# Studying the Manifold Structure of Alzheimer's Disease: A Deep Learning Approach Using Convolutional Autoencoders

Francisco J. Martinez-Murcia, Andres Ortiz, Juan-Manuel Gorriz, Javier Ramirez, and Diego Castillo-Barnes

***Abstract*—Many classical machine learning techniques have been used to explore Alzheimer's disease (AD), evolving from image decomposition techniques such as principal component analysis toward higher complexity, non-linear decomposition algorithms. With the arrival of the deep learning paradigm, it has become possible to extract high-level abstract features directly from MRI images that internally describe the distribution of data in low-dimensional manifolds. In this work, we try a new exploratory data analysis of AD based on deep convolutional autoencoders. We aim at finding links between cognitive symptoms and the underlying neurodegeneration process by fusing the information of neuropsychological test outcomes, diagnoses, and other clinical data with the imaging features extracted solely via a data-driven decomposition of MRI. The distribution of the extracted features in different combinations is then analyzed and visualized using regression and classification analysis, and the influence of each coordinate of the autoencoder manifold over the brain is estimated. The imaging-derived markers could then predict clinical variables with correlations above 0.6 in the case of neuropsychological evaluation variables such as the MMSE or the ADAS11 scores, achieving a classification accuracy over 80% for the diagnosis of AD.**

***Index Terms*—Alzheimer's disease, deep learning, convolutional autoencoder, manifold learning, data fusion.**

Manuscript received January 11, 2019; revised March 21, 2019 and April 24, 2019; accepted April 30, 2019. Date of publication June 17, 2019; date of current version January 6, 2020. This work was partly supported by the MINECO/ FEDER under the TEC2015-64718-R, RTI2018-098913-B-I00 and PGC2018-098813-B-C32 projects. The work of F. J. Martinez-Murcia was supported by the MICINN "Juan de la Cierva" Fellowship. This work of Data for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). This work of ADNI was supported by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. *(Corresponding author: Francisco J. Martinez-Murcia.)*

F. J. Martinez-Murcia and A. Ortiz are with the Department of Communications Engineering, University of Malaga, 29016 Malaga, Spain, and also with the Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), 18071 Granada, Spain (e-mail: fjesusmartinez@ugr.es; aortiz@ic.uma.es).

J. M. Gorriz, J. Ramirez, and D. Castillo-Barnes are with the Department of Signal Theory, Networking and Communications,University of Granada, 18071 Granada, Spain, and also with the DaSCI, 18071 Granada, Spain. (e-mail: gorriz@ugr.es; javierrp@ugr.es; diegoc@ugr.es).

Digital Object Identifier 10.1109/JBHI.2019.2914970

## I. Introduction

ALZHEIMER'S Disease (AD) is the most common type of neurodegenerative disease in the world, affecting more than 5% of the population in Europe [1] and with an incidence of 11.08 per 1000 person-years. With a yet unknown aetiology, current diagnosis of AD often depend on clinical history and the outcomes of widely extended neuropsychological tests such as the Mini-Mental State Exam (MMSE) that, according to recent studies [2], [3], may add confounding information to the procedure of diagnosis. Therefore, understanding the disease progression as well as studying and standardizing new disease markers is paramount [4].

Substantial advances in the technology used for neuroimaging can now track neurodegeneration even before the development of full-blown dementia [5]. However, the prodromal stages of AD are often confused with the cognitive decline associated with age or other diseases, in what is commonly known as Mild Cognitive Impairment (MCI). Longitudinal studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) [6], the Open Access Series of Imaging Studies (OASIS) [7] or the Dominantly Inherited Alzheimer Network (DIAN) [8] combine Magnetic Resonance Imaging (MRI), biological and cognitive markers to study these stages and observe which of the MCI affected subjects convert to AD.

These initiatives have largely contributed to the validation of computational analysis methodologies. Processing pipelines that combine spatial and intensity normalization and feature extraction, such as FreeSurfer [9] or Statistical Parametric Mapping (SPM) [10] are now commonplace. Structural and functional features are frequently used in differential diagnosis studies with great success [3], [4], [11], [12] and, together with the rise of Machine Learning (ML) methodologies, there are now complete systems that can recognize patterns associated with the diseases using automated feature extraction

[13]–[16], selection [17]–[20] and classification algorithms [21]–[25]. However, very few of them [13], [26] try to understand the relationship between test results, biomarkers, and structural changes in the brain, which are critical to validating diagnoses.

The revolutionary framework provided by Deep Learning (DL) architectures and the technological advances in computation have been key to biomedical image processing, as shown by different strategies using dense networks [27], two and three-dimensional convolutional neural networks [28], [29], residual nets [30], and many other examples [31], [32]. Among all of them, Autoencoders (AEs) [33], [34] have a very interesting feature: they can model the manifold of a dataset in a self-supervised way, which has successfully been applied to psychiatry [32], breast cancer [34] or AD [35]. Convolutional Autoencoders (CAEs) extends this architecture by using convolutional layers that can extract data-driven features directly from three-dimensional maps, making them optimal for image processing.

This work begins with the hypothesis that the underlying structure of an MRI image database is related to the progression of AD. By applying a CAE to model the dataset, we aim to predict clinical and neuropsychological scores as well as the progression of patients with cognitive impairment to dementia. This self-supervised decomposition of the original data could reveal an underlying manifold structure that correlates to other types of data such as biological markers (ApoE, Tau protein) or neuropsychological assessments, as well as a clear association to the traditional clusters: AD, MCI and Healthy Controls (CTLs). The main contributions of our work can be summarized as follows: (1) To the best of our knowledge, this approach is the first to directly apply–without prior extraction of features–deep convolutional feature representation methodologies to full MRI images in such a massive dataset. (2) Our approach not only provides classification into traditional AD clusters, but studies the relationship between the resulting manifold and a wide variety of clinical variables. (3) Of all deep image decomposition algorithms, it achieves the best performance in AD prognosis (AD conversion in MCI patients) with such large population. (4) It provides a novel model for visualizing the brain regions most influenced by each manifold coordinate that, when combined with neuropsychological tests, provides more information on how and where structural degeneration relates to the cognitive decline of dementia.

## II. Materials and Methods

### A. Dataset

*1) ADNI Dataset:* Data used in the preparation of this article were obtained from the ADNI database (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, other other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. In this article we have used a standard cohort: the "ADNI1:Complete 2Yr", containing a 2 year follow-up of the patients (every 6 and 12 months for MCI and CTL respectively) with assessments of cognitive function and MRI acquisitions (for further details visit www.adni-info.org). Information about MCI Stable (MCIS) and MCI Converters (MCIC), which respectively remain within MCI or convert to AD during the study will also be used. Since neurodegeneration is sadly not reversible, MCI-to-CTL and AD-to-MCI regression were considered possible mislabelling, and consequently excluded, resulting in a set of 2182 T1-weighted MRI images from 479 subjects (see demographics at Table I).

Other data from the dataset were also used; variables such as Age, Tau protein concentration in Cerebro-Spinal Fluid (TAU), presence of the Apolipoprotein E - $\varepsilon$4 (ApoE4) variant, the MMSE [6] or the Alzheimer's Disease ASsessment Scale (ADAS)-cognitive subscale (ADAS-cog, in 11 and 13-question variants, ADAS-11, ADAS-13 and ADAS-Q4) [36], and Clinical Dementia Ratio (CDRSB) will be used as output variables in the regression model.

*2) Data Preprocessing:* These images were spatially normalized to the standard TPM template using the SPM12 software [10] new Normalize with default parameters (non-rigid, preserve concentrations). In this work we will use Normalized (Norm) MRI images (no skull stripping), segmented Grey Matter (GM) and White Matter (WM) maps.

Further preprocessing has been applied to the MRI maps, in order to adapt them to the needs of the neural networks used. Due to memory and computational requirements, we have downsampled the input images by a factor between 1.5 and 3, to obtain final maps of size $64 \times 64 \times 64$. GM and WM maps are already in the intensity range [0, 1], but in the case of Norm images, we have to perform a intensity normalization procedure. To do so, we have used the so-called "integral normalization" [29] in which the whole array of intensities of the image $\mathbf{I}$ is divided by the average value of the whole image array $\bar{\mathbf{I}}$, obtaining a new map $\mathbf{I}' = \mathbf{I}/\bar{\mathbf{I}}$.

### B. Convolutional Autoencoders

In 2012, a major breakthrough in image analysis and processing occurred: the consolidation of Convolutional Neural Networks (CNNs) [27]–[29], [37]. Since then, its usage has dramatically increased, displacing the once-dominant dense-connected networks, such as the Multi-Layer Perceptron (MLP) [38]. The CNN paradigm introduces not only convolutional layers, but a whole new set of layers and algorithms that allow faster and more accurate models, among them the use of pooling, transpose convolutions or batch normalization [39], among others.

For its part, the AE architecture, a type of encoder-decoder architecture that minimizes the reconstruction error, have consistently been used for feature extraction and manifold modelling [33], [40]. These networks combine a encoder $f(x)$ and a decoder $g(x)$ network where the latter is typically the inverse of the former, connected via the Z-layer, a bottleneck fully-connected layer with $Z$ neurons. Given a set of samples $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^D\}$ of dimension $D$, the AE learns a low-dimensional representation of the dataset in the $Z$-layer $\mathbf{z}_i \in \mathbb{R}^Z$, that can be seen as the definition of a new manifold–the $Z$-manifold–and minimizes

TABLE I
DEMOGRAPHICS OF THE ADNI DATASET

| DX | Sex | # | Age | ADASQ4 | MMSE | ADAS13 | ADAS11 | CDRSB | TAU |
|---|---|---|---|---|---|---|---|---|---|
| CTL | Female | 83 | 76.43 [4.58] | 5.62 [3.00] | 2.43 [1.45] | 8.37 [4.11] | 5.62 [3.00] | 0.04 [0.14] | 236.10 [75.02] |
| | Male | 85 | 75.43 [5.55] | 6.24 [2.69] | 3.12 [1.73] | 9.85 [3.85] | 6.24 [2.69] | 0.01 [0.05] | 243.63 [80.17] |
| MCI | Female | 70 | 72.99 [7.43] | 10.80 [4.56] | 6.57 [2.64] | 18.35 [6.89] | 10.80 [4.56] | 1.50 [0.79] | 338.42 [126.75] |
| | Male | 142 | 75.62 [6.69] | 11.60 [4.26] | 6.11 [2.16] | 18.70 [6.04] | 11.60 [4.26] | 1.52 [0.86] | 290.37 [111.90] |
| AD | Female | 47 | 75.16 [7.75] | 18.63 [6.65] | 9.00 [1.18] | 29.40 [7.95] | 18.63 [6.65] | 4.54 [1.43] | 379.64 [138.62] |
| | Male | 52 | 75.26 [7.25] | 18.81 [5.61] | 8.63 [1.55] | 29.31 [6.92] | 18.81 [5.61] | 4.27 [1.67] | 318.84 [90.79] |

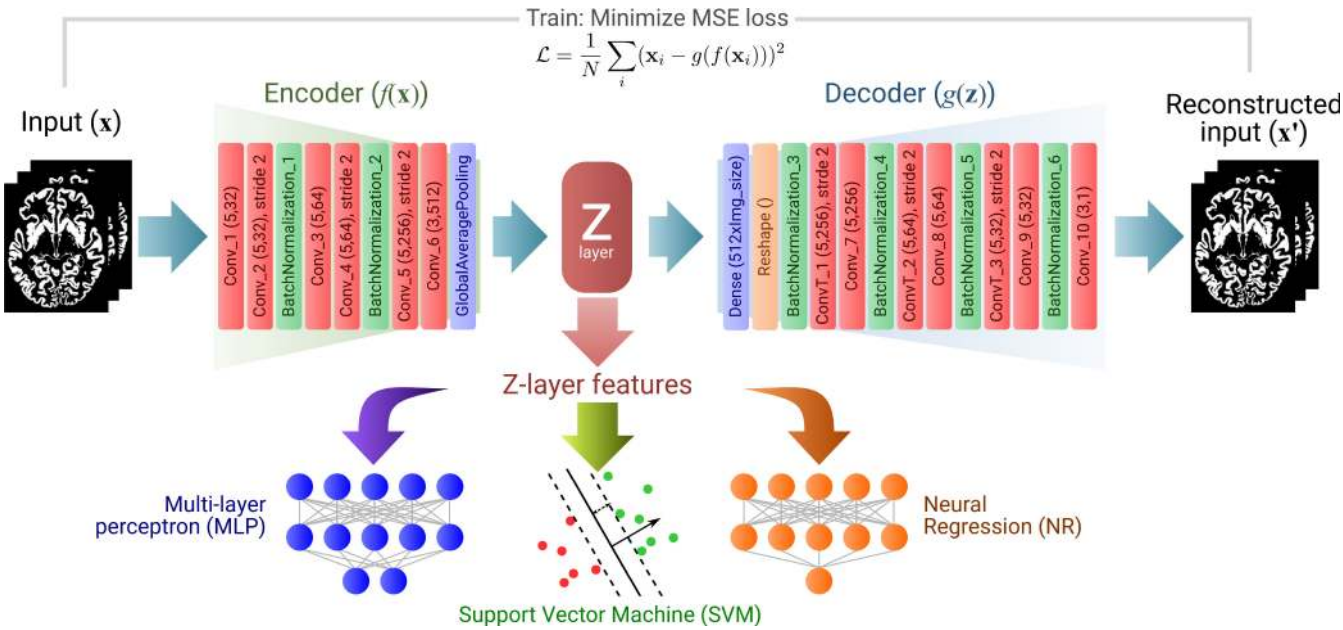



Fig. 1. Schema of the autoencoder architecture proposed in this work to model the underlying structure of the ADNI dataset, and the usage of the Z-layer features in classification and regression.

the reconstruction error between the input of the encoder $\mathbf{x}$ and the output of the decoder $\hat{\mathbf{x}} = g(f(\mathbf{x}))$. In this work, we use the Mean Squared Error (MSE) as reconstruction error between the input image $\mathbf{I}$ and the reconstructed image at the output $\hat{\mathbf{I}}_i = g(f(\mathbf{I}_i))$ (see Figure 1):

$$\mathcal{L} = \frac{1}{N}\sum_i (\mathbf{I}_i - g(f(\mathbf{I}_i)))^2 \tag{1}$$

Here, the output of each neuron at the $Z$-layer, or $Z$-features are computed for the $i$-th subject as:

$$\mathbf{z}_i = f(\mathbf{I}_i) \quad \rightarrow \quad \mathbf{z}_i \in \mathbb{R}^Z \tag{2}$$

and can be seen as the coordinates of each subject in the $Z$-manifold. In this work, they will be used as features in classification as well as to model the topology of the dataset.

In this work we use a convolutional AE, the CAE, particular type that uses all convolutional layers except for the $Z$-layer. This is a reasonable approach, since convolutional layers have many advantages over typical dense-connected networks, which include position-invariance–due to the combination of filters in the convolution and the pooling step–, or smaller memory requirements due to parameter sharing. It also allows us to use the volumetric image set as input $I = \{\mathbf{I}_i \in \mathbb{R}^{64\times64\times64}\}$. Most pooling layers, however, have been lately replaced by newer approaches in CNNs, such as the use of convolutions with stride, which provide similar results without loss of information [41].

Typical CAEs use dense layers at the end of the encoder and the beginning of the decoder to interface with the $Z$-layer. In this work, however, we replace the output of the encoder with a Global Average Pooling (GAP) layer [42], [43], which has proven to have less global overfitting and faster convergence while drastically reducing the number of parameters [42], achieving similar classification performance than other approaches but larger performance for localization of objects [44]. Following this approach, each of the neurons at the GAP layer will focus on a different cluster, allowing for a better discrimination of brain regions.

In all other cases, we have removed the dependency on pooling layers by using convolutions with stride ($S$) and convolution transpose layers. $S$ controls the overlapping of the receptive fields of each neuron (typically $S = 1$) but larger $S$ is analogous to the combination of convolution + pooling [41] without performance loss. In Figure 1, the layer name convention is: Conv_# $(P, K)$, and if $S \neq 1$, the stride is also presented.

Most layers used Rectifier Linear Unit (ReLU) activation:

$$f_{relu}(z) = \max(0, z) \tag{3}$$

except for the $Z$-layer, which used linear activation for a better modelling of the $Z$-manifold. Batch normalization was used to accelerate convergence, also acting as a regularizer [39].

The whole CAE was trained using the RMSprop optimization algorithm [45] (with learning rate 0.001, $\rho = 0.9$ and no decay or $\epsilon$). It took an average 4791.692 seconds with Standard Deviation (SD) of 27.46 in performing a whole 20-epoch training with the normalized $I = \{\mathbf{I}_i \in \mathbb{R}^{64\times64\times64}\}_i^{2182}$ image dataset on a Nvidia GTX 1080Ti GPU. We used the Keras library over Tensorflow with some custom modules.

### C. Evaluation

*1) Experimental Setup:* In this work we assess the relations between the underlying structure of the dataset, neuropsychological test scores, biological markers and the current diagnosis categories. To do so, we define two experiments:

- **Experiment 1: Classification** into traditional AD clusters under two scenarios: **AD vs CTL**, which includes all images labelled as CTL and AD; and **MCIC vs MCIS**, which uses the latest available image of each MCIS subject, as well as the latest image of each MCIC subject prior to its conversion to AD. This last scenario implies a greater difficulty of classification, given that the last acquisition of the MCIS is more affected by ageing. Different supervised learning algorithms were used:
  - A linear Support Vector Machine Classifier (SVC), as implemented in `LibLinear` [46]. SVC is a non-probabilistic binary classifier used in numerous recent works [13], [17], [19], [22], [24], [25], [47]. In this work we used it twice. First, as a baseline following the **Voxels-As-Features (VAF)** approach [48], where all voxel intensities in the image (more than 250,000) are used as features. Then, trained and tested on $\mathbf{z}$, in what we call **CAE-SVC**.
  - A neural classifier based on a MLP [38], with 2 hidden layers of 64 neurons and ReLU activations, trained and tested on $\mathbf{z}$ (**CAE-MLP**). The output layer contains 2 neurons with softmax activation. Dropout with $p = 0.5$ was added to prevent overfitting, and Batch Normalization for convergence.
- **Experiment 2: Prediction** of neuropsychological tests and other clinical variables using neural regression: a MLP with two hidden layers of 64 neurons. Again, we used ReLU activation, batch normalization and dropout with $p = 0.5$, and one neuron (linear activation) at the output layer.

Additionally, we provide visual aid for interpreting the manifold by plotting the subjects in the CAE space. The influence of each neuron is estimated via a linear model (see Sec. II-C3), and those most correlated to the clinical variables are studied in more depth.

*2) Performance Estimation:* For experiment 1, 10-fold stratified cross-validation was used to estimate the generalization ability of our methodology [49]. Mean and SD of the correct rate, sensitivity and specificity, and the total balanced accuracy and F1-score are presented. The actual error upper bound is estimated via the procedure found in [50] that, for any model $f(x)$, the difference between the actual error and the re resubstitution error is bounded by the actual risk $|P_{act}(f(x)) - P_{emp}(f(x))| \leq \gamma_{emp}$ for any significance $\eta < 0.05$. Therefore, the actual risk can be computed as:

$$\gamma_{emp} \leq \sqrt{\frac{1}{2l} \ln\left(\frac{N}{\eta}\right)} \tag{4}$$

where $l$ is the sample size and $N$ is defined [50] as:

$$N(l, Z) = 2 \sum_{k=0}^{Z-1} \binom{l-1}{k} \tag{5}$$

In experiment 2, the error between the original target variables $\mathbf{y}_i$ (neuropsychological scores or other clinical markers) and the model outputs $\hat{\mathbf{y}}_i$ was measured via the MSE $\left(\frac{1}{N}\sum_i(\hat{\mathbf{y}}_i - \mathbf{y}_i)^2\right)$. The Pearson's Correlation Coefficient (PCC) and the Coefficient of Determination (R2) were also used to estimate correlation and regression performance. The PCC measures the linear correlation between $\mathbf{y}$ and $\hat{\mathbf{y}}_i$:

$$PCC = \frac{\sum(\mathbf{y}_i - \mu_y)(\hat{\mathbf{y}}_i - \mu_{\hat{y}})}{\sqrt{\sum(\mathbf{y}_i - \mu_y)^2(\hat{\mathbf{y}}_i - \mu_{\hat{y}})^2}} \tag{6}$$

where $\mu_y$ and $\mu_{\hat{y}}$ are the means of all $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i$ respectively.

The R2 explains the proportion of variance in the target variables that is explained by the model outcomes $\hat{\mathbf{y}}_i$:

$$R2 = 1 - \frac{\sum_i(\mathbf{y}_i - \hat{\mathbf{y}}_i)^2}{\sum_i(\mathbf{y}_i - \mu_y)^2} \tag{7}$$

*3) Visualization of Area of Influences (AoIs):* We have developed a technique to visualize the AoIs of each neuron in the image space using a linear decomposition model similar to Principal Component Analysis (PCA). Since the MRI images are normalized in the image space, this assumption is not farfetched, and the AoIs can be approximated as a projection from a flattened–vectorized–image space $\bar{\mathbf{I}} = \{\bar{\mathbf{I}}_i \in \mathbb{R}^{262144}\}$ to the CAE space $\mathbf{Z} = \{\mathbf{z}_i \in \mathbb{R}^Z\}$, so that $\mathbf{Z} = \mathbf{W}\bar{\mathbf{I}}$, where $\mathbf{W}$ is the projection matrix. Assuming this model, $\mathbf{W}$ can be computed using the Moore Penrose inverse, or pseudo-inverse, of the data matrix $\bar{\mathbf{I}}^+$:

$$\mathbf{W} = \mathbf{Z}\bar{\mathbf{I}}^+ \tag{8}$$

In this model, the $i^{th}$ column of $\mathbf{W}$ contains the coefficients that transform an image to the $i^{th}$ coordinate of the Z-space. By displaying this unitless vector over a brain template, we can visualize the AoIs of each neuron–the basis of each coordinate of the Z-layer manifold–, and identify the brain regions under its influence.

## III. Results

### A. Experiment 1: Classification

In this experiment we evaluate the ability of the features extracted by the autoencoder at the z-layer when used to

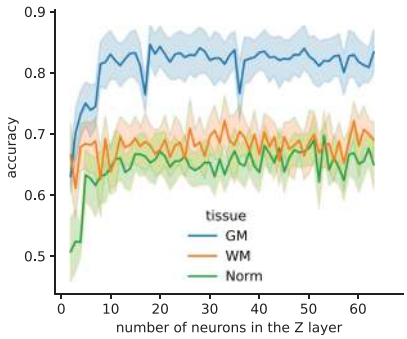



Fig. 2. Evolution of the accuracy (and its SD over the cross-validation folds) of the AD vs CTL setting when varying the number of neurons in the Z layer.

distinguish between classes in the dataset. First of all, in Figure 2 we explore how the performance of a binary classification (AD vs CTL) varies when changing the number of neurons $Z$ at the Z-layer for each tissue map. This is shown for the SVC classifier, although the for the MLP are similar.

We observe that, independently of the tissue map, the discrimination capacity of the autoencoder increases first with $Z$ and then it stabilizes after a certain $Z$. This occurs approximately at $Z = 20$ for GM, $Z = 5$ for WM and $Z = 18$ for Norm images. For comparing purposes, we will use $Z = 20$ throughout the rest of this section, in order to compare the performance of the resulting manifold.

The performance of the different scenarios for $Z = 20$ is shown at Table II. From this table, we observe that under all scenarios, GM and WM maps achieve better performance in SVC than the MLP. Best values in the AD vs CTL scenario are obtained, as expected, with the GM tissue maps, closely followed by WM. We can then confirm that the subjects in the Z space defined by the autoencoder form fairly separate clusters, obtaining precisions higher than 84%. The resubstitution error for GM is 0.139 which sets, according to [50], the upper bound for the actual error to 0.333, proving that the new space is representative of an underlying structure of the disease. Surprisingly, Norm images achieve very low correct rates with both the classifiers compared to VAF performance, which may be due to a larger heterogeneity in the source images as they include other tissues. The superior bounds for the WM and Norm actual error are 0.487 and 0.490 respectively.

### *B. Experiment 2: Regression Analysis*

In this experiment, we observe how the Z-features correlate with other data sources, such as clinical data. As in the previous experiment, Figure 3 shows how the performance (PCC and SD) varies with $Z$ for each of the tissue maps and the different clinical variables and cognitive scores used.

From Figure 3, we arrive at a similar conclusion: the PCC with these variables increases with n_comp until approximately $Z = 20$. Since performance at this point is reasonably optimal (despite local maximums in the surrounding area), we will also use this value to maintain consistency with Experiment 1. Regression performance for the variables ADAS-11, ADAS-13, ADAS-Q4, age, ApoE4, CDRSB, MMSE and TAU are provided at Table III.

We observe that the best-scoring tissue map is by far GM. It correlates better than any of the other maps with all clinical variables. The highest correlation (0.638 and R2 = 0.381) is with the ADAS-11, closely followed by other ADAS-cog measures, CDRSB and MMSE; in short, all variables related to the progression of AD and cognitive state except for ApoE4, which is hardly correlated to any map type. This proves a link between GM structure and cognitive state consistent with the literature. Highest correlation in WM is found with Age (0.396 and R2 = 0.141), followed by ADAS-cog, CDRSB and even MMSE. The case of Norm images is more particular, since the highest correlation is found again with Age, but very small correlations are found with the remaining variables.

## IV. Discussion

The CAE architecture used in this article has many advantages over other feature extraction methodologies. It is a complete data-driven approach; self-supervised so that the final result is guided only by the particularities of the data set, without prior assumptions, and very little pre-processing is required. It integrates a self-supervised recognition system (the convolutional layers) and an underlying logic (the dense layers) that, when combined, can model the manifold structure in a dataset. The non-linear nature of the neural networks [51] makes it possible to model more complex relationships without prior assumptions as in the case of kernel decomposition. In addition, by not requiring vectorization of each image, and thus maintaining the proximity restrictions of a matrix of three-dimensional images, it is more suitable for processing brain images, in contrast to other widely used decomposition methods such as PCA [29], [52].

The CAE trained with each tissue dataset (GM, WM and Norm) defines a new space to which the brain maps can be projected from their original image space. This allows us to explore the non-linear manifold structure contained within, and relate them to other variables used in clinical practice. We are going to focus on the combinations that allow for a better characterization of AD, as measured by the dementia scales, as well as other biological markers such as TAU.

First, in Figure 4, we display the distribution of the output of each neuron at the Z layer using a letter-value plot [53], a more detailed boxplot for large datasets. By analysing the most discriminant dataset, the one containing the tissue GM, it is possible to see differences between the images belonging to the traditional classes AD in most neurons. The greatest differences are located in neuron 14, followed by neurons 2, 3 and 10, which also show large differences in the mean. Therefore, we can assume that these neurons contain relevant information about the structure of the disease.

The high correlation of the Z-layer features with some clinical variables can be illustrated by plotting the values of the most discriminating neurons in GM, namely 14 and 3. Figure 5 shows how the different images are projected in the two-dimensional space of the two neurons, and how they relate to the ADAS-13

TABLE II
PERFORMANCE OF THE MODELS WITH DIFFERENT SCENARIO AND TISSUE MAPS (Z = 20). BEST SCORING COMBINATIONS ARE HIGHLIGHTED

| Scenario | Tissue | Model | Correct Rate | Sensitivity | Specificity | Bal. Accuracy | F1 (total) |
|---|---|---|---|---|---|---|---|
| AD vs CTL | GM | VAF | 0.837 [0.04] | 0.849 [0.08] | 0.837 [0.06] | 0.843 | 0.844 |
| | | **CAE-SVC** | **0.843 [0.06]** | **0.841 [0.07]** | **0.849 [0.07]** | **0.844** | **0.846** |
| | | CAE-MLP | 0.779 [0.07] | 0.751 [0.01] | 0.817 [0.08] | 0.777 | 0.784 |
| | WM | VAF | 0.772 [0.05] | 0.759 [0.13] | 0.783 [0.09] | 0.771 | 0.768 |
| | | **CAE-SVC** | **0.785 [0.05]** | **0.797 [0.09]** | **0.783 [0.07]** | **0.791** | **0.790** |
| | | CAE-MLP | 0.761 [0.06] | 0.753 [0.11] | 0.780 [0.08] | 0.763 | 0.766 |
| | Norm | **VAF** | **0.768 [0.05]** | **0.752 [0.09]** | **0.777 [0.07]** | **0.764** | **0.761** |
| | | CAE-SVC | 0.673 [0.04] | 0.683 [0.18] | 0.716 [0.06] | 0.669 | 0.680 |
| | | CAE-MLP | 0.629 [0.04] | 0.627 [0.09] | 0.668 [0.08] | 0.640 | 0.648 |
| MCI-S vs MCI-C | GM | VAF | 0.673 [0.10] | 0.582 [0.16] | 0.706 [0.14] | 0.644 | 0.621 |
| | | **CAE-SVC** | **0.696 [0.08]** | **0.628 [0.11]** | **0.744 [0.11]** | **0.666** | **0.686** |
| | | CAE-MLP | 0.640 [0.09] | 0.553 [0.11] | 0.713 [0.13] | 0.601 | 0.633 |
| | WM | VAF | 0.645 [0.09] | 0.567 [0.19] | 0.677 [0.15] | 0.622 | 0.599 |
| | | **CAE-SVC** | **0.715 [0.09]** | **0.693 [0.18]** | **0.726 [0.13]** | **0.704** | **0.709** |
| | | CAE-MLP | 0.710 [0.12] | 0.684 [0.20] | 0.755 [0.16] | 0.709 | 0.719 |
| | Norm | **VAF** | **0.682 [0.10]** | **0.660 [0.20]** | **0.694 [0.15]** | **0.677** | **0.671** |
| | | CAE-SVC | 0.561 [0.10] | 0.645 [0.06] | 0.532 [0.12] | 0.513 | 0.562 |
| | | CAE-MLP | 0.598 [0.08] | 0.505 [0.21] | 0.634 [0.16] | 0.539 | 0.569 |

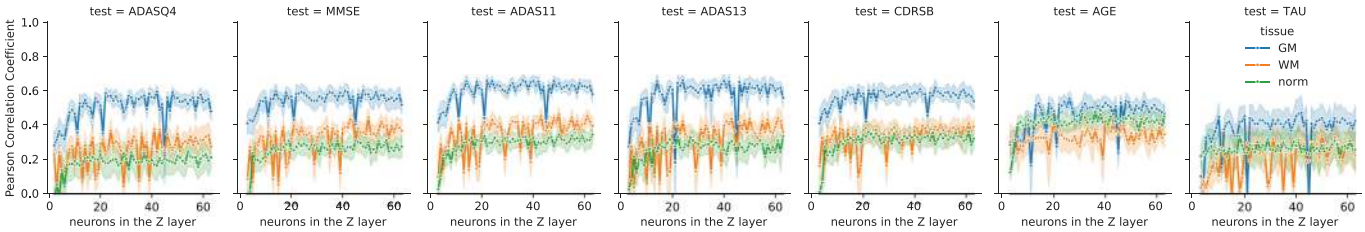



Fig. 3. Evolution of the PCC when varying the number of neurons in the Z layer.

TABLE III
REGRESSION RESULTS FOR DIFFERENT CLINICAL VARIABLES AND Z = 20. HIGHEST CORRELATIONS PER VARIABLE AND MAP ARE HIGHLIGHTED

| | | ADAS-11 | ADAS-13 | ADAS-Q4 | Age | ApoE-4 | CDRSB | MMSE | TAU |
|---|---|---|---|---|---|---|---|---|---|
| GM | MSE | 0.034 [0.01] | 0.036 [0.01] | 0.066 [0.01] | 0.004 [0.00] | 0.115 [0.01] | 0.045 [0.01] | 0.012 [0.00] | 0.029 [0.00] |
| | PCC | **0.638 [0.05]** | **0.624 [0.06]** | **0.571 [0.07]** | **0.547 [0.07]** | **0.171 [0.09]** | **0.591 [0.08]** | **0.585 [0.05]** | **0.449 [0.12]** |
| | R2 | 0.381 [0.06] | 0.361 [0.09] | 0.314 [0.07] | 0.273 [0.06] | 0.005 [0.06] | 0.337 [0.09] | 0.322 [0.05] | 0.175 [0.10] |
| WM | MSE | 0.048 [0.01] | 0.049 [0.00] | 0.090 [0.01] | 0.005 [0.00] | 0.120 [0.01] | 0.060 [0.01] | 0.016 [0.00] | 0.032 [0.01] |
| | PCC | 0.368 [0.07] | 0.391 [0.10] | 0.272 [0.14] | **0.396 [0.11]** | 0.039 [0.14] | 0.348 [0.07] | 0.333 [0.07] | 0.284 [0.13] |
| | R2 | 0.125 [0.04] | 0.141 [0.08] | 0.067 [0.10] | 0.141 [0.06] | -0.035 [0.06] | 0.114 [0.05] | 0.103 [0.05] | 0.074 [0.07] |
| Norm | MSE | 0.053 [0.01] | 0.054 [0.00] | 0.098 [0.01] | 0.005 [0.00] | 0.116 [0.01] | 0.067 [0.01] | 0.017 [0.00] | 0.034 [0.01] |
| | PCC | 0.193 [0.09] | 0.200 [0.06] | 0.134 [0.20] | **0.367 [0.10]** | 0.069 [0.15] | 0.202 [0.14] | 0.215 [0.16] | 0.183 [0.16] |
| | R2 | 0.028 [0.04] | 0.035 [0.03] | -0.013 [0.11] | 0.113 [0.05] | -0.005 [0.05] | 0.018 [0.10] | 0.024 [0.08] | 0.016 [0.09] |

score, colour-encoded. The image depicts larger ADAS scores at the top left corner, and lower values in the bottom right corner, which proves that there exist a direction in the data-driven Z-space related to the disease progression.

To explore the regions that better classify AD, we estimate the AoI of each neuron. Since it is the most correlated with disease progression, in Figure 6 we show the AoIs of the 14th coordinate of the $Z$-layer (neuron 14) trained with GM maps. That figure also depicts the PCC values obtained by that feature and different statistical tests.

Regions traditionally linked to the progression of AD are highlighted, such as parts of the temporal and parietal lobes [54], [55], as well as a shrinkage of the cerebellum, usually associated with age. This emphasizes the utility of the AoI in linking the underlying structural causes of neurodegeneration to the magnitude of a coordinate in the Z-manifold through other types of data, such as neuropsychological evaluations or other biological markers.

In Figure 7 we can check that other neurons, such as 7 and 17, are more correlated respectively with age (in which the effect is

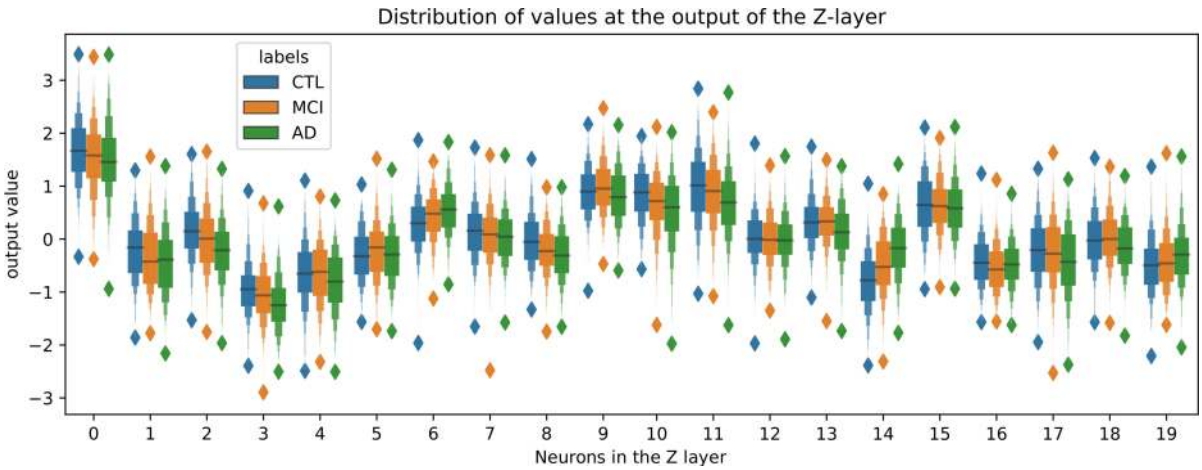



Fig. 4. Distribution of output values per class and neuron of the Z layer in the GM dataset.

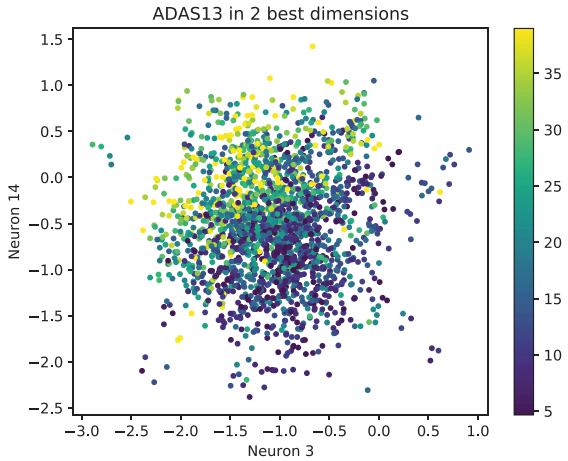



Fig. 5. Manifold distribution of activations of the Z-layer in relation to the ADAS-13 score of each subject and visit.

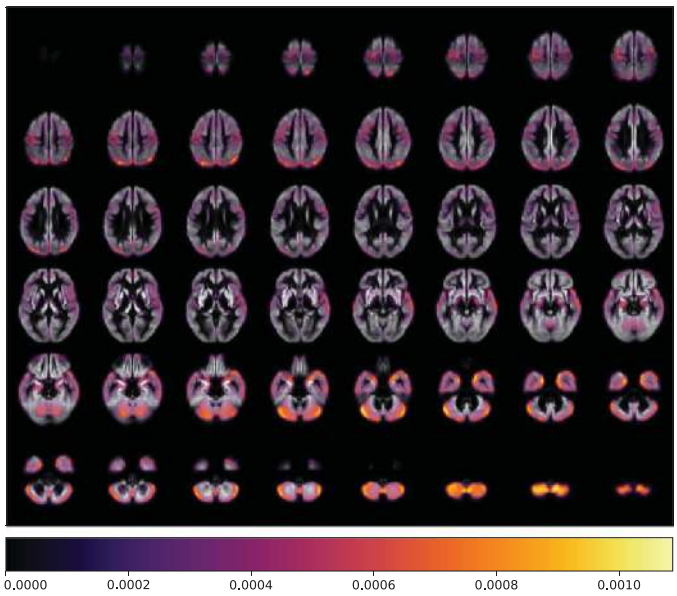



Fig. 6. AoI of neuron number 14 (in GM). PCCs (and uncorrected p-values) for the coordinates in this neuron; ADAS-13: 0.40 ($5 \cdot 10^{-80}$), ADAS-11: 0.39 ($3 \cdot 10^{-76}$), ADAS-Q4: 0.36 ($9 \cdot 10^{-62}$), MMSE: -0.40 ($4 \cdot 10^{-77}$), ApoE4: 0.18 ($7 \cdot 10^{-14}$), TAU: 0.06 (1.00), CDRSB: 0.38 ($3 \cdot 10^{-72}$), Age: 0.08 (0.20).

focused on the frontal and occipital lobes and the cerebellum) or even tau protein (also concentrated at the cerebellum, but also more spread over the parietal lobe, areas of the frontal lobe, the caudate nucleus and the thalamus), which has also been verified in the literature [54].

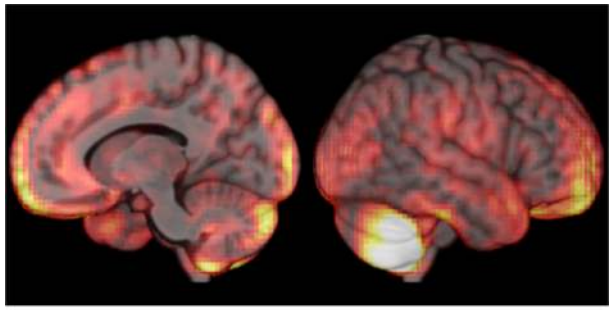
Therefore, if the different types of data are effectively related, shedding new light on the relationships between clinical and neuropsychological variables and self-driven decomposition of the CAE, it is possible to assume that this decomposition can also be used in classification, as demonstrated in Experiment 1. To contextualize the results of the classification we have used three recent works in the literature. Firstly, we used the methodology based on Spherical Brain Mapping (SBM) [13], a technique that projects a two-dimensional map from the three-dimensional MRI volumes, representing radial statistical texture values. In this approach, both classification and regression were used. Secondly, the approach proposed in [35], where a Stacked Autoencoder (SAE) was applied to regional GM tissue Low Level Features (LLFs), and the resulting features were used in classification. Finally, to the model in [26], where the authors proposed a PCA decomposition of GM MRI images, and used the extracted features to create a regression model that predicts MMSE yearly changes. These results are found at Table IV.

Notable differences between class were found when using the Z-layer features, leading to larger classification performance in the AD vs CTL scenario, especially when using GM tissue, but also with the segmented WM maps. We observe that our methodology improves the VAF baseline performance as well as the one obtained by the SAE features, which confirms that this is mainly due to the distribution of the self-extracted features of the CAE at the Z-layer, which is more optimal than the SAE. The actual upper error bounds for these tissues also

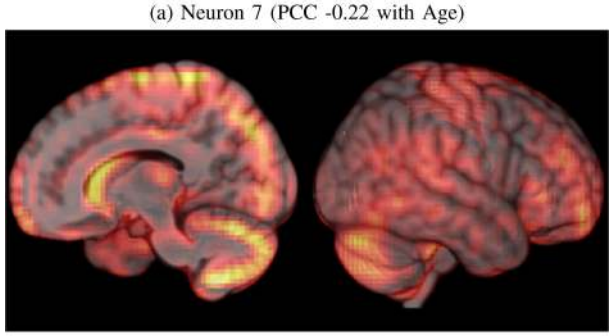
TABLE IV
COMPARISON BETWEEN OUR PROPOSED CAE-BASED CLASSIFICATION AND REGRESSION PERFORMANCE (ACCURACY AND PCC RESPECTIVELY) AND OTHER METHODOLOGIES FOR CLASSIFICATION [13], [35] AND REGRESSION [13], [26] IN THE LITERATURE. ALGORITHMS MARKED WITH A STAR SYMBOL USED A DIFFERENT DATASET. IF NOT STATED, GM FEATURES ARE USED

| Classification | VAF | CAE-SVC | CAE-MLP | SAE* [35] | LLF* [35] | LLF+SAE* [35] | SBM [13] |
|---|---|---|---|---|---|---|---|
| AD vs CTL | 0.837 [0.04] | 0.843 [0.06] | 0.779 [0.07] | 0.821 [0.03] | 0.890 [0.02] | 0.882 [0.02] | 0.923 [0.036] |
| MCI-C vs MCI-S | 0.673 [0.10] | 0.696 [0.08] | 0.640 [0.09] | 0.674 [0.02] | 0.736 [0.01] | 0.802 [0.02] | 0.719 [0.140] |

| Regression | CAE-GM | CAE-WM | SBM-GM [13] | SBM-WM [13] | PCA* [26] | PCA+MMSE* [26] |
|---|---|---|---|---|---|---|
| ADAS-11 | 0.638 [0.05] | 0.368 [0.07] | 0.644 | 0.526 | – | – |
| ADAS-13 | 0.624 [0.06] | 0.391 [0.10] | 0.709 | 0.604 | – | – |
| MMSE | 0.585 [0.05] | 0.333 [0.07] | 0.534 | 0.524 | 0.668 | 0.696 |



(a) Neuron 7 (PCC -0.22 with Age)



(b) Neuron 17 (PCC -0.39 with Tau)

Fig. 7. Three-dimensional representation of the overlapping of the AoIs of a couple of neurons that are more correlated with other clinical variables, such as Age (neuron 7) or Tau protein (neuron 17). PCC of these features and Age and Tau are respectively shown below each figure.

corroborate this. A similar behaviour is repeated with the MLP, which confirms that this is mainly due to the distribution of the self-extracting features of the CAE in the Z layer, and hardly depends on the classifier used. However, [35] reports higher accuracy for the LLF+SAE features. Here, it is important to note that while [13] uses a similarly big dataset, [35] is using a much smaller ADNI subset (202 subjects, approx. 50 per class). And furthermore, the images in this work were downsampled (see Section II-A2) by a factor of 3, which could lead to some information loss. However, our results are still good and improve some of the approaches in the bibliography, which means that the CAE features are significant for the differential diagnosis of AD.

A relevant contribution of this work is regression analysis (Experiment 2). Here we study the correlation between pure imaging characteristics and other clinical variables. The procedure has rarely been performed [13], [26], although it is crucial for validating diagnoses and developing a deeper understanding of the neurodegeneration process. In this regard, the largest PCC achieved by GM maps (0.638 and R2 = 0.381) with the ADAS-11 measure, similar to correlations with SBM features, hints at the probable link between the extracted CAE features and the cognitive state. PCC of our method with MMSE is also large, although smaller than the PCA-based [26]. However, these results must be taken with caution, since, as the authors themselves acknowledge, no separate test set or cross-validation has been used, and therefore, the performance might be overestimated. The differences between GM and WM (also with Norm) are larger in the case of regression. The is coherent with the literature, since AD is known to affect first GM and later progressing to WM [56].

There exist a big difference of performance when using the Norm maps. A more complex imaging, involving different information encoded in grey levels can have an impact over the performance, but this would also impact the baseline VAF. The use of integral normalization and the inclusion of the skull and other tissues can be a potential confounding effect. The latter is confirmed when looking at Figure 8, where we can observe that most of the AoIs are localized at the skull and Cerebro-Spinal Fluid (CSF). So, it is possible that the heterogeneity found in skull shape and size can be the main confusion factor for the CAE decomposition when applying the CAE to normalized images. Increasing $Z$ did not help to increase the performance either, as it can be seen at Figure 3. Therefore it is possible that other approaches, such as using skull-stripping, another intensity normalization procedure or changing the architecture (for example, to use denoising or sparse AEs) could improve the performance in the Norm images.

Overall, the demonstrated links between the data-driven decomposition performed by the CAE and different clinical and neuropsychological variables provides us with a new tool to study neurodegeneration in AD. The Z-manifold was useful in predicting classes and clinical variables, at the same time that the patterns and regions that influence the features are consistent with the well-known process of brain atrophy. The fact that features that are automatically extracted, without prior knowledge of the disease, correlate well with clinical examinations could be used to provide greater assistance in the diagnosis of dementia, as well as a deeper understanding of the structural processes of neurodegeneration and how they relate to cognitive impairment, paving the way for new sets of imaging biomarkers useful in clinical practice.

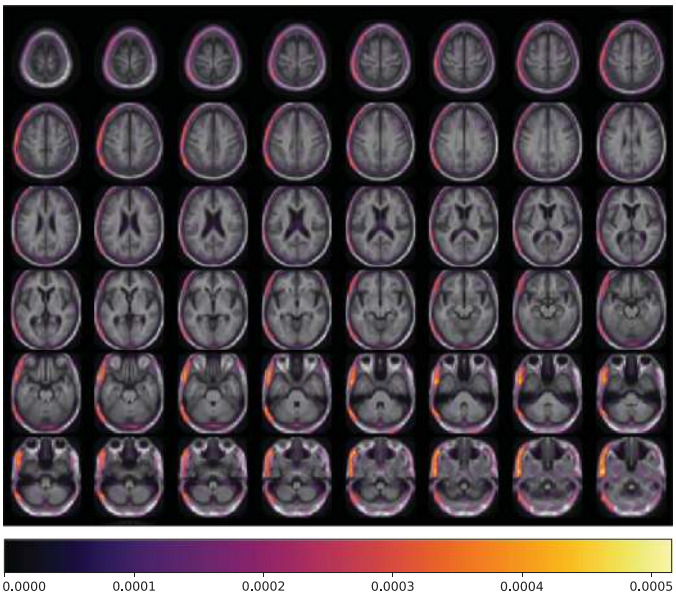



Fig. 8. AoI of neuron 0 in Norm maps. PCCs (and uncorrected p-values) for ADAS-13: 0.10 ($2 \cdot 10^{-6}$), ADAS-11: 0.11 ($1 \cdot 10^{-7}$), ADAS-Q4: 0.07 ($2 \cdot 10^{-3}$), MMSE: -0.11 ($4 \cdot 3^{-7}$), ApoE4: 0.01 (0.6), TAU: 0.07 (0.1), CDRSB: 0.12 ($1 \cdot 10^{-8}$), Age: -0.06 ($4 \cdot 10^{-3}$). Note that the AoIs are mainly placed outside the brain and mostly over the skull.

## V. Conclusion

Alzheimer's disease is the most common neurodegenerative disease in the world. While its underlying processes are beginning to be understood, new methodologies are needed to provide a new perspective on the disease and its diagnosis. The use of image biomarkers is key in current clinical practice, and numerous automatic diagnostic systems exist. The deep learning revolution has provided us with new tools that can automatically extract image characteristics without prior assumptions about the underlying process. In this paper we proposed a deep convolutional autoencoder (CAE) architecture, a tool that can perform an automatic non-linear decomposition of a very large dataset (over 2000 images). The data-driven features extracted with this technique have been largely related to other clinical and neuropsychological variables, showing large correlations (greater than 0.63) with clinical variables such as age, tau protein deposits and especially neuropsychological examinations. It has also been useful in the differential diagnosis of Alzheimer's disease with an accuracy of over 84%. The visualization of the areas of influence of each neuron, together with the correlation of these scores with clinical variables, provide information on the regions most affected by disease progression and how they are related to cognitive decline. Therefore, our CAE system can be used effectively to aid in the diagnosis of dementia, as well as to provide a new understanding of the relationships between structural damage and cognitive capacity, as measured by these neuropsychological tests, paving the way for new sets of imaging biomarkers useful in clinical practice.

## Acknowledgment

The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the USC.

## References

[1] H. Niu, I. Álvarez Álvarez, F. Guillén-Grima, and I. Aguinaga-Ontoso, "Prevalencia e incidencia de la enfermedad de Alzheimer en europa: metaanálisis," *Neurologia*, vol. 32, no. 8, 2017, Art. no. 523532.

[2] S. Z. Klekociuk, J. J. Summers, J. C. Vickers, and M. J. Summers, "Reducing false positive diagnoses in mild cognitive impairment: The importance of comprehensive neuropsychological assessment," *Eur. J. Neurol.*, vol. 21, no. 10, Jun. 2014, Art. no. 1330e83.

[3] G. H. Weissberger, J. V. Strong, K. B. Stefanidis, M. J. Summers, M. W. Bondi, and N. H. Stricker, "Diagnostic accuracy of memory measures in Alzheimer's dementia and mild cognitive impairment: A systematic review and meta-analysis," *Neuropsychol. Rev.*, vol. 27, pp. 354–388, Sep. 2017.

[4] G. B. Frisoni *et al.*, "Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers," *Lancet Neurol.*, vol. 16, no. 8, 2017, Art. no. 661676.

[5] B. Dubois *et al.*, "Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria," *Lancet Neurol.*, vol. 13, no. 6, pp. 614–629, 2014.

[6] M. W. Weiner *et al.*, "The Alzheimer's disease neuroimaging Initiative: A review of papers published since its inception," *Alzheimer's Dementia, J. Alzheimer's Assoc.*, vol. 8, no. 1, Feb. 2012, Art. no. S168. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22047634

[7] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies: Longitudinal MRI data in nondemented and demented older adults," *J. Cogn. Neurosci.*, vol. 22, no. 12, pp. 2677–2684, 2010.

[8] J. C. Morris *et al.*, "Developing an international network for Alzheimer research: The dominantly inherited Alzheimer network," *Clin. Investigation*, vol. 2, no. 10, pp. 975–984, 2012.

[9] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *Proc. Nat. Acad. Sci.*, vol. 97, no. 20, p. 11050–11055, Sep. 2000.

[10] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. New York, NY, USA: Academic, 2007.

[11] K. Adamczuk *et al.*, "Diagnostic value of cerebrospinal fluid a$\beta$ ratios in preclinical Alzheimer's disease," *Alzheimers Res Therapy*, vol. 7, no. 1, 2015, Art. no. 75.

[12] S. Adaszewski, J. Dukart, F. Kherif, R. Frackowiak, B. Draganski, and A. D. N. Initiative, "How early can we predict Alzheimer's disease using computational anatomy?" *Neurobiol. Aging*, vol. 34, no. 12, pp. 2815–2826, Dec. 2013.

[13] F. J. Martinez-Murcia *et al.*, "Assessing mild cognitive impairment progression using a spherical brain mapping of magnetic resonance imaging," *J. Alzheimer's Disease*, vol. 65, pp. 713–729, 2018.

[14] A. Ortiz *et al.*, "Automatic ROI selection in structural brain MRI using SOM 3D projection," *PLoS One*, vol. 9, no. 4, 2014, Art. no. e93851.

[15] I. A. Illan, J. M. Górriz, J. Ramírez, and A. Meyer-Base, "Spatial component analysis of MRI data for Alzheimer's disease diagnosis: A Bayesian network approach," *Frontiers Comput. Neurosci.*, vol. 8, 2014, Art. no. 156.

[16] P. Markiewicz, J. Matthews, J. Declerck, and K. Herholz, "Robustness of multivariate image analysis assessed by resampling techniques and applied to FDG-PET scans of patients with Alzheimer's disease," *NeuroImage*, vol. 46, pp. 472–485, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/B6WNP-4VFK7X3-3/2/e7833cb1d62f98e28326352e45981d00

[17] F. Polat *et al.*, "Computer based classification of MR scans in first time applicant Alzheimer patients," *Current Alzheimer Res.*, vol. 9, no. 7, pp. 789–794, Sep. 2012.

[18] J. M. Gorriz *et al.*, "A semi-supervised learning approach for model selection based on class-hypothesis testing," *Expert Syst. Appl.*, vol. 90, pp. 40–49, 2017.

[19] G. Fung and J. Stoeckel, "SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information," *Knowl. Inf. Syst.*, vol. 11, no. 2, pp. 243–258, 2007.

[20] F. Martínez-Murcia, J. Górriz, J. Ramírez, C. Puntonet, and D. Salas-González, "Computer aided diagnosis tool for Alzheimer's disease based on mann-whitney-wilcoxon U-test," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9676–9685, Aug. 2012.

[21] L. Khedher, J. Ramírez, J. Górriz, A. Brahim, and F. Segovia, "Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," *Neurocomputing*, vol. 151, Mar. 2015, Art. no. 139150.

[22] C. Montagne, A. Kodewitz, V. Vigneron, V. Giraud, and S. Lelandais, "3D local binary pattern for PET image classification by SVM, application to early Alzheimer disease diagnosis," in *Proc. 6th Int. Conf. Bio-Inspired Syst. Signal Process.,* Barcelona, Spain, Feb. 2013, Art. no. 145150.

[23] J. Ramírez *et al.*, "Ensemble of random forests one vs. rest classifiers for MCI and ad prediction using anova cortical and subcortical feature selection and partial least squares," *J. Neurosci. Methods*, vol. 302, pp. 47–57, 2018.

[24] C. Ecker *et al.*, "Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach," *NeuroImage*, vol. 49, no. 1, pp. 44–56, Jan. 2010.

[25] P. Vemuri *et al.*, "Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies," *NeuroImage*, vol. 39, no. 3, pp. 1186–1197, Feb. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/B6WNP-4PYGVWG-3/2/7b56e2e69c1947e636d2b2ab2d8317e0

[26] S. Duchesne, A. Caroli, C. Geroldi, G. B. Frisoni, and D. L. Collins, "Predicting clinical variable from MRI features: Application to MMSE in MCI," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer, 2005, pp. 392–399.

[27] A. Ortiz, J. Munilla, J. M. Gorriz, and J. Ramirez, "Ensembles of deep learning architectures for the early diagnosis of the Alzheimers disease," *Int. J. Neural Syst.*, vol. 26, no. 7, 2016, Art. no. 1650025.

[28] A. Payan and G. Montana, "Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks," 2015, *arXiv:1502.02506*.

[29] F. J. Martinez-Murcia, J. M. Górriz, J. Ramirez, and A. Ortiz, "Convolutional neural networks for neuroimaging in Parkinson's disease: Is preprocessing needed?" *Int. J. Neural Syst.*, vol. 2018, Art. no. 1850035.

[30] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[31] R. D. Hjelm, V. D. Calhoun, R. Salakhutdinov, E. A. Allen, T. Adali, and S. M. Plis, "Restricted boltzmann machines for neuroimaging: an application in identifying intrinsic networks," *NeuroImage*, vol. 96, pp. 245–260, 2014.

[32] S. Vieira, W. H. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neurosci. Biobehav. Rev.*, vol. 74, pp. 58–75, 2017.

[33] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proc. ICML Workshop Unsupervised Transfer Learn.*, 2012, pp. 37–50.

[34] J. Xu *et al.*, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.

[35] H.-I. Suk *et al.*, "Latent feature representation with stacked auto-encoder for ad/mci diagnosis," *Brain Struct. Function*, vol. 220, no. 2, pp. 841–859, 2015.

[36] S. J. Cano *et al.*, "The ADAS-cog in Alzheimer's disease clinical trials: Psychometric evaluation of the sum and its parts," *J. Neurol., Neurosurgery Psychiatry*, vol. 81, no. 12, pp. 1363–1368, Sep. 2010.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[38] D. W. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley, and B. W. Suter, "The multilayer perceptron as an approximation to a Bayes optimal discriminant function," *IEEE Trans. Neural Netw.*, vol. 1, no. 4, pp. 296–298, Dec. 1990.

[39] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learning*, Lille, France, vol. 37, Jul. 2015, pp. 448–456.

[40] A. Ortiz *et al.*, "Learning longitudinal MRI patterns by sice and deep learning: Assessing the Alzheimers disease progression," in *Proc. Ann. Conf. Med. Image Understanding Anal.*, 2017, pp. 413–424.

[41] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3857–3867.

[42] M. Lin, Q. Chen, and S. Yan, "Network in network," presented at the Int. Conf. Learn. Representations, 2013.

[43] S. Bell, C. Lawrence Zitnick, K. Bala, and R. Girshick, "Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2874–2883.

[44] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921–2929.

[45] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.

[46] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, Apr. 2011, Art. no. 27.

[47] J. M. Gorriz *et al.*, "Case-based statistical learning: A non parametric implementation with a conditional-error rate svm," *IEEE Access*, vol. 5, pp. 11468–11478, 2017.

[48] J. Stoeckel, N. Ayache, G. Malandain, P. M. Koulibaly, K. P. Ebmeier, and J. Darcourt, "Automatic classification of SPECT images of Alzheimer's disease patients and control subjects," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2004, vol. 3217, pp. 654–662.

[49] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *J. Artif. Intell.*, vol. 97, pp. 273–324, 1997.

[50] J. M. Górriz *et al.*, "A machine learning approach to reveal the NeuroPhenotypes of autisms," *Int. J. Neural Syst.*, vol. 29, Jan. 2019, Art. no. 1850058.

[51] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.

[52] F. J. Martinez-Murcia *et al.*, "Functional brain imaging synthesis based on image decomposition and kernel modeling: Application to neurodegenerative diseases," *Frontiers Neuroinform.*, vol. 11, no. 65, Nov. 2017.

[53] H. Hofmann, K. Kafadar, and H. Wickham, "Letter-value plots: Boxplots for large data," *J. Comput. Graphical Statist.*, vol. 26, no. 3, pp. 469–477, 2017.

[54] J. Zhang, C. Yu, G. Jiang, W. Liu, and L. Tong, "3D texture analysis on MRI images of Alzheimer's disease," *Brain Imag. Behav.*, vol. 6, no. 1, pp. 61–69, 2012.

[55] A. Martinez-Torteya, J. Rodriguez-Rojas, J. M. Celaya-Padilla, J. I. Galván-Tejada, V. Trevio, and J. Tamez-Pea, "Magnetization-prepared rapid acquisition with gradient echo magnetic resonance imaging signal and texture features for the prediction of mild cognitive impairment to Alzheimer's disease progression," *J. Med. Imag.*, vol. 1, no. 3, 2014, Art. no. 0310050.

[56] N. Villain *et al.*, "Sequential relationships between grey matter and white matter atrophy and brain metabolic abnormalities in early Alzheimer's disease," *Brain*, vol. 133, no. 11, p. 3301–3314, 2010.