POLITECNICO DI MILANO
School of Industrial and Information Engineering
Department of Electronics, Information and Bioengineering
Master of Science Degree in Computer Science and Engineering

Deep Generative Models for Predicting Alzheimer's Disease Progression from MR Data

AI & R Lab
Artificial Intelligence and Robotics Lab

Supervisor: Prof. Marcello Restelli
Co-supervisors: Dr. Abhijit Guha Roy, Dr. Federico Tombari

Master's Thesis by:
Diletta Milana, 850182

Academic Year 2017- 2018

To the fools who dream.

Abstract

Aim of this work is to predict the progression of Alzheimer's Disease (AD) in Structural Magnetic Resonance Imaging (sMRI) using Deep Generative methods. To the best of our knowledge, this is the first attempt to generate this progression using Deep Learning methods.

Alzheimer's disease is the most common cause of dementia worldwide, and this tendency is predicted to become even more marked in the next years, due the global aging of the population. While several therapies are currently being studied, they are mostly applied when patients experience the first symptoms of cognitive impairment, indicating that the disease is already in an advanced stage. A robust model able to predict the development of the disease and its influence on specific regions of the brain would guarantee higher chances to slow down, stop or even prevent the disease.

We use sMRI as the input data for this study for its being relatively cheap and non-invasive for the patient. We focus on the key regions for AD, namely hippocampus and ventricles, by extracting slices as well as 3D organs using a Fully Convolutional segmentation network. Both the 2D slices and the 3D shapes are then analysed using Convolutional Variational Autoencoders (CVAE) and Conditional Adversarial Autoencoders (CAA), integrating both supervised and unsupervised approaches. The Convolutional Variational Autoencoder is used to learn a manifold representation that encodes the most distinctive brain features and can be walked to progress them in terms of shape, size and morphological characteristics. The Conditional Adversarial Autoencoder, an integration of Autoencoders and Generative Adversarial Networks, is used to generate a progression in time of the input brain volumes. This progression is evaluated qualitatively on 2D slices and quantitatively on 3D shapes, in the latter case showing a statistically significant decrease in the shape of the hippocampus over time, that is more evident in AD subjects as opposed to NC, as confirmed by literature.

Estratto in Lingua Italiana

Obiettivo di questo lavoro di tesi è prevedere la progressione della malattia di Alzheimer in immagini di risonanza magnetica strutturale (sMRI) utilizzando modelli generativi *deep*. Al meglio della nostra conoscenza, questo è il primo tentativo di generare una progressione della malattia utilizzando metodi di Deep Learning.

La malattia di Alzheimer è la più comune causa di demenza, una tendenza destinata a diventare sempre più marcata nei prossimi anni per via dell'invecchiamento globale della popolazione. Diverse terapie sono attualmente in corso di sperimentazione, ma sono quasi sempre applicate a pazienti che mostrano già i primi sintomi di declino cognitivo, ovvero nei quali la malattia è ormai presente in stato avanzato. Un modello robusto in grado di predire lo sviluppo della malattia e la sua influenza su regioni specifiche del cervello potrebbe permettere di rallentare, fermare o prevenire la malattia.

Il problema di classificare una risonanza magnetica come Alzheimer (AD), Mild Cognitive Impariment (MCI) oppure Controllo Normale (NC) è già stato affrontato approfonditamente e con successo con tecniche di Machine Learning. Recentemente, anche metodi di Deep Learning sono stati utilizzati per gli stessi obiettivi, grazie alla loro abilità di produrre rappresentazioni di *feature* gerarchiche. D'altra parte, queste metodologie non sono ancora state applicate alla generazione di una progressione (o regressione) nel tempo a partire da un'immagine di risonanza magnetica per lo studio della malattia di Alzheimer.

La risonanza magnetica strutturale è stata scelta per questo studio poichè è meno costosa e meno invasiva per il paziente rispetto ad altri esami. Si può immaginare quindi che nel futuro questo tipo di esame possa essere utilizzato con maggiore frequenza, soprattutto nei pazienti con una marcata predisposizione alla malattia. In questo modo, potrebbe essere possibile tracciarne

lo sviluppo sin dalle origini, e agire in tempo con efficaci terapie.

Questo lavoro può essere diviso in due parti, in base alle differenti tecniche di pre-processing applicate ai dati. Un primo approccio utilizza *slice* 2D: vengono testate diverse tecniche per ottenere le *slice* più rilevanti dal punto di vista della diagnosi della malattia, poi estratte lungo i tre assi principali (assiale, sagittale, coronale). Poichè forme e volumi sono di importanza chiave per la diagnosi della malattia, si utilizza poi una tecnica di segmentazione Fully Convolutional per estrarre dall'intero volume cranico soltanto gli organi chiave, come ippocampo e ventricoli, sotto forma di mappe binarie 3D.

Le *slice* 2D e gli organi 3D sono poi analizzati utilizzando modelli generativi, per la loro capacità di apprendere rappresentazioni latenti che codifichino le feature più rilevanti. I metodi utilizzati sono Autoencoder Convoluzionali Variazionali (CVAE) e Autonecoder Condizionali Antagonisti (CAAE), integrando approcci supervisionati e non supervisionati.

Un CVAE viene allenato con l'obiettivo di imparare un manifold che incorpori le caratteristiche distintive della malattia, e che possa essere attraversato per mostrare l'evoluzione del cervello in termini di forma, dimensione e caratteristiche morfologiche. Dopo essere stato inizializzato con un allenamento non supervisionato, il modello viene quindi rifinito in maniera supervisionata, dimostrando così una correlazione tra classificazione e variabili latenti.

Un CAAE è stato poi utilizzando per generare una progressione nel tempo. Questo approccio integra l'autoencoder standard con una rete generativa avversaria, permettendo alla rete di apprendere un *manifold* che includa le feature più rilevanti estratte dai dati, e di generare immagini molto realistiche partendo da queste. Anche in questo caso, l'aggiunta di supervisione ha permesso alla rete di apprendere il legame tra immagini, età e classificazione, che ha poi permesso di generare e visualizzare una progressione nel tempo. La progressione è stata poi valutata qualitativamente nelle slice 2D e quantitavamente nelle forme 3D. In quest'ultimo caso la rete ha prodotto una decrescita statisticamente rilevante nella dimensione dell'ippocampo, risultato ancora più evidente nel caso di diagnosi AD, in maniera conforme alla letteratura precedente.

Contents

List of Figures

Acronyms

AD Alzheimer's Disease. 9, 15, 58, 62

AIBL Australian Imaging, Biomarker & Lifestyle flagship study of ageing. 58

CAAE Conditional Adversarial AutoEncoder. 11, 67, 72, 74

CSF CerebroSpinal Fluid. 19, 20, 45

CVAE Convolutional Variational AutoEncoder. 10, 67, 70, 74

GAN Generative Adversarial Network. 32

GM Grey Matter. 20, 45

HARP HARmonized Protocol. 58

KL Kullback-Leibler. 31

MCI Mild Cognitive Impairment. 9, 22, 58, 62

MMSE Mini-Mental State Examination. 10, 22, 62

NC Normal Control. 9, 58, 62

OASIS Open Access Series of Imaging Studies. 58

ReLU Rectifier Linear Unit. 35

sMRI Structural Magnetic Resonance Imaging. 19

SVM Support Vector Machine. 27

VAE Variational AutoEncoder. 32

WM White Matter. 20, 45

Acknowledgements

I would like to express my deep gratitude to Professor Marcello Restelli, for his continuous support, guidance, patience and trust.

From Technische Universität München, I am grateful to Professor Nassir Navab for the great opportunity of doing research at the chair of Computer Aided Medical Procedures. Special thanks to Dr. Federico Tombari for his perpetual availability, precious advice and vision and to Dr. Abhijit Guha Roy for his support and patience during these challenging months. Thank you for trying to teach me MATLAB and proper handwriting: I reached miserable results, but I promise I will not give up. I would also like to thank Yida Wang and Huseyin Coskun for their helpful hints.

Many thanks to Dr. Edoardo Barvas, Dr. Susanna Guttmann and Dr. Michele Sintini from San Marino Hospital, as well as Dr. Massimo Venturelli, from the University of Verona, for finding the time to provide us their support from the medical side. They have always shown forward thinking, open minds and pure interest towards the applications of Machine Learning to the study of Alzheimers disease, and I am extremely grateful for that.

Thank you to professor Elena Baralis from Politecnico di Torino for her support and interest.

A very warm thank you goes also to my friend Alessio Sollima, whose encouragements as well as accurate insights have constantly brought new energy to this thesis. I expect from him no less than a bright future as a doctor.

To the friends who have shared this journey with me, from Italy and all over the world: you made me miss home, and you made me feel at home away from home.

Finally, to my family, for introducing me to my first robot, Mr. Emilio, at age 3 and for capturing my bewildered reaction. But most of all, thank you for being the bravest and most relentless fighters I know.

Chapter 1

Introduction

The aim of this thesis is to predict the progression of Alzheimer's Disease in Structural Magnetic Resonance Imaging using Deep Generative methods. To the best of our knowledge, this is the first attempt to generate the progression of the disease using deep learning methods.

Alzheimer's disease is the most common cause of dementia worldwide, and this tendency is predicted to become even more marked in the next years, with the global aging of the population [16]. While several therapies are currently being studied, they are mostly tested when patients experience the first symptoms of cognitive impairment, indicating that the disease is already in an advanced stage. A robust model able to predict the development of the disease and its influence on specific regions of the brain would guarantee higher chances to slow down, stop or even prevent the disease.

The problem of classifying a brain scan into Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI) and Normal Control (NC) has already been tackled extensively and quite successfully by the machine learning community [9]. More recently, deep learning methods have also been applied to classification and feature extraction tasks, making extensive use of their ability to produce hierarchical feature representations. On the other hand, the idea of generating a progression (or regression) in time starting from an input image was never successfully applied to MRI scans for the study of AD, as opposed to other imaging fields [7, 8].

For this study, about 950 sMRI volumes have been gathered. Structural MRI is targeted as the input data type for its being relatively cheap and non-invasive for the patient. It is thus imaginable, in the future, to use this

type of exam in a more established pipeline, where patients that show a predisposition towards this type of dementia can undergo frequent examining sessions since a young age, in order to track effectively the development of the disease and act early on with proper treatments.

It must be noted however, that the progression of Alzheimer's disease depends on a number of different factors including, but not limited to, genetics, cardiovascular conditions, morphological characteristics and in general personal health history [17]. However, the sMRI scans used for this study are only accompanied by age, gender and Mini-Mental State Examination (MMSE) scores, which makes the problem of generating a progression in time even more challenging.

This work can be divided into two parallel tracks, based on different pre-processing techniques applied to the input volumes.

A first approach leverages 2D brain slices. Several techniques were attempted to find those slices that are most relevant with respect to the disease diagnosis, and to extract them along the three main axes (axial, sagittal, coronal). Following this approach, we extract slices from the centre of the brain and from those areas where the hippocampus, a key element in AD diagnosis [14], is maximally present.

Since the importance of shapes and volumes is widely recognised in order to understand the main characteristics of AD-affected brains, we focus also on the segmentation of sMRI scans using a Fully Convolutional neural network, resulting in 3D binary maps of the most relevant organs for AD diagnosis (such as hippocampus and ventricles).

We use these 2D slices and 3D shapes as inputs to generative models, for their ability to learn meaningful latent representations that encode the most relevant features, and to generate images with certain characteristics, or belonging to particular classes. To this end, we use a Convolutional Variational AutoEncoder (CVAE) and Conditional Adversarial AutoEncoder (CAAE). These models integrate both supervised and unsupervised approaches.

The CVAE [18] is able to learn a manifold that encoded the most distinctive brain features and be *walked* to progress them in terms of shape, size and morphological characteristics. This latent representation however is still unable to cluster the subjects into the three distinct classes based on

their diagnosis. For what concerns 2D slices, this behaviour can be explained from a theoretical standpoint considering the limitations of the automatic extraction of slices; from a 3D perspective, it must be noted that it is extremely hard, even for an expert neuroradiologist, to discriminate a diseased patient from a healthy one simply by analysing single shapes. When aided by the injection of supervised information, the CVAE is able to learn a more meaningful organization of the latent space, where labels and latent variables are correlated.

In order to produce a progression in time of the subjects, a CAAE [8] is implemented. This approach integrates the standard Autoencoder [19] with Generative Adversarial Networks [20]. Once again, the network first learns a manifold that can encode the most relevant features in the data, then learns to generate very realistic images starting from these encodings. In this case too, supervised information helps the network understand the link between imaging input, age and classification. This way, the network is able draw a progression in time, which is evaluated qualitatively on 2D slices and quantitatively on 3D shapes. For what concerns 2D slices, the network generates a progression where global cerebral volume is decreasing, grey matter becomes more marked and ventricles enlarged, but it still lacks general consistency in terms of individual, morphological characteristics and overall coherency in the progression. Results on 3D shapes are more encouraging: we obtain a statistically significant decrease in the shape of the hippocampus over time, that is more marked for AD patients with respect to NC subjects, which is confirmed by literature [14].

1.1 Overview

This work was done in collaboration with the chair for Computer Aided Medical Procedures at Technische Universität München, under the supervision of Dr. Federico Tombari and Dr. Abhijit Guha Roy. The remaining chapters are organised as follows.

Chapter 2 provides a brief overview of Alzheimer's Disease, clarifying its psychological as well as socio-economical impact, its neurological characteristics and the urgent need for a progression model.

Chapter 3 discusses the state-of-the-art Machine Learning and Deep Learning approaches for computer vision tasks, and in particular for Alzheimer's Disease.

Chapter 4 introduces the datasets used, explaining the characteristics of both images and labels.

Chapter 5 introduces the methodologies used for this analysis: Convolutional Variational Autoencoders and Conditional Adversarial Autoencoders. These architectures implemented following two different approaches: 2D and 3D. For each of them, we will also detail the techniques used to extract the slices and shapes from the original scans.

Chapter 6 is devoted to the results of the experiments carried out on the aforementioned methodologies. For each of them, we discuss the reconstruction capabilities of the autoencoders, the generated progression, and where useful additional metrics: accuracy of the supervised fine-tuning, dice score, clustering of the latent spaced (reduced by t-SNE [21]).

In Chapter 7 we draw relevant conclusions concerning the use of these methodologies for the study of Alzheimer's Disease progression, pointing out their contribution and mentioning possibile future improvements.

Chapter 2

Alzheimer's Disease

In this chapter, a brief introduction to Alzheimer's disease is presented. Starting with a general overview of its psychological as well as socio-economic impact, we will then focus on how the disease affects the brain, which features distinguish it from other pathologies, how it is diagnosed and why it is important to build a robust model able to predict its progression.

2.1 The pathological picture

AD is a degenerative brain disease and is the most common cause of dementia. This is mostly due to the worldwide aging of the population (Table 2.1), as the number of AD subjects older than 65 years of age is roughly expected to double between 1197 and 2050 [16]. Dementia affects about 14% of individuals of age 71 or older in the US and Alzheimer's is responsible for over 70% of all dementia cases [17].

Apart from its devastating effect on the diseased and their families, AD also represents an enormous economical burden for countries worldwide. In Europe for instance, about 7 Million individuals have been diagnosed with the disease, resulting in an approximate cost of 22000 €per patient per year. AD is the only disease amongst the 10 leading causes of deaths that is still incurable and unpreventable [16].

	65-74	75-84	>85
Individuals affected the Alzheimer's dementia	3%	17%	32%

Table 2.1: Alzheimer's Disease correlates with age (from [17]).

2.1.1 What Happens in the Brain

The causes behind the development of Alzheimer's disease are still debated, but useful insights can be found in [22] and [16]. According to most neuroscientists, the molecular causes of the disease take place at the point of connection between neurons, the *synapses*: this is where neurotransmitters are released. During ordinary inter-cell communication, neurons also release a small peptide called *amyloid-β*. This is normally cleared away and metabolised by *microglia* cells. Sometimes however, these peptides start to accumulate, until they form aggregates, or *amyloid plaques*, that tend to obstruct these synapses, thus interfering with the ordinary flow of information (Figure 2.1). While these aggregates appear with normal aging too, the ones triggering Alzheimer's start to appear 20 years before the clinical symptoms [16]. During what is called the *preclinical stage*, they tend to be more in number and to follow predictable patterns, starting from memory-related areas only to spread through other regions later.

When the presence of amyloid plaques reaches the tipping point, microglia cells become hyper-activated and attempt to overcome this obstruction by releasing chemicals, causing inflammation, cellular death and synapse destruction in return. These plaques, together with the tangles formed by a neural transport protein called *tau*, are believed to have a key role in blocking communication between neurons, preventing them from receiving the resources they need to survive, and eventually causing the symptoms of AD.

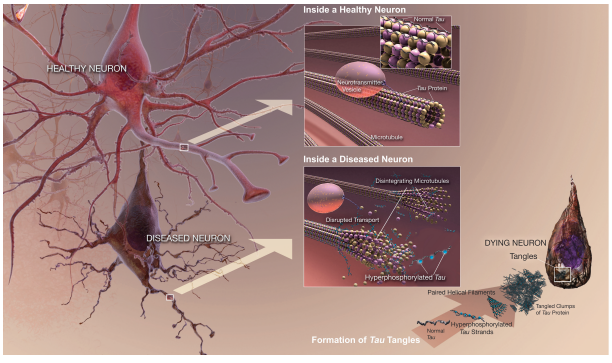


Figure 2.1: A visual comparison between healthy and diseased neurons (photo credit: National Institute on Aging/National Institutes of Health).

2.1.2 Drug Discovery

Current drug discovery is therefore focused on preventive medicine, in hopes to find a compound that will prevent or reduce amyloid plaque accumulation. Some of the reasons why most of the clinical trials to date have failed are mentioned in [16], and they include:

1. The long time period needed to check whether the treatment had any effect on the evolution of the disease;
2. The structure of the brain itself, which allows only very specialised small-molecule drugs to penetrate;
3. Most of all, the fact that these drugs were tested on patients that were already experiencing the cognitive symptoms of the disease: the plaques had already invaded the synapses, and the inflammation had already taken over.

The key to any effective treatment is thus detecting the development of the disease way before the tipping point is reached. This is why a robust progression model would make a strong difference in how (and how soon) new effective therapies are found and tested.

In the meantime, several non-pharmacologic therapies with the goal of reducing the behavioral symptoms are currently under trial.

2.1.3 Modifiable Risk Factors

It is true that no disease-modifying treatments have been discovered so far, and there are some important risk factors that are non-modifiable, such as genetic factors, low educational or occupational involvement, family history or traumatic brain injuries [17]. However, it is now very well known that there also exist several modifiable risk factors [16]: things that can be done to prevent the onset of the disease.

While traditionally AD was not linked to vascular dementia (such as stroke or similar), it was recently discovered that vascular difficulties might indeed play an important role in the disease [23]. Many of the risk factors for AD are recognised as risk factors for cardiovascular diseases too, and in general vascular function has been shown to be related to cognitive function. Moreover, it was found that strength training might mitigate the neural decline and improve neuroplasticity [24].

In addition to keeping cardiovascular risks under control and carrying out frequent and active exercise [25, 26], it is also very important to get enough quality sleep [27] and have a well-balanced and healthy diet. Moreover, it is of utmost importance to keep the brain active: promoting the birth of new synapses by learning new concepts, participating in discussions and keeping an active social life, reading books or even learning new languages[28], all contribute to the creation of the *cognitive reserve* that enables the brain to find alternative routes to replace its lost connections.

2.2 Diagnosis

The diagnosis of Alzheimer's Disease is carried out according to a number of factors. The current diagnostic standard for dementia in general is the Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition), and it distinguishes between two progressive phases of the disease: a mild, and a major neurocognitive impairment. However, other diagnostic criteria were developed by the National Institute on Aging (NIA) and the Alzheimer's Association (AA). These criteria for the first time acknowledged the diagnostic relevance of biomarkers:

1. Mesial temporal lobe atrophy on sMRI [17];
2. Posterior predominant hypometabolism on FluroDeoxyGlucose Positron Emission Tomography (FDG-PET) [17];
3. Levels of β-amyloid and tau in the cerebrospinal fluid and levels of certain groups of proteins in the blood.

While these biomarkers have the requirements of sensitivity, specificity and pathologic validity, the first two present a strong drawback: they cannot be uniquely associated to Alzheimer's Disease, as they are often also associated with normal aging [14, 29]. Moreover, FDG-PET abnormalities as well as hippocampal shrinkage are also associated to other types of dementia [30, 31]. The biomarkers that proved to be most successful at detect AD are amyloid-based ones: low CerebroSpinal Fluid (CSF) level of β-amyloid or positive amyloid PET scans [32, 33] are strong indicators of the disease.

The study of the progression of these biomarkers is of paramount importance because they:

1. Could identify the disease in its early stages, when there is still time for potential treatments to block the progression of the disease;

2. Allow potentially successful treatments to be tested and tracked early on, when the effects of the disease are not permanent yet;

3. Allow potential treatments to be tested on patients that show those brain changes that are specifically targeted by the treatment at hand.

In addition to these biomarkers, the complete diagnosis takes into account other factors, including:

1. Complete medical family history;

2. Cognitive conditions;

3. Physical and neurological exam.

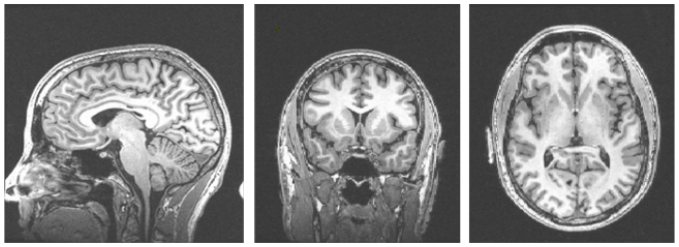
It must also be noted that complete certainty on the diagnosis of Alzheimer's disease can only be obtained with a histological exam *post mortem*.

2.2.1 The role of sMRI

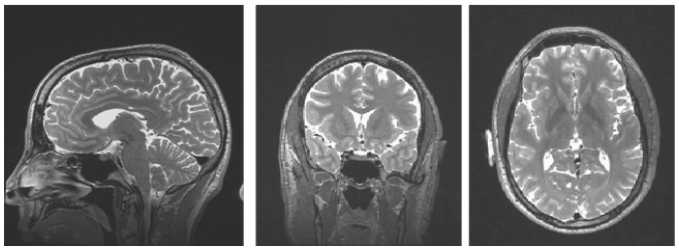
Structural Magnetic Resonance Imaging (sMRI) is therefore an important part of the diagnostic procedure. It provides static, anatomical information on the shape, size, and integrity of most structures of the brain [1]. Depending on the Repetition Time (TR) and Echo Time (TE) of the signal there are two main types of MRI:

- T1-weighted: short TR and short TE, shown in Figure 2.2a. Shows good contrast between Grey Matter (GM) shown in dark gray and White Matter (WM) shown in lighter gray. CSF is void of signal, shown in black;

- T2-weighted: long TR and long TE, shown in Figure 2.2b. Shows good contrast between CSF (bright) and brain tissue (dark).



(a) T1



(b) T2

Figure 2.2: MRI modalities (from [1]).

The importance of this examination in the workup of patients with dementia is two-fold.
First of all, it is used as one of the first steps when dealing with patients suffering from cognitive impairment in order to exclude possible nondegenerative lesions such as a slow-growing brain, subdural hematoma or tumors [34]. In these cases, an MRI would highlight ischemic changes that would point to further analysis or therapeutic treatment towards vascular risk or behavioral modifications [17].

Second of all, MRI can be used to provide hints on the possible cause of the patient's discomfort: as mentioned before, mesial temporal atrophy [35, 36], global brain atrophy or ventricle enlargement [37, 38], can act as a hint or a confirmation of the presence of Alzheimer's. When studying the effects of the disease on the brain, it is therefore important to focus on hippocampus and ventricles [14, 17, 39].

2.2.2 Mild Cognitive Impairment

In this context, MCI indicates a patient's condition characterised by mild yet noticeable cognitive difficulties that do not affect his or her ability to carry out daily activities independently [22]. The subject's cognitive abilities are usually evaluated using specific tests such as MMSE, which will also be used for the rest of this work (Section 4.2.2), the Rey Auditory Verbal Learning Test (RAVLT) and Clinical Dementia Rating Scale Sum of Boxes Scores (CDRSb).

The progression from MCI to Alzheimer's disease was often studied: an average 32% of individuals with MCI have developed Alzheimer's disease within a five-year time period in [40], and a meta-analysis of 41 studies showed that 38% of patients with MCI eventually developed dementia [41]. However, as studies have shown [16], the diagnosis of MCI involves a degree of subjectivity that may result in different conclusions being drawn by different experts. This is due to multiple reasons [16]:

- MCI is sometimes mistaken for an early stage of Alzheimer's disease or other forms of dementia. In the latter case, MCI might originate from causes that differ sharply from those of Alzheimer's;
- MCI can be mistakenly diagnosed also if the patient is taking medications that cause cognitive impairment as a side effect;
- MCI may eventually revert to normal conditions or remain stable;
- Any cognitive evaluation is necessarily subjective and is strongly dependent on the patient's psychiatric conditions, as they can interfere with his/her degree of collaboration and his/her ability to provide meaningful answers (for instance, in case of depression).

This general overview of the AD covered the most important characteristics of the disease and its diagnosis, while clarifying its socio-economical impact and the importance of focusing on the study of its progression.

These insights will be of great importance for the selection of the right machine learning techniques, which will be detailed in the next chapters. Our study starts in the next Chapter, with an overview of the Machine Learning techniques used in related work for the study of AD.

Chapter 3

State of the Art

In this chapter, the most relevant works concerning the use of machine learning techniques for the analysis of images, and in particular those used for AD diagnosis, will be discussed. First, a general introduction of the techniques applied to common vision tasks will be provided. Later, applications to medical imaging and neuroimaging, with particular focus on AD, will be discussed.

3.1 Machine Learning for Computer Vision

Image classification, segmentation, object tracking and detection are core tasks in computer vision. The reason why they are particularly hard is the semantic gap existing between what humans easily perceive (dogs, cats, trees...) and what algorithms see (nothing more than multi-dimensional arrays of bytes). The performance of these algorithms is strongly dependent on slight differences in pose, position, occlusion, illumination, background clutter, and on the many different appearances for the same objects.

The application of machine learning brought huge advancements in this field [42, 43]: using a data-driven approach, these algorithms are able to learn and extract complex patterns from the data. With the introduction of deep learning, these features are not handcrafted anymore, but automatically and well-scalably extracted by machines instead.
Most of the aforementioned tasks have thus become extremely accurate, almost approaching human accuracy [44].

In this section, a general introduction of machine learning methods for com-

puter vision tasks will be provided. Many of these studies will tackle the problem of image classification: while this is not the main task tackled in this work, they still provide extremely useful insights on the state-of-the-art techniques currently used to extract relevant features from images.

3.1.1 Classification

Traditionally, the task of image classification has been applied using various techniques on standard datasets (MNIST, CIFAR-10, StreetView House Numbers and ImageNet to name a few).

K-nearest neighbors [42] for instance, memorizes all the training data at training time and, at test time, checks which of the memorized images is closer (in terms of a distance metric of choice) to the test image, and assigns the same label. This way, the algorithm is learning boundaries that can separate the data points, based on the training data only. The main problems with this approach are the reduced speed at test time, the difficulty in finding a distance calculation that is informative enough for the task at hand and the curse of dimensionality [45].

Linear regression [42] can also be used for image classification [45]. It consists of a parametric classifier whose knowledge is stored in the parameter matrix W. Each row of this matrix can be thought as a template for one class. The input images are therefore stretched along a column vector, and the classifier then assigns a score to all classes by comparing the respective template with the input image doing an inner (or dot) product. The class that fits best will receive the highest score. If we indicate the input column vector as x_i and the output vector as y_i we can write the model equation as:

$$y_i = W x_i$$

.

Linear classification is therefore still learning boundaries, but in a high-dimensional space.

Linear regression can be better generalized to classification problems by applying a sigmoid (in the case of a binary classification) or a softmax (for multiclass problems) function to its output, resulting in a logistic regression [42]. This saturates the output to either 0 or 1, making it suitable to represent a probability: we can thus interpret the output of the model as a probability that the input image belongs to that class [19].

Another frequently used method is that of Support Vector Machine (SVM) [46, 47]. SVMs classify using the same linear function shown above: $w^T x + b$. If the result is positive, the sample is assigned to the positive class, while if it's negative it is assigned to the negative class. The *kernel trick*, contributed to making SVMs not only much more computationally efficient, but also able to learn models that are non-linear in the inputs using convex optimisation techniques, that are guaranteed to converge efficiently [19].

SVMs are a very powerful model, that reached the state of the art in many tasks [43]. In fact, before of the advent of deep learning most of the studies focusing on the classification of Alzheimer's disease (discussed in Section 3.2) leveraged this technique. Actually, deep learning slowly began to take off when it was demonstrated [48] that a neural network was able to outperform SVMs on the MNIST benchmark [19].
The next paragraphs will focus precisely on the type of neural networks that are currently mostly used on images: Convolutional Neural Networks.

The idea of an architecture able to resemble and possibly replicate the human vision dates back to the concept of perceptron [49]. The first application of Convolutional Neural Networks dates back to the '90: Yann LeCun's famous *LeNet* was able to accurately classify handwritten digits [2] with thw architecture shown in Figure 3.1.

This architecture is made of layers of different types (Convolutional, Fully Connected, Max pooling...) that apply successive differentiable transformations to the input image. Most layers are characterized by parameters that are learned during the training procedure (backpropagation), which is repeated for several iterations (epochs).

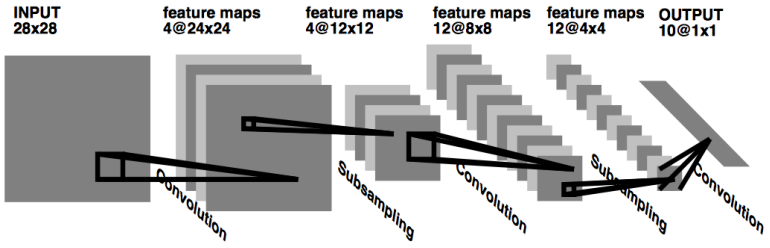


Figure 3.1: LeNet architecture (from [2]).

Distinctive layers in Convolutional Neural Networks are Convolutional layers, in which the parameters are a set of learnable filters [45]. These filters are of variable dimension in width and height, but they always extend through the full depth (represented by the channels) of the input. The neurons are thus connected only to local regions of the input, whose spatial extent is the filter size, often referred to as *receptive field*. These filters are, in fact, *convolved* across width and height of the input volume: this is basically a dot product between the filters and the input volume at every pixel. Sliding this filter along width and height of the image (and also depth, in the case of the 3D convolutions) results in activation maps that represent the response of every pixel to that filter. One filter corresponds to one activation map, and each of these filters will be activated when certain specific visual features are found in the input image. This can be very well seen in the AlexNet's activations (Figure 3.2), that focus especially on edges or blotches of colors, with various slopes. This will be especially of use in the next sections, when it will be compared with the convolutional activations applied to MRI scans.

Stacking these maps one after the other along the depth dimension produces the output tensor of the Convolutional layer. This tensor will be then fed into the next layer in the architecture, which used to be a Max pooling layer.

Other very frequently used layers in deep convolutional networks are:

1. Max pooling: the role of this layer is to reduce the size of the input tensor by selecting only the pixels with the highest activation. Recently, strided convolutions have started to be preferred instead [5];

2. Batch Normalisation [50]: addressing the problem of internal covariate shift by normalising the layer inputs at each mini-batch (thus acting as a regulariser);

3. Dropout [51]: addressing the problem of overfitting by randomly dropping neurons with probability p during the training, thus forcing the network to generalize and preventing neural co-adaptation..

In 2012, a deeper evolution of LeNet won the ImageNet Large Scale Visual Recognition Competition (ILSVRC) challenge: AlexNet [3], achieved a top-5 error of 16% on a dataset of approximately 14M images and 21K non-empty synsets. The model was composed of 7 layers, for a total of 60M parameters, was trained using stochastic gradient descent and leveraged data augmen-

tation and Dropout (Figure 3.3).

From that point on, the ImageNet challenge marked several fundamental milestones in the advancements of Convolutional Neural Networks [45]: in 2014, GoogLeNet [52] reduced drastically the number of trainable parameters (from 60M to 4M) by introducing the Inception module and substituting Fully Connected layers with Average Pooling layers. The same year, VGGNet was presented, and although it did not win the competition it showcased the importance of the depth of networks to reach good performances (it consisted of 16 Convolutional / Fully Connected layers for a total of 140M parameters).

In 2015, it was ResNet's turn [53], to win the ILSVR challenge thanks to skip connections and a heavy use of batch normalization. ResNets are currently the default choice in practice [45].

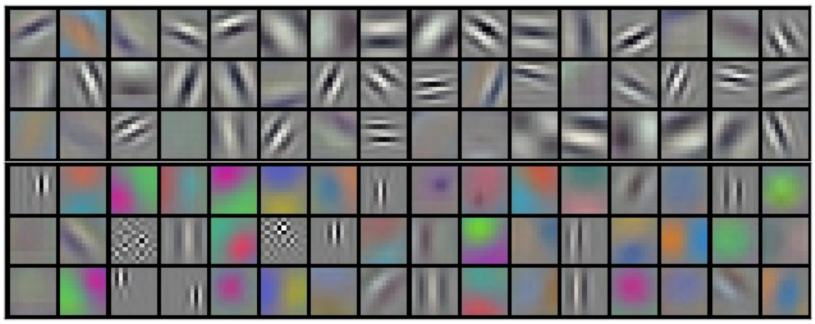


Figure 3.2: AlexNet activation maps (from [3]).

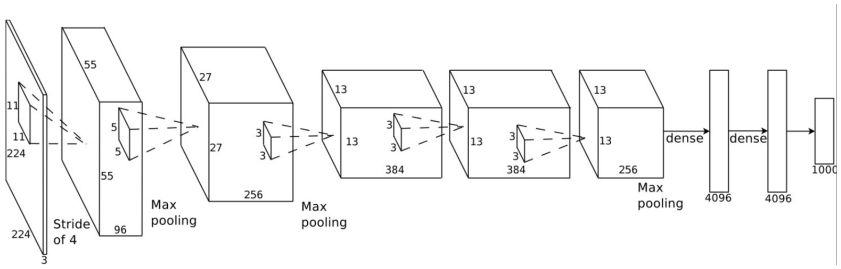


Figure 3.3: AlexNet architecture (from [3]).

3.1.2 Generative models and Representation Learning

The problem of learning a latent representation from the data in an unsupervised fashion has been drawing the attention of many reseachers in the machine learning community for a very long time, starting from Principal Component Analysis [54] to more recent works [55], [19]. It is also of particular interest for this work, in order to extract and analyse important features from AD-affected brains.

3.1.2.1 Autoencoders

The idea of a building model capable of learning an intermediate representation that outputs the exact reconstruction of its input dates back to early works by LeCun [56] and Hinton [57].

The idea is simple: the model is composed of two parts. The *encoder* will try to learn an encoding function $h = f(x)$ that will generate a latent representation starting from the input data x. Then, a *decoder* will learn a function $r = g(h)$ that will act on the encodings and try to reconstruct the original data x that generated them [19]. Since a model that learns an identity function is not particularly useful, the encoding is usually the bottleneck: its dimensionality is forced to be smaller than the original one. An autoencoder of this kind is called undercomplete [19]. This way, it is reasonable to assume that the autoencoder will use its limited parameters to learn only the fundamental features it needs to reconstruct the data, leaving out all the rest.

The autoencoder is trained to minimise a loss function that depends on the original and on the reconstructed data:

$$L(x, g(f(x))).$$

This loss function can vary, since it is what distinguishes one autoencoder from another. Among the regularising autoencoders, Sparse and Denoising are the most used ones: their characteristics encourage the model to learn a feature representation that not only reconstructs the data well, but has also small derivatives, is sparse and robust to noise.

Recently, the encoder's and the decoder's deterministic functions were generalised into stochastic mappings, where $p_{encoder}(h|x)$ and $p_{decoder} = (x|h)$ represent the learned posterior distributions [19]. They have thus become an active field of research amongst generative modeling techniques, for their

ability to model a latent space that is able to capture relevant features in the data.

3.1.2.2 Variational Autoencoder

Traditionally, finding a probability distribution meant defining a probability density: starting from a parametric family of densities that is assumed to fit the problem and then learning the optimal parameters by maximum likelihood. This approach amounts to minimising, asymptotically, the Kullback-Leibler (KL) divergence between the real data distribution and the parameterized distribution. However, when we are dealing with distributions supported by low dimensional manifolds, it is unlikely that the model and the true distributions' support have a non-negligible intersection, meaning that the KL divergence might be not defined or infinite [58].
What most works have done so far to overcome this problem is creating models that were able to directly generate samples following the learned parametric distribution (such as a generator network), since for many tasks this generative power turned out to be more important than knowing the numerical value of the density. One very interesting type of autoencoder with these characteristics is the Variational AutoEncoder (VAE) [18].

In addition to the normal characteristics of the autoencoder, the VAE not only minimises the reconstruction loss between original and generated images, but it also minimises the KL divergence, a distance measure between distributions. In this case, it measures the distance between the encoder distribution and a standard Normal distribution. If we indicate with x the input data, and with z the latent variables, we can define the former as variational posterior $q(z|x)$ and the latter as true posterior $p(z|x)$. Thus, the KL divergence is defined as:

$$\mathbf{E}_q[\ln q_\lambda(z|x)] - \mathbf{E}_q[\ln p(z|x)] = \mathbf{E}_q[\ln q_\lambda(z|x)] - \mathbf{E}_q[\ln p(z,x)] + \ln p(x).$$

This regularizing term aims at forcing a homogeneous distribution in the latent space, where samples of the same class are represented as close to each other in the latent space too. This is also a way of preventing the encoder from dedicating a specific region in the latent space to each training sample in the training set [59].

3.1.2.3 Generative Adversarial Networks

A rather recent advancement in generative methods came with the introduction of Generative Adversarial Network (GAN) [20]. In this framework, two networks are trained symultaneously: one of them is a Generator (which will now be addressed as G), and it tries to learn and reproduce the distribution of the input data, while the Discriminator (from now on D) tries to estimate the probability that a sample came from the real data rather than being artificially generated by G.

This training procedure resembles a two-player minimax game, with the following value function:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\ln D(x)] + \mathbb{E}_{z\sim p_z(z)}[\ln(1 - D(G(z)))].$$

Where x indicates the input data, and z the noise variables fed into the generator.
This principle found successful applications in vision tasks. In this case, G generates artificial images and tries to fool D into believing they are real. D on the other hand, during training will become more and more alert, in an attempt to reveal its competitors' trick and distinguish between fake and real images. The training procedure is thus forcing the generator to create images that are more and more realistic.
GANs are notoriously unstable and very hard to train [60]: they are very sensitive to the initialization and to architectural and hyper-parameters choices. In fact, it is even common to observe networks with similar architectures and hyper-parameters that exhibit very different behaviours. There is also very little theory trying to investigate from a theoretical perspective the unstable behaviour of GANs, as most recent works tend to focus on heuristics to make the training procedure more stable. However, multiple different approaches have been used, and those that have been taken into account for our study of AD are mentioned below.

One example is the Generative Latent Optimization (GLO) model [4], that tries to predict images from learnable noise, thus intuitively resembling an auto-encoder where the latent variables are not produced by an encoder but are free parameters to be learned during the training procedure. Therefore, the model is learning a meaningful organisation in the noise vectors by mapping one noise vector to each of the images in the dataset. Contrary to what an ordinary autoencoder would do, it is not only learning the parameters θ of a generator, g, but it is also jointly learning the optimal noise vector

z_i for each image x_i. The GLO is, therefore, trying to solve the following optimisation problem:

$$\min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} [\min_{z_i \in Z} l(g_\theta(z_i), x_i)].$$

The merit of this architecture is to produce an interpretable latent space while not suffering from the instability typically found in adversarial dynamics. Its efficacy was studied in several tasks, including linear interpolation in the noise space (shown in Figure 3.4 on two different datasets).

Another very recent evolution is the Wasserstein GAN [58]. This work's focus was on finding the proper metric to measure the distance between the real distribution of the data and the parametric one, generated by the network. The authors propose a new architecture that minimises the Earth Mover (EM) distance, making it possible to learn a probability distribution by gradient descent. The same paper also shows that since other distances (Total Variation, Kullback-Leibler, Jensen-Shannon) are not always continuous, they are unfeasible for gradient descent, and thus unfeasible to learn distributions supported by low dimensional manifolds.

To train a WGAN, the discriminator (critic) is first trained to optimality, which also prevents the mode collapsing problem, by sampling batches from the real data x and the priors z and then calculating the gradients with

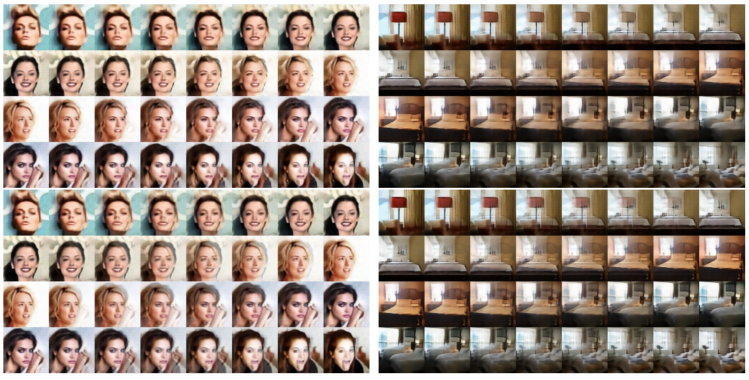


Figure 3.4: Walking along the latent space of two different datasets: celebrity faces with CelebA dataset on the left, rooms with LSUN dataset on the right (from [4]).

respect to w of the following cost function:

$$\frac{1}{m}\sum_{i=1}^{m} f_w(x_i) - \frac{1}{m}\sum_{i=1}^{m} f_w(g_\theta(z_i)),$$

where w and θ indicate the parameters of the discriminator and the generator respectively, and f is the discriminator's function. This cost function is therefore the difference between the output of the discriminator when fed with real and with fake images.

After a fixed number of iterations of the critic, the generator is trained by calculating the gradients with respect to θ of the following cost function:

$$-\frac{1}{m}\sum_{i=1}^{m} f_w(g_\theta(z_i)).$$

A different approach is used in [5], which introduces several guidelines for a more stable training of Deep Convolutional GANs. These include: strided convolutions instead of pooling layers, extensive use of Batch Normalisation, Rectifier Linear Unit (ReLU) activations and Leaky ReLU in G and D respectively, and elimination of Fully Connected layers by directly connecting the highest convolutional features to the input of G and the output of D.

In the same work, the convolutional layers extracted from the discriminator are used as a basis for a follow-up classifier training, as a way of evaluating the quality of the learned unsupervised representation. This approach outperformed the K-means baseline on CIFAR-10, reaching an accuracy of 82%, and reached the state of the art on the StreetView House Numbers dataset (SVHN) with a 22.48% test error. A similar technique will be applied also for the study of AD in Section 5.2.2.1.

The latent space was also investigated by walking along the manifold and checking whether it produced semantic changes in the generated images. This procedure was successful on the LSUN bedroom dataset, where a bedroom with no windows slowly morphs into a room with a big one (Figure 3.5, row 6) and where a TV turns into a window (Figure 3.5, row 10).

Finally, some *arithmetic operations* are performed on the latent features (Figure 3.6), to understand whether it encoded some sort of linear structure, and whether this was linked to a semantic meaning.

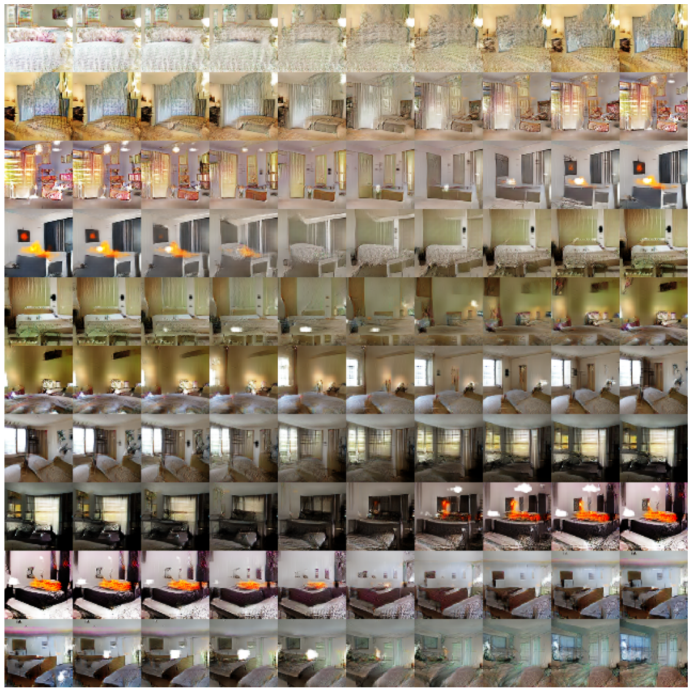


Figure 3.5: Walking along the latent space. A bedroom with no windows slowly morphs into a room with a big one in row 6; a bedroom with a TV turns into a window in row 10 (from [5]).

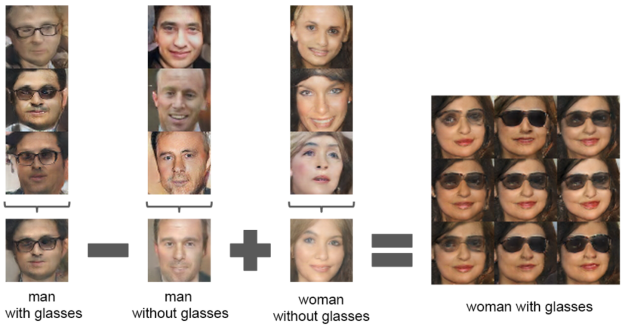


Figure 3.6: Arithmetic operations on the latent space (from [5]).

The idea of learning a latent representation that is semantically interpretable and that can be walked to produce meaningful variations in the original image is also investigated in [6] on face images. This latent representation (*graphics code*) is disentangled, meaning that changes in only a few of the latent features are mapped into realistic transformations [55]. Decomposing an image into variables that control pose, light and shape, means obtaining a fine-grained control over the changes in the objects in the image.

The proposed architecture is a Deep Convolutional Inverse Graphics Network (DC-IGN) and it is basically a Variational Autoencoder, with some changes: it is composed of an encoder, which captures the graphics code (latent features) from the input image, and a decoder, which learns a posterior distribution to reproduce the original image given the latent space.

Now, in order to obtain a disentangled latent space, the training data is split into mini-batches that correspond to changes in only one variable at a time. Similarly, other batches have certain variables being held fixed, while other face properties change, forcing the varying ones to learn features that describe identity and expression. The result is a rendering engine: Figure 3.7 shows what varying pose elevation and azimuth looks like on face images. Although this approach is not feasible for 3D sMRI volumes, because of the lack of data corresponding to single-variable variations in the evolution of AD, it is interesting to see how a Variational Autoencoder is used to reproduce different kinds of progressions in latent space.

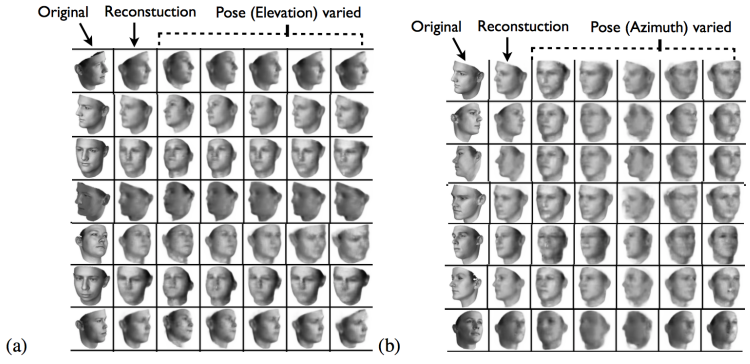


Figure 3.7: Walking along the latent space: varying pose elevation (a) and azimuth(b) on face images (from [6]).

When it comes to 3D input data, the problem of representation learning has been applied for the first time to point clouds in [7].
One very interesting aspect of this work is that generation and representation learning are completely decoupled during training. In fact, the architecture is constituted by two components: an Autoencoder in charge of creating the latent representation, and a GAN to generate new samples in the latent space. These components need not be trained simultaneously, which is a very nice property considering the challenges already mentioned in the finding the right hyper-parameters for GAN training. Moreover, this allowed more freedom in the selection of a domain-specific loss function for the latent space, since modeling it with an Autoencoder meant relying on a mature and stable tool.

The training procedure is as follows. First, the autoencoder is trained, then the GAN is trained, but instead of doing so directly on the raw point cloud, it operates on the latent space. Thus, the discriminator will try to detect the differences between the output of the generator (that is fed with samples from a fixed noise distribution) and that of the encoder. Finally, to go from the generated latent representation to a 3D point cloud that can be visualised, the output of the generator is sent to the decoder.

This *l-GAN* is shown in Figure 3.8. The autoencoder relies on 1D-Convolutional and Max pooling layers for the encoder (as the point clouds were pre-aligned following a lexicographic ordering of coordinates), and Fully Connected layers for the decoder. The structure of the GAN itself is fairly simple: the generator consists of 2 Fully Connected layers (with 128 and 512 neurons respectively), and the discriminator has of 3 Fully Connected layers (with 256 and 512 neurons respectively).

The evaluation of the model is two-fold. The latent representation is evaluated by exploiting its feature-extracting capabilities and applying them to supervised tasks. Walking along the latent space allows to see the different characteristics of the point clouds (Figure 3.9 shows this concept applied to chairs). At this task, it reaches the state of the art of the classification accuracy. The diversity generated by the GAN reaches the state of the art too based on two metrics: Minimum Matching Distance and Jensen Shannon Divergence.

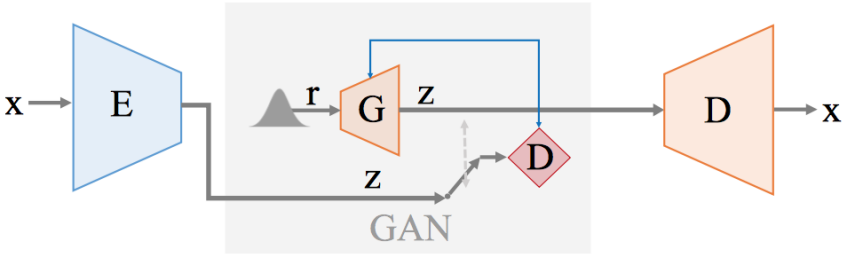


Figure 3.8: 3D GAN architecture (from [7]).

Figure 3.9: Walking along the latent space from 3D point clouds of chairs (from [7]).

Most of these representation learning techniques have not been applied yet to the study of sMRI scans for AD diagnosis, with some exceptions mentioned in Section 3.2. However, they can be particularly useful when trying to identify the best methodologies to generate a latent space of relevant features to predict the progression of AD in time, as we will show in Section 5.2.3.

3.1.3 Progression

While the works described so far tackled the idea of progression from the perspective of the latent space, thus trying to learn key features that can evolve producing visible variations in the generated images, other works focused on the progression generated by the aging process itself. A conditional adversarial autoencoder is used in [8] to produce a manifold that can be walked to obtain the younger or older version of a certain input image.

The architecture used is shown in Figure 3.10. It follows the idea of Generative Adversarial Networks, but in this case the Generator is a Conditional Autoencoder: it is composed of an encoder, that reduces the input image to an array of 50 latent features. Then a one-hot array containing information about the age corresponding to the input image is concatenated to the latent features vector. This augmented latent space is then sent as input to the decoder, which then tries to reconstruct the original image.

The architecture includes also two different discriminators: one of them forces the latent space to be distributed according to a prior distribution, which in this case is uniform since we want to populate the latent space without *holes*. The second encoder has the same goal as the one in traditional GANs (i.e., to learn to distinguish between artificial and real images), but in this case it receives as input the augmented latent space (i.e., the latent space from the encoder concatenated with the one-hot age vector). Moreover, in addition to the generator and discriminator losses, a *total variation* loss is minimised to remove ghosting artifacts and generate more realistic images. Given any test face, the network is thus able to make it younger or older, while preserving its characteristic features (Figure 3.11).

This architecture was trained on two datasets of, respectively, 55,000 images from 13,000 subjects, and 13,446 images from 2000 subjects. These numbers are nearly impossible to achieve when it comes to sMRI scans. Nevertheless, this work gave us the basis for the study of the progression of Alzheimer's Disease.

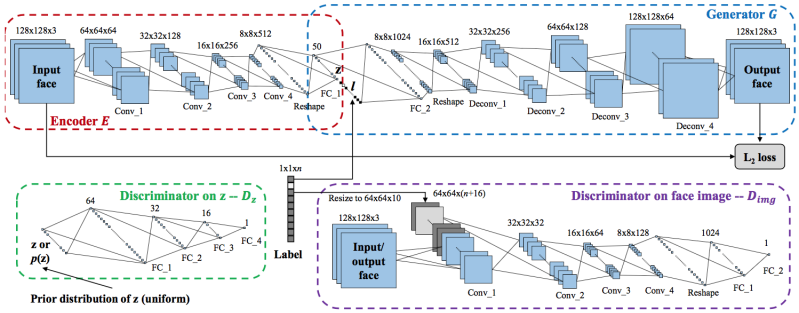


Figure 3.10: CAAE architecture (from [8]).

Figure 3.11: Age progression and regression using a CAAE (from [8]).

To the best of our knowledge, techniques of this kind have not been applied yet to study the progression of brains, and in particular with focus on AD development. However, the possibility of generating a progression in time starting from the brain scan of a patient at a certain age, would represent enormous improvements: researchers could study the effects of the disease on specific organs, or check the effects of test drugs on specific areas of the brain.

This section explored the state of art techniques in the analysis of images and their feature representations. The next sections will focus on the application of ML methods for the study of AD.

3.2 Machine Learning methods for AD

Machine learning techniques have been extensively employed in the study of Alzheimer's disease, mainly for what concerns the classification into three different categories: AD, MCI, NC.
Back in 2008, a study [61] already showed that SVMs were at least as good at distinguishing sporadic AD from normal aging and frontotemporal lobar degeneration (FTLD), as a team of radiologists. The study was carried out on three datasets, which consisted of:

1. 20 sporadic AD scans and 20 normal controls;
2. 18 sporadic AD scans and 19 FTLD;
3. 14 sporadic AD scans and 14 normal controls;

The SVM was fed grey matter segments extracted from MRI brain scans and normalized into the standard anatomical space, obtaining thousands of voxel that represented the amount of grey matter volume in that specific area. The SVM learned the difference between each pair of groups by finding its support vectors (i.e., its boundary) in the patients that were most difficult to diagnose (i.e., separate).

The results in Table 3.1 compare the performance of SVM classifier to that of six neuroradiologists with various degrees of experience. They were asked to make decisions mimicking the SVM method, so information usually available in ordinary clinical diagnosis (such as age) was not disclosed.

This study was among the first ones to highlight the potential of computer-aided diagnostic methods as a support where experts are scarce, or simply to reduce time and effort for the diagnosis.

A complete survey [9] identified more than 500 papers on the MRI-based prediction of brain diseases (including AD, Schizophrenia and depressive disorders) published between 1995 and 2015, and compared 112 studies on

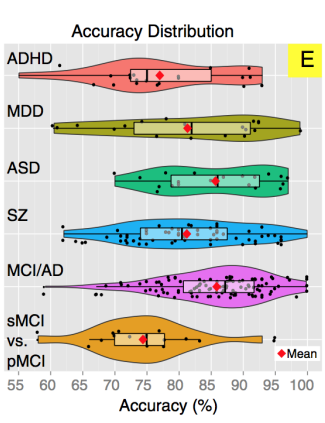
	SVM	Radiologists
AD vs Controls I	95%	65 - 95% — median 89%
AD vs Controls II	93%	80 - 90% — median 83%
AD vs FTLD	89%	63 - 83% — median 71%

Table 3.1: SVM vs radiologists

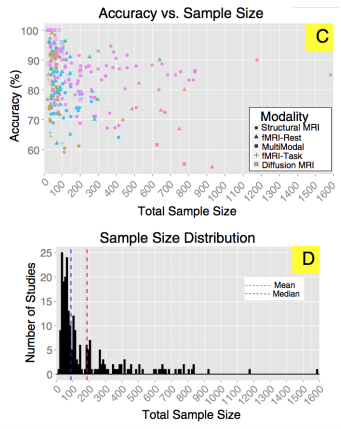
MCI/AD specifically. Figure 3.12 provides a very good picture of the studies on brain imaging using machine learning techniques: while the mean accuracy of MCI/AD classification is approximately 85%, the mean sample size is 200 and the median is 100 (for all kinds of dementia), which is a strong limitation in general, and even more so for any deep learning approach.

These studies used data from diffusion, functional or structural MRI. About half of them combined more than one input data type, while the other half focused on sMRI only. The most used classifiers include various kinds of SVM (59%), Linear Discriminant Analysis (20%) and Logistic Regression (10%). Moreover, most of these methods are characterized by the following phases: feature extraction according to different atlases or Regions of Interest (ROIs), feature selection or dimensionality reduction and finally classification. In some cases, particular emphasis was given to the process of finding ROIs, in others it lied in the classification itself.

Out of all these papers, we decided to focus our attention on the most recent ones, those based on sMRI only and those that resulted in the highest accuracies on larger sample sizes. Moreover, we tried to include an example for every relevant methodology. These works are discussed below.



(a) By disease



(b) By total sample size

Figure 3.12: Machine learning methods for AD, a survey: accuracy and sample size distribution (from [9]).

Calculating and extracting 2D slices from the initial 3D MRI scans was attempted by maximum inter-class variance based on grey-matter levels in [62]. Then, PCA was applied to generate 2D eigenbrains from these scans. Out of all these eigenbrains, the most relevant ones for classification purposes were selected using Welch's t-test on the 95% confidence interval, where the null hypothesis was that the eigenvalues of AD and NC had equal means. This test allowed to pick the eigenvalues of the most relevant eigenbrains as input for the classification, which was done by kernel SVMs. The training was done on 126 subjects. While different kernels were tested, the polynomial kernel proved to have a better accuracy at diagnosing AD (92.36 % $\pm$ 0.94) than the linear (91.47% $\pm$ 1.02) or the RBF kernel (86.71 % $\pm$ 1.93).

Another way of extracting features is that of atlas-based ROIs. In [63], for instance, the Desikan-Killany Cortical Atlas is used. Two kinds of features were gathered: morphological (such as mean cortical thickness values, cerebral cortical grey matter and cortical-associated WM volume) as well as correlative (in terms of similarity between different ROIs). Out of all these features, only the most discriminative ones with respect to AD were selected, using two different approaches. First, a filter-based strategy selected only the features whose p-values (from between-group t-test) were smaller than a certain threshold. Then, wrapper-based selection relied on Support Vector Machine Recursive Feature Elimination for further feature selection. Then, a multi-kernel SVM with RBF kernel was used to integrate the information coming from the different features, forming a mixed-kernel matrix. On a dataset of 598 subjects, this approach resulted in an accuracy of 92.35% for AD, 83.75% for MCI and 79.24% for AD-MCI. Moreover, the classifier was able to tell with 75.05% of accuracy whether an MCI patient would develop AD within 36 months or not.

Ensembles of different classifiers proved to be very successful in [64]. In this work, several pre-processing steps including motion and inhomogeneity correction, averaging, spatial normalization and brain surface extraction lead to the volume calculation of GM, WM, CSF and hippocampus. The latter in particular was segmented via ROI extraction, noise removal and trimming. These features were used as input for the final classification, which was performed through different classifiers: RBF-kernel SVM, MLP with 2 hidden layers with 3 neurons each and C4.5-based decision tree. These classifiers were first used separately, then in an ensemble fashion, where the final decision was determined by a majority voting mechanism. The training was performed on 416 brain scans, reaching an accuracy of 93.75% on AD

vs NC, using the combined features.

Another application of an ensemble of classifiers can be found in [10], where template-based clustering and multi-view learning is used for feature extraction (Figure 3.13). First, an affinity propagation clustering algorithm is used to group the subjects into different clusters, and the respective cluster centres are then used as templates. The feature representation is therefore multi-view, since there is one for every template. Feature selection is done using a mass-preserving shape transformation, which starts with brain segmentation that produces three tissues (GM, WM, CSF). Then, one tissue density map for each template is produced, reflecting the volumetric measurement. The millions of volumetric features produced needed to be reduced for the algorithm to train meaningfully (especially considering the small number of samples). In this case, the problem of dimensionality reduction cannot be solved by using pre-defined ROIs, because of the presence of different templates for different views. A segmentation algorithm is thus applied to do a regional grouping of these features in each template, resulting in different ROI partitions for every template. Regional features are aggregated for optimisation purposes and a sub-class clustering algorithm then identifies the most relevant classes in every template. Then, a multi-task feature selection algorithm removes those that are redundant or noisy. An SVM classifier is learned for every view, and an ensemble of these classifiers is in charge of the final diagnosis by majority voting.

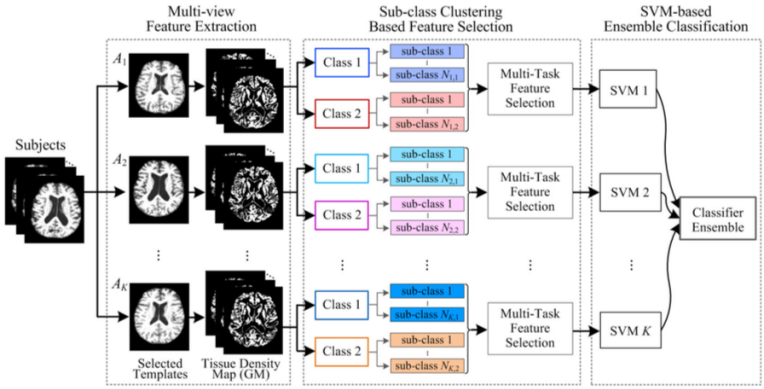


Figure 3.13: Flowchart of the method proposed in [10]: Multi-view feature extraction, clustering and feature selection, and finally SVM-based classification.

This algorithm results in 93.83% of accuracy on AD vs NC, 89.09% on progressive MCI vs NC and 80.90% on progressive MCI vs stable MCI.

The idea of associating neuroimaging features with cognitive scores in an attempt to better track the progression of AD was modeled as a multi-task problem in [65]. This study aimed at identifying a weight matrix W of shape $d \times mT$ (where d is the number of neuroimaging features, m the number of cognitive indicators, and T the timesteps) that is able to predict the cognitive score of the patient. In order to take into account the longitudinal information available, a longitudinal structured low-rank regression model was proposed. The dataset used comprised 385 subjects (56 AD, 181 MCI and 148 NC): for each of these patients, 7 different cognitive scores and Grey Matter volumes obtained from 93 identified ROIs were available. This method performed better than other regression models (in terms of Root Mean Square Error and Correlation Coefficient) at finding markers that could influence the cognitive score over the different time steps.

Similarly, a combination of MRI and neuropsychological parameters proved to work better than MRI alone in [66]. In this study, MRI was integrated with clinical variables including age and education, and with neuropsychological variables including MMSE, Rey Auditory Verbal Learning Test (RAVLT) and Clinical Dementia Rating Scale Sum of Boxes Scores (CDRSb). Using FreeSurfer [67], 115 cortical and subcortical regions were extracted and then integrated with the other variables. Step-wise linear regression is used for feature extraction, while Linear Discriminant Analysis (LDA) was used as a final classifier, to estimate the posterior probability for each sample to belong to one of the two classes in each two-way classification task. On a total of 385 subjects, the resulting accuracy was very high in NC vs AD (93.9%), NC versus late MCI (90.8%), early MCI vs AD (94.5%) and late MCI vs AD (90.1%). Overall, a combination of all these parameters reached an average accuracy of 87.6%, while using MRI or cognitive scores alone reached an overall accuracy of 71.2% and 85.3% respectively.

Voxel-based morphometric feature selection by Integer-Coded Genetic (ICG) algorithms was used in [68]. A self-adaptive Resource Allocation Network (SRAN) classifier operating with a sequential learning algorithm was also used to discard redundant samples and to select the most meaningful ones from the dataset.

Fractal multi-scale analysis at different scales, thus analysing signals at

different granularities, was used in [69] to estimate the Hurst's exponent of the MRI scans (previously converted to 1D signal). This stems from the assumption that a healthy brain's tissue shows more self-affinity (and therefore its pixel distribution shows more regularity) than that of a diseased brain. These Hurst's exponents were then grouped into a feature vector fed as input to an SVM classifier, resulting in an accuracy of 97.08% $\pm$ 0.05 on MCI vs NC, and an accuracy of 97.5% $\pm$ 0.04 on MCI vs AD.

The effectiveness of GM-based characteristics is proved in several works, including [70] and [71]. The former classified them using an SVM with bootstrap resampling. On a very small dataset (16 AD and 22 NC) this approach reached a 94.5% mean correct classification. The latter used GM together with WM as a starting point for segmentation. This procedure is done through the Statistical Parametric Mapping (SPM) software and models the intensity value distribution of the MRI. The final features are thus the probability values associated with the presence GM and WM in a given voxel. In order to identify the most relevant ROIs, an SVM is trained and feature selection is carried out using a wrapper approach, thus using the classifier's parameters as scores to select the features. This technique is thus recursively eliminating least-relevant features. The obtained ROIs turned out to focus on areas of pivotal importance in AD diagnosis, such as the hippocampal region. The training was performed on a dataset of 185 NC and 185 AD, and the overall accuracy on GM segmented images was of 94.32% on AD vs NC, while on WM segmented images it was of 95.14%.

While several studies had already tackled the study of Alzheimer's biomarkers using a network approach [72], a very recent work [11] approached the problem of detecting relevant features in MRI. This study leveraged multiplex networks to understand and highlight the drivers of the disease's propagation in each individual, as well as the differences in patterns between individuals. This study was too recent to be included in the aforementioned survey.

To understand the variations in the patterns of disease progression between individuals, they identifies patches, $3000mm^3$-wide segmented regions of the brain, and bound them by rectangular boxes. Each patient is modeled like an undirected weighted network, whose nodes are the brain patches and the edges are similary measures between them. A multiplex network [73], as shown in Figure 3.14, is therefore a family of weighted networks, each

of which represents a patient (layer). Each patient can thus be modeled easily using an adjacency matrix, and indicators like strength and inverse participation ratio can be used to study the distribution of weights within each layer, in an attempt to distinguish healthy from diseased patients.

The multiplex framework provides useful context information in the understanding of the most relevant AD features. Each network has the same number of nodes, and a set of links and respective weights calculated as Pearson's correlation between two nodes (patches). This type of architecture aims at facilitating inter-subject instead of group-wise characteristics.

Two Random Forest (RF) classifiers are employed to, respectively, do feature selection (from the initial 549, only 32 significant patches were kept) and to summarise the network measures in a unique atrophy score, which was then used as the basis for the RF classification. The study identified regions significantly related to AD, distinguishing AD from NC patients with an accuracy of 86% and MCI from NC patients with an accuracy of 84%.

In this section we provided a general overview on the machine learning methods for AD classification. Next, we will focus on deep learning.

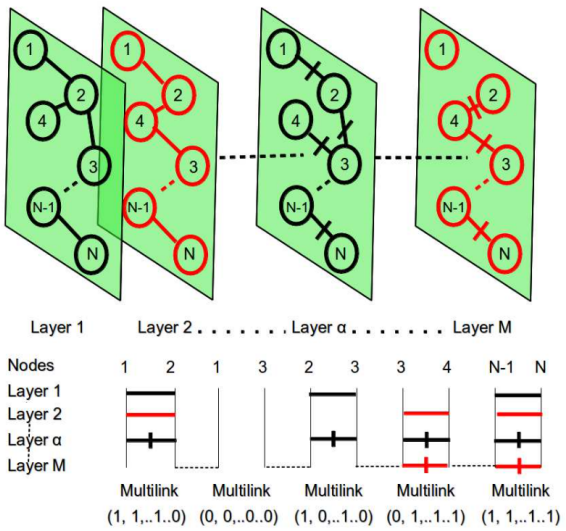


Figure 3.14: Multiplex network (from [11]).

3.2.1 Deep Learning for AD

In recent years, during what some call the *deep learning renaissance* [19], deeper and deeper architectures have been applied to several fields, including medical imaging. While this is extremely challenging with respect to the minimum dataset size that guarantees a proper learning (which is even more critical if we think of how expensive and time-consuming it is to produce labeled data in medical fields), it also allows to remove, more or less completely, the subjectivity and workload that comes with the hand-crafted selection of features. Additionally, deep learning allows to disentangle very intricate patterns in high dimensional data thanks to its typical deep hierarchy of layers, thus having potentially leading to new discoveries.

In this section, some of the most interesting studies leveraging these techniques will be presented.

A solution to the problem of small datasets was proposed in [74]. In this study, a dataset consisting of 188 AD, 399 MCI and 228 NC was used. After normalizing and pre-processing steps, 3D bounding boxes containing the hippocampus are extracted from each brain volume using the Automated Anatomical Labeling (AAL) atlas. But instead of using the 3D shape, five 2D slices were extracted and used as input to train a CNN with two Convolutional-Pooling-ReLU blocks, one Fully Connected layer and finally a Softmax classifier. Other data augmentation methods, including flipping and traslation, and class balancing techniques, were applied. The overall classification accuracy was 82.2% in AD vs NC, 66% in NC vs MCI and 62.5% in AD vs MCI.

As already noted in the previous section, integrating multiple sources of information is extremely important for the success of any learning algorithm. Low-dimensional biomarkers, such as MMSE or CSF measurements were combined with high-dimensional neuroimaging biomarkers, such as MRI or PET, in [75]. Combining these different data sources can be extremely advantageous, as they can help build a complete picture of the patient, just as they can make the procedure challenging: all biomarkers must be available for all data samples used for training, which may result in a reduction of the number of samples available for training. Moreover, being low-dimensional also implies that some biomarkers might be less sensitive to the real degree of cognitive impairment than high-dimensional ones. It is, in a way, a typical "curse of dimensionality" problem.

The study was carried out on a dataset of 331 subjects, of which 77 where NC, 169 MCI and 85 AD. The methodology used comprised a pipeline of 3 phases: first, a Stacked Autoencoder was trained on MRI and PET inputs. Stacked Autoencoders have been frequently used to extract hierarchical features from MRI and PET scans [13, 75], so the methodology is explained more in depth in Appendix A. Second, the model is fine-tuned to predict the low-dimensional scores (MMSE, CSF), by adding one linearly-activated layer on top of the previous network. Finally, a softmax layer is added to produce the final classification. The model reached what was the state of the art back then, with an overall accuracy of 90.11% on a two-way, AD-vs-NC classification, but significantly worse performance on three-way classification (59.19%).

Deep learning methods have also been used to learn a manifold from MRI scans in [12]. In this case, a Deep Belief Network (DBN) [19] made of multiple Restricted Bolzmann Machines (RBMs) [43] is used, for its power to extract patterns of similarity in images. Each RBM receives as input the output of the previous RBM: the first ones in this chain are Convolutional RBMs, and the hidden units are noisy ReLUs. Since these convolutions share the weights, they also have the advantage of reducing the overall number of parameters, which is a relevant problem considering that the model is fed with the entire 3D input volume of shape $128 \times 128 \times \times 128$. Sampling from this latent space produced visible variations in the generated images, that were varying in global size as well as specific organ-related features. This latent space also showed to be correlate with other parameters such as age, gender, MMSE. In addition to being able to work without the aprioristic definition of a similarity measure or proximity graphs, this approach is also the first example of representation learning applied to brain MRI scans.

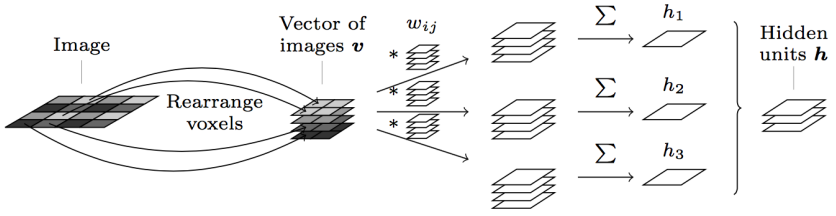


Figure 3.15: Deep Belief Network architecture (from [12]).

Other kinds of neuroimaging datasets have been used for the same purpose. In [76] PET images were used, while in [77] Arterial Spin Labeling (ASL), an fMRI technique, is used. In this case, a CNN composed of two Convolutional layers, two Pooling layers and one final Max pooling layer, is used to extract hierarchical, latent features from the images. To overcome the lack of labelled data, a semi-supervised pair-wise disease similarity learning based on the Kendall - Tau evaluation measure is added. Finally, the diagnosis prediction is carried out using a multi-nominal logistic regression.

fMRI-based volume maps of GM and WM were also used in [78] as input to a 3D CNN architecture with the objective of predicting brain age and detecting earlier signs of cognitive decline. Correlating neuroimaging with cognitive scores is of great importance for AD from the perspective of discovering new features as characteristics of the disease. The same does not hold from the perspective of generating a progression of the disease, since we already noted in 2.2 that the first cognitive symptoms appear only when the disease is already advanced.

A combination of unsupervised and supervised methods was employed to tackle the problem of AD diagnosis in [13], and it consisted of two steps.

First, a Convolutional Autoencoder (CAE) operating with 3D convolutions is trained to extract the fundamental features from the brain volumes. However, extracting the latent features through convolutions on the entire 3D volume is hardly successful, because of the enormous amount of learning parameters in such architectures, and because of the autoencoders' tendency to learn global features instead of local ones. To compensate for these issues, three Stacked Autoencoders (SAE) [79] are used, applying the hierarchical CAE from [80] to 3D voxel signals. The idea behind SAEs is provided in Appendix A.

Each autoencoder was trained minizing the mean squared reconstruction error using stochastic gradient descent. In each autoencoder, the encoder and decoder weights are tied by flipping over all dimensions, to reduce the number of parameters.

As a second step, the three encoder layers were stacked one right after the other, followed by three fully connected layers and a softmax layer, thus creating an overall 3D-CNN whose bottom layers were initialised with the weights obtained by unsupervised training. The three upper layers

were hence specialising in the three-way classification task, as opposed to the lower convolutional layers, focusing on extracting features related to biomarkers, such as shapes and volumes. The classification loss was minimised using Adadelta gradient descent, and the performance was evaluated on 210 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Figure 3.16 shows the complete architecture used for supervised fine-tuning.This classification approach outperforms previous works on the same tasks, reaching an accuracy of 89% on three-way classification, 97.6% on AD vs NC, 95% on AD vs MCI and 90.8% on MCI vs NC.

The idea of extracting low-level features using an autoencoder, and then fine-tuning the obtained model using a supervised approach proved to be extremely useful also in the present study (Section 5.2.2.1), when applied to specific organs of the brain that are key for AD diagnosis, including hippocampus and ventricles.

As this overview shows, many works have already applied deep learning methods to the study of AD, using different types of pre-processing and pipelines. However, deep generative models have not been applied successfully to the task of studying the progression of AD on sMRI scans yet.

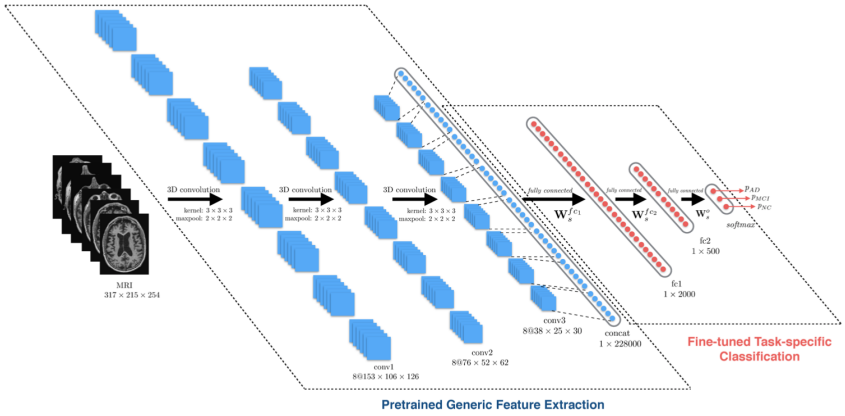


Figure 3.16: 3D SAE architecture (from [13]).

Chapter 4

Data

This work is focused on three datasets: HARP, AIBL, OASIS. Each dataset associates 3D neuroimaging volumes with various labels, that include age, gender and cognitive scores. In this chapter both the images and the labels will be presented and described.

4.1 Structural MRI

This study will focus on Structural Magnetic Resonance Imaging (sMRI), for its being free of ionizing radiation exposure (thus less invasive) and more affordable compared to other examinations. Therefore, it is plausible to imagine that patients at risk might be monitored more frequently and from a young age, having a change to track the disease earlier and allowing progression models to become more robust and accurate.

A general introduction to this type examination can be found in Section 2.2.1. In this study will focus on T1 only.

4.2 Datasets

One of the first and foremost issues encountered in any machine learning model is the collection of data. Three different datasets were gathered for this study: HARP, OASIS, AIBL.

- HARP: created by the European Alzheimer's Disease Consortium together with the Alzheimer's Disease Neuroimaging Initiative (ADNI). It was provided as HARmonized Protocol (HARP) for manual hippocampal segmentation from MRI [81], and it consists of 131 volumes.

- AIBL: created by the Australian Imaging, Biomarker & Lifestyle flagship study of ageing (AIBL), aimed at discovering the underlying factors that influence the development of AD (including biomarkers, health and lifestyle). The data comes from two different centres, Perth and Melbourne, and it consists of 586 volumes.

- OASIS: the Open Access Series of Imaging Studies (OASIS) was created by the Washington University Alzheimer's Disease Research Center, Dr. Randy Buckner at the Howard Hughes Medical Institute (HHMI) at Harvard University, the Neuroinformatics Research Group (NRG) at Washington University School of Medicine, and the Biomedical Informatics Research Network (BIRN). It consists of 233 volumes.

All datasets include samples from the three classes: AD, MCI, NC, for a total of of 950 volumes. The last two datasets however, suffer from a non-negligible unbalace towards the normal control class, as shown in the Figure 4.1.

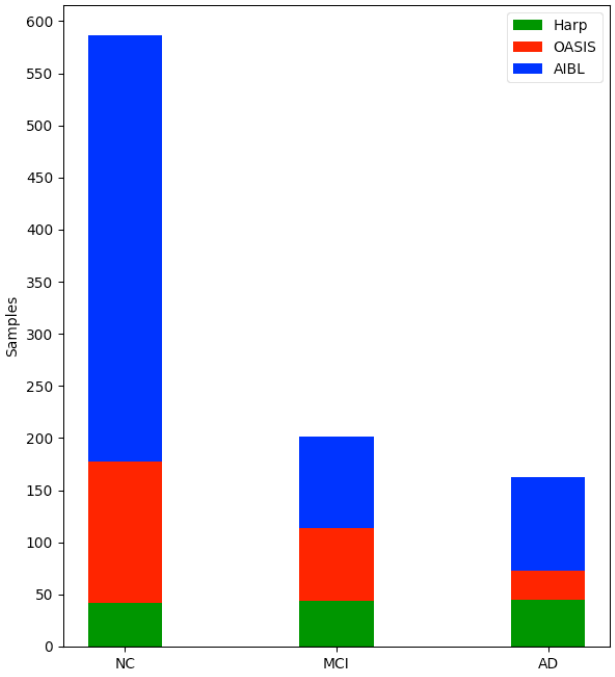


Figure 4.1: Dataset distribution by class.

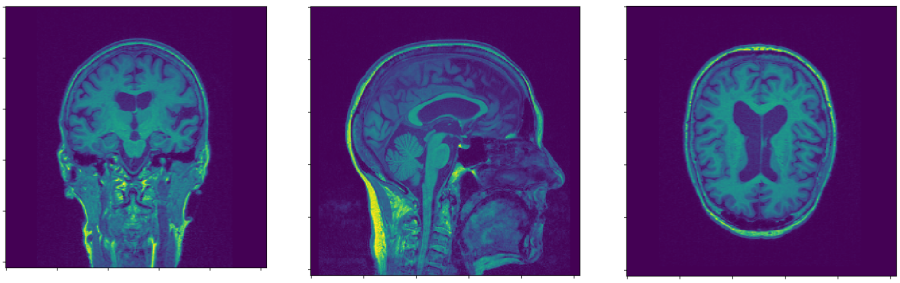
4.2.1 Images

For every image, the three perspectives shown in Figure 4.2 are taken into account: coronal, sagittal, axial.

Since the volumes are gathered from three different datasets, each one with more or less different facilities, different machines and therefore different acquisition steps, they all appear quite different from each other. Moreover, since the focus of this work is to study the most relevant features of the brain, it is important to remove from the volumes any extra feature that could act as a distraction for the network, such as the structure of the skull. This can be seen for instance in the third row of Figure 4.3a.

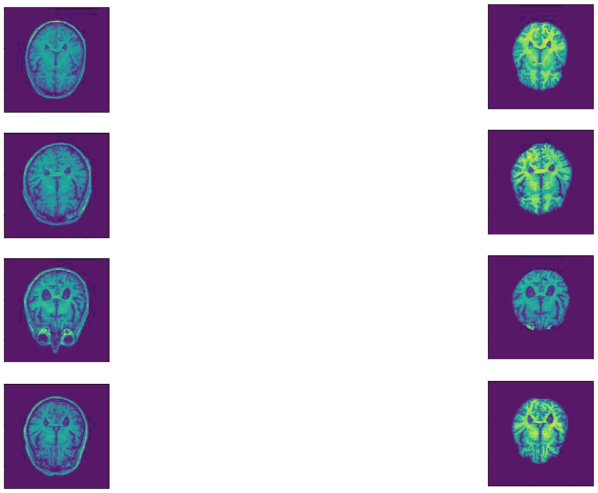
To overcome these problems, we skull-stripped the volumes using the software FreeSurfer [67], to obtain a cleaned version of the volumes. In particular, this procedure eliminated the facial skeleton, the frontal bone, the occipital bone, the parietal bones and the temporal bones. While this preprocessing step turned out to be fundamental for the success of various models used in this work, it has also shown (Figure 4.4b) to produce slight differences in the images, generating a set of images that are not entirely homogeneous in shape. Therefore, for some tasks the normalised dataset lead to better results, while for others the original one seemed to work better.

Figures 4.3 and 4.4 show the difference between the original (on the left) and the skull-stripped images (on the right) from an axial and a coronal point of view respectively. This comparison shows how skull-stripping eliminates those features that are of little importance when trying to extract Alzheimer's-related features. From a coronal perspective, skull-stripping also eliminates the structure of the neck.



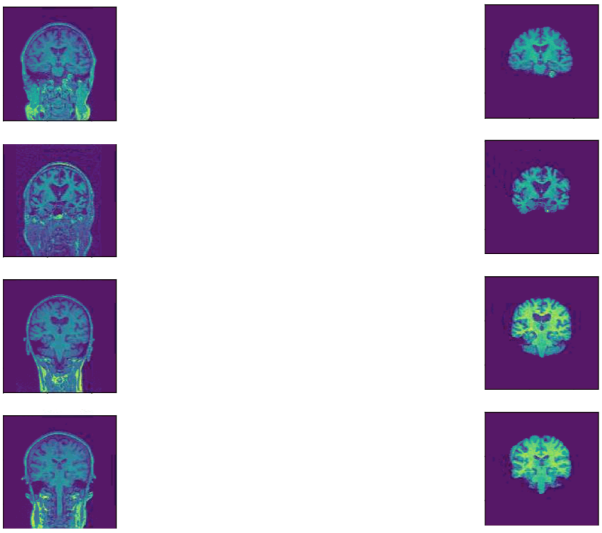
(a) Coronal *(b) Sagittal* *(c) Axial*

Figure 4.2: Brain axes: coronal (a), sagittal (b), axial (c).



(a) Original Images *(b) Skull-stripped Images*

Figure 4.3: A comparison between original and skull-stripped images - axial view.



(a) Original Images *(b) Skull-stripped Image*

Figure 4.4: A comparison between original and skull-stripped images - coronal view.

4.2.2 Labels

For every patient, a number of parameters are provided for each dataset:

- Age: later divided into bins of 10 or 5 years from 50 to 100 years of age. Smaller bins are used in the most critical years for Alzheimer's;
- Gender: female and male
- Diagnosis (Classification): AD, MCI, NC;
- MMSE on a scale 0-30, where 30 indicates a degree of impairment questionably significant, and 0 a severe impairment, as shown in Table 4.1. Even though this score is normally included in standard AD datasets and frequently used in most machine learning studies focusing on medical imaging (Section 3), this type of examination is currently considered limited to allow a meaningful cognitive evaluation of the patient. After an in-depth discussion with Dr. Michele Sintini, from San Marino Hospital, it became clear that not only the score is strongly dependent on the patient's attitude, mood, or psychiatric real-time conditions and on the physician's subjectivity, but it is also possible for patients of high-level education with dementia to outperform patients of low-level education without dementia [77]. There have been cases where a highly-confident biomarkers-based AD diagnosis was associated with high MMSE scores.

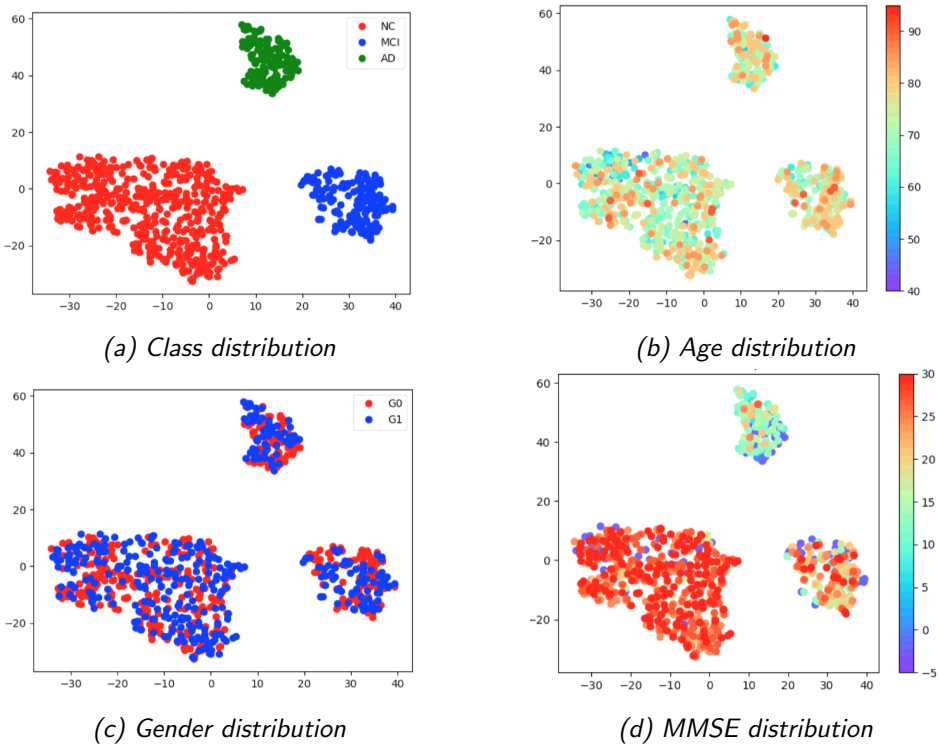
Overall, these labels are not enough to build a clear and complete picture of the patient to generate a completely robust progression, as the diagnosis of AD, in particular for older patients, can be strongly influenced by multiple factors and intertwined causes (as pointed out by Dr. Sintini).

MMSE score	**30-25**	**24-20**	**19-10**	**9-0**
Condition	May be NC	Early AD	Moderate AD	Severe AD

Table 4.1: MMSE score explanation

Figure 4.5 takes advantage of a t-SNE visualisation [21] obtained with one of the models used in Section 5.1.2, in order to provide a qualitative visualisation of the distribution of labels across the datasets. Every point in the figure represents a patient but, for the purpose of this section, the specific location of the patients in the 2D space is irrelevant: what matters is their membership to one of the three clusters shown in Figure 4.5a. By comparing these clusters with the three remaining scatter plots, it is possibile to get an idea of the age, gender and mmse distribution across the three classes.

In particularly, the youngest patients are more frequently associated with the NC class. For what concerns gender, females tend to be more frequently diagnosed Alzheimer's disease, which is confirmed by the literature[16]. The MMSE score is highly correlated with the class label: patients with the highest score belong to the NC class or to the MCI class, whereas the most part of patients in the AD class show very low MMSE scores. MMSE scores included in the interval $[-5, 0]$ just indicate that the examination score was not available for that patient.



(a) Class distribution

(b) Age distribution

(c) Gender distribution

(d) MMSE distribution

Figure 4.5: Dataset distribution by class, age, gender and MMSE.

For what concerns longitudinal information, which already proved to be a very important tool in the study of AD progression (Section 3), it is only available for some patients belonging to the AIBL dataset. For each of them, age, MMSE and diagnosis corresponding to three visits are provided: baseline, 18 months after the baseline and 36 months after the baseline. However, using this longitudinal information is problematic, because it only tracks about two years of the patient's life: if the patient already has the disease, it is too late to grasp any useful information on its previous conditions; if the patient is healthy, it is too early to know whether he/she will develop the disease or not. Moreover, it is unknown what kinds of therapies, physical or cognitive changes the patient has been going through, and this makes it very hard to interpret the progression in a meaningful manner.

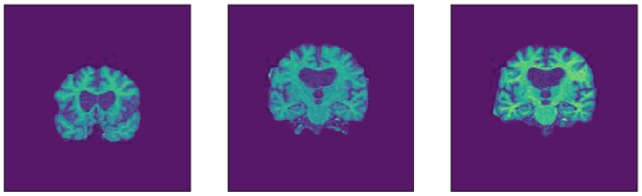
Some progressions are shown in Image 4.6. Each line concerns one patient: the first on the left is the baseline and last one on the right is the +36 months visit. The captions state diagnosis, age and MMSE score for the beginning and the end of the progression.

At this point, it is interesting to make some comparisons. At a first glance, the patient in Figure 4.6a shows significantly more enlarged ventricles compared to those of the patient in Figure 4.6b but while the first is healthy, the latter is not. Unsurprisingly, the latter is also significantly younger. It must also be noted that the enlargement in ventricles size is particularly from a longitudinal perspective, that is, studying the same patient along several years.

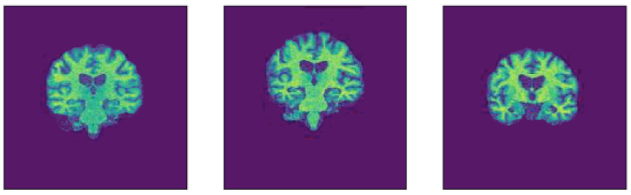
The progression of the patient in Figure 4.6c shows a marked enlargement of the ventricles and marked relevance of the grey matter, which is consistent with the AD diagnosis.

The patients in Figure 4.6d and 4.6e highlight some of the issues mentioned above: the former shows an improvement in the MMSE score; the latter shows a change in diagnosis from NC to AD while the MMSE score has not changed (probably for the same reason above). Figure 4.6e also highlights a relevant issue with slicing that will be discussed in Section 5.1.1.

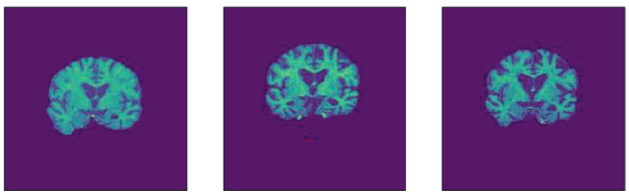
For all these reasons, unless specifically stated, the longitudinal information will be removed and all visits of each patient will be considered separately, as independent patients, in order to gather more data.



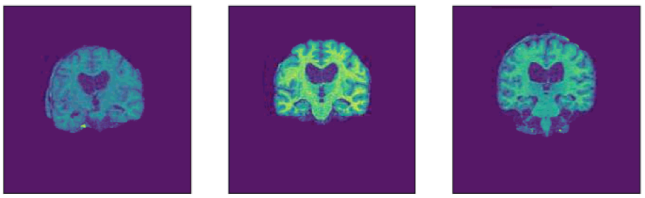
(a) Patient 722, NC | 75 | 30 - NC | 78 | 28



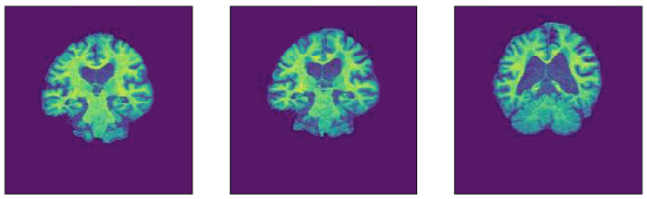
(b) Patient 1139, AD | 55 | 21 - AD | 58 | 12



(c) Patient 1102, AD | 71 | 21 - AD | 74 | 11



(d) Patient 518, NC | 67 | 28 - NC | 70 | 30



(e) Patient 28, NC | 80 | 26 - AD | 83 | 26

Figure 4.6: Progression in time of coronal slices using longitudinal information.

Chapter 5

Methodology

In this chapter, we will start by discussing the methodologies chosen to pre-process the input MRI volumes, that produced 2D slices and 3D organ shapes. These will then serve as input data types for two different architectures: a Convolutional Variational Autoencoder, used to extract latent features and walk along the generated latent space, and a Conditional Adversarial Autoencoder, used to generate progressions in time.

5.1 2D Approach

In this section, we will discuss the slicing techniques used and the 2D CVAE and CAAE models trained on them.

5.1.1 Slicing methods

One of the most simple and fundamental approaches in computer vision is tackling 2D images. When dealing with brain volumes however, relevant challenges come up. To begin with, not all patients are completely still during an MRI scan, and not all of them adopt exactly the same pose. Some skulls might, therefore, result slightly inclined or rotated with respect to others. Moreover, brains might present morphological differences between each other, in terms of shapes and sizes, because of those small characteristics that differentiate every human being from the others. In ordinary clinical practice, neuroradiologists normally check multiple slices in the MRI volume until they find the ones that allow the best view of the key regions. Therefore, extreme caution must be taken in order to extract comparable slices in a fully automated fashion from the different skulls. To this end, we consider two approaches.

5.1.1.1 Approach 1: Slices with maximum hippocampus coverage

This approach produces three slices (one for each view shown in Figure 4.2) centred on the areas where the hippocampus, extracted using the segmentation technique described in Section 5.2.1, is maximally present, leveraging the effects of the hippocampal atrophy as a strong biomarker for AD. The procedure is as follows:

1. The generated segmentation map is converted to a binary map only highlighting the voxels corresponding to the hippocampus structure;
2. The generated binary map is smoothened by morphological denoising (using a disk-shaped structuring element of radius 1) to eliminate any stray erroneous predictions;
3. The projected area of the hippocampus along the three principal axes is estimated from the smoothened binary map;
4. The slice corresponding to the maximum projected area (i.e. having the highest visual field of the hippocampus) is selected for training.

This approach did not provide sufficiently homogeneous results. A marked difference is explicitly visible from both the sagittal and the axial views, as clearly visible from the sample slices in Figure 5.1.

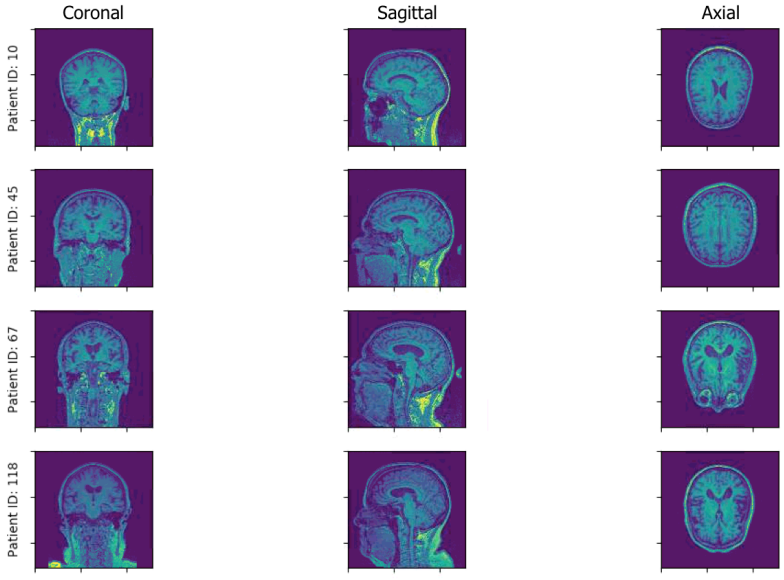


Figure 5.1: Slices with maximum hippocampus coverage: each row is a different test patients, each column a different perspective (coronal, sagittal, axial).

5.1.1.2 Approach 2: Center slices

To overcome these issues, a different approach was attempted: we localize the centre of each brain from each view and then extract the corresponding slice, resulting once again in one slice for each perspective. In order to find the very centre, the brain must be extracted from the complete 3D scan, and the centre must be localized independently from the position and orientation of the subject. The procedure is as follows:

1. The segmentation map is converted to a binary mask that highlights the brain tissues from the background. This way, we obtain the exact cubical region within the $256 \times 256 \times 256$ volume that contains only the brain tissues;

2. By mapping the corresponding indices, similar cubical region can be extracted from the normalized scan as well;

3. The centre slices are extracted from this extracted cubical region along the three principal axes. These slices are zero-padded to give consistent dimension across different scans.

This approach clearly leads to more homogeneous and thus comparable slices, as shown in Figure 5.2. Therefore, these slices can guarantee better results for the 2D models that will be presented next.

Since these volumes were drawn from different datasets, our models might tend to cluster the subjects by dataset (which would be the easiest clustering of all), based on the small differences mentioned at the beginning of Section 5.1.1 and in Section 4.2.1. For these reasons, other normalising procedures have been taken into account. In particular, before being sent into the network, the slices are thus further normalised by forcing the values in each pixel to be between 0 and 1.

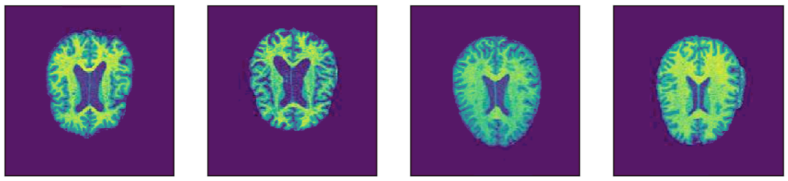


Figure 5.2: Center slices from four different test patients from axial perspective.

5.1.2 Convolutional Variational Autoencoder

To start our analysis of the the key features of AD and their evolution, we implemented a CVAE. This model is intended to combine the distinctive features of convolutional networks, that allow the creation of hierarchical features for the interpretation of images, with the idea of learning a latent distribution at the basis of variational inference. The merit of this architecture, as outlined in Section 3.1.2.2, is that of learning a latent representation that encodes useful information about the characteristics of the brain, radically reducing the dimensionality of the inputs but still keeping the most important features. The theoretical hope that is being investigated here, is whether there exists a manifold where the three classes of interest (AD, MCI and NC) have enough common features within them that allow a clean clustering of their elements. If this were the case, in order to have an idea about the diagnosis of a given test volume, based uniquely on the sMRI scan, we could send it through the encoder and check which cluster is the closest. What we are most interested in is thus to understand how samples from different classes are mapped into this manifold, and how the features change when we walk along it.

The reconstruction capabilities of the network are not the main concern here. However, they will be taken into account to have an idea about the features that the network is focusing on, even though a well-known drawback of VAEs is, in fact, the risk of generating blurry images, which could be an intrinsic effect of the maximum likelihood approach (but this is still debated [19]).

The structure of the network will be detailed later. At a very high level, it has an encoder block that receives a slice as an input, and through strided convolutions reduces the initial $[256 \times 256]$ image into an array of latent features of smaller dimension. The second block is a decoder, whose purpose is to reconstruct the original image starting from the latent features.

Because of the high dimensionality of the latent space, a practical dimensionality-reduction method is needed to visualize the results. In the following, t-SNE [21] will be used.

Starting from a dataset in a high-dimensional space (in this case the latent space), this technique returns a new, lower-dimensional representation, as close as possible to the original. This is done using a non-convex objective function, which is minimized by gradient descent. This is particularly useful

in all those situations where the dimensionality of the original data is too high to allow comprehensive visualization and understanding.

It was observed [82], that the relative size of clusters and the distance between clusters may not be meaningful in the corresponding t-SNE representations. Moreover, an important parameter of this algorithm is the perplexity: it is a function of Shannon entropy, and is conceptually close to the concept of the number of nearest neighbors K used in similar algorithms. The higher the perplexity, the more uniformly distributed the data points will look in the manifold. Since there is no best perplexity value a priori, different combinations will be used and the best results will be provided.

5.1.3 Conditional Adversarial Autoencoder

In the previous section, the CVAE was used with the objective of creating a manifold that could be walked to see the evolution of important brain features in the generated images, and where the different classes would cluster by common features.

Now, we focus our attention on the generation of a progression (or regression) in time: for this objective we use an integration of Generative Adversarial Networks and Autoencoders.

While a VAE learns to predict the posterior distribution over the latent variables, a Generative Adversarial Network (Section 3.1.2.3) shapes the output distribution of the network using backpropagation [83]. Just like a VAE however, a GAN will learn a latent representation that encodes the most relevant features in the input images[5]. For this reason, a combination of a GAN with a Conditional Autoencoder will be used: the result is a Conditional Adversarial AutoEncoder (CAAE). It is worthwhile to point out several distinctive features of the architecture:

1. The generator is modeled as an autoencoder, taking inspiration from [8]. The decoder, instead of receiving as input noise, is fed with the encodings generated by an encoder (concatenated with additional information, as explained in point 2), which in turn is fed directly with the input images. This approach is used because we want to generate artificial brains whose morphological characteristics are very close to the real ones.

2. The GAN is conditioned on the age and on the class corresponding

to the input volume, by concatenation of the encodings with a one-hot vector before sending them as inputs to the decoder and to the discriminator. For what concerns the latter, this one-hot vector is tiled and reshaped in order to match the first layer's activations' shape. This way, we are adding some supervision to the training procedure, in hopes of visualizing a progression in these volumes that is function of individual, class and age;

3. Both the encoder, the decoder and the discriminator use strided convolutions, differently from [7], allowing the network to learn its own downsampling. No batch normalisation is used, as we have experienced its tendency to produce blurred features;

4. No Total Variation (TV) loss, nor uniforming discriminator were included, as opposed to [8], to facilitate training.

5.1.4 Limitations of the 2D approach

Both architectures proved to have the potential to learn and model important characteristics of the disease. However, they are strongly dependent on the homogeneity of the input slices. While the second slicing approach (Section 5.1.1.2) in combination with normalised volumes produced better results, it still lacks robustness to morphological variations of the different brains. Moreover, considering the relatively small and time-limited dataset available, it is hard to validate quantitatively the progression of the disease.

Instead of focusing on different slicing approaches (as seen in Section 3), we decided to focus our attention on the 3D shapes, and in particular, those that are most relevant for the disease. This type of data has the strong theoretical foundations and the potential to be much better quantitatively evaluated than the 2D approach.

5.2 3D Approach

In this section, we will start by describing the segmentation approach used to extract the 3D shapes of key organs of the brain, and we will then focus on the 3D CVAE and 3D CAAE architectures used to analyse them.

5.2.1 Brain Segmentation

As mentioned in Section 2.2, characteristic features of Alzheimer's disease progression in MRI scans are hippocampal atrophy and ventricular enlargement. This can be seen from Figure 5.3, where the volume size of the right and left hippocampus as well as right and left, superior, inferior and occipital ventricles are compared for AD and NC subjects. An approach to segment these organs, in order to focus successive learning algorithms on them specifically, is discussed in the next section.

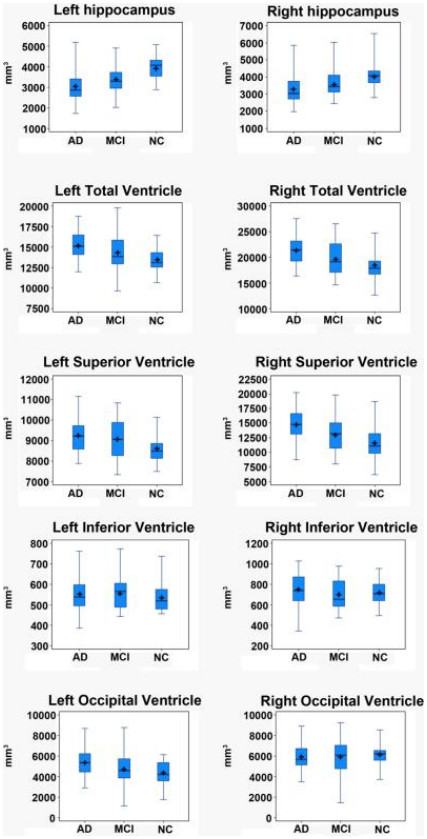


Figure 5.3: hippocampus and ventricles volumes in AD and NC (from[14]).

5.2.1.1 Segmenting via a Fully Convolutional neural network

Focusing on specific 3D regions of the brain, instead of processing the entire volume in one pass, facilitates the training on any deep architecture. Thus, we segment the original MRI volumes to obtain the shapes corresponding to the key organs and train neural networks on them only. For this purpose, a deep encoder-decoder segmentation network [15] is used to segment the MRI volumes. The architecture is showed in Figure 5.4.

This Fully Convolutional neural network (F-CNN), extends the U-Net [84], an architecture that was already used to tackle segmentation problems for its ability to segment on all image pixels without the need to be fed patches previously extracted from the image. This end-to-end training approach exploits efficiently the information from the context and is much faster than traditional techniques.
It is constituted by three main components: an encoder, a decoder and a softmax classifier. The encoder has three Convolutional - Batch Normalisation - ReLU - Max pooling blocks. The decoder is made of an Unpooling layer, a Skip-connection concatenation from the encoder, a Convolutional layer (with 7x7 kernels), a Batch Normalisation layer and finally a ReLU activation. It is interesting to note that Unpooling, which substitutes the Upsampling of the original U-Net, is based on the Max pooling operation during the encoding phase, when the indices corresponding to the maximum activations are stored to aid the Unpooling operation. For what concerns skip connections, those are added to provide a direct path for the gradients to flow and to provide more contextual information. Finally, the classifier consists of a 1x1 Convolutional layer and a Softmax layer.

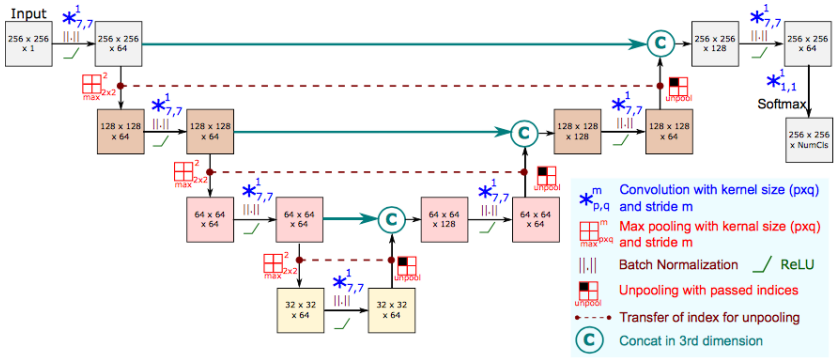


Figure 5.4: Fully Convolutional segmentation network architecture (from [15]).

The network is trained by minimisation of this loss:

$$\sum_x w(x)g_l(x)log(p_l(x)) - \frac{2\sum_x p_l(x)g_l(x)}{\sum_x p_l^2(x) + \sum_x g_l^2(x)}$$

where $p_l(x)$ is the estimated probability that pixel x belongs to the class l, $g_l(x)$ is the ground truth probability and w is a weighting coefficient. This loss is composed of two parts: a logistic loss (left term) weighted to tackle the class imbalance via median frequency balancing [85], in charge of the classification performance, and a dice loss, in charge of the reconstruction accuracy.
To overcome the scarcity of labelled data available for the training, the network is first trained on auxiliary labels generated by widely used FreeSurfer segmentation routine [67]. Now that the parameters have been meaningfully initialised, the network is fine-tuned on labelled data.

This segmentation network was applied to our three datasets: Figure 5.5 shows the results on one volume and the reconstruction of the respective hippocampus.

After the organs were extracted in the form of binary maps, padding was applied by creating 3D bounding boxes. The size of these boxes was set to the smallest size that allowed the complete containment of the biggest hippocampus. This procedure allowed the creation of a homogeneous dataset to be fed into the network.

The corresponding volume measures (in mm^3) were plotted in Figures 5.6, 5.7, 5.9 and 5.10 and compared to the results found in literature [14, 17, 39].

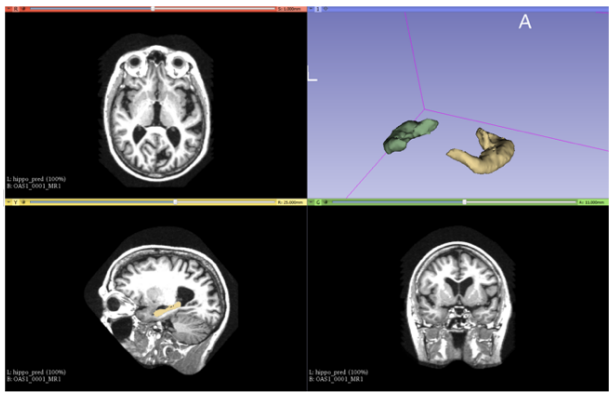
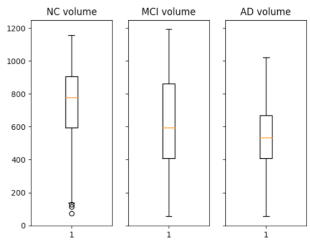
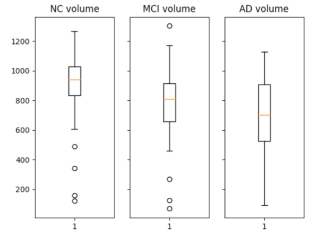


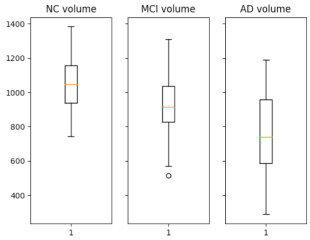
Figure 5.5: Segmented hippocampus, views obtained via FreeSurfer.



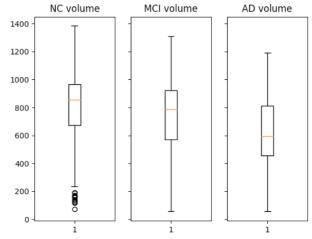
(a) AIBL Dataset



(b) OASIS Dataset

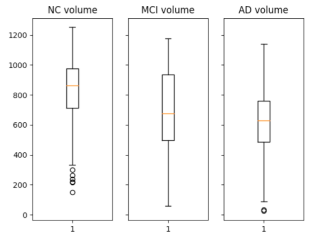


(c) HARP Dataset

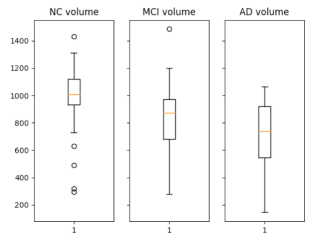


(d) All Datasets

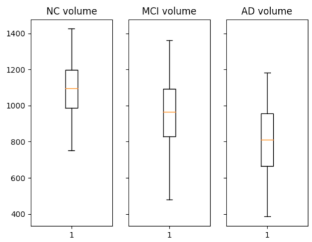
Figure 5.6: Right hippocampus volume: comparative boxplots for AD, MCI and NC.



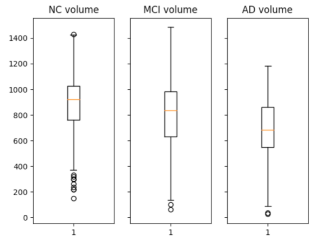
(a) AIBL Dataset



(b) OASIS Dataset

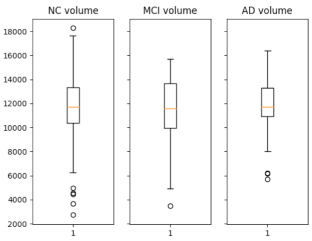


(c) HARP Dataset

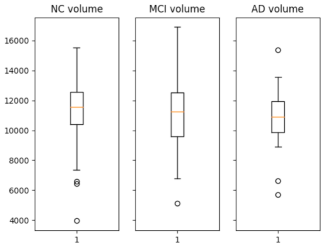


(d) All Datasets

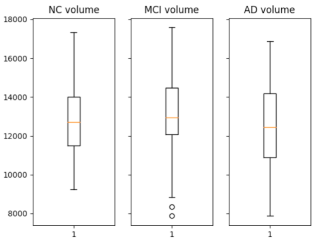
Figure 5.7: Left hippocampus volume: comparative boxplots for AD, MCI and NC.



(a) AIBL Dataset



(b) OASIS Dataset



(c) HARP Dataset

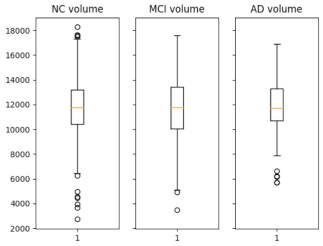
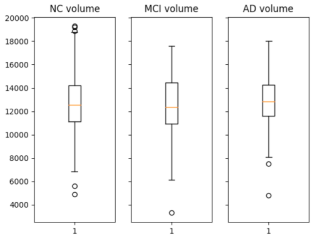
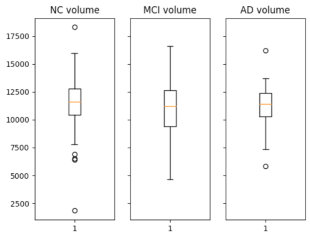


Figure 5.8: All Datasets

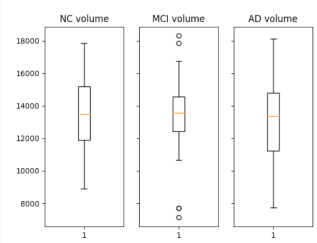
Figure 5.9: Right Ventricle volume: comparative boxplots for AD, MCI and NC.



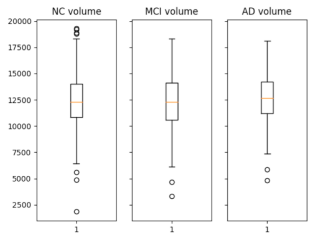
(a) AIBL Dataset



(b) OASIS Dataset



(c) HARP Dataset



(d) All Datasets

Figure 5.10: Left Ventricle volume: comparative boxplots for AD, MCI and NC.

The plottings on the hippocampus confirmed the results from the literature. Figure 5.6 and 5.7 show that the hippocampus in AD subjects is smaller, compared to NC or MCI patients. However, the same results could not be achieved on the ventricles: across the three different classes, there does not appear to be a sharp difference in terms of size, while from literature [14] we expect it to be larger in AD subjects (Figures 5.10 and 5.9). This is also true for the ventricles boxplots from medical literature (Figure 5.3) for what concerns right and left inferior as well as occipital ventricles. On the other hand, rght and left total ventricles, show stronger differences, mainly by the effect of the superior ventricles.

To verify the entity of the problem, the same boxplots were also plotted on the segmentation maps obtained by FreeSurfer [67], and no strong difference between NC and AD ventricles were shown. The reasons why the total ventricle boxplots obtained after segmentation fail to show the same effect may lie in the segmenting algorithms, whose 2D-based approach fails to reproduce the same differences faithfully.

To further investigate this problem, and the segmentation approach in general, we studied gender-based differences in important organs for AD diagnosis. The results are described in the next Section.

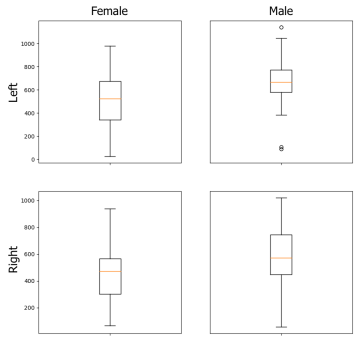
5.2.1.2 Gender-based differences

The segmented binary maps can be key to more studies for Alzheimer's diagnosis, as they allow to segment full 3D sMRI in less than 10 seconds into many organs of interest [15].

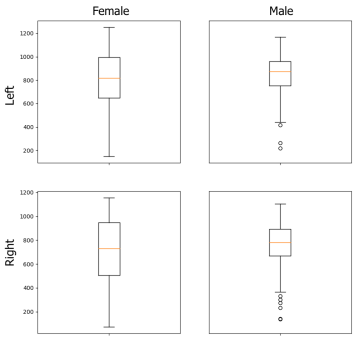
To further investigate this technique, we decided to analyse gender-based differences in NC and AD patients for what concerns the volume of hippocampus, amygdala, and lateral ventricles. In fact, this problem has not been frequently tackled in literature and remains controversial to some extent.

In order to facilitate this analysis, and to factor out possible between-dataset differences we take into account patients from one dataset only, AIBL. The volume was calculated in mm^3 for each binary map and for each interesting organ (right/left hippocampus, right/left Amygdala, right/left ventricles).

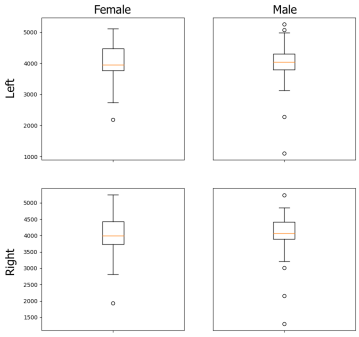
The resulting boxplots are shown in Figure 5.11.



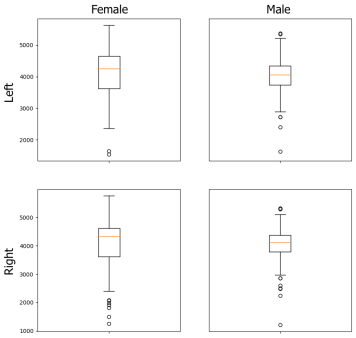
(a) Hippocampus - AD



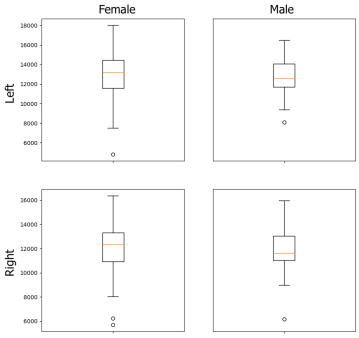
(b) Hippocampus - NC



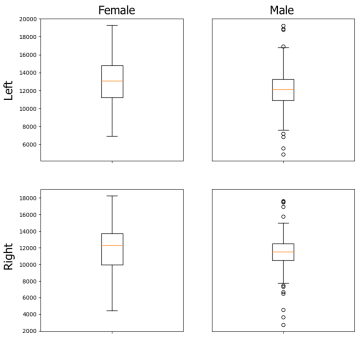
(c) Amygdala - AD



(d) Amygdala - NC



(e) Ventricles - AD



(f) Ventricles - NC

Figure 5.11: Gender-comparative boxplots on hippocampus, amygdala and ventricles volumes in AD, MCI and NC subjects.

These results show that AD patients have more prominent gender-based differences in both right and left hippocampus, which is in agreement with the literature [86]. In particular, 5.11a shows that male AD subjects present a bigger hippocampus compared to female AD subjects, and the same holds also for NC patients (Figure 5.11b). In general, by comparing these two figures one can clearly see that in AD patients (Figure 5.11a) the hippocampus tends to be smaller than in NC patients 5.11b.

The same does not hold however for other organs.
This is particularly interesting for what concerns amygdala. In fact, while some studies stated that it might be a strong biomarker of AD, and its volume a relevant measure for the early diagnosis of the disease, the general opinion about this topic is still controversial [87]. In this study, no relevant gender-based differences were found in AD patients for what concerns Amygdala. Some have pointed out that many of the studies supporting that theory did not focus on the initial phases of the disease [87]. Another reason behind this might be that the psychiatric symptoms might not be originating from morphological but rather hormonal alterations instead, as previously theorised in [88]. More generally, it can be related to a generalized lack of understanding of the factors associated with the atrophy of Amygdala in AD patients.

5.2.2 3D Convolutional Variational Autoencoder

To process these 3D organs, we design a 3D version of both the CVAE and the CAAE architectures already presented. Instead of 2D convolutions, this new Variational Autoencoder applies 3D Convolutions to the segmented binary maps. The architecture includes also ReLU activations, Dropout layers and a final Sigmoid layer.
In addition to the standard loss function used in Variational Autoencoders, which is composed of a reconstruction loss and a KL divergence, we attempted to force a better organisation of the latent space by introducing also a clustering loss, following the approach suggested in [89]. This method is inspired by K-means and it introduces nebula anchors to overcome two main limitations of Variational Autoencoders, namely tendency to overfitting and convergence to local optimum. This approach is not discussed further as it did not improve our results significantly.

Because of the presence of encouraging evidence coming from the study of hippocampus volumes shown in previous sections, more techniques were applied in an attempt to produce a more meaningful clustering of the subjects. One approach was that of applying supervised tuning after the ordinary VAE unsupervised training. This approach will be discussed in the next section.

5.2.2.1 Supervised Finetuning

After the unsupervised training of the VAE is completed, supervised information is injected by extracting the encoder and attaching three Fully Connected layers on top of it, the latter of which had three neurons (corresponding to the three main classes, NC, MCI and AD) and included a sigmoid activation function. This is a rather common approach when one wants to evaluate a latent representation [7, 13]. Being initialised with the previous weights, the new network is thus fine-tuned, in hopes of finding more meaningful clusters thanks to the supervised information. The latent space is thus compared before and after the fine-tuning using t-SNE, and accuracy of the classification is also measured as a general indicator of the success of the procedure.

Since the input data type is substantially different from the 2D images shown before, additional pre-processing has to be applied for a more effective training. This will be discussed in the next section.

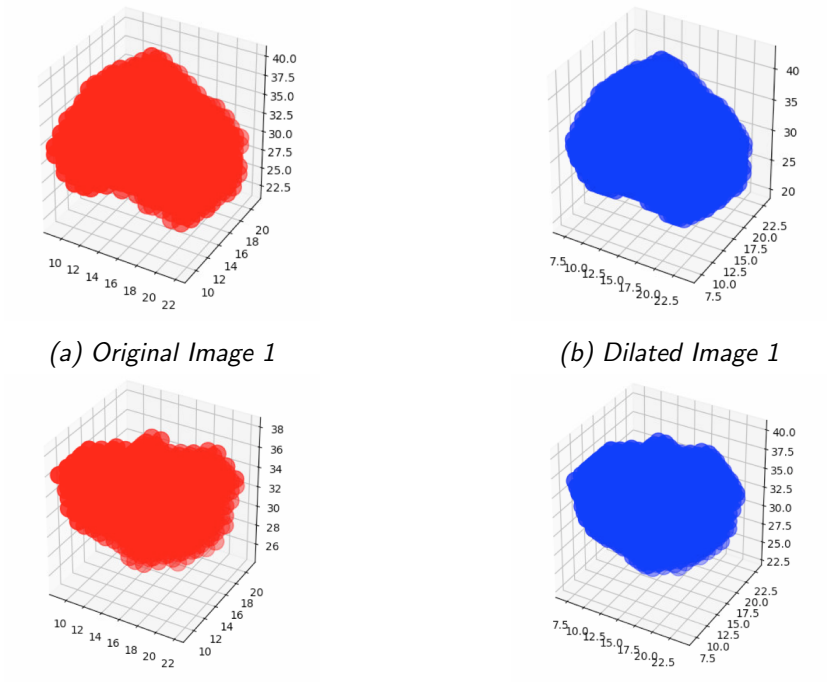
5.2.2.2 Gradient and Dilation on Binary Maps

To enhance the encoding and decoding ability of the autoencoder in correspondence of the edges and corners of the organs, which frequently represent key features in understanding their conditions, we apply image processing techniques to the binary maps.

First, we calculate the gradients on the binary maps along all three dimensions, and we sum the three obtained maps into one.

Then, we apply dilation, a morphological transformation that uses a structuring element to add pixels to the edges detected in an image. Every pixel in the output map is calculated as the maximum among all the input pixel's neighbors. In this case, since the input image is a binary, it is enough that one of the neighbors is set to 1, to have the output pixel set to 1 too.

The results of these operations on two test volumes are shown in Figure 5.12.



(a) Original Image 1 *(b) Dilated Image 1*

(c) Original Image 2 *(d) Dilated Image 2*

Figure 5.12: Effects of dilation on 3D binary maps.

To account for the class unbalance showed in Figure 4.1, we use median frequency balancing [85, 90], which results in over-represented classes having a smaller weight and under-represent classes having bigger weights.
During the supervised fine-tuning, the cross-entropy loss is multiplied by an array of weights of the same size. Each of the weights in this array depends on the frequency of the class to which the data point belongs, and in particular:

$$f_{nc} = \frac{count_{nc}}{count_{total}},$$
$$f_{mci} = \frac{count_{mci}}{count_{total}},$$
$$f_{ad} = \frac{count_{ad}}{count_{total}},$$

$$w_{nc} = \frac{median(f_{nc}, f_{mci}, f_{ad})}{f_{nc}},$$
$$w_{mci} = \frac{median(f_{nc}, f_{mci}, f_{ad})}{f_{mci}},$$
$$w_{ad} = \frac{median(f_{nc}, f_{mci}, f_{ad})}{f_{ad}}.$$

To evaluate the reconstruction ability of the network, the dice score [91] between the original binary map p and the generated binary map g was calculated as:

$$Dice(p, g) = \frac{2\sum[p(:) \times g(:)]}{\sum[p(:)] + \sum[p(:)] + \epsilon}.$$

Where $p(:)$ and $g(:)$ respectively indicate that the images must be flattened into one-dimensional arrays.

5.2.3 3D Conditional Adversarial Autoencoder

For the same reasons explained in 5.1.3, that is, to analyse the progression in time of the disease, a Conditional Adversarial Autoencoder was applied to the binary maps corresponding the key brain organs. The 3D architecture used in this case is similar to the one used on 2D images, with the exception

of the 3D Convolutional layers.

The only previous work in literature that used a 3D adaptation of Generative Adversarial Networks on point clouds was found in [7], and yet it is very different from the current model, for the following reasons:

1. In [7] the GAN and the Autoencoder are completely separated architectures (even though they are fed with each other's inputs or outputs), while in our architecture the autoencoder coincides with the adversarial generator;

2. In [7] Autoencoder and GAN are trained separately, while we traine them simultaneously;

3. The generator is fed with noise in [7], while here it is fed with the latent representation produced by the encoder;

4. The GAN in [7] is rather simple, including only Fully Connected layers, while here both the generator (thus the autoencoder) and the discriminator are convolutional.

In the next chapter, the experiments conducted on the 2D as well as 3D methodologies will be presented and explained.

Chapter 6

Experiments

In this chapter, we focus on the experimental results of the models presented in Chapter 5 on both 2D slices and 3D shapes. For what concerns the CVAE, we will first check the reconstruction capabilities of the networks, evaluating both how realistic the generated slices or shapes look and the dice score between original and reconstructed images; then, we will evaluate the expressiveness and the distribution of the latent space, as well as its reduction to lower dimensionality using t-SNE. For the CAAE, we will study its ability to generate a progression in time of the slices and the shapes.

6.1 Tools

We use Tensorflow [92] and selected python packages (numpy, scipy, sklearn, matplotlib) for all architectures. Training is done on a Nvidia's Tesla k40 GPU. All image processing techniques (Section 5.1.1) are implemented in MATLAB and all statistical tests in R.

6.2 2D Approach

In this section, we presents the results obtained on 2D slices with the Convolutional Variational Autoencoder for the latent representation and with the Conditional Adversarial Autoencoder for the progression.

6.2.1 Convolutional Variational Autoencoder

The architecture of our Convolutional Variational Autoencoder implemented as obtained from Tensorboard is shown in Figure 6.2. The encoder is composed of three Convolutional layers, each of which has stride two and kernel

size $5x5$. In fact, as explained in [5], strided convolutions allow the network to learn its own spatial downsampling. As the height and width of the images are progressively reduced along these layers, the number of channels is incremented. This can be seen as a way to compensate for the loss of information caused by the strided convolutions. After the output of the last convolution is flattened, it then flows into two Fully Connected layers: the mean and variance used for the reparameterization trick [18]. After this reparameterisation, two intermediate fully connected layers and one reshaping layer are needed to move from a 1D array to a 2D image. Three Transpose-Convolutional layers are used to reconstruct the image, thus upsampling it instead of downsampling it as done before. The latent space has dimension 150. The network is trained using RMSprop [93] with a learning rate of 0.002.

For what concerns the evaluation of the results, the quality of image reconstruction of the CVAE is qualitatively measured, even though it is not expected to achieve particular accuracy, given the intrinstic characteristics of the VAE. The axial perspective is used in this case for its view of the ventricles and their variability. The network proved to be able to reconstruct sufficiently well the images from the test set: in Figure 6.1 the original images belonging to different patients are on the left, whereas the rightmost column shows the images generated by the decoder fed with the latent features produced by the encoder on the images on the left.

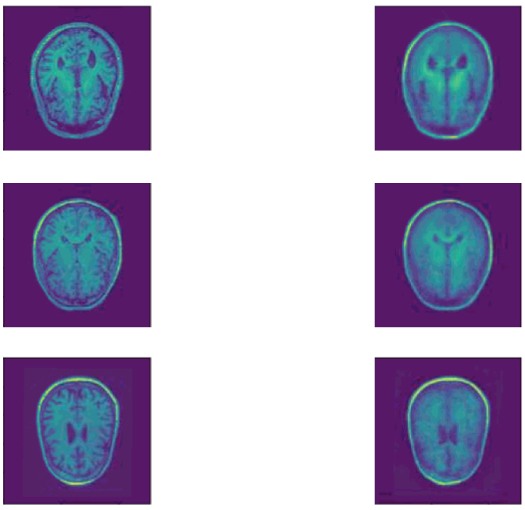


Figure 6.1: 2D CVAE: Original (left) and reconstructed (right) images.

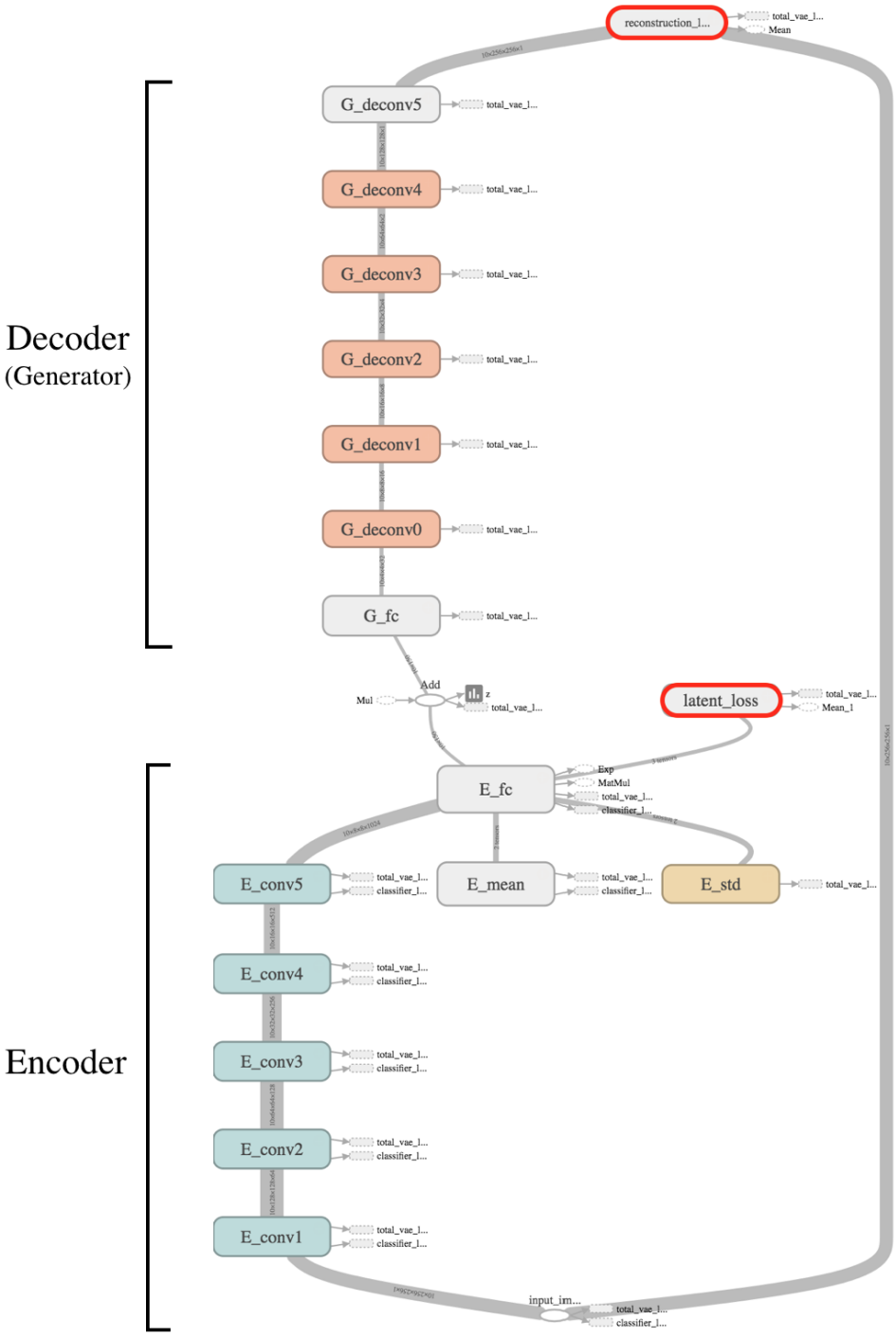


Figure 6.2: 2D CVAE: architecture.

One can also *walk* the latent space, to see whether the latent variables are actually encoding meaningful features. The images reconstructed by the generator, starting from changes in the latent variables, are then plotted and compared as in [12]. Figure 6.3 shows an interesting change in the size and shape of the ventricles, that is however still corrupted by the structure of the skull (in particular in the last row). Figure 6.4 shows the t-SNE components (obtained with *perplexity* set to 30) calculated on the latent space.

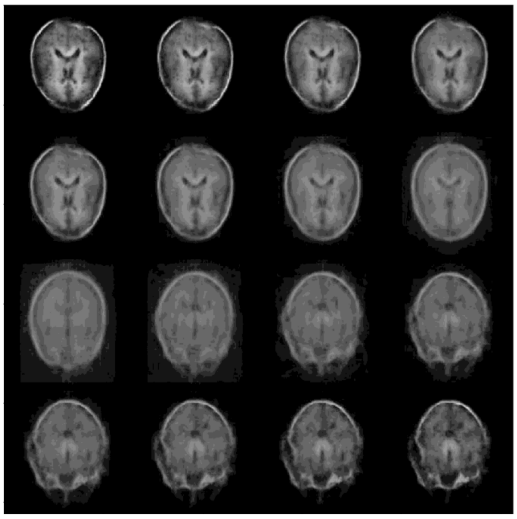
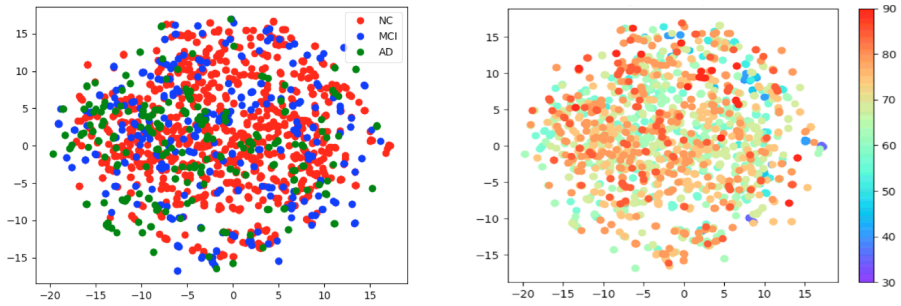


Figure 6.3: 2D CVAE: Walking along the latent space, reduced by t-SNE (all images are generated).

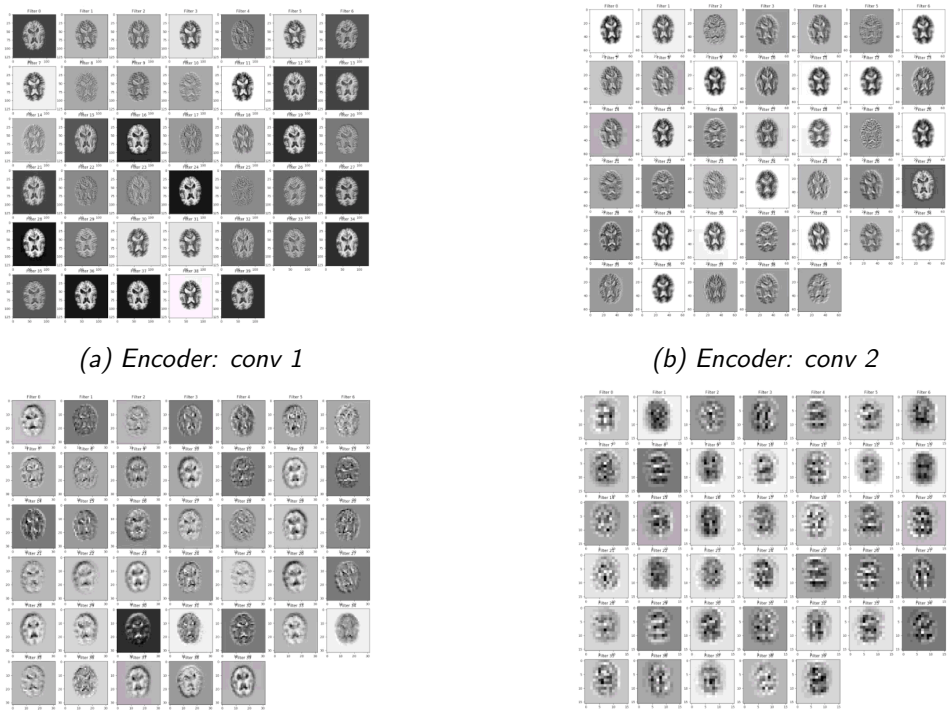


(a) By class *(b) By age*

Figure 6.4: 2D CVAE: class(a) and age (b) scatter plots from the latent space reduced by t-SNE. AD subjects seem to be more concentrated around the centre-left corner.

As predicted, the network struggled to cluster the images into classes (NC, MCI, AD) as can be clearly seen in Figure 6.4a. This is reasonable, considering that the input data is not only not skull-stripped, but is also taken from a perspective that does not facilitate the classification. However, analogous tests using skull-stripped or coronal slices did not produce better results. While it is true that there are no clear boundaries that separate the three classes, the AD class distribution seems to be mostly located around the left bottom corner of the plot. The same manifold can be seen also in Figure 6.4b, where the data points are labeled with age information. This second plot helps to clarify from a theoretical perspective the behaviour of the plots: the learner does not seem to be able to distinguish between aging and Alzheimer's.

In order to gain a more detailed understanding of what is going on inside the convolutional layers of the network, the network was elicited with test images (and compared to those in Figure 3.2). The resulting activation maps appeared to give great emphasis to the ventricles (Figure 6.5).



(a) Encoder: conv 1

(b) Encoder: conv 2

(c) Encoder: conv 3

(d) Encoder: conv 4

Figure 6.5: 2D CVAE: activation maps from four convolutional layers.

6.2.2 Conditional Adversarial Autoencoder

The Conditional Adversarial Autoencoder described in Section 5.1.3 is trained using the architecture shown in Figure 6.6 with the objecting of generating a progression in time of the brain. It is composed of a generator and a discriminator: the former includes an encoder and a decoder. The encoder is composed of 6 Convolutional layers, while the decoder is composed of 8 Transpose-Convolutional layers. The decoder receives as input the latent variables concatenated with the age and diagnosis information of each volume as a 1D vector. The discriminator includes 4 Convolutional and 2 Fully Connected layers. It receives as input both the artificial and the real images, and it is provided with the age and class information, that are first reshaped into a 3D tensor and then concatenated to the output of the first Convolutional layer.

All convolutions have stride 2 and kernel size 5. The network is trained using Adam [94] optimiser, which can be seen as a variation of the combination of RMSProp [93] and momentum [19], and tries to minimise two losses: a generator loss, constituted by the sum of the reconstruction loss and the ordinary GAN generator loss, and a discriminator loss.

The first experiment was carried out using one centre slice for every volume. This experiment turned out to be unsuccessful as it was affected by mode collapse, a rather common issue encountered in GANs training [95], [58]. In this case, the generator has learned to generate one single artificial image that the discriminator was not able to detect. This image incorporated the most common features found in all images: the network has found a cheap way of tricking the discriminator, and is not wasting resources in trying to learn more rare features. This is very clear in Figure 6.7, which shows that identical images are generated by the network when fed by 10 different test images. The original images are not shown here, since a comparison would not make much sense for the reasons just mentioned.

As mentioned earlier, GANs are notoriously very hard to train, as slight changes in the hyperparameters can result in huge changes in the network's performance and reconstruction capability. In this case, modifications in the learning rate, optimiser, frequency of discriminator training as opposed to generator training (as suggested in [58]) did not result in any visible improvement. Including Batch Normalisation after the Convolutional layers proved unsuccessful too, as well as a Wasserstein-style [58] training procedure.

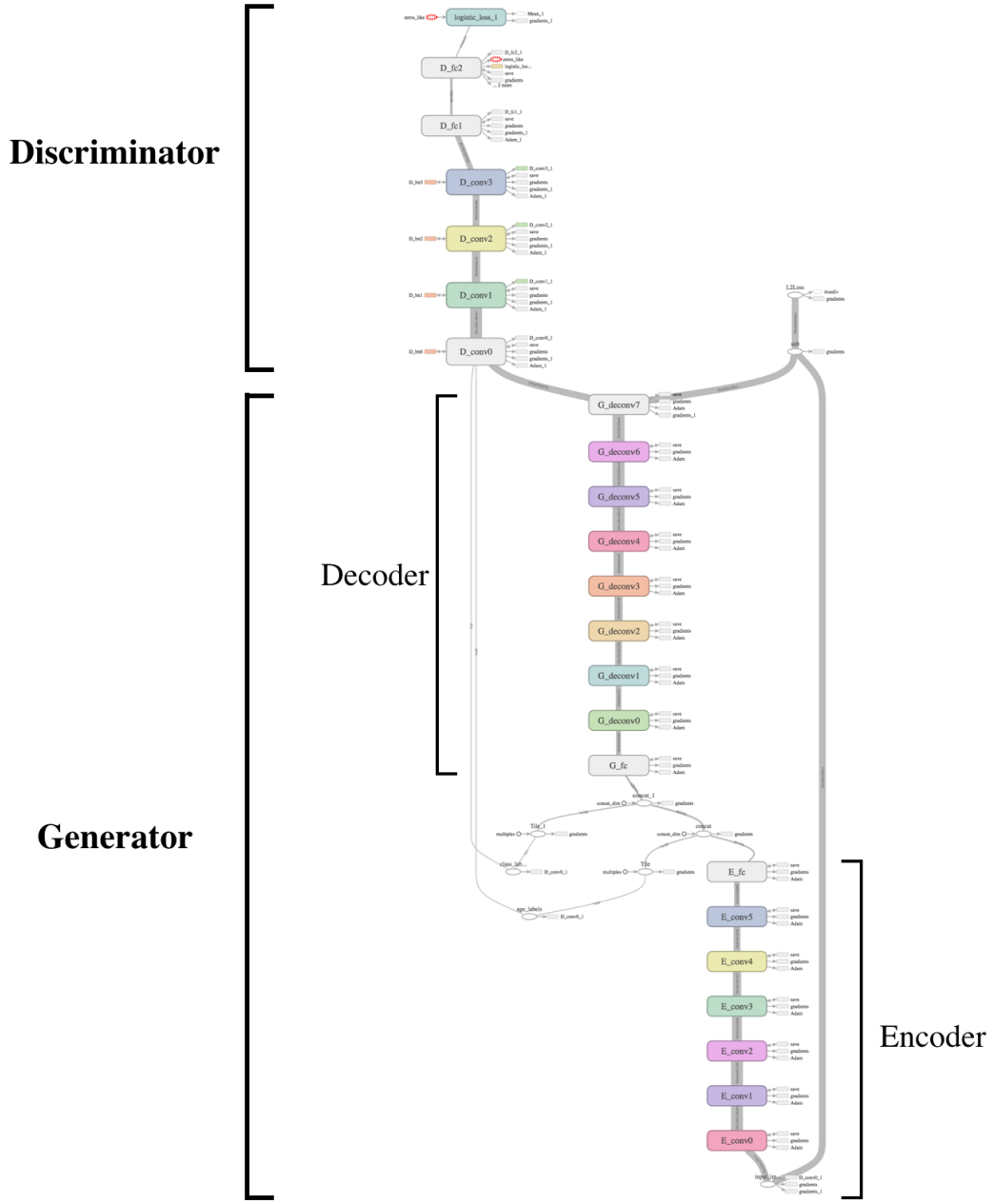


Figure 6.6: 2D CAAE: architecture.

The real cause behind the poor performance of the network turned out to be the size of the dataset (950 slices), too small for a network of this size. Therefore, instead of feeding only one slice per volume, we feed four centre slices of each volume directly into the network, thus augmenting the dataset by a factor of 4. After about 100 epochs of training, the network is able to generate images showing a relatively realistic artificial MRI scans. We compare the original images to the reconstructed ones in Figure 6.8, where the two versions are located in the same position in 6.8a and 6.8b respectively.

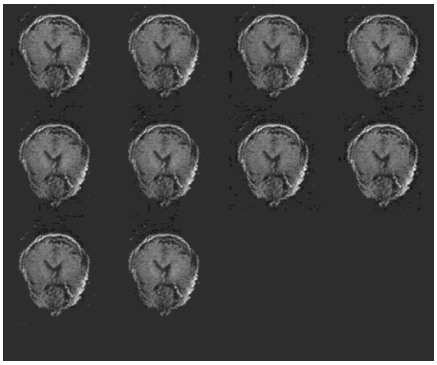
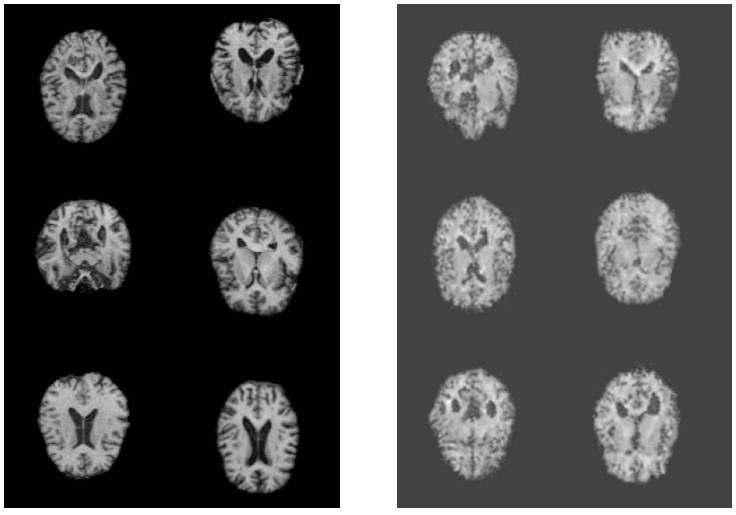


Figure 6.7: 2D CAAE: the problem of mode collapse detected during training. All slices are generated, and they look almost equal.



(a) Original Images *(b) Generated Images*

Figure 6.8: 2D CAAE axial reconstruction: original and reconstructed images are located in the same positions in (a) and (b) respectively.

Since during training class and age information are fed into the network too, one can use this information to check whether the model has recognised any pattern in the progression of the disease. To do this, we pick a test image and send it through the network five times, each time with a different age label. Only one of these labels corresponds to the real age of the image. As for the diagnosis, for this test we keep the same label that is associated with the input: the purpose of this test is to see whether the network can grasp relevant differences in the progression of AD, MCI or NC subjects, and then reproduce them clearly in order to allow an expert eye to recognise them. The network is thus being asked to reconstruct how that patient, with that diagnosis, would look like at each age.

In Figure 6.9 each column represents a different progression and the rows evolve from youngest age (first row) to the oldest (last row). Each row represents a different age interval, from 50 to 100 years old.

To understand whether these progressions have the characteristics of real progressions, we asked for expert feedback from Dr. Barvas and Dr. Sintini, Ospedale di San Marino. While this does not count as a rigorous validation procedure, it still allowed us to understand whether the network was learning something meaningful or not. The observations that follow were provided by our experts, and account for a necessary approximation, considering also the low resolution (256×256) of the images.

First of all, the images themselves were considered realistic for being completely artificial. Moreover, some of the progressions, specifically the second and third columns, were judged realistic-looking, even if with limitations. One of the most relevant issues is that skull shape should not be varying visibly from one age-bin to the next, while this seems to happen quite explicitly in the first column, for instance. Moreover, some progressions look quite similar, so the network seems to have learned one "standard" type of progression that is not, at least visibly, dependent on the individual, morphologic features of each brain.

Overall, we can conclude that the network learned a progression that includes important characteristics: a general reduction in the cerebral volume on a cortical level, more clear wrinkles and larger ventricles. For what concerns AD progression, it is very hard from this perspective to judge whether these progressions are actually encoding the difference between an AD progression and a NC one for instance.

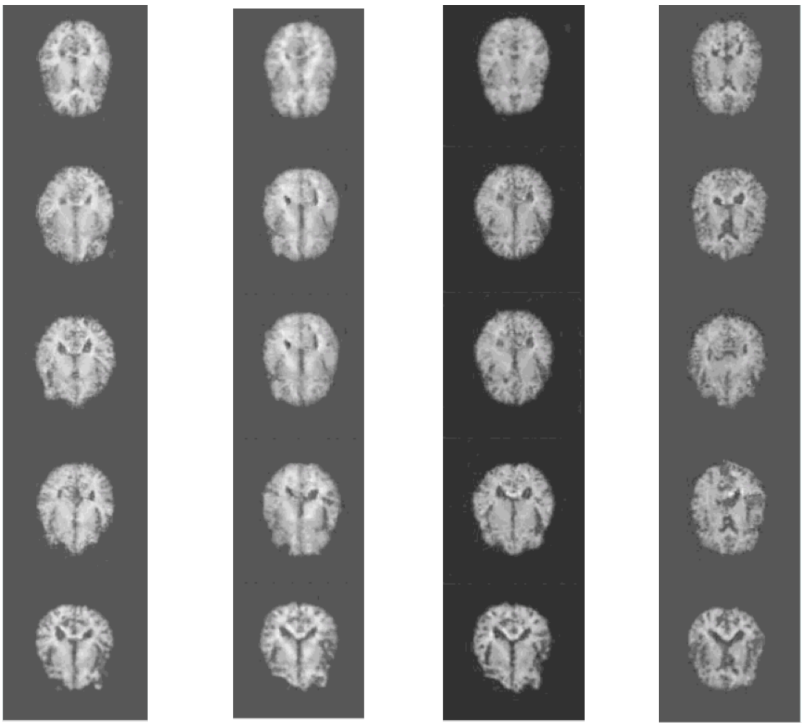


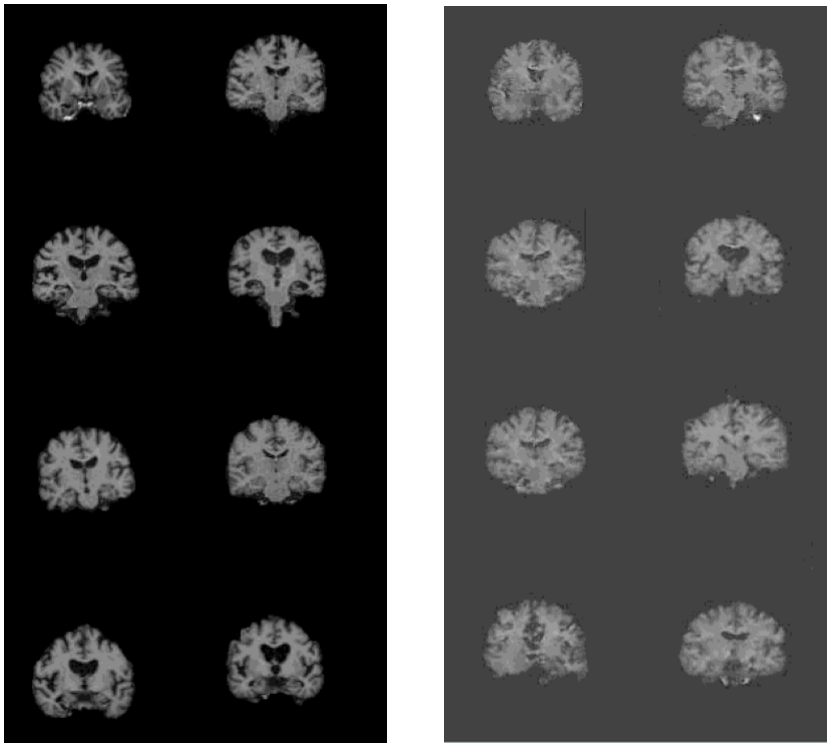
Figure 6.9: 2D CAAE: axial progression on four test images.

For this reason, following the suggestion of [39] we chose a coronal perspective, as it is perpendicular to the hippocampus and temporal horn axis. We therefore applied the same procedure to coronal slices.

Figure 6.10 shows the reconstruction ability of the network: once again, for every slice, we present the original and the reconstructed version in the same position in Figures 6.10a and 6.10b respectively. We can notice how this time the reconstruction is much more similar to the input image.
Figure 6.11 shows the progression of a test subject from youngest (left) to oldest (right). In this case, the progression seems more coherent, in particular in its extremes: the ventricles are enlarging as the patient gets older. Moreover, the grey matter becomes increasingly marked, which is not only a sign of Alzheimer's but also a sign of aging. However, this progression still presents flaws, as some of the slices reproduced by the network seem to belong to slightly different positions in the brain.

Overall, as Dr. Barvas noted, on one hand the global cerebral volume is reducing, specifically concerning the temporal lobes, whose wrinkles become

more visible; on the other hand however, the skull shape is still not entirely preserved. In fact, some steps in the progressions still seem to suffer from glitches in the skull stripping process (for instance in the third row), from a lack of data in specific age ranges and from the disadvantages of automatically picking slices following a suboptimal 2D slicing approach.



(a) Original Images *(b) Generated Images*

Figure 6.10: 2D CAAE coronal reconstruction: original and reconstructed images are located in the same positions in (a) and (b) respectively.

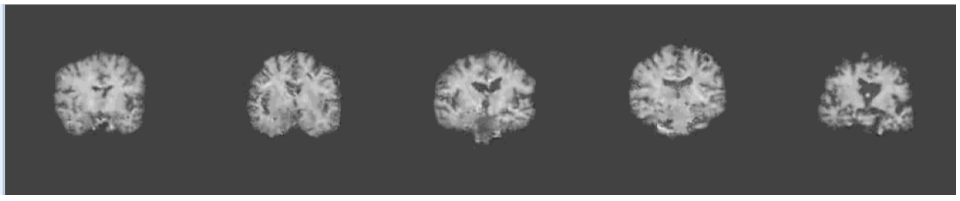


Figure 6.11: 2D CAAE: coronal progression on four different test images.

6.3 3D Approach

In this section, we present in detail the 3D versions of the CVAE and the CAAE architectures discussed in the previous chapters. Moreover, we discuss the experimental results obtained in terms of reconstruction capability, clustering, dice score and classification accuracy. We will also study the progression of these subjects along the latent space and through time, asking experts to evaluate this progression.

We focused our experiments mainly on the key regions for Alzheimer's Disease: right and left hippocampus, right and left ventricles. The best results were obtained on the hippocampus, whose segmentation already proved to be rather robust in Section 5.2.1.1, so the next sections will focus on this organ only.

6.3.1 3D Convolutional Variational Autoencoder

The CVAE architecture, obtained using Tensorboard, is shown in Figure 6.12.

The encoder is made of 2 Convolutional and 3 Fully Connected layers, two of which are responsible respectively for the mean and the standard deviation (used in the reparameterization trick [18]). The decoder is composed of one Fully Connected and 4 Transpose-Convolutional layers. The green node that connects the encoder to the decoder is in charge of the parameterization trick, which prevents the variational component of the architecture from impeding the backpropagation through the entire network. The latent space has dimension 150.

More details concerning the motivation behind the increasing number of channels, the use of strides as opposed to pooling layers and the reparameterization trick are presented in Section 6.2.1 and are thus not repeated here.

Once the unsupervised training is completed, the decoder is unplugged, and on the top of the encoder a classifier network is attached. This classifier includes 3 Fully Connected layers, each followed by Dropout layers with a *keep probability* of 0.5. All Convolutional layers are strided and have a kernel size of 5.

Figure 6.12 also shows in red the two losses minimised by the CVAE, namely the latent loss (KL divergence) and the reconstruction loss (L2 loss).

The autoencoder is trained using RMSprop [93] with a learning rate of 0.001, while the classifier is trained using Adam [94] with the same learning rate.

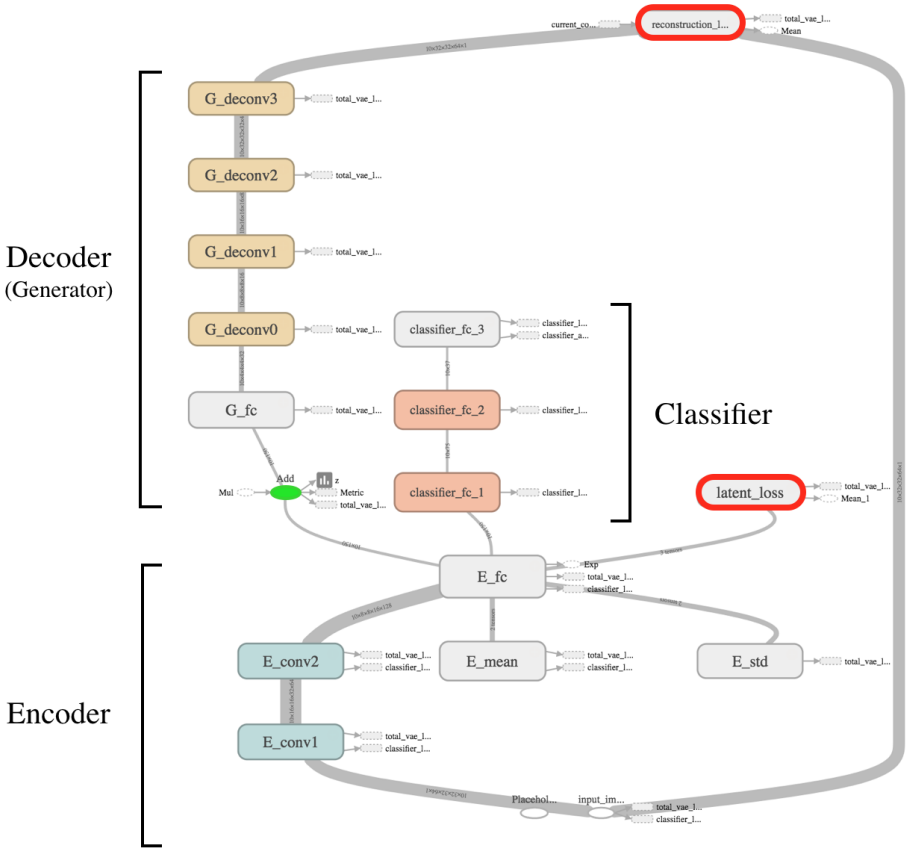
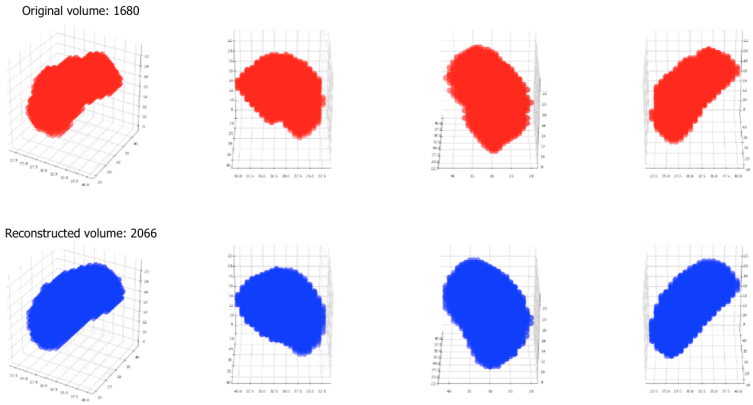
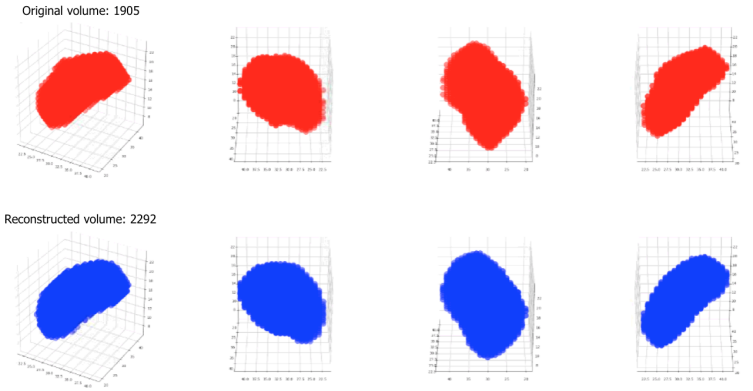


Figure 6.12: 3D CVAE architecture.

Figure 6.13 shows the reconstruction performance of the VAE on two test images, each of which is shown from four different points of view (along each row). The red binary maps represent the originals, while the blue are the reconstructed ones. These binary maps are plotted as scatter plots, in an attempt to visualise sharp edges or corners without approximation. In both cases, the reconstruction images are smaller in volume with respect to the original ones, but the dice scores are high enough to prove a meaningful reconstruction: 0.87 and 0.86 respectively.



(a) Test patient 1: dice score 0.87



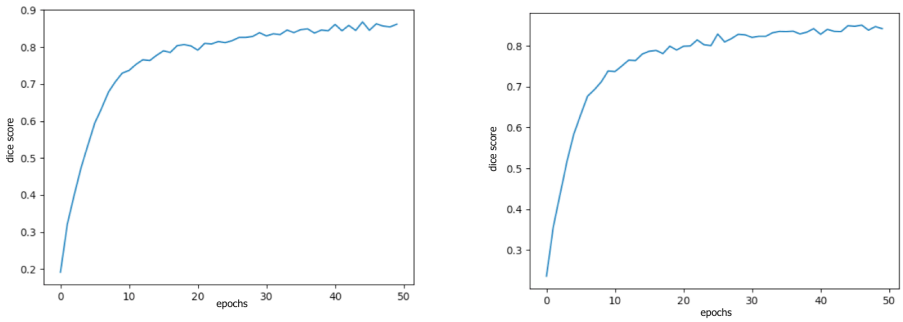
(b) Test patient 2: dice score 0.86

Figure 6.13: 3D CVAE: hippocampus reconstruction on two test patients.

To evaluate quantitatively the accuracy of reconstruction, the dice score was plotted along all training epochs on both training and test set (Figure 6.14). It should be noted that a dice score of 0.8 is generally considered very good performance for these types of problems [15].

After the supervised fine-tuning, we apply t-SNE to reduce the latent space to two dimensions for better visualisation. The scatter plots show a more meaningful correlation with the corresponding labels, as shown in the figures below. While the training set is obviously perfectly clustered, the test set shows a concentration of AD patients close to the lower left corner, with limited exceptions. In the same area, however, other MCI and even a few NC patients are located. The reason behind this distribution is most likely the fact that it is uniquely based on the hippocampus: in some of these cases, the network's confusion can be explained with aging (Figure 6.16), while in other cases it might be due to morphological characteristics or to other effects that are not visible from MRI (Section 2.2).

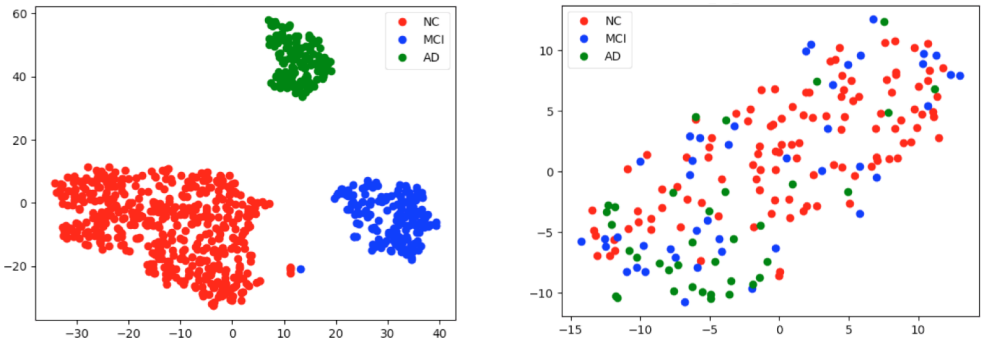
Overall, this scatter plot is much more meaningful compared to the one showed in Figure 6.4. The Silhouette score [96] could not provide a good measure of this improvement, since the three clusters are still overlapping (the value of the score would have been around 0). We therefore follow the approach showed in [12] and use the Pearson's correlation to check whether there exists a correlation between the two t-SNE components (x and y) and the labels. The results presented in Table 6.1, show a very low p-value for class and age (in bold), indicating a relevant correlation. The p-value is also much bigger for gender and membership to a dataset, which indicates that the network is dedicating its resources to meaningful features.



(a) Train set: 0.87

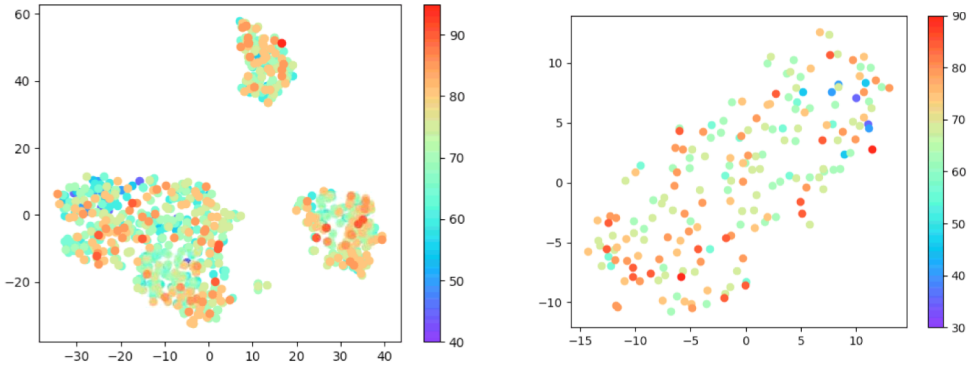
(b) Test set: 0.83

Figure 6.14: 3D CVAE: dice score.



(a) Train set *(b) Test set*

Figure 6.15: 3D CVAE: class scatter plots.



(a) Train set *(b) Test set*

Figure 6.16: 3D CVAE: age scatter plots.

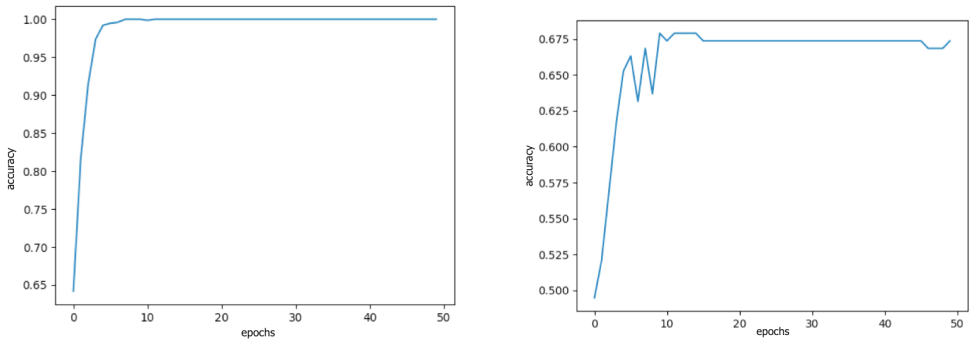
	Statistic (x)	p-value (x)	Statistic (y)	p-value(y)
Class	-0.347	**9.268e-07**	-0.527	**5.504e-15**
Age	-0.287	**5.896e-05**	-0.310	**1.354e-05**
Gender	-0.135	**0.0625**	0.007	0.921
Dataset	0.022	0.766	-0.057	0.4354
MMSE	0.095	0.191	0.204	**0.005**

Table 6.1: Pearson correlation between the two t-SNE components (x and y) and the labels.

Finally, to further test whether the extracted latent dimensions are somehow correlated with the respective classes, we also measure the classification performance (in Figure 6.17). This classification accuracy is solely reached based on the hippocampus, which is by far not the only relevant factor for AD diagnosis (Section 2.2). The final accuracy on the test set is 67.5% which is still subject to strong overfitting (as we can see from the accuracy of the training set). While this result is clearly not optimised, nor comparable to state of the art results, it still shows a correlation between the latent features and the diagnosis.

On the trained and fine-tuned latent space, the longitudinal information present in the dataset AIBL (see Section 4.2.2) was used to see whether it could provide any insight on the trajectory followed by the development of the disease. Once the latent space has been synthesized by t-SNE, it can be used to plot the progression of the 123 patients for whom longitudinal information is available (baseline, +18 months, +36 months).

In Figure 6.18 only a subset of 5 patients is shown for clarity reasons. The darkest extreme of the line represents the situation of the patient at baseline, while the lightest represents the situation thirty-six months after. The annotations, unfortunately not particularly readable, describe starting and ending conditions of the patients (age, MMSE score, diagnosis) .



(a) Train set (b) Test set

Figure 6.17: 3D CVAE: classification accuracy.

This plot is undeniably difficult to disentangle, mostly because of the very limited information available for each patient. For instance, patient 518 starts from an MMSE score of 28 at the baseline only to present an MMSE score of 30, only 36 months later: the cognitive abilities are thus improving, instead of declining. This is not a unique case in this dataset, as other patients have shown the same behaviour. Similar cases need much more information on the clinical conditions of the patient in order to be interpreted: what can be supposed in here is that there has been a relevant change in the mood or psychiatric conditions of the patient, or a different physician with a slightly different perception examined the patient.
This plot will not be examined further, since its results do not provide more insights.

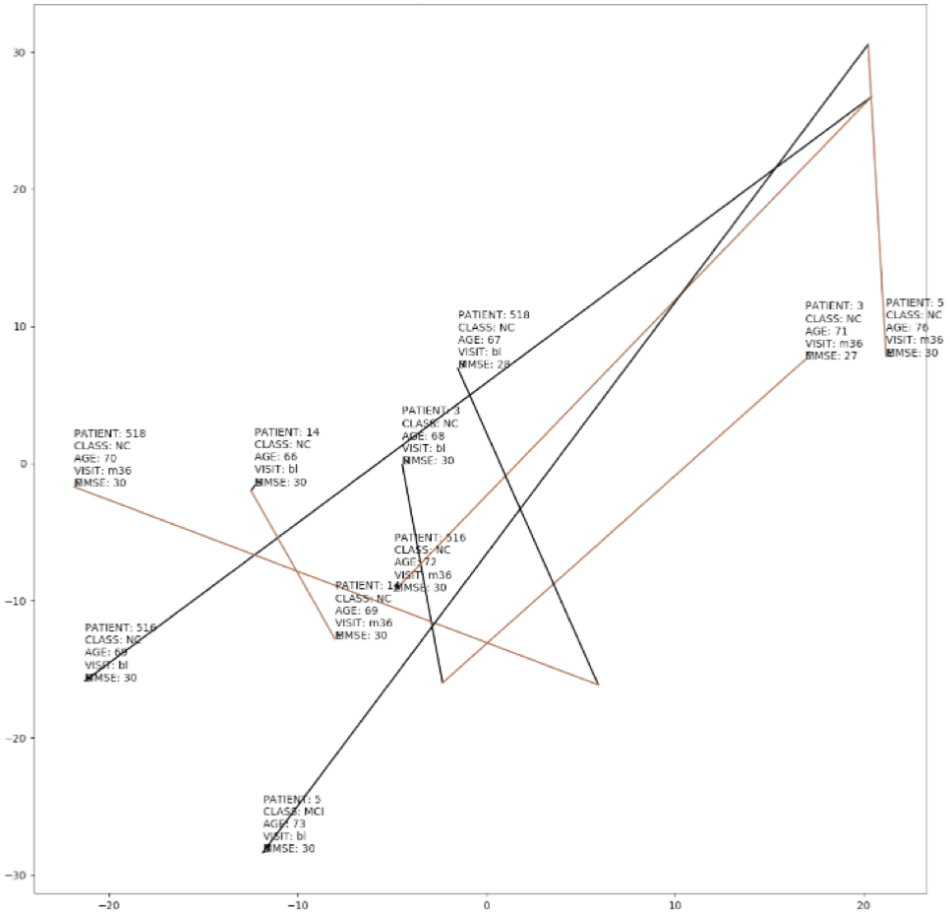


Figure 6.18: 3D CVAE: progression on five test patients.

6.3.2 3D Conditional Adversarial Autoencoder

The 3D architecture explained in Section 5.2.3 was implemented as an extension of 2D version used in Section 6.2.2, this time leveraging 3D Convolutions and binary inputs. Just as with the previous CVAE, the network was first evaluated on the test set by its reconstruction capabilities (Figure 6.19), where the red images represent the originals and the blue the generated ones. The dice score was also plotted in Figure 6.20b.

Then, the progression is tested following the same methodology from (Section 6.2.2) and shown in Figure 6.21: every row represents the reconstruction of the hippocampus from 4 different points of view. Again, the rows progress from the youngest age (first row) to the oldest age (last row).
Differently from before however, this progression can be quantitatively evaluated. In fact, not only we are expecting a reduction in the volume of the hippocampus as the patient ages, but this trend is also expected to be more marked in AD patients than in NC patients.

Hence, the volume decrease from the beginning (youngest age) to the end (oldest age) of the progression was calculated in terms of percentage of the initial volume. It is important to note that both volumes considered for this calculation are generated volumes (never the original ones), in order to factor out any generator-related trend in the reconstruction. It must also be noted that the absolute value *per se* should be taken into account only with caution, as the decreasing trend is the most relevant aspect here.

Two types of tests were carried out on the test set. First, a one-sample t-test was performed to check whether a decrease in volume through time was statistically significant. This test considered all patients in the test set independently from their diagnosis: it was assumed that the diagnosis was kept constant throughout the progression. The H1 hypothesis was therefore that the mean percentage volume decrease was greater than 0.

Second, a Welch Two-Samples test was carried out on two sets of percentages: the first one, was based on the progression obtained when AD diagnosis was assigned to all samples from the test set, the second when NC diagnosis was assigned instead. The H1 hypothesis was in this case that the mean decrease with AD diagnosis is greater than with NC diagnosis. The low p-value confirms that this is the case.
Results for the right hippocampus are shown in Table 6.2.

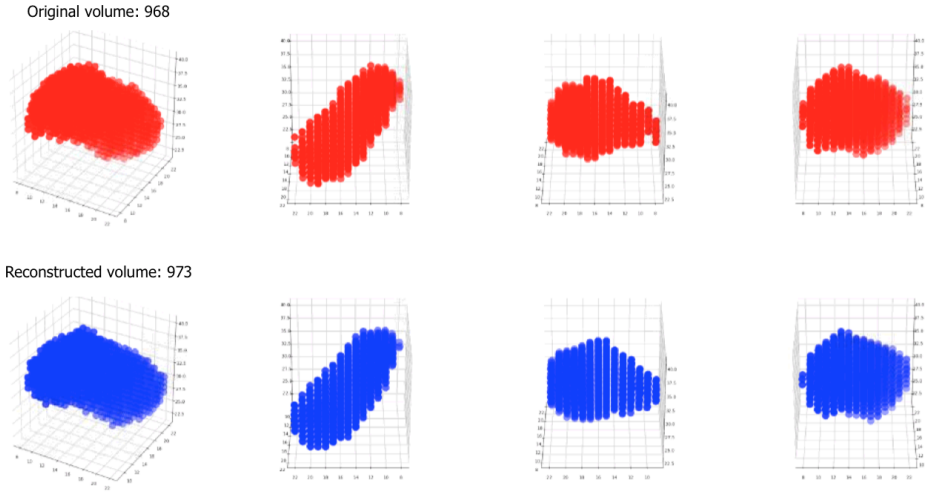
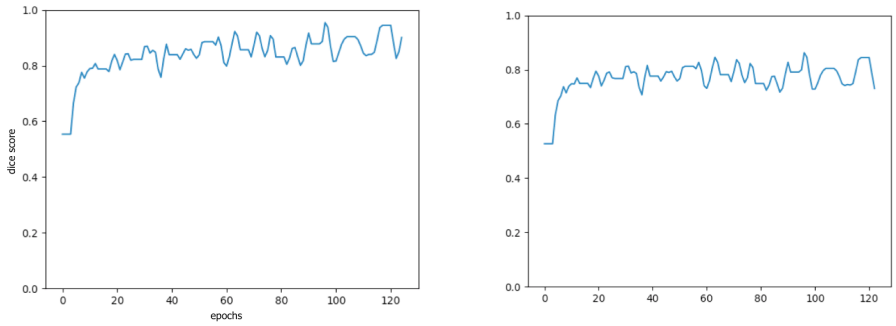


Figure 6.19: 3D CAAE: right hippocampus reconstruction on a test patient. Dice score: 0.9.



(a) Train set: 0.87 *(b) Test set: 0.83*

Figure 6.20: 3D CVAE: dice score.

Test	One Sample t-test	Welch Two Sample t-test
H1	mean percentage decrease >0	difference in mean (AD vs NC)>0
t-statistic	5.399	1.7915
p-value	2.754e-07	0.03746
Mean decrease	3.96%	AD = 4.79% NC = 3.60%

Table 6.2: 3D CAAE: t-tests on AD and NC decrease percentage.

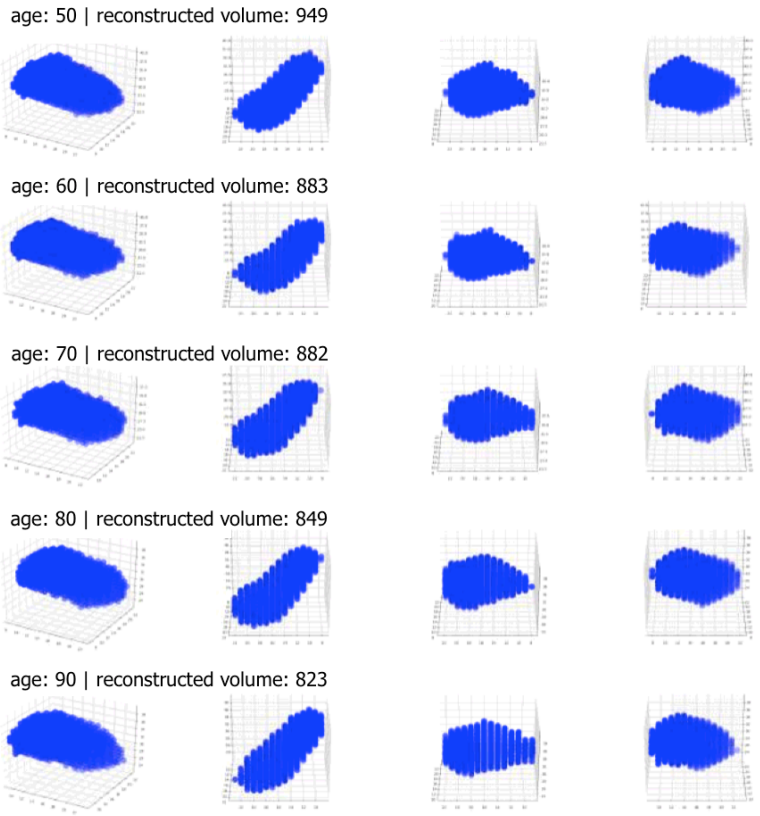


Figure 6.21: 3D CAAE: right hippocampus progression in time.

Chapter 7

Conclusions and Future Work

Starting from Structural MRI scans, we studied both 2D slices as well as 3D AD-related organs. For what concerns slices, they were obtained using various slicing approaches, while the organs were extracted using a Fully Convolutional network. We then used Convolutional Variational Autoencoders and Conditional Adversarial Autoencoders to extract fundamental features for AD diagnosis, to learn a manifold that can be walked to morph the inputs in terms of shape, volume and morphological characteristics, and finally to generate a progression and a regression of the volumes through time. Both these models integrate supervised and unsupervised approaches.

While the 2D approach presented relevant limitations, mainly due to the slicing techniques that were not able to extract a completely homogeneous and AD-relevant set of slices, it still produced interesting latent features and meaningful progressions. On the other hand, the study of 3D shapes proved to be produce a more meaningful latent space, which correlates better the the labels, and allowed a more quantifiable progression in time. In particular, the progression in time generated by the network showed a decrease in the volume of the hippocampus that is statistically relevant. This decrease is more marked in AD than in NC subjects, which is confirmed by literature.

Overall, the results presented in this study are very encouraging from the perspective of using deep generative methods to study the progression of Alzheimer's Disease. Several improvements are already worth mentioning.

The 2D slicing approaches can be made more elaborate and efficient,

thanks to the useful input (concerning for instance the location of important organs in the brain) provided by the segmentation. This methodology was already attempted in Section 5.1.1.1, but many variations and improvements are still possible.

Improved slicing combined with Convolutional layers, activation maps and supervised labels can provide useful information concerning the features found in the images that most frequently trigger a certain classification result. This could be a very important tool in the hands of neuroradiologists to potentially discover new insights about the disease.

The segmentation algorithm used in this work is extremely fast, and can thus be used to extract many different organs very efficiently. Hence, the same algorithms applied in this work can be applied to different organs that are not yet associated with the disease, in order to study their behaviour and discover potential new links with the disease. These organs can, for instance, also be fed to different networks that are trained in parallel, following an ensemble approach: integrating different organs, thus different viewpoints from the same brain, that are fed in a simplified yet essential format, can help the training figure out (and focus on) the most important features of Alzheimer's Disease.

The Conditional Adversarial Autoencoder is a very promising approach to the generation of progressions of images or shapes. Future work could include more information in the conditioning part of the network (such as gender, cognitive tests, blood examination or CSF values) for a more accurate prediction.

For most of these advancements to be robust and meaningful, gathering more data, especially from AD patients and through a longer period of time, is of paramount importance.

Bibliography

[1] B. Birur, N. V. Kraguljac, R. C. Shelton, and A. C. Lahti. Brain structure, function, and neurochemistry in schizophrenia and bipolar disorder - a systematic review of the magnetic resonance neuroimaging literature. *NPJ Schizophrenia, 3, 15*, 2017.

[2] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. *NIPS*, 1989.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems (NIPS), 1097-1105*, 2012.

[4] P. Bojanowski, A. Joulin, D. Lopez-Paz, and A. Szlam. Optimizing the Latent Space of Generative Networks. *arXiv:1707.05776v1 [stat.ML]*, 2017.

[5] A. Radford, L. Metz, and S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv:1511.06434v2 [cs.LG]*, 2015.

[6] T. D. Kulkarni, W.F . Whitney, P. Kohli, and J. B. Tenenbaum. Deep Convolutional Inverse Graphics Network. *Advances in Neural Information Processing Systems 28 (NIPS)*, 2015.

[7] A. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas. Representation Learning and Adversarial Generation of 3D Point Clouds. *arXiv:1707.02392v1 [cs.CV]*, 2017.

[8] Z. Zhang, Y. Song, and H. Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[9] M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage; 145(Pt B):137-165*, 2016.

[10] M. Liu, E. Zhang, D. Adeli-Mosabbeb, and D. Shen. Inherent Structure Based Multi-view Learning with Multi-template Feature Representation for Alzheimer's Disease Diagnosis. *IEEE Trans Biomedical Engineering; 63(7): 14731482*, 2016.

[11] N. Amoroso, M. La Rocca, s: Bruno, T. Maggipinto, A. Monaco, R. Bellotti, and S. Tangaro. Brain structural connectivity atrophy in Alzheimer's disease.

[12] R. Brosch, T. andTam and Initiative for the Alzheimer's Disease Neuroimaging. Manifold learning of brain MRIs by deep learning. *MICCAI 16 (Pt 2):633-40*, 2013.

[13] E. Hosseini-Asl, R. Keynton, and A. El-Baz. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. *IEEE International Conference on Image Processing (ICIP)*, 2016.

[14] L. G. Apostolova, A. E. Green, S. Babakchanian, K. S. Hwang, Y.-Y. Chou, A. W. Toga, and P. M. Thompson. Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment and Alzheimer's disease. *Alzheimer's Disease and Associated Disorders, 26(1), 17 - 27*, 2012.

[15] A. G. Roy et al. Error Corrective Boosting for Learning Fully Convolutional Networks with Limited Data. *Proceedings Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017 - 20th International Conference*, 2017.

[16] 2017 Alzheimer's Disease Facts and Figures. *Alzheimer's Disease Association*, 2017.

[17] L. G. Apostolova. Alzheimer's Disease. *Continuum: Lifelong Learning in Neurology 22. 2 Dementia, 419-434*, 2016.

[18] D. P. Kingma and M. Welling. Auto-Encoding Variational Bayes. *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, 2014.

[19] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[20] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. *arXiv:1406.2661v1 [stat.ML]*, 2014.

[21] L. Van der Maaten and G. Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research 9 2579-2605*, 2008.

[22] What is Alzheimer's? *Alzheimer's Disease Association*, 2017.

[23] R. S. Richardson. Vascular factors associated with healthy ageing: new evidence in the brain and muscles. *Convegno 'Il ruolo dello stile di vita in un invecchiamento di successo' - Fondazione ONLUS Mons. A. Mazzali - Mantova*, 2017.

[24] E. Wang. Impact of strength training on the ageing neuromuscolar system. *Convegno 'Il ruolo dello stile di vita in un invecchiamento di successo' - Fondazione ONLUS Mons. A. Mazzali - Mantova*, 2017.

[25] A rare Success Against Alzheimer's. *Scientific America*, 2017.

[26] A. Solomon, F. Mangialasche, E. Richard, S. Andrieu, D. A. Bennett, M. Breteler, and M. Kivipelto. Advances in the prevention of Alzheimer's disease and dementia. *Journal of Internal Medicine, 275(3), 229250*, 2014.

[27] How Sleep Clears the Brain. *National Institutes of Health*, 2013.

[28] What nuns are teaching us about Alzheimer's. *Alzheimer's Association*, 2017.

[29] M. J. De Leon, A. E. George, J. Golomb, et al. Frequency of hippocampal formation atrophy in normal aging and Alzheimer's disease. *Neurobiol Aging 18(1):1Y11*, 1997.

[30] C. Cerami, P. A. Della Rosa, G. Magnani, et al. Brain metabolic maps in Mild Cognitive Impairment predict heterogeneity of progression to dementia. *Neuroimage Clin 7:187Y194*, 2014.

[31] C. Zarow, M. W. Weiner, W. G. Ellis, and H. C. Chui. Prevalence, laterality, and comorbidity of hippocampal sclerosis in an autopsy sample. *Brain Behav; 2(4):435Y442*, 2012.

[32] M. Faull, S. Y. Ching, A. I. Jarmolowicz, et al. Comparison of two methods for the analysis of CSF A-beta and tau in the diagnosis of Alzheimer's disease. *Am J Neurodegener Dis 3(3):143Y151*, 2014.

[33] T. G. Beach, J. A. Schneider, L. I. Sue, et al. Theoretical impact of Florbetapir (18F) amyloid imaging on diagnosis of Alzheimer's dementia and detection of preclinical cortical amyloid. *J Neuropathol Exp Neurol 73(10):948Y953*, 2014.

[34] D. S. Knopman, S. T. DeKosky, J. L. Cummings, et al. Practice parameter: diagnosis of dementia (an evidence-based review). *Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 56(9): 1143Y1153*, 2001.

[35] L. G. Apostolova, R. A. Dutton, I. D. Dinov, et al. Conversion of mild cognitive impairment to Alzheimer's disease predicted by hippocampal atrophy maps. *Arch Neurol; 63(5): 693Y699*, 2009.

[36] C. R. Jr Jack, M. M. Shiung, J. L. Gunter, et al. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology 62(2):591Y600*, 2004.

[37] L. G. Apostolova, C. A. Steiner, G. G. Akopyan, et al. Three-dimensional gray matter atrophy mapping in mild cognitive impairment and mild Alzheimer's disease. *Arch Neurol; 64(10): 1489Y1495*, 2007.

[38] P. M. Thompson, K. M. Hayashi, G. de Zubicaray, et al. Dynamics of gray matter loss in Alzheimer's disease. *J Neurosci; 23(3): 994Y1005*, 2003.

[39] G. Tedeschi, S. Cirillo, and C. Caltagirone. Le Neuroimmagini delle Demenze. *Critical Medicine Publishing, ch.10*, 2005.

[40] A. Ward, S. Tardi, C. Dye, and H. M. Arrighi. Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: A systematic review of the literature. *Dement Geriatr Cogn Disord Extra;3:320-32.*, 2013.

[41] A. J. Mitchell and M. Shiri-Feshki. Rate of progression of mild cognitive impairment to dementia: Meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand:119:252-65*, 2009.

[42] C. M. Bishop. Pattern Recognition and Machine Learning. *Springer*, 2006.

[43] K. P. Murphy. *Machine Learning - A Probabilistic Perspective*. Adaptive Computation and Machine Learning. MIT Press, 2012.

[44] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *arXiv:1409.0575v3 [cs.CV]*, 2015.

[45] A. Karpathy, Fei-Fei Li, and J. Johnson. *CS231n: Convolutional Neural Networks for Visual Recognition.* Stanford University.

[46] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory, pp. 144-152*, 1992.

[47] C. Cortes and V. Vapnik. Support vector networks. *Machine Learning, 20, 273-297*, 1995.

[48] G. E. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for Deep Belief Nets. *Neural Computation, 18, 1527-1554*, 2006.

[49] F. Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Cornell Aeronautical Laboratory, Psychological Review, v65, No. 6, pp. 386 - 408.*, 1958.

[50] S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167v3 [cs.LG]*, 2015.

[51] Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research 15, 1929-1958*, 2014.

[52] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.

[53] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 206.

[54] Pearson, k. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 1901.

[55] Y. Bengio, A. Courville, and P. Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 35, Issue: 8)*, 2013.

[56] Y. LeCun. Modéles connexionistes de l'apprentissage. *Ph.D. thesis, Universit de Paris VI*, 1987.

[57] G. E. Hinton and R. S. Zemel. Autoencoders, minimum description length, and Helmholtz free energy. *NIPS*, 1993.

[58] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. *arXiv:1701.07875 [stat.ML]*, 2017.

[59] Carl Doersch. Tutorial on Variational Autoencoders. *arXiv:1606.05908v2 [stat.ML]*, 2016.

[60] M. Arjovsky and L. Bottou. Towards principled methods for training generative adversarial networks. *International Conference on Learning Representations. Under Review.*, 2017.

[61] S. Kloeppel, C. M. Stonnington, J. Barnes, F. Chen, C. Chu, C. D. Good, I. Mader, L. A. Mitchell, A.C. Patel, C. C. Roberts, N. C. Fox, C. R. Jr Jack, J. Ashburner, and R. S. Frackowiak. Accuracy of dementia diagnosisca direct comparison between radiologists and a computerized method. *Brain - 131, 2969-2974*, 2008.

[62] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, and T. F. Yuan. Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. *Comput Neuroscience; 9:66.*, 2015.

[63] C. Y. Wee, P. T. Yap, D. Shen, and Alzheimer's Disease Neuroimaging Initiative. Prediction of Alzheimer'ss Disease and Mild Cognitive Impairment Using Baseline Cortical Morphological Abnormality Patterns. *Human Brain Mapping; 34(12)*, 2013.

[64] S. Farhan, M. A. Fahiem, and H. Tauseef. An Ensemble-of-Classifiers Based Approach for Early Diagnosis of Alzheimer's Disease: Classification Using Structural Features of Brain Images. *Computational and Mathematical Methods in Medicine*, 2014.

[65] X. Wang, D. Shen, and H. Huang. Prediction of Memory Impairment with MRI Data: A Longitudinal Study of Alzheimer's Disease. *Medical Image Computing and Computer-Assisted Intervention; 9900:273-281.*, 2016.

[66] M. Goryawala, Q. Zhou, W. Barker, R. Loewenstein, D. A. Duara, and M. Adjouadi. Inclusion of Neuropsychological Scores in Atrophy Models Improves Diagnostic Classification of Alzheimer's Disease and Mild

Cognitive Impairment. *Computational Intelligence and Neuroscience*, 2015.

[67] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, and A. Montillo. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron, 33(3), pp. 341-55*, 2002.

[68] B. S. Mahanand, S. Suresh, N. Sundararajan, and M. Aswatha Kumar. Identification of brain regions responsible for Alzheimer's disease using a Self-adaptive Resource Allocation Network. *Neural Networks; 32:313-22*, 2012.

[69] S. Lahmiri and M. Boukadoum. New approach for automatic classification of Alzheimer's disease, mild cognitive impairment and healthy brain magnetic resonance images. *Healthcare Technology Letters*, 2014.

[70] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Plgrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehricy, and H. Benali. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology; 51(2):73-83*, 2009.

[71] A. R. Hidalgo-Munoz, J. Ramirez, J.M. Gorriz, and P. Padilla. Regions of interest computed by SVM wrapped method for Alzheimer's disease examination from segmented MRI. *Frontiers in Aging Neuroscience, Volume 6, Article 20*, 2014.

[72] B. M. Tijms, C. Moeller, H. Vrenken, A. M. Wink, W. de Haan, W. M. van der Flier, and F. Barkhof. Single-Subject Grey Matter Graphs in Alzheimer's Disease. *PLoS ONE, 8(3), e58921*, 2013.

[73] G. Menichetti, D. Remondini, P. Panzarasa, R. J. Mondrago, and G. Bianconi. Weighted Multiplex Networks. *arXiv:1312.6720v1 [physics.soc-ph].*

[74] K. Aderghal, J. Benois-Pineau, and A. Karim. Classification of sMRI for Alzheimer's disease Diagnosis with CNN: Single Siamese Networks with 2D+ϵ Approach and Fusion on ADNI. *Association for Computer Machinery (ACM)*, 2017.

[75] Siqi Liu, Sidong Liu, W. Cai, H. Che, R. Kikinis, and M. J. Fulham. Multi-Modal Neuroimaging Feature Learning for Multi-Class Diagno-

sis of Alzheimer's Disease. *IEEE Trans Biomed Engineering; 62(4): 11321140.*, 2015.

[76] H. Choi and K. H. Jin. Predicting Cognitive Decline with Deep Learning of Brain Metabolism and Amyloid Imaging. *arXiv:1704.06033v1 [cs.CV]*, 2017.

[77] W. Huang, J. Zeng, C. Wan, et al. Image-based dementia disease diagnosis via deep low-resource pair-wise learning. *Multimed Tools Applications*, 2017.

[78] S. Sarraf and G. Tofighi. Classification of Alzheimer's Disease using fMRI Data and Deep Learning Convolutional Neural Networks. *arXiv:1603.08631v1 [cs.CV]*, 2016.

[79] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy Layer-Wise Training of Deep Networks. *Advances in Neural Information Processing Systems (NIPS), pages 153-160*, 2006.

[80] J. Masci et al. Stacked convoltional auto-encoders for hierarchical feature extraction. *ICANN, pp.52-59*, 2011.

[81] M. Boccardi et al. Training labels for hippocampal segmentation based on the EADC-ADNI harmonized hippocampal protocol. 2015.

[82] M. Wattenberg, F. Vigas, and I. Johnson. How to use t-sne effectively. *Distill*, 2016.

[83] Adversarial Autoencoders. *International Conferece of Learning Representations*, 2016.

[84] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *Proceedings MICCAI, Springer, pp. 234-241*, 2015.

[85] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015.

[86] E. Cavedo, M. Pievani, S. Boccardi, M. Galluzzi, M. Bocchetta, M. Bonetti, P. M. Thompson, and G. B. Frisoni. Medial temporal atrophy in early and late-onset Alzheimer's disease. *Neurobiol Aging ;35(9):2004-12*, 2014.

[87] Y. Klein-Koerkamp, R. A. Heckemann, K. T. Ramdeen, O. Moreaud, S. Keignart, A. Krainik, A. Hammers, M. Baciu, P. Hot, and the Alzheimer's disease Neuroimaging Initiative. Amygdalar atrophy in early Alzheimer's disease. *Curr Alzheimer Res. 11(3):239-52*, 2014.

[88] M. Venturelli, A. Sollima, E. Cé, E. Limonta, A. V. Bisconti, A. Brasioli, E. Muti, and F. Esposito. Effectiveness of Exercise- and Cognitive-Based Treatments on Salivary Cortisol Levels and Sundowning Syndrome Symptoms in Patients with Alzheimer's Disease. *Alzheimer's Disease - 53(4):1631-40*, 2016.

[89] Anonymous CVPR submission. Variational Coder with Hyper Clusters trained from Sparse Metric Learning. 2016.

[90] D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. *ICCV, pp. 26502658*, 2015.

[91] F. Milletari, N. Navab, and S. A. Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *IEEE 3DV, pp.565-571*, 2016.

[92] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, C. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[93] T. Tieleman and G. Hinton. Lecture 6.5-rmsprop: Divide the Gradient by a Running Average of Its Recent Magnitude. *COURSERA: Neural Networks for Machine Learning, 4, 26-31.*, 2012.

[94] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980 [cs.LG]*, 2014.

[95] I. Goodfellow. NIPS 2016 Tutorial: Generative Adversarial Networks. 2016.

[96] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, Volume 20, Pages 53-65*, 1987.

Appendix A

Stacked Auto Encoders

The idea behind SAEs is shown in Figure A.1: multiple autoencoders are trained one after the other, and each autoencoder encodes and decodes the higher-level, latent features produced by the previous autoencoder. Each layers can thus be seen as a higher-level representation of the previous one [13] that is the one that it's trying to reconstruct.

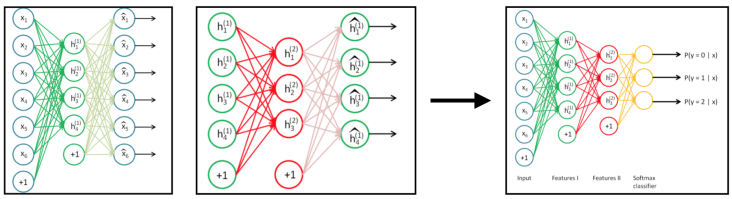


Figure A.1: Steps in the SAE training

Platone, le idee.

Questo luogo sopraceleste nessuno dei poeti di quaggiù ha mai cantato ne mai canterà degnamente. Ma è così, perchè bisogna ben avere il coraggio di dire la verità, soprattutto quando si parla della verità. Infatti è la sostanza che è realmente, priva di colore, senza figura e intangibile, e che può essere contemplata solo dal pilota dell'anima, dall'intelletto, ed è l'oggetto proprio del genere della vera scienza che occupa questo luogo.

Fedro, Platone
Tradotto da Giuseppe Cambiano
Utet, 1981