

Ensembling shallow siamese architectures to assess functional asymmetry in Alzheimer's disease progression

Juan E Arco^{a,b,c} (jearco@ugr.es), Andrés Ortiz^{b,c} (aortiz@ic.uma.es), Diego Castillo-Barnes^{b,c} (diegoc@ugr.es), Juan M. Górriz^{a,c} (gorriz@ugr.es), Javier Ramírez^{a,c} (javierrp@ugr.es), for the Alzheimer's Disease Neuroimaging Initiative

^a Department of Signal Theory, Networking and Communications, University of Granada, 18010 Granada, Spain

^b Department of Communications Engineering, University of Malaga, 29010 Malaga, Spain

^c DaSCI Institute, University of Granada, Spain

Corresponding Author:

Juan E Arco

Department of Signal Theory, Networking and Communications, University of Granada, 18010 Granada, Spain

Email: jearco@ugr.es

Ensembling shallow siamese architectures to assess functional asymmetry in Alzheimer's disease progression

Juan E Arco^{a,b,c,*}, Andrés Ortiz^{b,c}, Diego Castillo-Barnes^{a,c}, Juan M. Górriz^{a,c}, Javier Ramírez^{a,c}

^a*Department of Signal Theory, Networking and Communications, University of Granada, 18010 Granada, Spain*

^b*Department of Communications Engineering, University of Malaga, 29010 Malaga, Spain*

^c*DaSCI Institute, University of Granada, Spain*

Abstract

The development of methods based on artificial intelligence for the classification of medical imaging is widespread. Given the high dimensionality of this type of images, it is imperative to use the information contained in relevant regions for further classification. This information can be derived from the morphology of the region of interest, in terms of measurements such as area, perimeter, etc. However, the performance of the classification system strongly depends on the correct selection of the type of information employed. We propose in this work an alternative for evaluating differences between brain regions that relies on the basis of Siamese neural networks. Initially, brain scans are delimited by an anatomical atlas. Next, each pair of regions of interest is then entered into a Siamese network, which is formed by relating the distance between the two individual outputs and the corresponding label. Features are extracted from the embeddings of the final linear layer. Finally, the classification is performed by combining the characteristics of each pair of regions into an ensemble architecture. Performance was assessed by determining how asymmetry between the right and left hemispheres changes during progressive brain degeneration, from mild cognitive impairment to severe atrophy associated with Alzheimer's disease

*Corresponding author

(AD). Our method discriminates with an accuracy of 98.95% between controls and AD patients, and most important, it predicts the cognitive decline in patients suffering from mild cognitive impairment that will develop AD before it occurs with an accuracy of 78.41%. These results demonstrate the applicability of our proposal in the study of a wide range of pathologies.

Keywords: Deep learning; Siamese network; Alzheimer’s disease; Asymmetry; PET; Computer-aided diagnosis.

1. Introduction

Nowadays, the use of medical imaging is vital for the diagnosis of a wide range of diseases. The boost in the resolution of these images in recent times has revealed information that helps clinicians in this task, although there are situations in which different pathologies present very similar symptoms. In these contexts, the use of methods based on artificial intelligence (AI) can be especially important. When applying to neuroimaging, these models assist in the identification of patterns that are highly relevant for the detection of neurological diseases, as many previous works have successfully shown. For example, they have demonstrated to be effective in the study of Alzheimer’s disease (AD [1, 2, 3, 4, 5]. These works employed information from structural or functional images that allows the distinction between controls and AD patients, as well as evaluating the progression of dementia from its initial stages. Similarly, these approaches have also detected changes associated with Parkinson’s disease [6, 7, 8, 9, 10]. They often use DaTSCAN neuroimaging because of their ability to quantify the spatial distribution of dopaminergic transporters in the brain [11, 12, 13, 14, 15].

Traditionally, classification frameworks focus only on the brain regions that are commonly affected by the disease instead of searching for information patterns throughout the whole brain. This reduces the number of features to be processed, which partially solves the curse of dimensionality present in most statistical classifications [16]. Once the brain region is delimited, the easiest ap-

proach is to use the intensity of the voxels as predictors of the different classes to be modelled, using them as inputs of the classifier. Nevertheless, differences
25 in intensity could not explain the patient’s cognitive state in contexts where these differences are subtle [17]. Another alternative is to define features that characterize the brain region of patients who suffer from the pathology under study. These features can be based on the morphology of the region, so that information is expressed in variables such as the area, roundness, perimeter, etc.
30 Although previous research has validated this approach [18, 19, 20, 21, 22], the main drawback is that features that are relevant in the study of a disease can be uninformative in a different one. This is especially problematic in the diagnosis of rare diseases where the information in the initial stages is limited, and it is not clear which features could be relevant.

35 One possibility is to employ feature selection methods to identify from the total number of variables only those that are different between groups. This means that it is possible to extract a subset of features (based on morphological analysis, as introduced earlier) and then select only those that are statistically significant [23, 24, 25]. Thus, features that do not exceed the statistical thresh-
40 old are not used for classification. This approach would alleviate the curse of dimensionality problem while guaranteeing that variables are differential between the groups evaluated (e.g. Alzheimer’s disease and controls). Previous works have shown a good performance [26, 27, 28], but the main concern is that these tests are not appropriate when differences between regions are multivariate instead of univariate. Thus, it seems of vital importance to find a general
45 framework that does not rely on specific features whose relevance depends on the context, in addition to address the curse of dimensionality problem without using statistical tests whose assumptions are not always met.

In this work, we address the curse of dimensionality by proposing an en-
50 semble framework that iteratively evaluates differences between pair of brain regions, which considerably reduces the amount of information that is simultaneously processed. Differences between regions are not computed by defining a number of variables whose statistical significance is evaluated. On the contrary,

we employ a Siamese neural network whose inputs are pairs of brain regions
 55 to be compared. These regions are obtained from the brain subdivisions pro-
 posed by a neuroanatomical atlas, so that we focus on one region and the inputs
 of the network are the parts of this region contained in the left and the right
 hemisphere. The model is trained by relating the similarity between the latent
 space of each pair of regions (evaluated in terms of the Hinge function) and the
 60 corresponding diagnostic label. The information retrieved from the embeddings
 provided by the final linear layer is then combined for each pair of regions, lead-
 ing to the aforementioned ensemble approach. We evaluated this method using
 Positron Emission Tomography (PET) images for studying the development of
 Alzheimer’s disease. Specifically, we hypothesize that differential degeneration
 65 is of great importance, instead of degeneration itself, as a crucial point for the
 study of the development of Alzheimer’s disease. All brain regions are not
 affected at the same time and with the same severity in all the stages of the dis-
 ease, so evaluating the differential degeneration in the left and right hemispheres
 can provide vital information. Our approach successfully identifies functional
 70 asymmetries between regions of both hemispheres, demonstrating its relevance
 in the study of the progression of this pathology. Figure 1 shows a schematic
 diagram of the framework proposed in this work. A more detailed description
 of each stage will be provided in next sections.

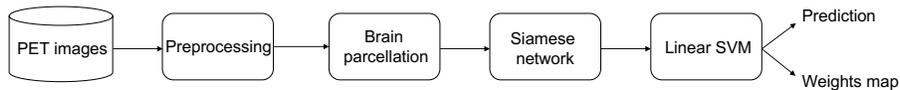


Figure 1: General scheme of the whole idea of the paper.

The organization of the work is explained as follows. Section 2 describes
 75 previous works that used siamese architectures or ensemble frameworks in the
 analysis of medical imaging. In Section 3, the methods developed for this work
 are described, from the architecture of the siamese network proposed to the
 classification stage. Afterwards, we describe in Section 4 the database used
 for evaluating the performance of our method and the experiments conducted.

80 Section 5 summarizes the results obtained, whereas a discussion of the implications of these results is contained in Section 6. Finally, Section 7 includes the conclusions of the study and a description of future lines of research.

2. Related works

Siamese networks have been used in a wide range of contexts such as visual tracking [29, 30, 31], signature recognition [32, 33, 34], anomaly detection [35, 36, 37] and speech signal processing [38, 39, 40]. This technique has shown an excellent performance handling with heterogeneous data [41, 42, 43], including the evaluation of medical images [44, 45, 46]. [47] proposed a solution based on convolutional siamese networks to evaluate the state of the patient at different stages of the disease. Specifically, they used this kind of networks for measuring changes within a longitudinal study, revealing a high effectiveness in classification of retinal scans and the evaluation of knee radiographs. We can also find models focused on the diagnosis of COVID-19. For example, [48] developed a diagnostic method for speeding up the analysis of chest X-ray (CXR) images. To do so, they employed an encoder based on an VGG-16 architecture to extract a latent representation of the input space while avoiding any potential overfitting. After that, the nature of siamese networks led to image classification based on the similarity between the input images, which was characterized by using a contrastive loss function. This model yielded a 95.6% of accuracy when diagnosing COVID-19 from healthy patients, achieving an excellent performance even in scenarios with a reduced number of samples.

Regarding ensemble architectures, they have been successfully used in different contexts. In fact, an optimal combination of information from different sources is vital for obtaining a superior performance than when individual information is used [49]. In [50], authors developed a deep convolutional neural network approach able to fuse several information sources to detect and classify abnormalities in mammographic scans. First, features from regions of interests were fused in the convolutional block, whereas the decision of three different

classifiers (based on SVM and Random Forest) were then combined. Results
110 demonstrated a substantial increase in performance in the combined approach
compared with the one obtained individually by each classifier. In [3], we can
find another example of an ensemble classification framework for the diagnosis
of AD. Specifically, authors combined MRI data from several sessions of a lon-
gitudinal study with the results of neuropsychological tests in order to predict
115 the appearance of AD. Results showed the importance of effectively computing
how each individual source contributes to the classification decision, assigning
a higher weight to most informative modalities while penalizing those with a
lower accuracy. In the same line, other works have also attempted to combine
different models to increase the reliability of the results. For example, in [51] au-
120 thors trained 300 models independently in a classification context to identify the
informative patterns associated with autism. These outputs were then averaged
within a cross-validation scheme, leading to a stable and accurate result. The
use of ensemble architectures has shown a boost in performance when combined
with neural networks [52], especially in image classification [53]. [54] strategi-
125 cally combined basic models to classify functional magnetic resonance imaging,
whereas [55] fused features extracted from local binary patterns to improve the
classification of hyperspectral imaging.

3. Methodology

3.1. Regions parcellation

130 The first step when trying to evaluate differences between a couple of re-
gions is to properly delimitate them. Although our proposal could be used in
a wide range of scenarios, we focus on the analysis of brain images. Thus, the
initial step is to delimitate the anatomical brain regions. Previous studies have
demonstrated how brain parcellations may alter final classification results [56].
135 One important point is to determine the number of regions the brain should
be divided into since this affects to the size of those regions and their location.
A high number of subdivisions can enhance spatial precision, but reduce the

ability to find informative patterns. On the other hand, dividing the brain into a low number of regions means that they have a large size. Thus, although only
140 part of a region contains relevant information, the whole one would be mark
as informative. In order to strike a balance between brain parcellations and
spatial accuracy, we employed the Automated Anatomical Labelling (AAL) atlas,
which divides the brain into 116 regions, an intermediate number compared
with other atlases that are publicly available [57, 58]. Figure 2 depicts a visual
145 representation of the brain parcellations proposed by the AAL atlas.

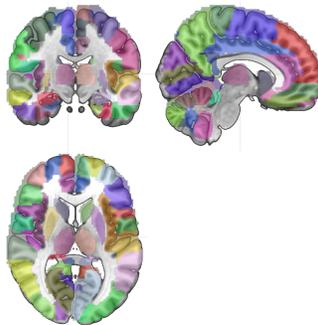


Figure 2: Representation of the 116 regions delimited by the AAL atlas.

3.2. Siamese neural network

Siamese networks were introduced in the 1990s as part of a signature verification system [59]. They are based on the combination of two neural networks with identical architecture, i.e parameters are the same and weights are commonly
150 shared. During training, each input is individually processed as a common feed-
forward network, which means that the information only follows one direction.
In short, neurons of each layer process the inputs and send the output to the
neurons of the following layer. Since weights are shared by the two networks,
they are updated at the same time following an error back-propagation process.
155 Each individual network receives one input and produces one output in its final
layer. A crucial point of this architecture is that similarity between the two
outputs is evaluated by a distance measure, which is then employed to assign a

final label to each input data. This output can be interpreted as the semantic difference between the projected representation of the inputs [60].

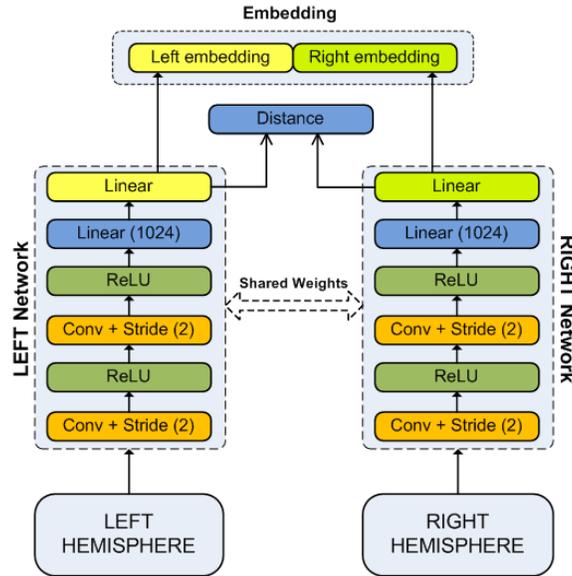


Figure 3: Scheme of the siamese neural network designed in this work. The left and right parts of each region defined by the AAL atlas are entered into each branch of the siamese network. The convolutional layers are employed as feature extractors, so that the output of the full convolutional layer of each branch comprises a representation in a lower dimension of the brain region. During the training process, the idea is to find a relationship between the differences between both regions and the diagnostic label. Once the training finishes, the network quantifies the asymmetry between the left and the right subregions and outputs the embeddings in a lower dimensionality, which are then combined with the latent spaces of the rest of the regions of the brain within an ensemble framework to predict the diagnosis of the patient.

160 Although siamese architectures have been used for assessing similarity between two samples, they can also be employed as a previous step in a classification tasks. Figure 3 summarizes the architecture used in this work. Specifically, we employed PET images from the dataset described in Section 4.1 as inputs of the network. The goal in this scenario was to evaluate whether asymmetry between brain regions from left and right hemispheres may be important
 165 in the study of the development of AD. Thus, we hypothesize that there is a relationship between the stage in the development of this disease and functional

differences between the left and right hemisphere of the brain. We evaluated differences from the 116 parcellations contained in the AAL atlas, which leads
170 to 58 pairs of regions (left and right part) that are iteratively entered into the siamese network. During training, the Hinge function [61] is used to compute the loss as a measure that quantifies the distance between the two outputs of the siamese network, as follows:

$$l(y) = \begin{cases} 0 & t \cdot y \geq 1 \\ 1 - t \cdot y & \text{otherwise} \end{cases} \quad (1)$$

where $t = \{-1, 1\}$ denotes the actual label and y corresponds to the output of
175 the linear layers.

3.3. Classification

Embeddings extracted for each pair of regions of the AAL atlas are combined and used as input features of a linear classification algorithm. Linear classifiers provide a weight map that quantifies the contribution of each individual feature
180 to the decision function. In our case, we used the weights to rank the asymmetry of each couple of regions in order to quantify their relevance in a certain classification context. Moreover, we employed an SVM classifier with a linear kernel whose decision function is based on maximizing the geometrical margin between the two classes. The classification rule f , can be specified by a pair of
185 (\mathbf{w}, \mathbf{x}) , as follows:

$$f(\mathbf{x}_i) = \langle \mathbf{w}, \mathbf{x}_i \rangle + b \quad (2)$$

where \mathbf{w} and \mathbf{x}_i are the weight and the feature vector, respectively, and b is the error term. A sample x would be classified as positive or negative depending on if $f(x) > 0$ or if $f(x) < 0$. The resulting decision function is based on a linear

rule obtained by solving the optimisation problem described in [62] :

$$\begin{aligned} \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_i \xi_i & \quad \text{subject to} \\ y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i & \quad \forall_i \xi_i \geq 0 \quad \forall_i \end{aligned} \quad (3)$$

190 where C denotes the penalty for misclassification. The solution of the optimisation problem is given as follows:

$$\mathbf{w} = \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i \quad (4)$$

after applying the Lagrangian multipliers. Substituting the value of \mathbf{w} in Equation 2, the decision function can be rewritten in its dual form, as follows :

$$f(\mathbf{x}_i) = \sum_{i=1}^n \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \quad (5)$$

where α_i and b are the coefficients to be learnt from the examples and $K(\mathbf{x}, \mathbf{x}_i)$ 195 is the kernel function employed to characterize the similarity between samples \mathbf{x} and \mathbf{x}_i .

During the training process, we employed an L1-regularization to enforce sparsity. This means that features associated with most of regions are automatically discarded (set to zero), while a few remaining ones are non-zeros, 200 which means that a bunch of extremely informative regions guide the decision of the classifier [63]. Once the classifier was trained, we extracted the weight maps associated with the embeddings used as inputs. The dimensionality of the embeddings is derived from the number of neurons in the output layer of the siamese network. We set a final layer of 40 neurons, and the resulting feature 205 vector was built as the concatenation of the embeddings of the subregions (left and right parts of the region), so that the input feature vector had a size of 80. This means that weight maps had exactly this size, but the aim was to know the relevance of the region as a whole, and not the contribution of each of the 80 values. To address this issue, we computed a normalized weight for each couple 210 of regions of the atlas as the absolute value of the embeddings extracted from

these subregions. Besides, we normalized this resulting value by the size of the region in order to penalize those regions that are extremely large. Equation 6 provides a mathematical summarization of this computation:

$$NW_{ROI} = \frac{\sum_{v \in ROI} \|W_v\|}{m_{ROI}} \quad (6)$$

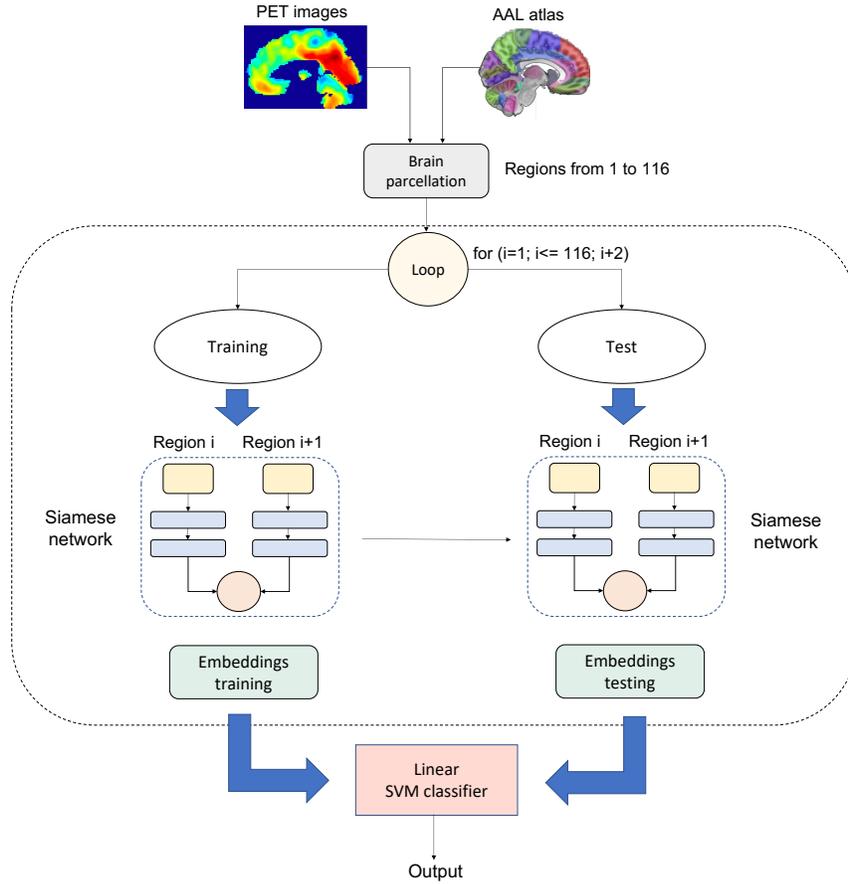


Figure 4: Scheme of the classification framework proposed in this work. PET images are divided into training and test subsets in order to guarantee the independence between the images used to build the model and those employed to test the performance. Regions defined by the neuroanatomical atlas are iteratively entered into the siamese neural network in pairs. Once the network is trained, the resulting embeddings are combined for all the regions of the brain and used to train a linear SVM classifier. The embeddings associated with the patients from the test sample are used to check the generalization ability of the model, leading to the final prediction.

with v representing the index of an embedding in the weight map, W_v is its weight
215 and m_{ROI} is the number of voxels contained in each ROI. Thus, the normalized
weight (NW_{ROI}) is a score that represents the amount of information contained
in a specific brain region. A large value means that the embeddings associated
with the ROI have had a large contribution to the classification model. A scheme
of the classification method proposed in this work is shown in Figure 4.

220 4. Evaluating asymmetry in Alzheimer’s disease

4.1. Database description

The data used in the preparation of this paper were obtained from the
Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).
The ADNI was launched in 2003 by the National Institute on Aging (NIA), the
225 National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food
and Drug Administration (FDA), private pharmaceutical companies and non-
profit organizations, as a \$60 million, 5-year public-private partnership. 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
230 measure the progression of Mild Cognitive Impairment (MCI) and AD. Determination
of sensitive and specific markers of very early AD progression is intended
to aid researchers and clinicians to develop new treatments and monitor their
effectiveness, as well as lessen the time and cost of clinical trials. The Principal
Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center
235 and University of California, San Francisco. ADNI is the result of efforts of
many co-investigators from a broad range of academic institutions and private
corporations, and subjects have been recruited from over 50 sites across the U.S.
and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has
been followed by ADNI-GO and ADNI-2. To date these three protocols have
240 recruited over 1500 adults, ages 55-90, to participate in the research, consist-
ing of cognitively normal older individuals, people with early or late MCI, and
people with early AD. The follow up duration of each group is specified in the

protocols for ADNI-1, ADNI-2, and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-
 245 date information, see www.adni-info.org. All procedures performed during the acquisition of the data by ADNI were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards (see [64] for more details.)

250 The experiments conducted in this work employed 18F-FDG PET data from 241 patients, consisting of 70 suffering from Alzheimer’s disease (AD), 39 Mild Cognitive impairment that in future sessions of the longitudinal study convert to AD (MCIc), 64 MCI patients that remain stable in all sessions (MCI_s) and 68 Controls (CTL). Demographic data (gender and age) in addition to the Mini
 255 Mental State Examination scores (MMSE) of the patients in the database are summarized in Table 1, whereas a representation of the PET images for each individual pathology in the database is shown in Figure 5.

Table 1: Patient’s demographics for Alzheimer’s disease (AD), mild cognitive impairment converter (MCIc) and stable (MCI_s), and controls (CTL).

| Diagnosis | Number | Gender (M/F) | Age | MMSE |
|------------------|--------|--------------|-------------|-------------|
| AD | 70 | 46/24 | 75.26 ±7.53 | 23.22 ±2.19 |
| MCIc | 39 | 28/11 | 74.50 ±7.05 | 26.76 ±1.74 |
| MCI _s | 64 | 42/22 | 75.03 ±7.72 | 27.18 ±2.53 |
| CTL | 68 | 43/25 | 75.87 ±5.02 | 29.11 ±0.99 |

4.2. Image preprocessing

Images were first spatially normalized according to a PET template using
 260 SPM12 [65]. This allows to modify the shape and size of the brain of each individual into a standard template in order to establish a correspondence between them, allowing the comparison between brains from different patients. After that, it is also necessary to apply an intensity normalization to standardize the intensity values in the images of the different subjects. Thus, the intensity of
 265 the images was normalized to a value, I_{max} , which was obtained by averaging

the 0.1% highest voxel intensities that exceeded a specific threshold [66]. This threshold was fixed to the 10th bin intensity value of a 50 bins intensity histogram, so that all voxels that did not surpass the threshold were discarded and considered as background, as they do not contain any relevant information but
 270 can add noise and artifacts [67].

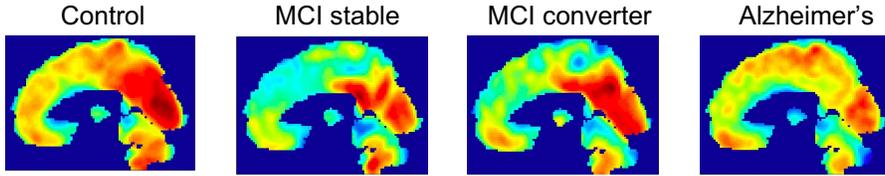


Figure 5: Slice of a PET image for the different stages of the AD development: control, MCI (stable or converter) and AD.

4.3. Performance evaluation

We employed a 5-fold stratified cross-validation scheme [68] in order to guarantee that there was approximately the same percentage of images of each class in each individual fold. The total number of images were divided into five different groups: four of them were used to train the model and the remaining
 275 one was used to estimate the generalization ability of the model. These subdivisions were performed within an iterative process, so that images contained in each subdivision were employed once to test the performance of the model. Performance was evaluated in terms of the following metrics derived from the
 280 confusion matrix:

$$Bal\ Acc = \frac{1}{2} \left(\frac{TP}{P} + \frac{TN}{N} \right) \quad Prec = \frac{TP}{TP+FP}$$

$$Sens = \frac{TP}{TP+FN} \quad Spec = \frac{TN}{TN+FP}$$

$$F1 - score = \frac{2 \times Prec \times Sens}{Prec + Sens} \quad AUC = \frac{1}{2} \left(\frac{TP}{P} + \frac{TN}{N} \right)$$

In the CTL *vs* AD classification, T_P refers to the number of patients correctly

classified as AD (true positives), T_N corresponds to the number of controls properly identified (true negatives), F_P quantifies the number of controls labelled as AD (false positives), whereas F_N refers to the number of AD patients incor-
 285 rectly classified as controls. Besides, the area under the ROC curve (AUC) is employed as an additional measure for evaluating the ability of the model to identify the different classes [69, 70].

After computing performance, it is crucial to assess the statistical significance of the results. To do so, we employed a non-parametric test based on
 290 permutations [71]. Labels associated with each individual image were shuffled and ensemble classification was performed. This process was repeated 500 times in order to build an empirical distribution of the accuracies. The probability of occurrence of a certain accuracy was assessed by comparing the scores obtained when training the classifier with the correct labels and the empirical
 295 distribution. The associated p -value is computed as follows:

$$p = \frac{1 + n}{N} \quad (7)$$

where n is the number of occurrences that exceeds the actual accuracy and N is the number of permutations performed to build the empirical distribution. Thus, a result is considered significant if it does not surpass the significance threshold, which is widely established as $p = 0.05$.

300 4.4. *Experimental setup*

The experiments conducted in this work can be summarized as the evaluation of the functional asymmetry at different stages of Alzheimer’s disease:

- **AD vs CTL:** identification of asymmetry between regions from both hemispheres in order to detect whether AD modifies them in a non-
 305 homogeneous way.
- **MCIs vs CTL:** in this context, the evaluation is performed in an early stage of the degeneration process.

- **MCIc vs CTL:** similar to the previous context, but in this case the asymmetry is evaluated in MCI patients that will develop AD in the future.
- 310 • **MCIc vs MCIs:** the hardest case, since the diagnosis is the same for both groups. The aim is to evaluate the existence of asymmetry in some regions that predicts the conversion of MCI patients versus those that remain stable.

The framework employed in all the experiments relies on the use of siamese
 315 networks for feature extraction and an ensemble of linear SVM classifiers for classification purposes. The final value of the cost parameter, $C = 1$, was selected as the optimum value by using a grid search within a 5-Fold cross-validation scheme. This regularization parameter controls the trades off between the misclassification of training samples and the simplicity of the decision
 320 surface. Besides, we employed an early-stopping procedure during the training of the siamese network in order to control the overfitting of the model [72]. Specifically, we used part of the training data as a validation sample in order to finish the training procedure when the validation loss stopped decreasing.

We employed custom code written in Python 3.6, in addition to a number
 325 of libraries such as Torch 1.7.1, Numpy 1.19.5 and Scikit-Learn 1.0. All the experiments were carried out on a computing cluster with high performance: two Intel[®] Xeon[®] E5-2630 node 2.40GHz processors, with 10 cores per processor; one Nvidia Geforce RTX3090 with 128 GB DDR6 memory. Besides, the total RAM memory capacity of the system is 128 GB. Moreover, the average execution
 330 time was 8 hours and 12 minutes for the AD *vs* CTL classification context, which is the one with the highest number of images to process.

Table 2 summarizes recent works focused on the automatic identification of AD, including the performance that they obtained. Besides, we included the best results obtained by the method proposed in this work. It is important to
 335 note that the comparison between the different approaches is not a straightforward task. First, the aim of our method is not classification itself, but a better understanding of the regions affected in different stages of AD through

Table 2: Performance obtained by previous works and by our method in the identification of Alzheimer’s disease.

| Research work | Dataset | Method | Classification context | Results (%) |
|-------------------|-------------------|------------------------------|---------------------------------|-------------------|
| [73] | 138 MRI and PET | Deep belief networks | AD <i>vs</i> CTL | AUC = 0.95 |
| [67] | 138 MRI and PET | Sparse representation + SVM | AD <i>vs</i> CTL | Acc = 92.00 |
| [74] | 400 MRI | FLS-TWSVM | AD <i>vs</i> CTL | Acc = 97.15 |
| [75] | 818 MRI | Broad Learning System | AD <i>vs</i> CTL | Acc = 91.83 |
| [76] | 200 MRI | AlexNet | AD <i>vs</i> CTL | Acc = 95.00 |
| [77] | 500 MRI | ResNet-50 | AD <i>vs</i> CTL | Acc = 86.67 |
| [78] | 479 PET | Autoencoder | AD <i>vs</i> CTL | Acc = 88.73 |
| [79] | 511 MRI | Log-Gabor filters | AD <i>vs</i> CTL | AUC = 0.90 |
| [80] | 250 diffusion MRI | Elastic Net | AD <i>vs</i> CTL | AUC = 0.90 |
| [81] | 416 MRI | Wavelet + SVM | AD <i>vs</i> CTL | AUC = 0.90 |
| [82] | 97 SPECT | PCA + SVM | AD <i>vs</i> CTL | Acc = 89.69 |
| [83] | 246 PET | Recurrent Neural Network | MCI <i>vs</i> CTL | Acc = 83.90 |
| [84] | 416 MRI | CNN | AD <i>vs</i> CTL | Acc = 97.75 |
| [85] | 6400 MRI | LPQNET | Dementia <i>vs</i> CTL | Acc = 99.62 |
| [86] | 769 MRI | AlzVNet | AD <i>vs</i> MCI <i>vs</i> CTL | Acc = 98.26 |
| [87] | 99 MRI and PET | Adaptive Similarity Learning | AD <i>vs</i> MCI <i>vs</i> MCIc | Acc = 69.41 |
| Our method | 138 PET | Siamese network | AD <i>vs</i> CTL | AUC = 0.98 |
| Our method | 107 PET | Siamese network | MCIc <i>vs</i> CTL | AUC = 0.98 |
| Our method | 132 PET | Siamese network | MCIc <i>vs</i> MCIc | AUC = 0.90 |
| Our method | 103 PET | Siamese network | MCIc <i>vs</i> MCIc | AUC = 0.91 |

the quantification of asymmetry between brain regions. Second, many of these studies employ images from the ADNI database, but some of them use private datasets. Since performance is clearly affected by the images employed, the comparison between different works should be done as an informative way, and not as a categorical ranking of the different works.

5. Results

Table 3: Performance of the proposed method in the different classification scenarios.

| Classification | Bal Acc (%) | Sens (%) | Spec (%) | Prec (%) | AUC | F1-score (%) |
|---------------------|-------------|-------------|-------------|-------------|------------|--------------|
| AD <i>vs</i> CTL | 98.95 ±1.27 | 97.88 ±1.31 | 98.02 ±1.36 | 97.56 ±1.41 | 0.98 ±0.17 | 98.67 ±1.21 |
| MCIc <i>vs</i> CTL | 94.01±3.07 | 96.92±6.15 | 91.10±3.25 | 91.27±2.35 | 0.98 ±0.55 | 93.89±3.22 |
| MCIc <i>vs</i> CTL | 80.67 ±8.45 | 78.91±7.26 | 91.52±5.43 | 83.31±10.45 | 0.90 ±0.47 | 82.03±12.59 |
| MCIc <i>vs</i> MCIc | 78.41±9.19 | 95.38±3.76 | 61.43±12.30 | 81.02±7.48 | 0.91 ±0.44 | 87.40±4.61 |

We first explore the performance obtained by the proposed method in the different classification contexts, as summarized in Table 3. We can see that there is a clear relationship between the performance and the difficulty of the classification scenario. The maximum accuracy, 98.95%, is obtained when distinguishing between AD and controls, the context where differences between the two groups are maximum. After that, the second maximum performance

350 corresponds to the context in which differences between both groups are also high (MCIc vs CTL, Acc = 94.01%), since MCI patients will convert to AD in future sessions of the longitudinal study. Results show a decrease in performance for the two contexts where the diagnosis of patients of the two groups is more similar. Specifically, our method provides a 80.67% of accuracy in MCIs vs CTL, with a similar performance (78.41%) in MCIc vs MCIs. The results associated with this last context demonstrates a relationship between an asymmetric degeneration of the brain and a high probability of developing AD in the future.

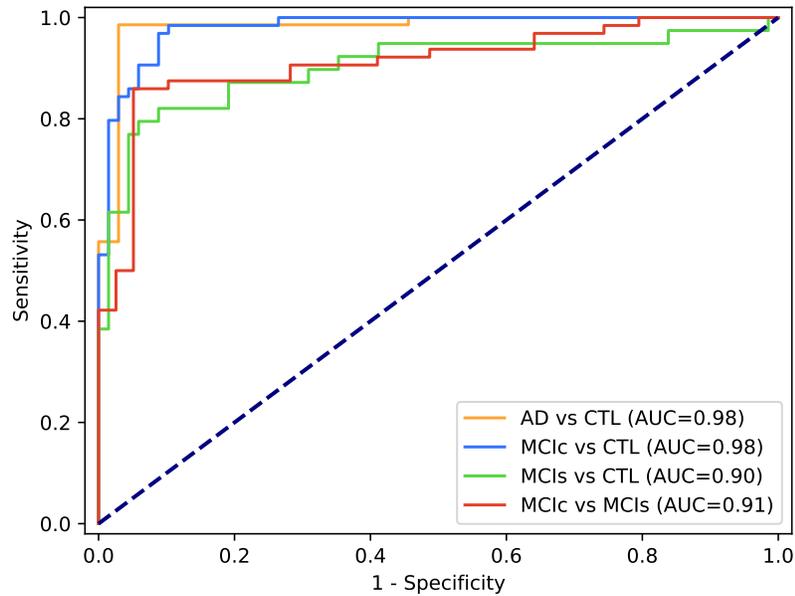


Figure 6: ROC curves obtained for the different classification contexts.

360 A similar behavior occurs when referring to the area under the ROC curve, as Figure 6 shows. Differences between controls and AD patients are higher than any of the other classifications, leading to the largest diagnostic power (AUC=0.98), whereas distinguishing MCI stable and controls/MCI converters lead to the lowest AUC (0.91 and 0.90, respectively). However, it is remarkable

the high performance obtained in all scenarios, demonstrating the usefulness of
 365 the proposed method. Figure 7 shows the empirical distributions obtained in the
 four classification contexts, in addition to the actual accuracy after classifying
 using the correct labels. It can be seen that results are clearly significant,
 validating our findings from a statistical standpoint.

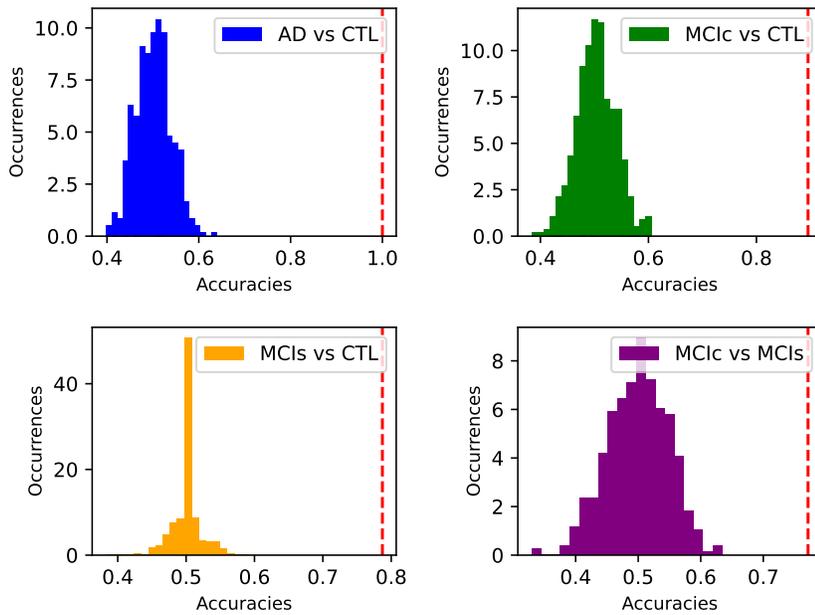


Figure 7: Empirical distribution of the classification accuracy obtained after performing per-
 mutations. The vertical red lines represent the accuracy obtained when using the correct
 labels.

Table 4: Most relevant regions in each classification scenario.

| Classification | Regions |
|----------------|--|
| AD vs CTL | Superior frontal gyrus, Middle frontal gyrus, Hippocampus, Inferior occipital gyrus. |
| MCIC vs CTL | Posterior cingulate gyrus, Caudate, Middle cingulate gyrus, Gyrus rectus, Superior frontal gyrus. |
| MCIs vs CTL | Inferior occipital gyrus, Rectus gyrus, Middle cingulate gyrus, Posterior cingulate gyrus, Precentral gyrus. |
| MCIC vs MCIs | Inferior occipital gyrus, Middle cingulate gyrus, Rectus gyrus, Superior frontal gyrus, Heschl's gyrus. |

Figure 8 provides a map of the regions with a higher contribution to the
 370 classification decision in the four contexts evaluated. These regions are also
 summarized in Table 4, ranked according to the associated weight in order to
 compare their influence in classification. This representation allows a visual

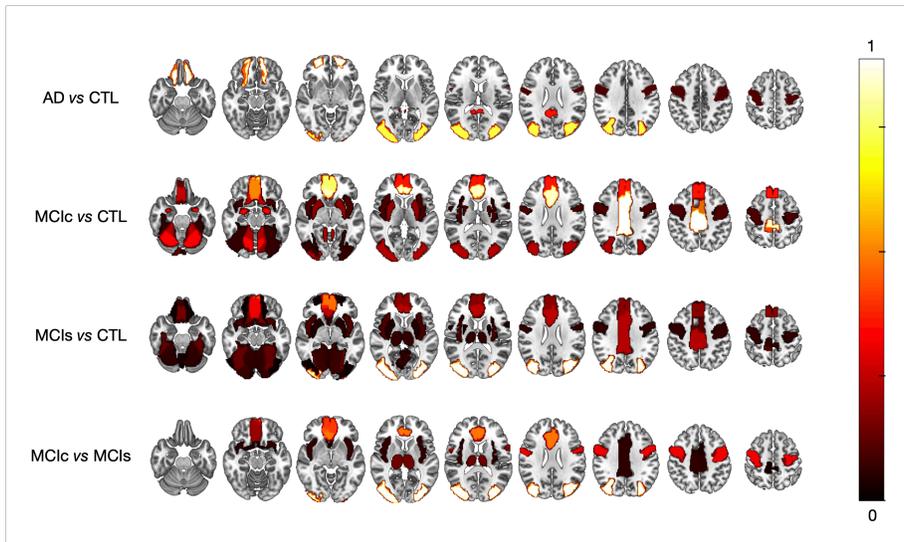


Figure 8: Weights associated with the asymmetric regions that guide the classifier’s decision in the different scenarios.

identification of the regions whose asymmetry allows to differentiate between the different cognitive states. Most importantly, changes in the regions map
 375 across the four classification contexts manifest the atrophy of brain regions in all the stages of the degeneration process associated with AD. Further discussion and its implications are covered in Section 6.

Figure 9 shows a two-dimensional visualization of the embeddings associated with the different stages of AD. The representation provides a clear separation
 380 into clusters between the different classes, especially between AD y CTL. This separation is also high between MCI patients, but some of the MCIc have more similar embeddings to MCIc than to the rest of the MCIc class. This explains the decrease in performance obtained in the MCIc *vs* MCIc classification context. This figure also includes ApoE genetic information from the different
 385 subjects regarding the three genotypes more related to AD. Previous studies have demonstrated that subjects with two copies of the allele 3 (ApoE- ϵ 3) have a high likelihood of developing AD, whereas the risk factor is much higher in subjects with two copies of the allele-4 (ApoE- ϵ 4). On the other hand, subjects

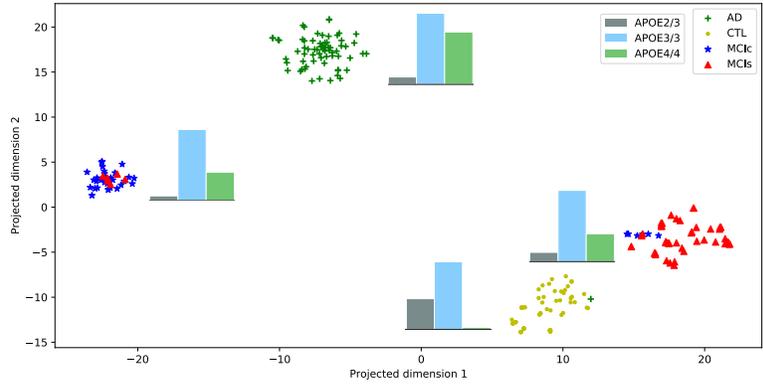


Figure 9: Two-dimensional representation of the embeddings associated with each pathology and distribution of the genetic information for the different stages of AD progression.

with genotype ApoE 2,3 are considered protected against AD. The bar charts
 390 included in the figure corroborates that most of ApOE-4 subjects belong to the
 AD group, whereas the prevalence in the controls is minimum (only one per-
 son). Conversely, subjects with genotype ApoE 2,3 are mostly located in the
 CTL group, whereas the presence of these people in the MCIc and AD groups
 considerably decreases.

395 **6. Discussion**

In this work, we propose a method based on siamese architecture to evaluate
 functional differences in medical imaging. This approach relies on the union of
 two convolutional neural networks with identical configuration, i.e. their pa-
 rameters are the same and weights are commonly shared. The siamese network
 400 has two inputs, which corresponds to the brain regions to be compared. After
 training the network, the embeddings from the last linear layer of the archi-
 tecture are extracted. This process is contained within an iterative scheme in
 which differences between left and right hemispheres are evaluated for all regions
 provided by an anatomical atlas. Thus, the embeddings of each pair of regions
 405 are then combined via an ensemble in order to quantify their predictive power.

We evaluated the performance of this approach in terms of different metrics and studied how informative functional asymmetry is in the progression of AD.

The high performance shown by the method proposed in this work when handling functional data demonstrates its usefulness for studying the spatial patterns associated with the progression of AD. Although accuracy is extremely high (98.95% when differentiating between AD and CTL), it is worth mentioning that classification *per se* is not the main point of this work. Instead, the identification of the regions that are asymmetrically affected by the progression of AD at different stages of the disease. With reference to this spatial information, the regions with a large asymmetry when distinguishing controls from AD patients are the superior frontal gyrus, the middle frontal gyrus and the hippocampus, which its role in AD has been reported in previous studies [88, 89, 90]. These regions are slightly different in the conversion from MCI to AD, where the posterior cingulate gyrus, the caudate and the middle cingulate gyrus play a crucial role [91, 92, 93]. Finally, an incipient asymmetry is found in the inferior occipital gyrus, the middle cingulate gyrus and the gyrus rectus, evidencing the importance of these regions in an early stage of the disease [94, 95, 96].

It is worth understanding that these results have not been derived from the region as a whole, but from the asymmetry between the left and right parts of each region. For example, hippocampus had been previously linked to the degeneration of brain, and even identified as informative by intelligent systems. The difference between these previous results and our findings is that we demonstrate the importance of asymmetry in the evaluation of the AD. Our proposal divides each brain region into two subregions: the left and the right one. Let imagine a region that is extremely affected by the disease, but the atrophy is similar both in the left and right one. In this case, our method would not marked as informative this region, since the classification performed by our approach is based on differences in atrophy, and not atrophy itself. This is highly relevant for the development of computer-aided diagnosis (CAD) systems to help clinicians in the early diagnosis of this disease.

We have developed a robust tool for evaluating differences in medical images that is extremely useful in a wide range of pathologies. It is also important to note that the design of our method eliminates the need of *a priori* defining features that are subsequently entered into a classifier to evaluate their informativeness. This is remarkably relevant since there is a crucial relationship between the election of the features and performance. Region A can differ from Region B both in its eccentricity and in the average intensity value of the voxels that contains. However, if the input feature to a classifier is the size of both regions, the algorithm will probably not find any differences between them despite they exist. The training of the classification system proposed in this work is based on the difference between the regions itself, not the differences in a bunch of features previously computed. This assures that the system is specifically adjusted to detect real differences between the regions.

7. Conclusion

In this work, we present a method based on siamese architecture to evaluate functional differences in medical imaging. These differences can be seen as a measure of asymmetry between the brain regions evaluated, and are then assessed in a specific classification scenario to quantify their relevance. Once the network is trained, vectors from the final linear layer of the two branches of the siamese network are extracted. These embeddings are then used as input of a SVM classifier with linear kernel, yielding an accuracy of 98.95 % when distinguishing between controls and AD patients. Our findings reveal the suitability of the method proposed for the study of the development of AD. Moreover, this work paves the way to future research not only in brain imaging but also for its application to other biomedical signals. Additionally, the inclusion of other measures for the computation of the similarity between two samples could lead to a more realistic estimation of brain asymmetry. Finally, our results demonstrate that differences between two regions can be even more relevant than the study of an isolated region, given the importance of the asymmetrical deterioration in

neurological disorders.

Funding

This work was supported by projects PGC2018-098813-B-C32 and RTI2018-098913-B100 (Spanish “Ministerio de Ciencia, Innovación y Universidades”),
470 UMA20-FEDERJA-086, A-TIC-080-UGR18 and P20 00525 (Consejería de economía
y conocimiento, Junta de Andalucía) and by European Regional Development
Funds (ERDF); and by Spanish ”Ministerio de Universidades” through Margarita-
Salas grant to J.E. Arco.

References

- 475 [1] J. E. Arco, J. Ramírez, C. G. Puntonet, J. M. Górriz, M. Ruz, Improving
short-term prediction from MCI to AD by applying Searchlight analysis, in:
2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI),
2016, pp. 10–13. doi:10.1109/ISBI.2016.7493199.
- [2] L. J. C. de Mendonça, R. J. Ferrari, Alzheimer’s disease classification based
480 on graph kernel svms constructed with 3d texture features extracted from
mr images, *Expert Systems with Applications* 211 (2023) 118633. doi:
<https://doi.org/10.1016/j.eswa.2022.118633>.
- [3] J. E. Arco, J. Ramírez, J. M. Górriz, M. Ruz, Data fusion based on Search-
light analysis for the prediction of Alzheimer’s disease, *Expert Systems with*
485 *Applications* 185 (2021) 115549. doi:[https://doi.org/10.1016/j.eswa.](https://doi.org/10.1016/j.eswa.2021.115549)
2021.115549.
- [4] B. Lei, E. Liang, M. Yang, P. Yang, F. Zhou, E.-L. Tan, Y. Lei, C.-M. Liu,
T. Wang, X. Xiao, S. Wang, Predicting clinical scores for alzheimer’s disease
490 based on joint and deep learning, *Expert Systems with Applications* 187
(2022) 115966. doi:<https://doi.org/10.1016/j.eswa.2021.115966>.

- [5] Z. Gao, Y. Wang, M. Huang, J. Luo, S. Tang, A kernel-free fuzzy reduced quadratic surface ν -support vector machine with applications, *Applied Soft Computing* 127 (2022) 109390. doi:<https://doi.org/10.1016/j.asoc.2022.109390>.
- 495 [6] D. Castillo-Barnes, F. J. Martínez-Murcia, A. Ortiz, D. Salas-Gonzalez, J. Ramírez, J. M. Górriz, Morphological characterization of functional brain imaging by isosurface analysis in parkinson's disease, *International Journal of Neural Systems* 30 (09) (2020) 2050044. doi:[10.1142/S0129065720500446](https://doi.org/10.1142/S0129065720500446).
- 500 [7] J. Barbero-Gómez, P.-A. Gutiérrez, V.-M. Vargas, J.-A. Vallejo-Casas, C. Hervás-Martínez, An ordinal cnn approach for the assessment of neurological damage in parkinson's disease patients, *Expert Systems with Applications* 182 (2021) 115271. doi:<https://doi.org/10.1016/j.eswa.2021.115271>.
- 505 [8] J. E. Arco, A. Ortiz, D. Castillo-Barnes, J. M. Górriz, J. Ramírez, Quantifying inter-hemispheric differences in Parkinson's disease using siamese networks, in: J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, H. Adeli (Eds.), *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications*, Springer International Publishing, 2022, pp. 156–165.
- 510 [9] J. E. Arco, A. Ortiz, J. Ramírez, F. J. Martínez-Murcia, Y.-D. Zhang, J. M. Górriz, Uncertainty-driven ensembles of multi-scale deep architectures for image classification, *Information Fusion* 89 (2023) 53–65. doi:<https://doi.org/10.1016/j.inffus.2022.08.010>.
- 515 [10] B. Vidya, S. P, Gait based parkinson's disease diagnosis and severity rating using multi-class support vector machine, *Applied Soft Computing* 113 (2021) 107939. doi:<https://doi.org/10.1016/j.asoc.2021.107939>.
- [11] A. Rojas, J. Górriz, J. Ramírez, I. Illán, F. Martínez-Murcia, A. Ortiz,

- 520 M. Gómez Río, M. Moreno-Caballero, Application of empirical mode decomposition (emd) on datscan spect images to explore parkinson disease, *Expert Systems with Applications* 40 (7) (2013) 2756–2766. doi:<https://doi.org/10.1016/j.eswa.2012.11.017>.
- [12] R. Prashanth, S. Dutta Roy, P. K. Mandal, S. Ghosh, Automatic classification and prediction models for early parkinson’s disease diagnosis from spect imaging, *Expert Systems with Applications* 41 (7) (2014) 3333–3342. doi:<https://doi.org/10.1016/j.eswa.2013.11.031>.
525
- [13] D. Castillo-Barnes, C. Jimenez-Mesa, F. J. Martinez-Murcia, D. Salas-Gonzalez, J. Ramírez, J. M. Górriz, Quantifying differences between affine and nonlinear spatial normalization of FP-CIT SPECT images, *International Journal of Neural Systems* 32 (05) (mar 2022). doi:[10.1142/S0129065722500198](https://doi.org/10.1142/S0129065722500198).
530
- [14] A. Brahim, J. Górriz, J. Ramírez, L. Khedher, Intensity normalization of datscan spect imaging using a model-based clustering approach, *Applied Soft Computing* 37 (2015) 234–244. doi:<https://doi.org/10.1016/j.asoc.2015.08.030>.
535
- [15] C. Li, J. Zhou, D. Wang, X. Li, S. Jiang, Y. Zhang, Z. Wen, G. Wang, F. Yan, M. Chen, Amide proton transfer imaging of alzheimer’s disease and parkinson’s disease, *Magnetic Resonance Letters* (2022). doi:<https://doi.org/10.1016/j.mrl.2022.10.002>.
- 540 [16] F. J. Martinez-Murcia, J. M. Górriz, J. Ramírez, Feature Extraction, John Wiley & Sons, Ltd, 2017, pp. 1–9. doi:<https://doi.org/10.1002/047134608X.W5506.pub2>.
- [17] J. E. Arco, C. González-García, P. Díaz-Gutiérrez, J. Ramírez, M. Ruz, Influence of activation pattern estimates and statistical significance tests in fMRI decoding analysis, *Journal of Neuroscience Methods* 308 (2018) 248–260.
545

- [18] C. C. Gross, A. Schulte-Mecklenbeck, L. Madireddy, M. Pawlitzki, C. Strip-
pel, S. Räuber, J. Krämer, L. Rolfes, T. Ruck, C. Beuker, A. Schmidt-
Pogoda, L. Lohmann, T. Schneider-Hohendorf, T. Hahn, N. Schwab,
550 J. Minnerup, N. Melzer, L. Klotz, S. G. Meuth, G. Meyer zu Hörste,
S. E. Baranzini, H. Wiendl, Classification of neurological diseases using
multi-dimensional CSF analysis, *Brain* 144 (9) (2021) 2625–2634. doi:
10.1093/brain/awab147.
- [19] F. Segovia, J. M. Górriz, J. Ramírez, F. J. Martínez-Murcia, D. Castillo-
555 Barnes, Assisted diagnosis of parkinsonism based on the striatal mor-
phology, *International Journal of Neural Systems* 29 (09) (2019) 1950011.
doi:10.1142/S0129065719500114.
- [20] Álvaro S. Hervella, J. Rouco, J. Novo, M. Ortega, End-to-end multi-task
learning for simultaneous optic disc and cup segmentation and glaucoma
560 classification in eye fundus images, *Applied Soft Computing* 116 (2022)
108347. doi:https://doi.org/10.1016/j.asoc.2021.108347.
- [21] Y. Li, T. Chu, Y. Liu, H. Zhang, F. Dong, Q. Gai, Y. Shi, H. Ma, F. Zhao,
K. Che, N. Mao, H. Xie, Classification of major depression disorder via
565 using minimum spanning tree of individual high-order morphological brain
network, *Journal of Affective Disorders* 323 (2023) 10–20. doi:https:
//doi.org/10.1016/j.jad.2022.11.029.
- [22] P. S.-P. Hung, A. Noorani, J. Y. Zhang, S. Tohyama, N. Laperriere, K. D.
Davis, D. J. Mikulis, F. Rudzicz, M. Hodaie, Regional brain morphology
570 predicts pain relief in trigeminal neuralgia, *NeuroImage: Clinical* 31 (2021)
102706. doi:https://doi.org/10.1016/j.nicl.2021.102706.
- [23] Student, The probable error of a mean, *Biometrika* (1908) 1–25.
- [24] M. Neuhäuser, *Wilcoxon–Mann–Whitney Test*, Springer Berlin Hei-
delberg, Berlin, Heidelberg, 2011, pp. 1656–1658. doi:10.1007/
978-3-642-04898-2_615.

- 575 [25] T. Sathies Kumar, C. Arun, P. Ezhumalai, An approach for brain tumor detection using optimal feature selection and optimized deep belief network, *Biomedical Signal Processing and Control* 73 (2022) 103440. doi:<https://doi.org/10.1016/j.bspc.2021.103440>.
- [26] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, Machine learning
580 framework for early MRI-based Alzheimer’s conversion prediction in MCI subjects, *NeuroImage* 104 (2015) 398–412.
- [27] M. Termenon, M. Graña, A two stage sequential ensemble applied to the classification of Alzheimer’s disease based on MRI features, *Neural Processing Letters* 35 (2012) 1–12. doi:[10.1007/s11063-011-9200-2](https://doi.org/10.1007/s11063-011-9200-2).
- 585 [28] M. Ganaie, M. Tanveer, Fuzzy least squares projection twin support vector machines for class imbalance learning, *Applied Soft Computing* 113 (2021) 107933. doi:<https://doi.org/10.1016/j.asoc.2021.107933>.
- [29] J. Zhuang, Y. Dong, H. Bai, Ensemble learning with siamese networks for visual tracking, *Neurocomputing* 464 (2021) 497–506. doi:<https://doi.org/10.1016/j.neucom.2021.08.025>.
590
- [30] B. Li, J. Yan, W. Wu, Z. Zhu, X. Hu, High performance visual tracking with siamese region proposal network, in: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018*, pp. 8971–8980. doi:[10.1109/CVPR.2018.00935](https://doi.org/10.1109/CVPR.2018.00935).
- 595 [31] J. Zhang, M. Miao, H. Zhang, J. Wang, Y. Zhao, Z. Chen, J. Qiao, Object semantic-guided graph attention feature fusion network for siamese visual tracking, *Journal of Visual Communication and Image Representation* 90 (2023) 103705. doi:<https://doi.org/10.1016/j.jvcir.2022.103705>.
- [32] S. Ghosh, S. Ghosh, P. Kumar, E. Scheme, P. P. Roy, A novel spatio-temporal siamese network for 3D signature recognition, *Pattern Recognition Letters* 144 (2021) 13–20. doi:<https://doi.org/10.1016/j.patrec.2021.01.012>.
600

- [33] C. S. Vorugunti, P. Mukherjee, D. S. Guru, V. Pulabaigari, OSVNet: Convolutional siamese network for writer independent online signature verification, 2019 International Conference on Document Analysis and Recognition (ICDAR) (2019) 1470–1475.
- [34] S. Matsuo, X. Wu, G. Atarsaikhan, A. Kimura, K. Kashino, B. K. Iwana, S. Uchida, Deep attentive time warping, Pattern Recognition (2022) 109201doi:<https://doi.org/10.1016/j.patcog.2022.109201>.
- [35] W. Rao, Y. Qu, L. Gao, X. Sun, Y. Wu, B. Zhang, Transferable network with siamese architecture for anomaly detection in hyperspectral images, International Journal of Applied Earth Observation and Geoinformation 106 (2022) 102669. doi:<https://doi.org/10.1016/j.jag.2021.102669>.
- [36] J. Wang, Y. Jia, D. Wang, W. Xiao, Z. Wang, Weighted IForest and siamese GRU on small sample anomaly detection in healthcare, Computer Methods and Programs in Biomedicine 218 (2022) 106706. doi:<https://doi.org/10.1016/j.cmpb.2022.106706>.
- [37] S. M. Nagarajan, G. G. Deverajan, A. K. Bashir, R. P. Mahapatra, M. S. Al-Numay, Iadf-cps: Intelligent anomaly detection framework towards cyber physical systems, Computer Communications 188 (2022) 81–89. doi:<https://doi.org/10.1016/j.comcom.2022.02.022>.
- [38] P. Manocha, R. Badlani, A. Kumar, A. Shah, B. Elizalde, B. Raj, Content-based representations of audio using siamese neural networks, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2018) 3136–3140.
- [39] Y. Zhang, B. Pardo, Z. Duan, Siamese style convolutional neural networks for sound search by vocal imitation, IEEE/ACM Transactions on Audio, Speech, and Language Processing 27 (2) (2019) 429–441. doi:[10.1109/TASLP.2018.2868428](https://doi.org/10.1109/TASLP.2018.2868428).

- 630 [40] S. Ramoji, P. Krishnan, S. Ganapathy, Plda inspired siamese networks for speaker verification, *Computer Speech & Language* 76 (2022) 101383. doi:<https://doi.org/10.1016/j.csl.2022.101383>.
- [41] X. Zhang, P. Ye, S. Peng, J. Liu, G. Xiao, Dsiammft: An RGB-T fusion tracking method via dynamic siamese networks using multi-layer feature fusion, *Signal Processing: Image Communication* 84 (2020) 115756. doi:
635 <https://doi.org/10.1016/j.image.2019.115756>.
- [42] Y. Zhang, L. Man Po, M. Liu, Y. A. Ur Rehman, W. Ou, Y. Zhao, Data-level information enhancement: Motion-patch-based siamese convolutional neural networks for human activity recognition in videos, *Expert Syst. Appl.* 147 (2020). doi:
640 [10.1016/j.eswa.2020.113203](https://doi.org/10.1016/j.eswa.2020.113203).
- [43] C. Zhang, Y. Feng, L. Hu, D. Tapete, L. Pan, Z. Liang, F. Cigna, P. Yue, A domain adaptation neural network for change detection with heterogeneous optical and sar remote sensing images, *International Journal of Applied Earth Observation and Geoinformation* 109 (2022) 102769. doi:
645 <https://doi.org/10.1016/j.jag.2022.102769>.
- [44] Y. Liu, X. Chen, J. Cheng, H. Peng, A medical image fusion method based on convolutional neural networks, in: *2017 20th International Conference on Information Fusion (Fusion)*, 2017, pp. 1–7. doi:
8009769. [10.23919/ICIF.2017.8009769](https://doi.org/10.23919/ICIF.2017.8009769).
- 650 [45] A. Sabeeh Yousif, Z. Omar, U. Ullah Sheikh, An improved approach for medical image fusion using sparse representation and siamese convolutional neural network, *Biomedical Signal Processing and Control* 72 (2022) 103357. doi:
<https://doi.org/10.1016/j.bspc.2021.103357>.
- [46] S. Ahuja, B. K. Panigrahi, N. Dey, A. Taneja, T. K. Gandhi, Mcs-net: Multi-class siamese network for severity of covid-19 infection classification
655 from lung ct scan slices, *Applied Soft Computing* 131 (2022) 109683. doi:
<https://doi.org/10.1016/j.asoc.2022.109683>.

- [47] M. Li, K. Chang, B. Bearce, C. Chang, A. Huang, J. Campbell, J. Brown, P. Singh, K. Hoebel, D. Erdogmus, S. Ioannidis, W. Palmer, M. Chiang, J. Kalpathy-Cramer, Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging, *npj Digital Medicine* 3 (12 2020). doi:10.1038/s41746-020-0255-1.
- 660
- [48] M. Shorfuzzaman, M. S. Hossain, Metacovid: A siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients, *Pattern Recognition* 113 (2021) 107700. doi:https://doi.org/10.1016/j.patcog.2020.107700.
- 665
- [49] D. Castillo-Barnes, J. Ramírez, F. Segovia, F. J. Martínez-Murcia, D. Salas-Gonzalez, J. M. Górriz, Robust ensemble classification methodology for i123-ioflupane SPECT images and multiple heterogeneous biomarkers in the diagnosis of Parkinson's disease, *Frontiers in Neuroinformatics* 12 (2018) 53. doi:10.3389/fninf.2018.00053.
- 670
- [50] I. U. Haq, H. Ali, H. Y. Wang, C. Lei, H. Ali, Feature fusion and ensemble learning-based cnn model for mammographic image classification, *Journal of King Saud University - Computer and Information Sciences* (2022). doi:https://doi.org/10.1016/j.jksuci.2022.03.023.
- 675
- [51] M. Leming, J. M. Górriz, J. Suckling, Ensemble Deep Learning on large, mixed-site fMRI datasets in autism and other tasks, *International Journal of Neural Systems* 30 (07) (2020) 2050012. doi:10.1142/S0129065720500124.
- 680
- [52] L. Deng, J. C. Platt, Ensemble deep learning for speech recognition, in: *Proc. Interspeech 2014*, 2014, pp. 1915–1919. doi:10.21437/Interspeech.2014-433.
- [53] A. Khoder, F. Dornaika, Ensemble learning via feature selection and multiple transformed subsets: Application to image classification, *Applied Soft Computing* 113 (2021) 108006. doi:https://doi.org/10.1016/j.asoc.2021.108006.
- 685

- [54] A. O. Abuassba, Z. Dezheng, H. Ali, F. Zhang, K. Ali, Classification with ensembles and case study on functional magnetic resonance imaging, *Digital Communications and Networks* 8 (1) (2022) 80–86. doi:<https://doi.org/10.1016/j.dcan.2021.03.004>.
690 URL <https://www.sciencedirect.com/science/article/pii/S2352864821000158>
- [55] J. Cheng, Y. Xu, L. Kong, Hyperspectral imaging classification based on LBP feature extraction and multimodel ensemble learning, *Computers & Electrical Engineering* 92 (2021) 107199. doi:<https://doi.org/10.1016/j.compeleceng.2021.107199>.
695
- [56] J. E. Arco, P. Díaz-Gutiérrez, J. Ramírez, M. Ruz, Atlas-based classification algorithms for identification of informative brain regions in fMRI, *Neuroinformatics* 18 (2019) 219–236.
- [57] M. Joliot, G. Jobard, M. Naveau, N. Delcroix, L. Petit, L. Zago, F. Crivello, E. Mellet, B. Mazoyer, N. Tzourio-Mazoyer, Aicha: An atlas of intrinsic connectivity of homotopic areas, *Journal of Neuroscience Methods* 254 (2015) 46–59. doi:<https://doi.org/10.1016/j.jneumeth.2015.07.013>.
700
- [58] L. Fan, H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A. R. Laird, P. T. Fox, S. B. Eickhoff, C. Yu, T. Jiang, The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture, *Cerebral Cortex* 26 (8) (2016) 3508–3526. doi:[10.1093/cercor/bhw157](https://doi.org/10.1093/cercor/bhw157).
705
- [59] J. BROMLEY, J. W. BENTZ, L. BOTTOU, I. GUYON, Y. LECUN, C. MOORE, E. SÄCKINGER, R. SHAH, Signature verification using a “siamese” time delay neural network, *International Journal of Pattern Recognition and Artificial Intelligence* 07 (04) (1993) 669–688. doi:[10.1142/S0218001493000339](https://doi.org/10.1142/S0218001493000339).
710
- [60] D. Chicco, Siamese neural networks: An overview, *Methods in molecular biology* 2190 (2021) 73–94.
715

- [61] K. Crammer, Y. Singer, On the algorithmic implementation of multiclass kernel-based vector machines, *J. Mach. Learn. Res.* 2 (2002) 265–292.
- [62] B. Boser, I. Guyon, V. Vapnik, A training algorithm for optimal margin classifier, *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory* 5 (08 1996). doi:10.1145/130385.130401.
- [63] J. Schrouff, J. M. Monteiro, L. C. L. Portugal, M. J. Rosa, C. Phillips, J. M. Miranda, Embedding anatomical or functional knowledge in whole-brain multiple kernel learning models, *Neuroinformatics* 16 (2017) 117 – 143.
- [64] T. A. D. N. I. (ADNI), Ethics (2022).
URL <https://adni.loni.usc.edu>
- [65] Wellcome Centre for Human Neuroimaging, Statistical Parametrical Mapping, <https://www.fil.ion.ucl.ac.uk/spm/software/spm12> (2018).
- [66] I. A. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. M. López, F. Segovia, R. Chaves, M. Gómez-Rio, C. G. Puntonet, 18F-FDG PET imaging analysis for computer aided Alzheimer’s diagnosis, *Inf. Sci.* 181 (4) (2011) 903–916. doi:10.1016/j.ins.2010.10.027.
- [67] A. Ortiz, F. Lozano, J. Górriz, J. Ramirez, F. Martinez-Murcia, A. D. N. Initiative, Discriminative sparse features for Alzheimer’s disease diagnosis using multimodal image data, *Current Alzheimer Research* 15 (1) (2018) 67–79. doi:10.2174/1567205014666170922101135.
- [68] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI’95*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995, p. 1137–1143.
- [69] J. N. Mandrekar, Receiver operating characteristic curve in diagnostic test assessment, *Journal of Thoracic Oncology* 5 (9) (2010) 1315 – 1316. doi: <https://doi.org/10.1097/JTO.0b013e3181ec173d>.

- [70] K. Hajian-Tilaki, Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation, *Caspian journal of internal medicine* 4 (2013) 627–635.
- [71] P. Golland, B. Fischl, Permutation tests for classification: Towards statistical significance in image-based studies, in: C. Taylor, J. A. Noble (Eds.), *Information Processing in Medical Imaging*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 330–341.
- [72] Y. Yao, L. Rosasco, A. Caponnetto, On early stopping in gradient descent learning, *Constructive Approximation* 26 (2007) 289–315.
- [73] A. Ortiz, J. Munilla, J. M. Górriz, J. Ramírez, Ensembles of deep learning architectures for the early diagnosis of the Alzheimer’s disease, *International Journal of Neural Systems* 26 (07) (2016) 1650025. doi:10.1142/S0129065716500258.
- [74] R. Sharma, T. Goel, M. Tanveer, R. Murugan, FDN-ADNet: Fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer’s disease using the sagittal plane of MRI scans, *Applied Soft Computing* 115 (2022) 108099. doi:<https://doi.org/10.1016/j.asoc.2021.108099>.
- [75] R. Han, Z. Liu, C. P. Chen, Multi-scale 3D convolution feature-based broad learning system for Alzheimer’s disease diagnosis via MRI images, *Applied Soft Computing* 120 (2022) 108660. doi:<https://doi.org/10.1016/j.asoc.2022.108660>.
- [76] D. Chitradevi, S. Prabha, Analysis of brain sub regions using optimization techniques and deep learning method in Alzheimer disease, *Applied Soft Computing* 86 (2020) 105857. doi:<https://doi.org/10.1016/j.asoc.2019.105857>.
- [77] R. Sharma, T. Goel, M. Tanveer, S. Dwivedi, R. Murugan, FAF-DRVFL: Fuzzy activation function based deep random vector functional links net-

- work for early diagnosis of Alzheimer disease, *Applied Soft Computing* 106 (2021) 107371. doi:<https://doi.org/10.1016/j.asoc.2021.107371>.
- [78] P. M. Tuan, T.-L. Phan, M. Adel, E. Guedj, N. L. Trung, Autoencoder-based feature ranking for alzheimer disease classification using PET image, *Machine Learning with Applications* 6 (2021) 100184. doi:<https://doi.org/10.1016/j.mlwa.2021.100184>.
775
- [79] K. M. Poloni, I. A. Duarte de Oliveira, R. Tam, R. J. Ferrari, Brain MR image classification for Alzheimer’s disease diagnosis using structural hippocampal asymmetrical attributes from directional 3-D log-Gabor filter responses, *Neurocomputing* 419 (2021) 126–135. doi:<https://doi.org/10.1016/j.neucom.2020.07.102>.
780
- [80] T. M. Schouten, M. Koini, F. de Vos, S. Seiler, M. de Rooij, A. Lechner, R. Schmidt, M. van den Heuvel, J. van der Grond, S. A. Rombouts, Individual classification of Alzheimer’s disease with diffusion magnetic resonance imaging, *NeuroImage* 152 (2017) 476–481. doi:<https://doi.org/10.1016/j.neuroimage.2017.03.025>.
785
- [81] K. Hu, Y. Wang, K. Chen, L. Hou, X. Zhang, Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis, *Neurocomputing* 175 (2016) 132–145. doi:<https://doi.org/10.1016/j.neucom.2015.10.043>.
790
- [82] J. Górriz, F. Segovia, J. Ramírez, A. Lassl, D. Salas-Gonzalez, GMM based SPECT image classification for the diagnosis of Alzheimer’s disease, *Applied Soft Computing* 11 (2) (2011) 2313–2325. doi:<https://doi.org/10.1016/j.asoc.2010.08.012>.
- [83] M. Liu, D. Cheng, W. Yan, A. D. N. I. , Classification of Alzheimer’s disease by combination of convolutional and recurrent neural networks using FDG-PET images, *Frontiers in Neuroinformatics* 12 (2018). doi:[10.3389/fninf.2018.00035](https://doi.org/10.3389/fninf.2018.00035).
795

- [84] E. Hussain, M. Hasan, S. Hassan, T. Azmi, M. Rahman, M. Parvez, Deep
800 learning based binary classification for Alzheimer’s disease detection using
brain MRI images, in: 2020 15th IEEE Conference on Industrial Electron-
ics and Applications (ICIEA), 2020, pp. 1115–1120, 15th IEEE Confer-
ence on Industrial Electronics and Applications : ICIEA 2020 ; Conference
date: 09-11-2020 Through 13-11-2020. doi:10.1109/ICIEA48937.2020.
805 9248213.
- [85] E. Kaplan, S. Dogan, T. Tuncer, M. Baygin, E. Altunisik, Feed-forward
LPQNet based automatic Alzheimer’s disease detection model, *Computers
in Biology and Medicine* 137 (2021) 104828. doi:https://doi.org/10.
1016/j.combiomed.2021.104828.
- 810 [86] N. Goenka, S. Tiwari, Alzvnet: A volumetric convolutional neural net-
work for multiclass classification of alzheimer’s disease through multiple
neuroimaging computational approaches, *Biomedical Signal Processing and
Control* 74 (2022) 103500. doi:https://doi.org/10.1016/j.bspc.2022.
103500.
- 815 [87] Y. Shi, C. Zu, M. Hong, L. Zhou, L. Wang, X. Wu, J. Zhou, D. Zhang,
Y. Wang, ASMFS: Adaptive-similarity-based multi-modality feature selec-
tion for classification of Alzheimer’s disease, *Pattern Recognition* 126 (2022)
108566. doi:https://doi.org/10.1016/j.patcog.2022.108566.
- [88] A. Cajanus, E. Solje, J. Koikkalainen, J. Lötjönen, N.-M. Suhonen, I. Hal-
820 likainen, R. Vanninen, P. Hartikainen, M. de Marco, A. Venneri, H. Soini-
nen, A. M. Remes, A. Hall, The association between distinct frontal
brain volumes and behavioral symptoms in mild cognitive impairment,
Alzheimer’s disease, and frontotemporal dementia, *Frontiers in Neurology*
10 (2019). doi:10.3389/fneur.2019.01059.
- 825 [89] P. T. Nelson, D. W. Dickson, J. Q. Trojanowski, C. R. Jack, P. A.
Boyle, K. Arfanakis, R. Rademakers, I. Alafuzoff, J. Attems, C. Brayne,
I. T. S. Coyle-Gilchrist, H. C. Chui, D. W. Fardo, M. E. Flanagan,

- G. Halliday, S. R. K. Hokkanen, S. Hunter, G. A. Jicha, Y. Katsumata, C. H. Kawas, C. D. Keene, G. G. Kovacs, W. A. Kukull, A. I. Levey, N. Makinejad, T. J. Montine, S. Murayama, M. E. Murray, S. Nag, R. A. Rissman, W. W. Seeley, R. A. Sperling, C. L. White III, L. Yu, J. A. Schneider, Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report, *Brain* 142 (6) (2019) 1503–1527. doi:10.1093/brain/awz099.
- 830
- [90] W. Jaroudi, J. Garami, S. Garrido, M. Hornberger, S. Keri, A. A. Moustafa, Factors underlying cognitive decline in old age and Alzheimer’s disease: the role of the hippocampus, *Reviews in the Neurosciences* 28 (7) (2017) 705–714. doi:doi:10.1515/revneuro-2016-0086.
- 835
- [91] E. Yu, Z. Liao, D. Mao, Q. Zhang, G.-J. Ji, Y. Li, Z. Ding, Directed functional connectivity of posterior cingulate cortex and whole brain in Alzheimer’s disease and mild cognitive impairment, *Current Alzheimer research* 13 (12 2016). doi:10.2174/1567205013666161201201000.
- 840
- [92] F. Novellino, M. E. López, M. G. Vaccaro, Y. Miguel, M. L. Delgado, F. Maestu, Association between hippocampus, thalamus, and caudate in mild cognitive impairment APOE ϵ 4 carriers: A structural covariance MRI study, *Frontiers in Neurology* 10 (2019). doi:10.3389/fneur.2019.01303.
- 845
- [93] D. W. Kang, S.-M. Wang, Y. H. Um, H.-R. Na, N.-Y. Kim, C. U. Lee, H. K. Lim, Distinctive association of the functional connectivity of the posterior cingulate cortex on memory performances in early and late amnesic mild cognitive impairment patients, *Frontiers in Aging Neuroscience* 13 (2021). doi:10.3389/fnagi.2021.696735.
- 850
- [94] G. J. Schnellbacher, F. Hoffstaedter, S. B. Eickhoff, S. Caspers, T. Nickl-Jockschat, P. T. Fox, A. R. Laird, J. B. Schulz, K. Reetz, I. Dogan, Functional characterization of atrophy patterns related to cognitive impairment, *Frontiers in Neurology* 11 (2020). doi:10.3389/fneur.2020.00018.
- 855

- [95] H. Liu, D. Liu, K. Li, X. Xue, X. Ma, Q. Bu, J. Ma, Z. Pan, L. Zhou, Microstructural changes in the cingulate gyrus of patients with mild cognitive impairment induced by cerebral small vessel disease, *Neurological Research* 43 (2021) 1–9. doi:10.1080/01616412.2021.1910903.
- 860 [96] A. Cajanus, E. Solje, J. Koikkalainen, J. Lötjönen, N.-M. Suhonen, I. Hallikainen, R. Vanninen, P. Hartikainen, M. De Marco, A. Venneri, H. Soininen, A. Remes, A. Hall, The association between distinct frontal brain volumes and behavioral symptoms in mild cognitive impairment, Alzheimer’s disease, and frontotemporal dementia, *Frontiers in Neurology* 10 (10 2019).
865 doi:10.3389/fneur.2019.01059.