA7; APOA2; BIN1; APOE
Regulation of amide metabolic process	GO:0034248	0.00128	PICALM; RPS27L; PTK2B; ABCA7; BIN1; APOE; CELF1; CLU
Positive regulation of endocytosis	GO:0045807	0.00147	FCER1G; ABCA7; PLCG2; APOA2; BIN1; APOE; CLU
Plasma lipoprotein particle assembly	GO:0034377	0.00436	APOC1; ABCA7; APOA2; APOE
Regulation of endocytosis	GO:0030100	0.00453	PICALM; APOC1; FCER1G; ABCA7; PLCG2; APOA2; BIN1; APOE; CLU
Sterol transport	GO:0015918	0.00519	APOC1; ABCA7; APOA2; APOE; CLU
Organic hydroxy compound transport	GO:0015850	0.00519	APOC1; ABCA7; APOA2; APOE; CLU
Cholesterol transport	GO:0030301	0.00519	APOC1; ABCA7; APOA2; APOE; CLU
Protein-lipid complex organization	GO:0071825	0.00519	APOC1; ABCA7; APOA2; BIN1; APOE
Regulation of lipid metabolic process	GO:0019216	0.00841	LACTB; APOC1; ABCA7; PLCG2; APOA2; EPHX2; APOE
Negative regulation of amide metabolic process	GO:0034249	0.01440	ABCA7; BIN1; APOE; CELF1; CLU
Negative regulation of amyloid precursor protein catabolic process	GO:1902992	0.01440	PICALM; ABCA7; BIN1; APOE; CLU
Monoatomic cation homeostasis	GO:0055080	0.01440	PICALM; TMEM106B; SLC24A4; TSPOAP1; APOE
Intracellular monoatomic cation homeostasis	GO:0030003	0.01440	PICALM; TMEM106B; SLC24A4; TSPOAP1; APOE

Table 5: Most significant biological processes for MCInc patients obtained from the analysis of the most relevant SNPs with positive IG attributions. In this Table are shown the biological processes, their raw p-value, and the overlap genes.

| Biological processes	GO term	p-value	Overlap genes
T cell activation	GO:0042110	0.00695	FCER1G; TREML2; SPI1; RELB
Lymphocyte activation involved in immune response	GO:0002285	0.00695	FCER1G; PLCG2; ERCC1; RELB
Negative regulation of hydrolase activity	GO:0051346	0.00695	PICALM; APOA2; BIN1; CR1
T cell differentiation	GO:0030217	0.00836	FCER1G; SPI1; RELB
Negative regulation of peptidase activity	GO:0010466	0.00836	PICALM; BIN1; CR1
Negative regulation of endopeptidase activity	GO:0010951	0.00836	PICALM; BIN1; CR1
Adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	GO:0002460	0.01170	FCER1G; CR1L; HLA-DRB1; ERCC1; CR1; RELB
Immune effector process	GO:0002252	0.01800	FCER1G; CR1L; PLCG2; ACE; ERCC1; CR1; RELB
Leukocyte mediated immunity	GO:0002443	0.02020	FCER1G; CR1L; ACE; ERCC1; CR1
Male gonad development	GO:0008584	0.02850	ACE; ERCC1; PLEKHA1
Male sex differentiation	GO:0046661	0.02850	ACE; ERCC1; PLEKHA1
Development of primary male sexual characteristics	GO:0046546	0.02850	ACE; ERCC1; PLEKHA1
Antigen processing and presentation of peptide antigen	GO:0048002	0.02850	FCER1G; ACE; HLA-DRB1
Lymphocyte activation	GO:0046649	0.03080	FCER1G; TREML2; PLCG2; ERCC1; SPI1; RELB
Positive regulation of lymphocyte activation	GO:0051251	0.03400	BLOC1S3; HLA-DRB1; SMARCD3; CR1; SPI1
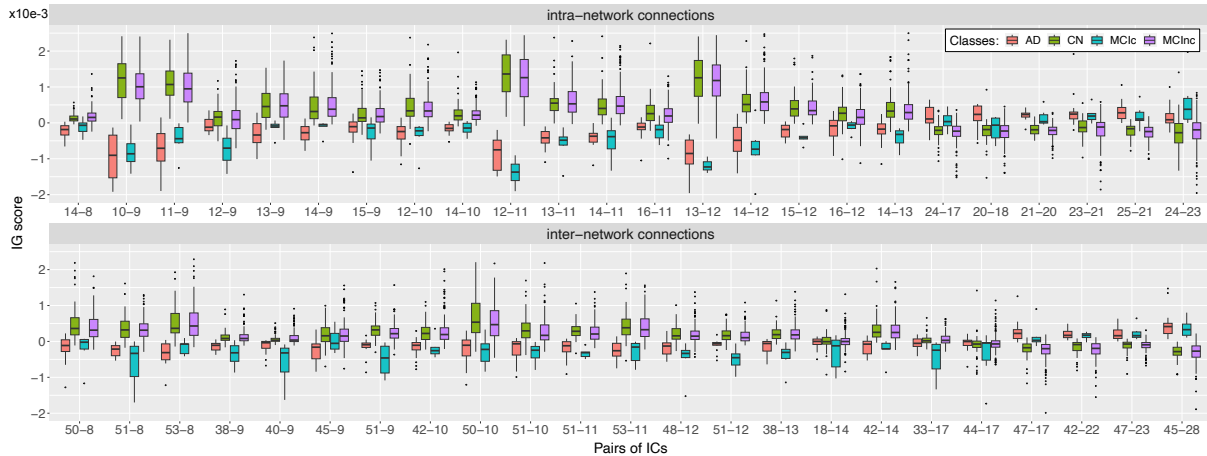
Figure 1: Boxplots that show the distributions of the subjects CN, AD, MCInc, and MCIc in the most important connections based on the fMRI IG score.