UNIVERSITÀ POLITECNICA DELLE MARCHE

FACULTY OF ENGINEERING

Master's Degree in Biomedical Engineering

COMPUTERIZED DIAGNOSIS OF ALZHEIMER'S DESEASE FROM FUNCTIONAL MAGNETIC RESONANCE

DIAGNOSI COMPUTERIZZATA DELLA MALATTIA DI ALZHEIMER DA RISONANZA MAGNETICA FUNZIONALE

Supervisor:

Prof. Laura Burattini

Co-supervisors:

Dr. Agnese Sbrollini

Dr. Selene Tomassini

Candidate:

Francesca Antolini

ACADEMIC YEAR: 2021/2022

Abstract

Background and objective: AD is the most common kind of dementia affecting millions of people worldwide. Its severity and social impact is given by the consequences related to this pathology, affecting patient and family lives. Several studies in the last decades have focused their attention of the understanding of this disease and proposing faster and more accurate diagnosis. The purpose of this thesis is to explore this pathology and the state of the art of the Machine Learning (ML) algorithms implied in this field. Moreover, starting from the research done in a recently-published study where they propose a 3D framework called Brain-on-Cloud for the identification of AD from structural MRI (sMRI) scans, the main purpose of this thesis is to verify the effectiveness of Brain-on-Cloud in pursuing the same classification task by taking into account a different data type from sMRI, precisely functional magnetic resonance images (fMRI).

Methods: In order to have a solid background for the understanding of the disease, the anatomy and physiology of the brain are analysed in detail, being the most complex organ in the human body. AD consequences in the brain are explored and include tissue damage and impairment of neuronal communication, leading to cognitive and physical issues, which worsen in time as a degenerative pathology. The diagnostic techniques in use for AD are described, focusing the attention on Magnetic Resonance Imaging (MRI) principles and application. After the study of the basic principles of ML and its use in medicine, the state of the art of its application in AD diagnosis is explored. To the aim of verifying the effectiveness of Brain-on-Cloud on fMRI images, the first experiment of the reference study is implemented, including the basic steps of the image pre-processing: intensity normalization and image cropping. Moreover, the model hyper-parameters are investigated, the best hyper-parameter combination is selected and used, and the results of the first experiment of Brain-on-Cloud by using both sMRI and fMRI scans are compared and discussed.

Results: The best combination of hyper-parameters obtained is the following: batch size of 10, learning rate equal to 0.005 and dropout rates of 0.5 and 0.6. With the implementation of this set of hyper-parameters, a mean area under the curve of about 0.86 is obtained and is comparable with the one in the reference study. Moreover, this model is able to identify

AD with an accuracy of 77.45%, sensitivity of 81.08 ±% specificity of 73.73 % and F1-score of 77.95 %.

Conclusion: In this thesis only a preliminary evaluation is made on fMRI scans, implementing the first experiment performed in the reference study. The promising results obtained open the perspective of the use of this kind of images as input data of Brain-on-Cloud for AD identification, obtaining comparable or even better performances with respect to the already tested structural magnetic resonance images (sMRI). The outcome of this research also highlights the importance of hyper-parameter tuning as necessary tool for the improvement of the model.

Index

Introduction

Alzheimer's disease (AD) is the most common cause of dementia affecting millions of people worldwide. It is a neurodegenerative disease leading to the impairment of cognitive functions progressing from a mild cognitive impairment (MCI) to a severe state. Given the social impact of this disease, several research has been conducted in the last years in order to identify this pathology, especially in the early stages where the symptoms are still mild and the brain damages are lower. This work was carried out with the purpose of understanding the AD characteristics and issues and propose the application of a 3D framework called Brain-on-Cloud, on functional magnetic resonance images (fMRI).

AD is still not well understood, as well as the human brain itself, which is the most complex organ in the human body. Therefore, the first chapter resumes the principal concepts of the anatomy and physiology of the human brain, in order to provide the basic knowledge for the understanding of the disease characteristics. In fact, AD as a type of dementia, is characterized by damages in the brain tissue affecting the communication between neurons leading to the impairment of cognitive, motor and physical functions. In particular, in AD this damage starts in the hippocampus, a region responsible for memory and learning, explaining the first stages of the disease where the patient presents memory issues. Given the degenerative nature of this disease, the conditions of the patient worsen progressively until the complete inability to live independently. There is not a definitive cure for AD but only some treatments for slowing down the degenerative process.

Together with the analysis of the clinical history and cognitive tests, the main source for the diagnoses of AD is neuroimaging. Several techniques are in use for this purpose, such as Computed Tomography (CT), Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI). For this study, the attention is focused on the use of MRI, an imaging technique based on the analysis of the behaviour of the different tissues under a magnetic field, able to provide both structural and functional information.

Several studies in the last decades have considered the importance of the identification of AD through the use of Machine Learning (ML) algorithms able to assist the clinician in the diagnostic phase.

ML is characterized by the use of algorithms able to learn from the provided data without the necessity of predefined rules. An application of ML is Deep learning (DL), which is widely used in the clinical practise since it is more effective when facing big datasets, because of the higher complexity of the algorithms. Classification algorithms are the most used in the AD diagnoses, based on the labelling of the provided data into classes of interest. Among them, Convolutional Neural Networks (CNN) are DL kind of algorithms widely implemented in the AD application that are able to provide high accuracy in the identification of the disease. A recent study proposed the "Brain-on-Cloud" 3D neural network tested on structural magnetic resonance images (sMRI) and it is interesting to explore its application on different input data, such as fMRI, in order to analyse the behaviour of the hyper-parameters governing the model. From this comparison should be also possible the evaluation of the presence of an imaging modality more informative than the other for AD identification.

I. Anatomy and physiology of the brain

1.1 Central nervous system

The human nervous system can be divided in two different parts which work in synergy to control all the voluntary and automatic actions of the body: the peripheral nervous system (PNS) and the central nervous system (CNS).

The peripheral nervous system is the bridge connecting the central nervous system with the whole body, collecting information from environment and organs to be sent to the CNS and at the same time delivering to periphery the commands coming from the CNS.

The central nervous system is the nerve centre where all the decisions are made and this chapter is focused on it. The central nervous system is made of two main components, the brain and the spinal cord.

The brain is the organ, protected by the skull, where the information are processed and the responses are sent to the rest of the body. The spinal cord is found in the vertebral column and is responsible for delivering motor commands from the brain to the PNS, while sensory information coming from the PNS begin to be processed and then delivered to the brain.

For the purpose of this study, the following sections will focus on the structure and function of the human brain.

1.1.1 Brain

The brain is the most complex organ of the human body. Scientists started to study its conformation in the nineteenth century and nowadays some mechanisms are still not completely understood. The brain is deputed to the control of movements, emotions, homeostasis regulation, language and memory. Its functioning is dependent on the 100 trillions of connections, the synapses that is able to create among the cells composing

it, the neurons. It is possible to analyse the brain in a structural and in a functional point of view, considering its anatomy and the activities related to each part.

Regions

The human brain can be divided into four main regions that can be seen in figure 1: cerebrum, cerebellum, diencephalon and brainstem.

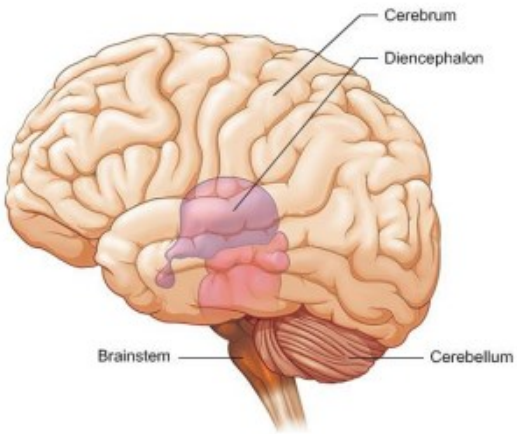


Figure 1: Subdivision of the brain structure into four main regions: cerebrum, cerebellum, diencephalon and brainstem [45].

The cerebrum is the largest part of the brain, accounting for 83% of the total mass. It controls skeletal muscles, thinking, memory, personality, and intelligence. It can be divided into two cerebral hemispheres, left and right. Each of them, looking at the coronal section in figure 2, consist of an outer cerebral cortex (grey matter) and an inner region of white matter. The cerebral cortex is the predominant portion of the cerebrum, which has a large surface, and it is folded to fit inside the skull. The folds are characterized by ridges of cortical tissue, called gyri, separated by grooves, called sulci [1].

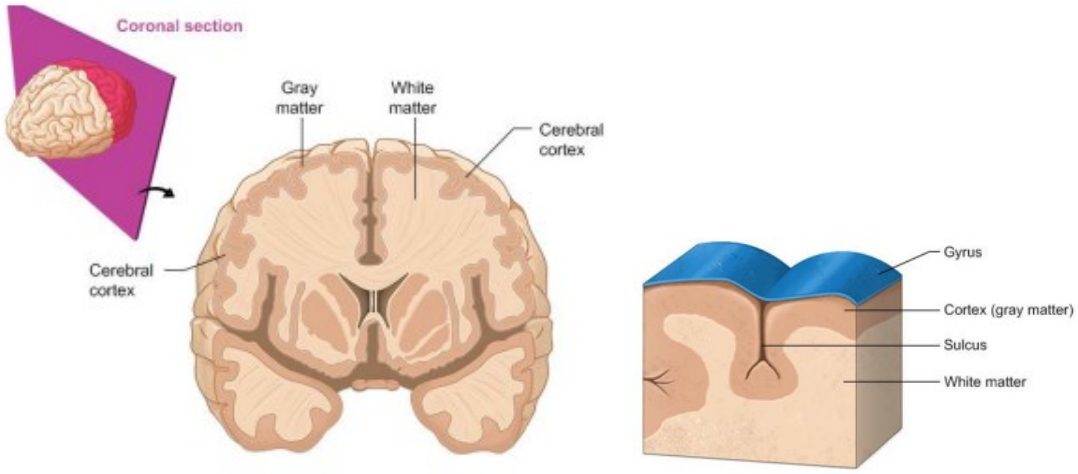


Figure 2: Coronal section of the cerebrum and detailed zoom of the folding of the cerebral cortex [45].

The cerebral cortex has three types of regions: sensory, motor, and association. Sensory areas deal with the perception of sensory information, motor areas govern the execution of voluntary movements, and association areas combine complicated activities [45,1].

Each side of the body is primarily under the control of the opposite hemisphere of the cerebral cortex, but conscious behaviour necessitates the involvement of the entire cortex, not just one functional region.

The cerebellum, also called little brain, is the second largest portion of the brain after the cerebrum representing the 11% of the total brain mass and is folded in a similar fashion achieving a higher amount of surface and, consequently, of neurons. It is made of three main lobes: anterior, posterior and flocculonodular (figure 3). The cerebellum function is to control and smooth muscular contractions and ensure balance and to do so it is connected to the other parts of the brain through peduncles. The superior peduncle links it with the midbrain allowing the coordination of leg and arm movements. The inferior peduncle ensures the connection between the cerebellum and the medulla, which receive information from the proprioceptors of the body in order to control balance and posture. The link with the pons is possible thanks to the medial peduncle, enabling the communication from the pons to the cerebellum in order to warn it about the beginning of voluntary contractions [2]. Moreover, the cerebellum is constantly in communication with the cerebral cortex in both ways, being the bridge between the

motor areas of the brain and the spinal cord, and consequently the muscle fibres. It receive motor commands from the brain and impulses from the muscles, eventually comparing the two information and adjusting the movement [3]. To do so, the cerebellum is divided into right and left hemisphere regulating respectively the right and the left part of the body, unlike the cerebral cortex where there is an opposite control.

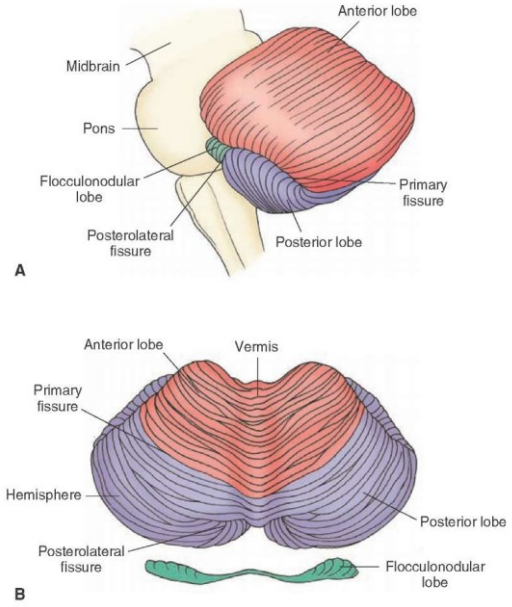


Figure 3: Representation of the cerebellum in a lateral (A) and upper (B) view [46].

The diencephalon is embedded in the centre of the brain surrounded by the cerebrum (figure 4). The main parts included in this region are thalamus and hypothalamus and together are deputed to the regulation of the autonomic nervous system. The thalamus is made of two bulges connected by a bridge, it is the relay of the brain collecting information from the senses (touch, sight, taste and hearing) and delivering it to the cerebral cortex. It senses pressure, pain and temperature through the receptors throughout the body and send the impulse to the appropriate portion of the brain. It also plays a role in the regulation of the sleep. The hypothalamus is one of the smallest parts of the brain but is crucial for the maintaining of the homeostasis. In fact, its main function is to control the smooth muscles, present in the organs of the autonomic nervous system such as vessel and digestive system, thus regulating the heartbeat, the body temperature at surface and the contractions of the muscles during the passage of food, as well as blood pressure, appetite and thirst. The hypothalamus also represent

the connection between the nervous system and the endocrine system, sensing changes in the body and stimulating the production or the inhibition of hormones. Finally, it has a role in the expression of the emotions, giving a physical response to them producing changes in the body [2,3].

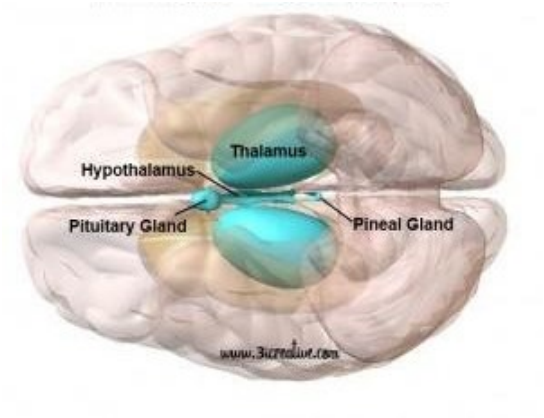


Figure 4: Representation of the diencephalon [4].

The brainstem is located on top of the spinal cord and is made of three main parts: midbrain, pons and medulla oblongata (figure 5). The midbrain is the upper portion of the brainstem and contains the nerve tracks connecting the motor areas to the spinal cord. It also integrate sensory information coming from eyes and ears controlling eyeball and pupil movements and responses to auditory stimuli, but also make fine adjustments to the movements and control reflexes of eyes, head and trunk. Then, the pons is the link of the thalamus with the cerebellum and medulla. The medulla is deputed to the stimulation of the majority of the involuntary actions such as breathing, digestion, swallowing and heart beating. Here, all the nerves coming from the spinal cord are directed to the brain, but first are subjected to decussation, which is the crossover of the nerves from the right side of the body to the left side of the brain and vice versa [2,3].

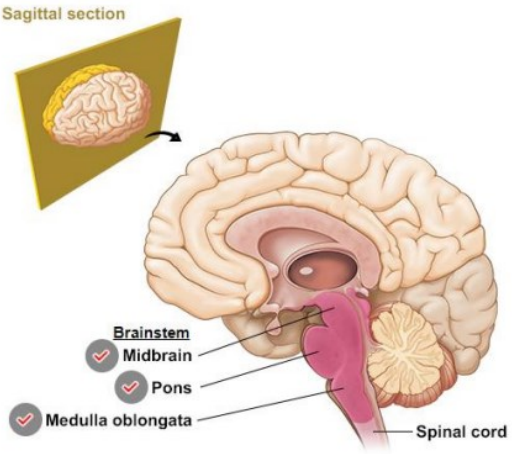


Figure 5: Sagittal section of the brainstem structure [45].

Lobes and functional areas

Looking at the sulci, there are some important landmarks that help dividing the cerebral cortex in particular regions. The deepest groove defines the two hemispheres, left and right, going from the front to back of the head. Each hemisphere is further divided into four lobes: frontal, parietal, temporal, and occipital by the central sulcus, which extends from the centre of the brain outward dividing the frontal lobe from the parietal lobe, and the lateral sulcus, known as Sylvian fissure, located lower on the hemispheres and toward the back of the head (Figure 6).

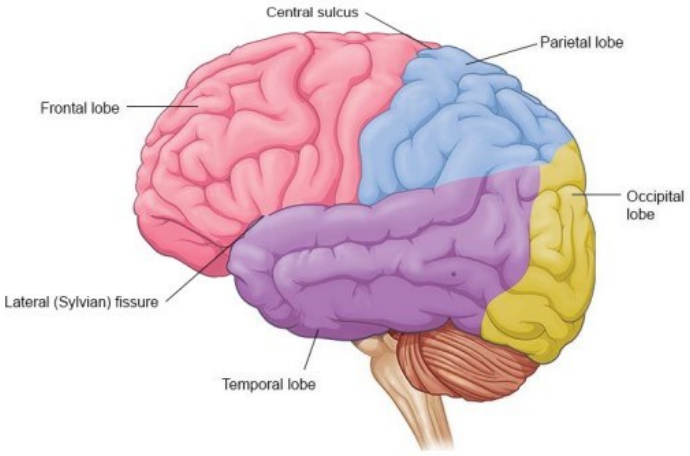


Figure 6: Lobe subdivision of the cerebral cortex [45].

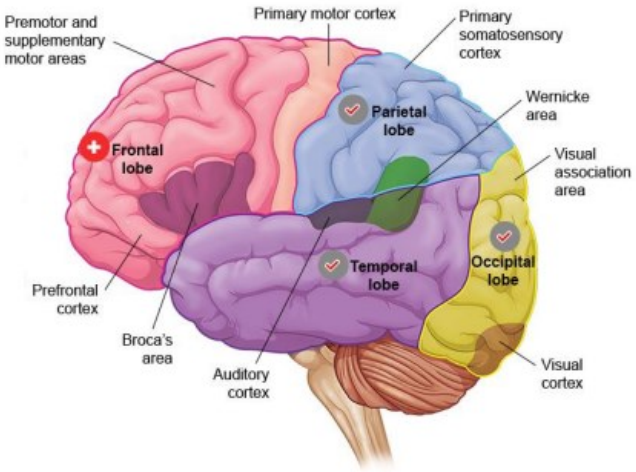


Figure 7: Functional areas, which characterize the brain lobes [45].

Each lobe has specialized regions or functional areas, as seen in figure 7, that are designed for particular purposes. The remaining parts of the lobe are referred to as "association" cortex because their duty is to combine data from various functional regions and provide an appropriate response [3].

The frontal lobe is responsible for executive functions, memory, problem-solving and production of language [45,2]. Particular portions of the frontal lobe predominantly carries each of these functions out. The prefrontal cortex is involved with executive function, memory and personality, is also in charge of maintaining focused attention and suppressing improper behaviour [3]. The production of language, as the motor performance related to it, is mainly controlled by the Broca's area, located in the frontal lobe of the left hemisphere. The most central portion of the frontal lobe just before the central sulcus is called precentral gyrus, where the primary motor area is found. Here the voluntary movement is controlled, indeed each muscle group or limb is represented by a group of neurons in this region of the cortex, which maps the body's moveable components.

Right behind the central sulcus begins the parietal lobe with the first gyrus called postcentral gyrus representing the primary somatosensory area. Similarly to the primary motor area, the body parts are mapped in this region where information from the sensory receptors of the body are collected and processed. Moreover, the parietal lobe houses the spatial orientation system and contributes to language understanding. About

this, the receptive language area, called Wernicke area, is partially located in the parietal lobe and is responsible to understand and interpret words, both written and spoken. It has a nerve connection with the Broca's area in order to comprehend and then produce a language response [3].

The other part of the Wernicke area is contained in the temporal lobe, where is possible to find also the primary auditory cortex which collects sensory information from the receptors on the cochlea. The temporal lobe is also involved in learning and memorizing together with the frontal lobe.

Lastly, the visual cortex is located in the occipital lobe, also referred to as the visual processing region. The occipital lobe takes information from the retina, much like the parietal lobe, and then use prior visual experiences to interpret and recognize the inputs, thanks to the coordination between the visual cortex and the visual association area [45, 2].

1.1.2 Neuron

The cellular composition of the human nervous system started to be studied in the beginning of the nineteenth century together with the discovery of the cell as building block of the organism. At first, given the complexity of structure of the single nerve cell, scientists believed in a reticular theory, where all the cells were connected to each other forming a continuous network. Then, with the evolving technologies, this theory was replaced with the neuron doctrine, understanding that the neurons are single cells in communication between them. Actually, nowadays, the nervous system's cellular components are divided in two classes: neurons and glial or neuroglial cells [1]. Glial cells have a role of support for the neurons even though they represent a major part of the brain cells, accounting for the 90% of them. Among their functions they produce myelin, which is a substance sometimes present around the neuron able to speed up the neurotransmission. Moreover, they provide nutrients and oxygen to the neuron and destroy pathogens. They could also represent a physical support on which neurons can arrange [5].

Regarding the neurons, their primary role is to receive and transmit information through electrochemical signals and there are 100 billion neurons in an adult brain. The structure of a neuron can be summarized by three main parts: soma, dendrite and axon. The soma, or cell body, is where the nucleus is located and proteins are synthetized, similarly to other cells. At least one branch detach from the soma and they are called neurites which can be classified as dendrite or axon. If the neurite collects information from other cells and delivers them to the soma, it can be considered dendrite. As seen in figure 8 dendrites have a branched shape able to receive information from the axon terminals of other neurons. The axon, instead, is usually a long filament transmitting information coming from the soma to other neurons or muscle cells, through the branched ending characterised by the axon terminals.

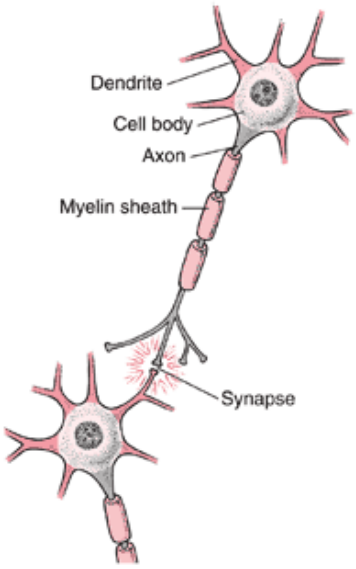


Figure 8: Neuron structure [6].

A sheath of myelin, produced by glial cells, which acts as an electrical insulator and speeds the signal conduction, surrounds most of the cases, axons and long dendrites. Axons from different neurons can be found bounded together inside connective tissue forming a structure called nerve [7].

Neuron classification

Given the different functions performed by the neurons, there is a great variety of structures and shapes with which is possible to find them. For this reason, some

classifications can be considered in order to describe this variety based on different criteria.

A structural classification can be made if is taken into account the number of neurites, dendrites and axons, present in the neuron and is possible to distinguish unipolar, bipolar and multipolar neurons. The unipolar neurons are composed by a single filament exiting the soma (figure 9 a.), which then divides into two branches, one directed to the CNS called central process and the other directed to the periphery known as peripheral process. The ending of the peripheral process can be considered dendrite since delivers information to the soma. Bipolar neurons have two processes going in opposite directions (figure 9 b.), one dendrite and one axon; the former receive stimulus from the environment and the latter conduct action potential to the CNS. They are usually found in eyes, nose and hears. Lastly, the multipolar neurons are made of many dendrites and a single axon (figure 9 c.); they are typical of the CNS and mostly are motor neurons [47].

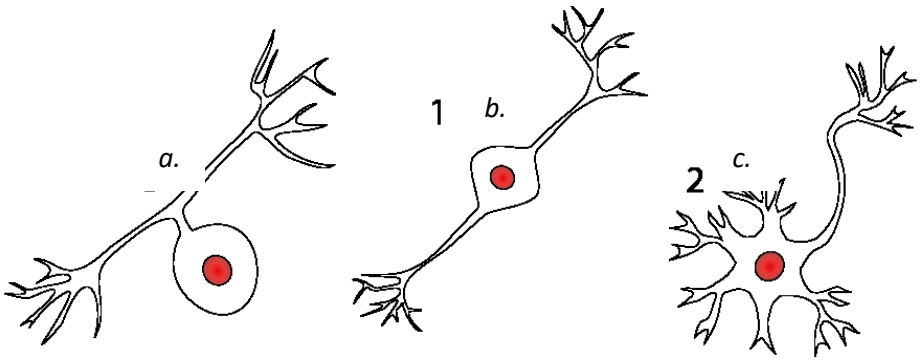


Figure 9: Different kind of neurons based on the structure: a. unipolar neurons; b. bipolar neurons; c. multipolar neurons [47].

A functional classification can also be made according to the task that the neuron has to carry on and in this case is possible to talk about sensory neurons, motor neurons and interneurons. Sensory neurons are deputed to send environmental information from outside and inside the body to the brain. The majority of them are unipolar neurons so the dendrites are located in the organs and collect signals to be carried to the CNS. Motor neurons work in the opposite direction, delivering commands from the CNS to the effectors of the body. As said before, most of the motor neurons have a multipolar shape. The interneurons are found only in the CNS and are often multipolar; their task is to receive, store and process signals from the sensory neurons and produce a

response. They are called interneurons because represent a sort of connection between different parts of the CNS [47].

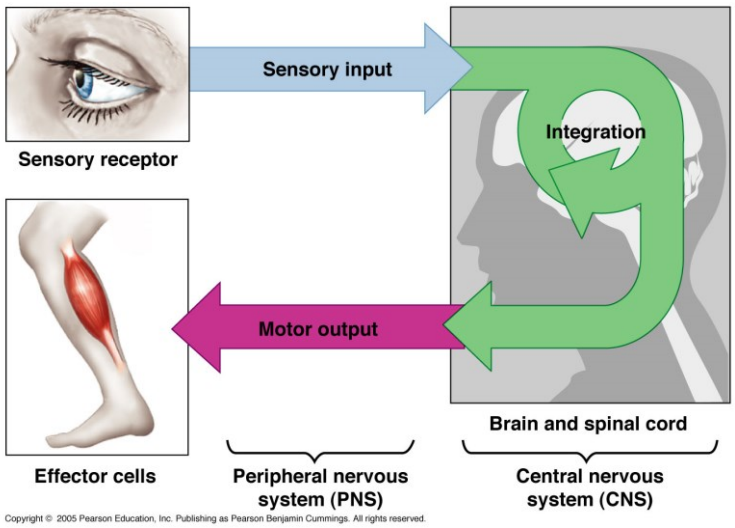


Figure 10: Schematic representation of the mechanism which distinguish motor and sensory neurons [48].

In a directional point of view, neurons can be divided into afferent and efferent: afferent if the impulse goes from the PNS to the CNS, such as the sensory neurons; efferent if the impulse travels from the CNS to the PNS, such as the motor neurons.

Neurotransmission

Action potentials, which are transient changes in membrane potential, are the means by which nerve impulses are communicated and transferred. Neurons are surrounded by a semipermeable membrane, which do not allow the exchange of ions between inside and outside the cell, preventing diffusion. The Na-K pump pumps sodium and potassium against their concentration gradients to avoid concentration equilibration. Indeed, due to a layer of negatively charged ions on the inner cell membrane and a layer of positively charged ions on the outer cell membrane, neurons have a resting membrane potential of about -70 mV.

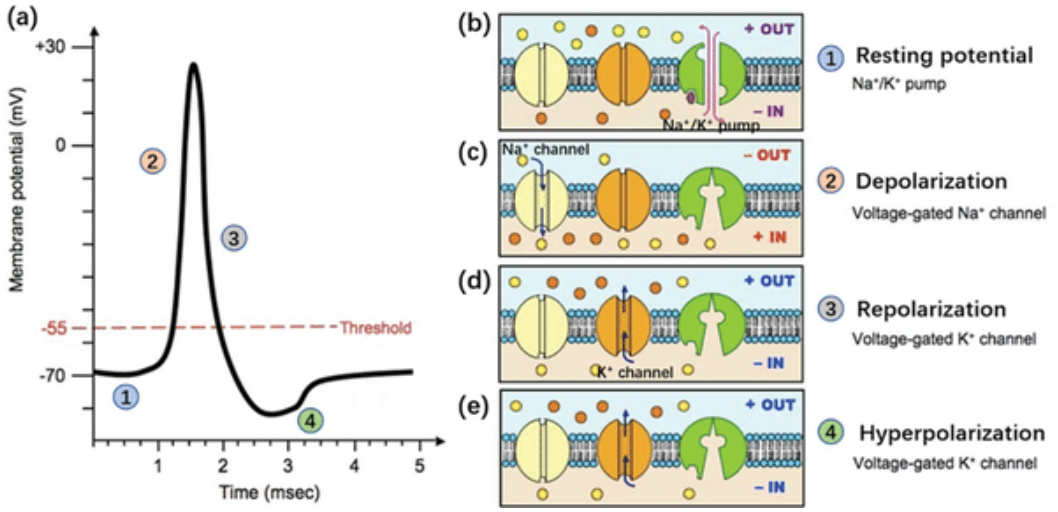


Figure 11: Representation of the Action Potential of the neuron (a) and of the phases that characterize it (b,c,d,e) [8].

The action potential can be divided into four main phases that are depicted in figure 11. At first the neuron's membrane is at the resting potential, but an electrical perturbation can be able to open the voltage-gated sodium channels, causing the action potential to become less negative. If the impulse is high enough to let the potential reach a threshold of about 55 mV, the action potential take place and the membrane quickly depolarizes (phase 2). After the overshoot, where the membrane potential overcome the value of 0 mV and become positive, sodium channels begin to close and potassium voltage-gated channels begin to open, trying to restore the resting potential. In the repolarization phase the potential rapidly fall returning negative. Before reaching again the resting potential, the membrane is subjected to a hyperpolarization due to the delayed closing of the potassium channels, making the potential more negative than necessary. After that, voltage-gated channels are subjected to a refractory period, where no kind of impulse could stimulate again an action potential. The amplitude of the action potential is always the same and is not proportional to the entity of the stimulus; in fact, it is considered an all-or-none event [7].

Once generated the action potential must be conducted along the axon of the neuron. This is possible because when the sodium ions enters the membrane and are then rejected, the negative adjacent portion of the axon attracts them, stimulating the opening of the voltage-gated channels and start the depolarization. In the case of an axon surrounded by myelin (figure 12), the conduction is more rapid because it decreases the capacitance of the membrane, so the capacity to store charge. In this way,

the impulse is spread faster along the axon, but at the same time, the presence of the myelin does not allow the passage of ions across the membrane and consequently the formation of a new action potential. The action potential is then generated in correspondence of the nodes between two myelin sheaths, called nodes of Ranvier and then propagated rapidly to the next one thanks to the myelin sheath.

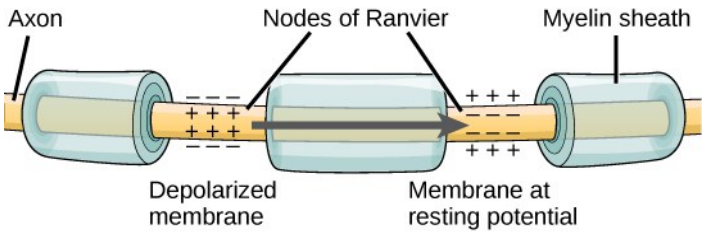


Figure 12: Schematic representation of the myelinated axon and the Nodes of Ranvier between the myelin sheets [49].

Once the impulse reaches the end of the neuron, the process by which the information is delivered to the next one is called synaptic transmission. Two kind of synapses are present in the human nervous system: electrical and chemical (figure 13). The electrical synapse can take place only if the two neurons involved are in physical contact. The transmission is ensured by gap junctions which connects the two inner parts of the neurons enabling a rapid bidirectional flow of ions. This kind of synapse, even if it is the most rapid, is less common because usually is necessary a unidirectional transmission of information [7,49].

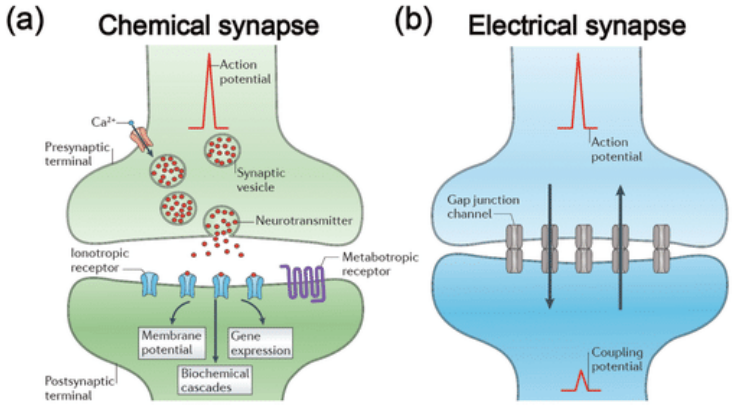


Figure 13: Two kind of synapse: (a) chemical synapse and (b) electrical synapse [9].

The chemical synapse takes place when the cells are not in touch and are separated by a synaptic space. The transmitting side of the synapse is called presynaptic neuron and here the presence of the action potential stimulate the entrance of calcium inside the membrane. This leads to the releasing of neurotransmitters from inside their vesicles to the synaptic space. Neurotransmitters reaches the receiving neuron, called postsynaptic neuron, and bind with the receptors, stimulating or inhibiting the cell.

II. Alzheimer's disease

2.1 Overview and historical background

Alzheimer's disease (AD) is the most common cause of dementia, affecting the 60-70% of the 50 million people worldwide presenting dementia [50]. In general, dementia is caused by damage of the brain tissue and this affects the communication between neurons, leading to problems in the cognitive functions but also motor and physical activities. Regarding AD, this damage affects firstly the hippocampus, caused by modification of some protein levels. Since this area of the brain is responsible for memory and learning, the first stages of AD present impairments in these cognitive functions [10,11]. It is a neurodegenerative disease, characterized by different stages worsening progressively from a mild loss of functions to a complete inability to live independently. The high number of clinical cases and the expected growth in the next years, have generated significant interest among the researchers, looking for a full understanding of this disease, its early diagnose and the possible treatments.

The study of AD began in the early twentieth century, when Alois Alzheimer, a German neurologist, in 1907 described the symptoms of a 51-year woman, Auguste Deter. Her family described its behaviour as a person with memory problems, difficulties in reading and speaking and lack of understanding. When Auguste died after five years, Alzheimer performed the autopsy, founding neuritic plaques, neurofibrillary tangles and amyloid angiopathy. These landmarks are the typical evidence of the disease that a few year after took his name, Alzheimer's disease. Starting from his discoveries later in the twentieth century, studies on dementia and in particular AD were carried out. In 1968 an important study showed for the first time how strongly were correlated the physiological changes in the brain and the clinical condition of patients with AD. [12]

2.2 Stages of Alzheimer's disease

As previously mentioned, AD is a degenerative disease that progressively worsen and this process can take decades. The course of AD is characterised by different stages

describing the severity of the conditions of the brain and consequently the behaviour of the patient. In the graph in figure 14 is possible to appreciate this progression. The preclinical Alzheimer's Disease is an asymptomatic phase where only little changes in the brain physiology are noticeable such as abnormal values in

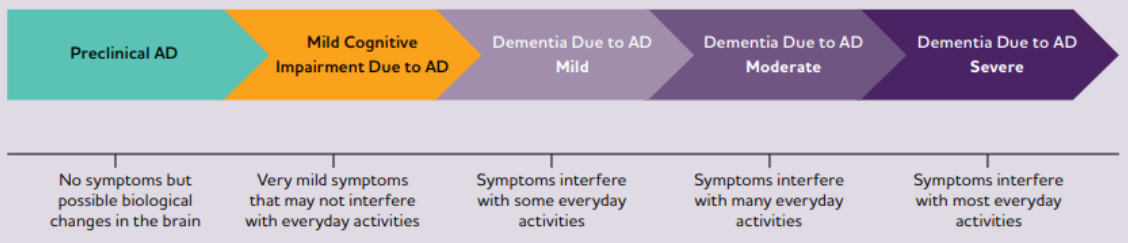


Figure 14: Graph of the progression of the Alzheimer's disease divided in different stages based on the severity of the symptoms [51].

beta-amyloids and tau protein and a decreased glucose metabolism. In this phase patients do not experience any symptom because brain is still able to compensate. Much research is directed on the study of early diagnose of AD, because an accurate recognition of AD-related changes in this phase of the disease could help handling it promptly. The second phase is the mild cognitive impairment due to AD, where symptoms begin to show up even if very mild. In this condition, brain biomarkers are recognisable together with some problems in memory language and thinking, even if they do not affect the everyday life. In some cases, this condition does not worsen or even gets better, but in one third of the cases, the patient develops dementia in the next five years.

Going further along the progression of AD, the dementia due to AD is the most severe stage of the disease and it is characterized by three main phases, where symptoms are more evident since the first one and brain damage keep worsening. In the mild stage the patient can experience episodes of difficulties in handling complex tasks which require an intense cognitive effort or it takes more time to perform simple activities. In this case he is still independent in the everyday life, but could need assistance during some particular tasks. In the moderate stage, the memory and language related abilities begin to decrease, indeed the patient may start to not recognize relatives. Also personality may be affected, with mood changes and agitation. This is the phase, which often last longer. In the severe stage the patient become much more dependent,

requiring assistance 24 hours a day. He starts having problems in communication being able to speak only simple phrases and often is unable to stand a meaningful conversation. The brain damage in this stage is spread in motor areas and the ones of the autonomic nervous system, resulting in difficulties in motor activities and swallowing. The patient is constricted in bad so other complications can arise such as body infection and consequent organ failure [51].

2.3 Risk factors

The occurrence of AD can be caused by various factors or also the combination of multiple conditions. The main risk factors, which can increase the probability of the presence of this disease, can be represented by age, genetics, education and coexisting problems.

Age is the most relevant one but not sufficient to explain the presence of AD, unlike normal aging. AD may be of two different kind related on the time of appearance of symptoms: late-onset AD (LOAD) which manifests after 65 years old and familiar AD (FAD), which is less common and is characterized by the appearance of symptoms in a younger age [13]. The percentage of elderly people affected by LOAD changes with the increasing of the age: between 65-74 years olds the percentage of AD patients is 5%, 13.1% of people between 75 and 84 years old and 33.2% of people older than 85 years old [51].

Regarding the genetic factor, several studies have proved a relationship between AD and the presence of a particular gene the APOE, which is involved in the transport of cholesterol in the blood. It may be inherited by the parents the APOE gene in one of three different forms, e2 e3 or e4. The presence of APOE-e4 increases the probability of developing AD, while APOE-e2 decreases them; e-3 is a neutral form of APOE. In addition, mutation in certain chromosomes, such as the one concerning the Down syndrome, may increase the chances and the appearance of symptoms before 65 years old.

Education also could play a role in the prevention of AD; some studies indeed show that people with more years of education are at lower risk. A theory explaining that is the

presence of more synaptic connection in the well-educated patients, compensating the neuronal loss characterizing the disease.

There is also a correlation between AD and other pathological conditions such as cardiovascular diseases. In this case, damaged blood vessels may affect the progression of AD, causing a decreased blood flow in the brain. The prevention of cardiovascular diseases may delay the appearance of AD symptoms. Another example is the type 2 diabetes, which is proved to increase the risk for AD [51].

2.4 Diagnosis

Diagnosis of AD is quite challenging, its easy state the presence of dementia but the complications arise when determining the causes of it. Knowing the patient's medical history is the first stage in making a diagnosis. The doctor will then ascertain what symptoms are present, when they first appeared, and how they have changed over time. The medical history of the family is also important. Also the conduction of a Mini Mental State Examination (MMSE) may be helpful, which consist of a series of questions asked from the doctor in order to understand his conditions. This test is repeated during the years in order to supervise the degenerative progression. The doctor will do a physical examination, which will include blood and urine testing. By doing this, it is possible to rule out additional dementia-causing factors, including hormonal imbalance, vitamin deficiencies, and urogenital infections. Brain scans can also be used to rule out infections, traumatic brain damage, malignancies, and cerebrovascular accidents.

Scans are used also to check the presence of the typical biomarkers of AD. The increasing level of tau or neurofibrillary tangles can be found inside the neurons blocking the passage of nutrients and other molecules useful for the metabolism of the neuron. The other biomarker is the presence of beta-amyloid plaques (plaques surrounded by an abnormal protein called beta-amyloid and cellular material), which interrupt the synaptic communication between neurons. The consequences of this changes are the death of neuronal cells and thus could be also lead to the presence of debris and inflammation of the brain tissue. For this reason the brain of an AD patient appears shrunk with respect to the healthy one (Figure 15) [15].

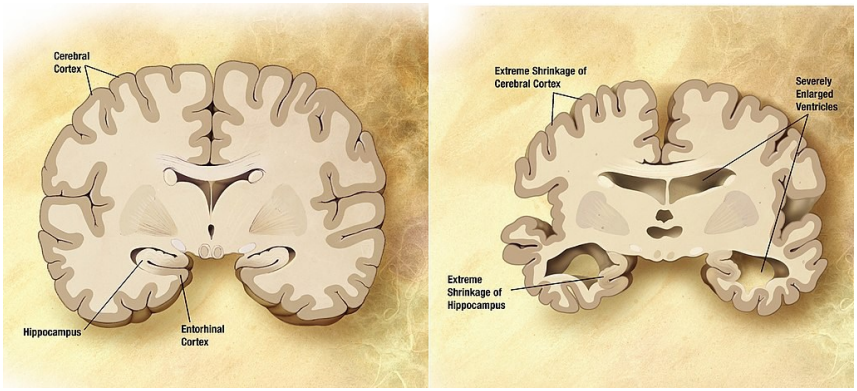


Figure 15: Comparison between healthy and AD typical brain [14].

Some studies have identified another biomarker of AD present in the cerebrospinal fluid (CSF), a liquid circulating around the brain under the skull reaching all the cavities which brings oxygen, glucose and other nutrients to the brain tissue and has a role of protection against physical and chemical injuries. The CSF levels of the protein beta-amyloid is 100 greater compared with the blood plasma, so is easier to be measured. At the same time, the tau protein is released by the neurons affected by neurofibrillary degeneration. For this reason, in an AD patient is possible to appreciate in the CSF a reduced level of beta-amyloid and an increased tau; the combination of this two evidence can discriminate between healthy and AD affected brain [16].

2.5 Treatment

There is no definitive cure for the Alzheimer's disease, but some treatments both pharmacological and non-pharmacological can be considered for slowing the disease progression. Among the non-pharmacological ones, lifestyle changes and interventions can slow down the cognitive decline. For example balanced diet, physical exercise, social activities and cognitive training can decrease the risk of AD. Even if these factors do not affect the pathology itself, they are useful in the progression of the disease.

Two kinds of pharmacological treatments have to be considered: drugs directed to the improvement of the cognitive symptoms, also affecting the behavioural aspect; drugs targeting the psychiatric symptoms such as aggressiveness and agitation. For the cognitive symptoms treatment acetylcholinesterase inhibitors are a solution. These drugs are capable of inhibiting the enzyme deputed in the degradation of acetylcholine,

resulting in an enhanced level of this neurotransmitter, crucial for the communication between neurons. In this way, it slightly compensate the neuronal loss, leading to an improvement in memory, attention and awareness. The inhibitors currently in use are donepezil, galantamine, and rivastigmine and are manly prescribed during the mild or moderate stage on the disease. Another solution used in the moderate or severe stage is the memantine. This drug act blocking the glutamate receptors (NMDA). Glutamate is a neurotransmitter involved in memory and learning processes, but with the presence of AD, evidence encountered high levels of glutamate in the brain. Excessive stimulation by glutamate is toxic for the neuron, causing neuron loss. For that, memantine is a valid treatment for delaying the worsening of symptoms. It can be prescribed alone or in combination with an acetylcholinesterase inhibitor.

Regarding the psychiatric symptoms, antipsychotics and anxiolytics can be prescribed, if the patient present agitation, paranoid, anxiety aggressiveness and hallucinations.

2.6 Prognosis

The estimated prevalence of dementia in the world is of 50 million people according to a report of the Alzheimer's Disease International of 2018 and are supposed to triple within 2050. Moreover, AD is a long duration disease considering that some stages can last for several years. Some studies directed on the estimation of the duration of the dementia, show that could be relatively short. In fact, in US was reported a survival time of 3-4 years and in Europe of 6 years after the diagnoses of dementia. However, if previous stages are considered, the duration may drastically increase. Considering a 70 year old patient, the preclinical stage duration is estimated to be 10 years, the mild cognitive impairment stage other 4 years and the dementia stage is estimated to last 6 years, in accordance with the aforementioned estimate.

III. Standard imaging techniques for Alzheimer's disease identification

3.1 Imaging techniques

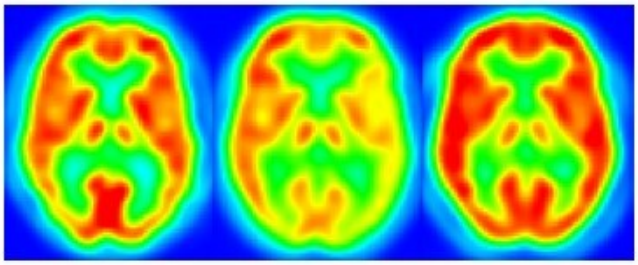
Neuroimaging is an important tool for the identification of AD and for having a better understanding of the progression of the disease. More and more researchers have started studying its accuracy in the early diagnose of AD. The techniques that may be used for the detection of the described biomarkers and structural abnormalities of the brain are Computed tomography (CT), Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI).

CT scans provide a detailed three dimensional overview of the brain structure, bones and soft tissue. The technology behind the CT scan is based on the acquisition of a series of slices of the brain that are then reconstructed in a 3D image. This procedure can be used for the identification of AD, also for rejecting other causes of the symptoms. It is less used than the other techniques because its accuracy is restricted only in the later stages of the disease, where tau tangles and beta-amyloid plaques can be recognized.

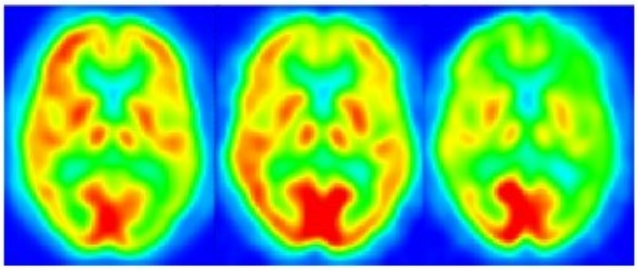
PET is based on the use of radiotracers injected in the human body in order to assess the functional abilities of the interested organ. The radiotracer travels in the body reaching the region of interest and when the molecule to which is bonded is metabolized by the organism it emits positrons. They are detected by the PET scan and, based on the amount of emitted positrons, a 3D image is reconstructed, showing the activity of that region. In this way is able to provide information about metabolism, blood flow and neuronal communication. Two kind of tracers are the most useful for the identification of AD: Fluorodeoxyglucose and Pittsburgh compound B (PiB).

Fluorodeoxyglucose is a molecule of glucose labelled with ^{18}F. Tracking the distribution of this molecule in the brain tissue is possible to have an idea of the glucose metabolism in the brain. In AD patients, the glucose metabolism appears to be reduced than normal

in the temporo-parietal region as shown in figure 16. These metabolic changes are consistently present both in the later stages of AD and in the mild cognitive impairment stage, with a higher reduction in the first case. Evidence showed a first reduction in the posterior regions of the brain, appreciable in the mild stage, while the frontal lobe is affected only in the later stages [17].



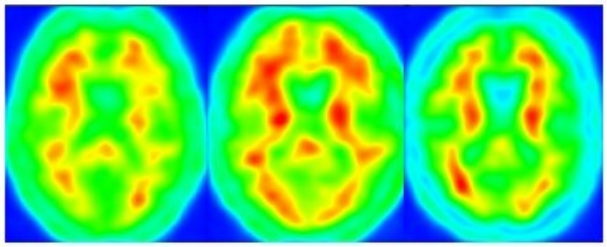
(a) Healthy individuals



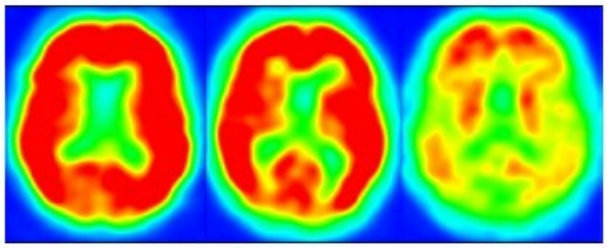
(b) AD patients

Figure 16: Comparison between healthy and AD brains in a Fluorodeoxyglucose PET test. Red regions represent the glucose metabolism [17].

The other tracer used, the PiB, is a ^{11}C labelled thioflavin-T derivative. It is able to identify amyloid plaques and bind to them, giving an overview of their distribution. AD patients present a great concentration of amyloid plaques, especially in the white matter, if compared with a healthy brain (figure 17). This presence is also appreciable in the mild stage even if is less evident. The response of this exam is positive or negative, but is not a sufficient parameter to diagnose AD. Sometimes, in fact, some positives appeared in patients without AD. However, it may indicate some individuals who will develop the disease in the future [17].



(a) Healthy individuals



(b) AD patients

Figure 17: Comparison between healthy and AD brain in a Pittsburgh compound B (PiB) PET test. The red regions highlight the amyloid plaques present in the brain [17].

MRI is one of the most used techniques in the AD environment, because it is able to detect both anatomical and functional changes of the brain through the use of structural MRI or functional MRI (fMRI) respectively. Its functioning is based on the hydrogen atoms properties of the brain cells stimulated by a magnetic field. The radiofrequencies emitted by the atoms allow to distinguish among the different tissues present in the brain. The principles at the basis of MRI are explained more precisely in the next paragraph. Structural MRI can provide a quantification of the atrophy and tissue changes due to AD, which consequently lead to communication alterations visible in fMRI.

Cerebral atrophy, characterized by the reduction of tissue volume, in AD is progressively more aggressive and extends, during the worsening of the disease, from certain regions of the brain to the others, allowing the discrimination of AD with other cerebral diseases. In fact, MRI scans have highlighted the presence of atrophy in the medial temporal lobe in the first place, followed by hippocampus, amygdala and parahippocampus (figure 18). Then, it extends in the temporal neocortex and all neocortical association areas in the later stages of the disease. Atrophy is related to the neuronal death, in fact MRI evidence are coherent with the biological changes such as the presence of tau tangles in the neurons. The limitation of this imaging technique is that is not able to detect these biomarkers (tau tangles and amyloid plaques) lacking specificity for the diagnoses.

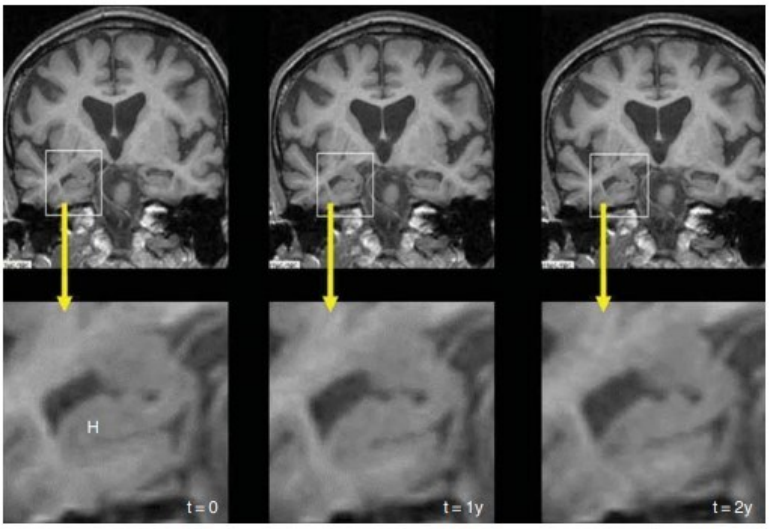


Figure 18: Hippocampal atrophy (H) observed from a MRI scan of the brain in three different AD stages: initial memory complaints, mild stage and severe stage [17].

Regarding the fMRI, it is the non-invasive way to understand the brain activity. Through the measurement of the blood oxygen level-dependent (BOLD) it is able to give a quantification of the blood flow, blood volume and oxygen exchange. This exam may be performed in two different ways: during the performance of a cognitive task, in order to check particular brain functional areas, or at rest, analysing the neuronal functional connectivity (fc-MRI). The brain region involved in the functional impairment follow the anatomical damage seen in the structural MRI. A decreased activity of the hippocampus is registered as well as a similar condition in the medial temporal lobe during the mild cognitive impairment stage. In addition, an increased activity in the prefrontal cortex is appreciated in the AD patients as a compensation for the connection loss in other regions [18].

3.2 Magnetic resonance imaging

Magnetic Resonance Imaging (MRI): is a non-invasive imaging technique that produces 3D images. It maps the distribution of hydrogen atoms in the tissues, providing a detailed discrimination between healthy and damaged tissues. The characteristic of this imaging technique is that allows the analysis of both brain structure and function in the same scanner, thanks to its ability to detect anatomical features but also flowing liquids like blood.

MRI uses magnetic fields and radio-frequencies, rather than ionizing radiations used in other techniques, such as x-rays and CT, providing a harmless solution for neuroimaging. The magnetic field strength is measured in Tesla (T) and the majority of the MRI systems in the clinical practise are between 1.5-3 T.

3.2.1 Basic principles

MRI relies on the magnetic properties of the hydrogen nucleus to produce images, contained in the body tissues and water. A single proton composes the hydrogen nucleus, which is a spinning charged particle producing a magnetic field, the magnetic momentum. In a normal condition protons are oriented randomly generating a net magnetization equal to zero. With the application of a strong magnetic field, protons align parallel (low energy state) or anti-parallel (high energy state) to the field B_0, with a net result called net magnetic vector (M) in the direction of B_0 (figure 19).

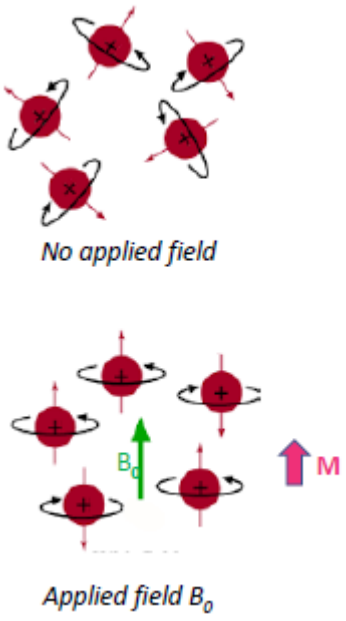


Figure 19: Behaviour of the hydrogen protons in a normal condition (top) and with the application of a magnetic field. M is the net magnetic vector.

The protons spin around the long axis of the primary magnetic field is called precession (figure 20) and has an angular frequency called Larmor frequency, specified in equation 1.

$$f = \gamma B_0 \quad (1)$$

where γ is the magnetron ratio.

This frequency changes in proportion to the applied magnetic field. Moreover, when protons process together they are defined as in phase, while if they spin separately they are defined out of phase (figure 21).

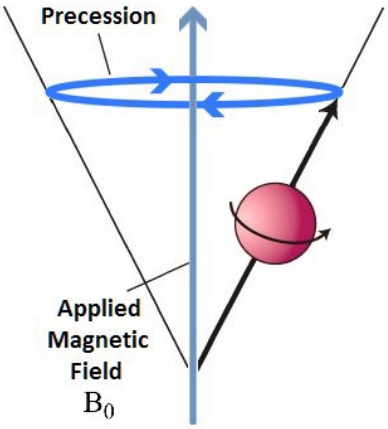


Figure 20: Sketch of the precession of the proton around the magnetic field B_0.

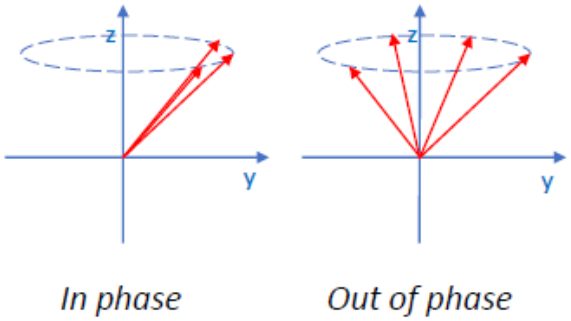


Figure 21: The red arrows represent the magnetic vectors of the single particles. At right they spin together around the applied magnetic field (along z axis) and are in phase; at left they spin randomly so are defined out of phase.

The excitation energy (E) that a proton needs to excite a nucleus from the lower state to the higher state is defined by the Planck's equation (equation 2).

$$E = hw_0 \tag{2}$$

where h is Planck's constant divided by 2π.

This energy is provided by a radio frequency (RF) pulse that generate a second magnetic field which results in a disturbance of the proton alignment.

The RF pulse is applied with the same precession frequency, producing two effects:

a) Some low energy parallel protons flip to high-energy state, decreasing the longitudinal magnetization (figure 22 a.).

b) Protons start to synchronize and process in phase (figure 22 b.).

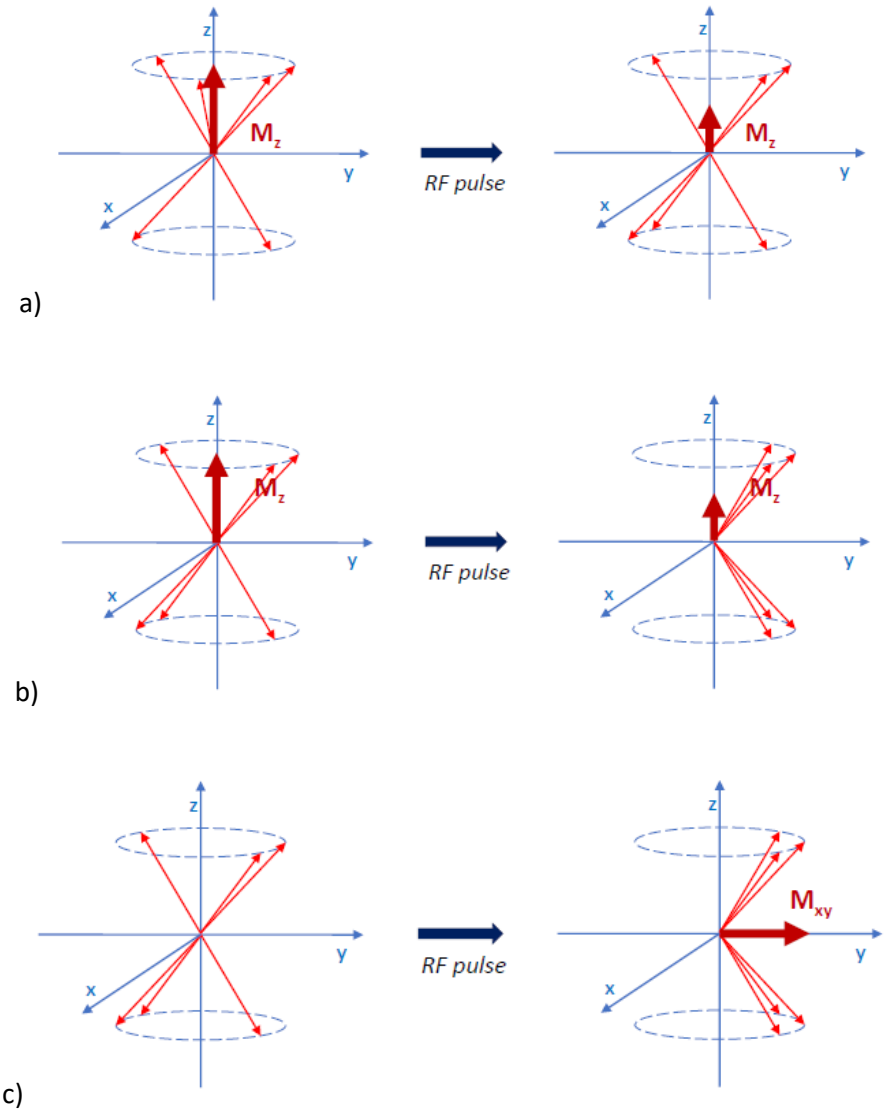


Figure 22: Behaviour of the protons due to the application of the radiofrequency (RF) pulse. a) decreasing of longitudinal magnetization (M_Z). b) in phase precession. c) net magnetic vector in the transverse plane.

These effects lead the net magnetic vector to turn towards the transverse plane, that is a right angle to the primary magnetic field; this is known as Transverse Magnetization (figure 22 c.).

After the RF pulse, protons tend to return in the initial equilibrium state through a process called relaxation, which produces a signal captured by the same RF coils and used to create images.

The proton relaxation can be measured in two directions (figure 23):

1. Longitudinal axis (z): T_1 relaxation
2. Transverse axis (xy): T_2 relaxation

Figure 23: Sketch of the two kind of measurements of the protons relaxation.

Regarding the T_1 relaxation, after the RF pulse, several protons flip back to the low energy state, parallel to the primary magnetic field giving up the energy to the surroundings, the lattice. This result in changings to the longitudinal relaxation T_1, also called spin-lattice relaxation.

On a plot of the magnetization over time (figure 24), magnetization increases with time, following the trend defined by the function in equation 3.

$$M_z = M_0(1 - e^{-\frac{t}{T_1}}) \quad (3)$$

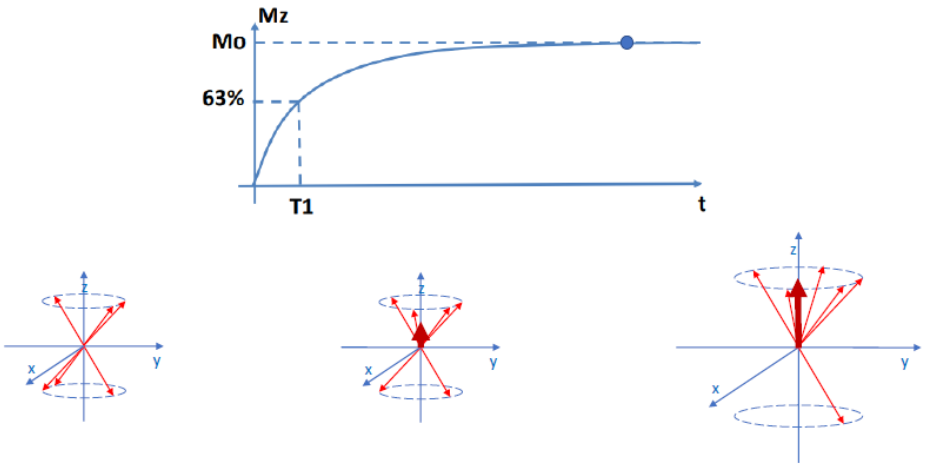


Figure 24: Magnetization in the T_1 relaxation.

The T_1 relaxation time will varies depending on tissue composition and structure and it is measured on the 63% of the M_0, the saturation value.

In figure 25, can be appreciated some examples of different tissues and their typical relaxation; water molecules move rapidly and do not move into the low energy state quickly, so T1 relaxation takes longer.

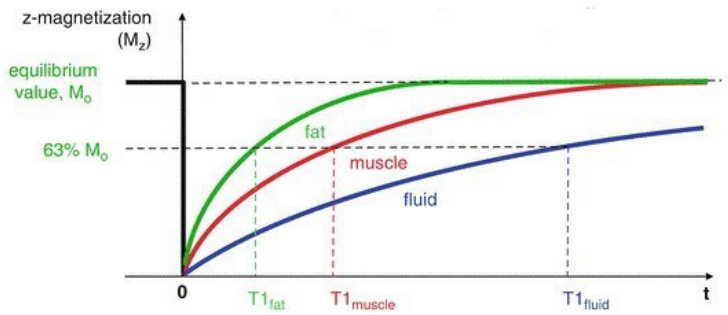


Figure 25: Examples of T_1 relaxation on different tissues.

Regarding the T_2 relaxation, after the RF pulse protons that where in phase begin to diphase, out of the Larmor frequency, in the transverse axis. This result in a reduction of the transverse magnetization, known as spin-spin relaxation.

Plotting transverse magnetization in the x-y plane versus time (figure 26), transverse magnetization decreases over time following the trend in equation 4.

$$M_{xy} = M_0 e^{-\frac{t}{T_2}} \quad (4)$$

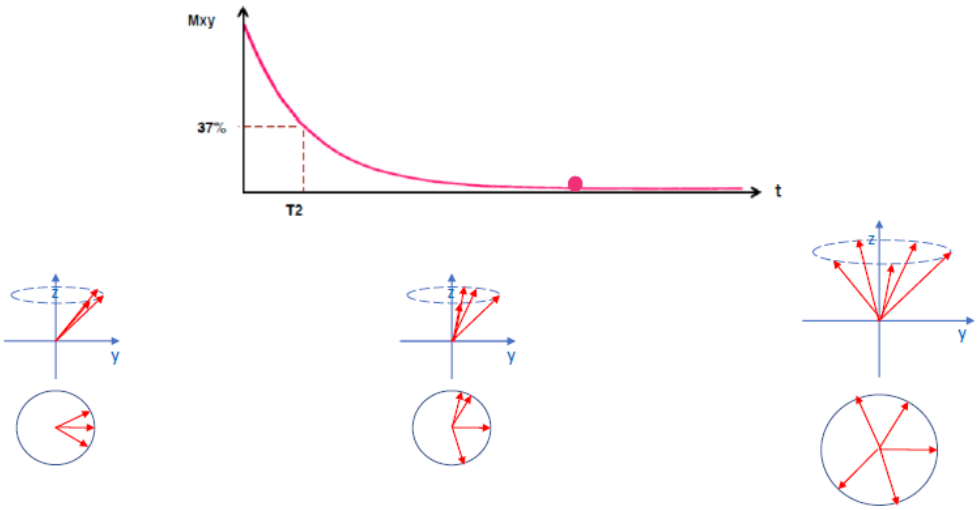


Figure 26: Magnetization in the T_2 relaxation.

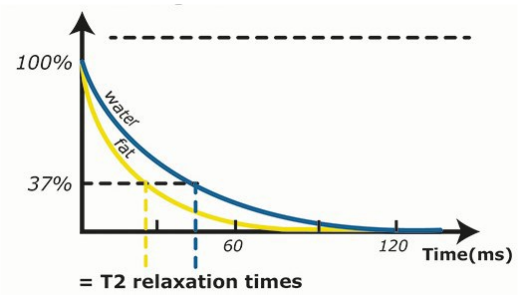


Figure 27: Comparison between T_2 relaxation in water and fat tissue.

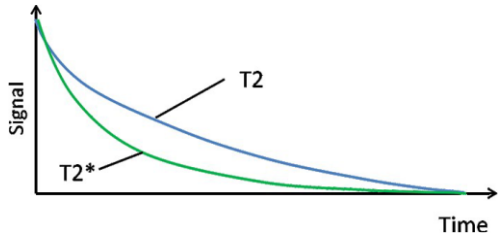


Figure 28: Difference between T_2 and T_2^* relaxation.

Also in this case the relaxation is different among tissues compositions as can be appreciated in figure 27 and it calculated on the 37% of M_0. Is possible to notice that water has a longer relaxation time, given by the quicker movement of the molecules resulting in a minor field inhomogeneity. In reality the spins diphase much quicker then T_2 because of inhomogeneity in the magnetic field B_0, in fact usually is considered

another parameter, defined as T_2^*, which takes into account these inhomogeneity and the molecular interaction (figure 28).

The Net magnetic vector is the sum of longitudinal and transverse magnetization and spirals around the z axis. The changing magnetic moment of the net magnetic vector, results in Free Induction Decay (FID), an electrical signal received by the RF coils of the system. This signal has the shape shown in figure 29, it develops in the transverse plane and reduces as the net magnetic vector moves closer to the z axis.

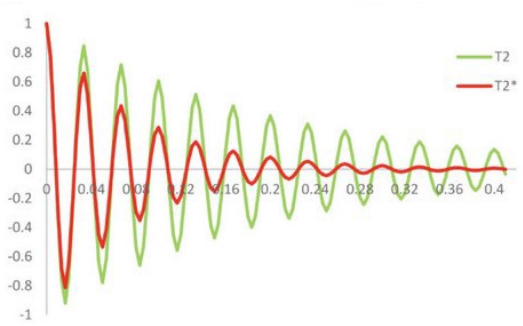


Figure 29: Free induction decay (FID) on the transvers axis. Comparison between T_2 and T_2^* relaxations.

3.2.2 Image contrast

The quality of the image that can be obtained thanks to the principle explained in the previous section, depends on two main parameters related to the pulse sequence:

- Repetition time (TR), which is the time between two consecutive RF-pulses.
- Echo time (TE), which is the time calculated between the centre of the RF-pulse and the centre of the echo produced by the tissues

By modifying the duration of TR and TE is possible to observe different signals between different tissues. Each part of the body has different T_1, T_2 and proton density (PD) and the quality of images depends on these parameters. The aim is to improve and maximize the contrast between different kind of tissues focusing on the “weight” of each parameter: is possible to control which characteristic need to be emphasized.

Four combinations of TR and TE values can be considered (figure 30):

- Short TR/Short TE: T_1-weighting
- Long TR/Short TE: PD-weighting
- Long TR/Long TE: T_2-weighting
- Short TR/Long TE: not used

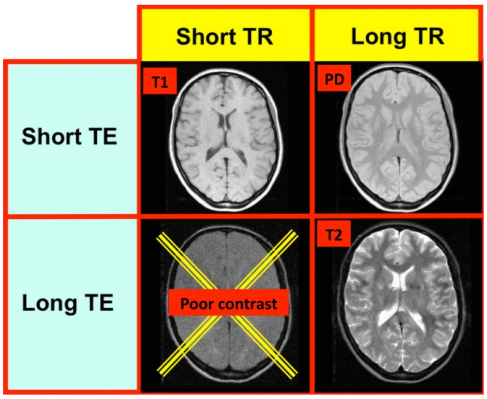


Figure 30: The four possible combinations of parameters from which is possible to obtain different image contrasts.

The measured signal is approximately the one given by equation 5.

$$M_0(1 - e^{-\frac{TR}{T_1}})e^{-\frac{TE}{T_2}} \quad (5)$$

When TE is short the ratio TE/T_2 tends to 0, so the T_2-weighting term $e^{-\frac{TE}{T_2}}$ tends to e^{-0} so to 1 and T_2 effects largely disappear. Instead when TR is long compared to T_2, the importance of the exponential weighting term increases. Considering the signals generated by two tissues with different T_2 values (plotted in figure 31), TE changes can result in contrast variation between the two tissues. When TE is short, the echo occurs when there has been little time for T_2-decay and the tissues are not differentiated, so there is a poor contrast but a high recorded MRI signal. When TE is long, the differences in signal decay between the two tissues become more visible, so there is a good contrast but a low recorded NMR signal.

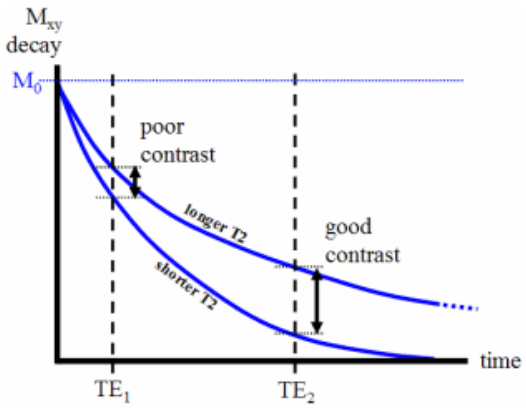


Figure 31: Contrast variation between two tissues related to the echo time (TE).

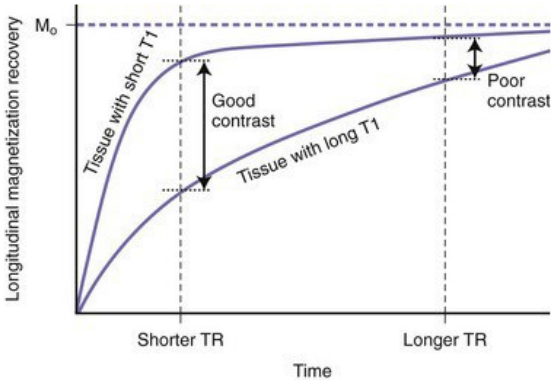


Figure 32: Contrast variation between two tissues related to the repetition time (TR).

When TR is long compared to T_1, the T_1-weighting term $e^{-\frac{TR}{T_1}}$ tends to 0, so T_1 effects disappear. Considering the signals generated by two tissues with different T_1 values (figure 32), contrast variation can be appreciated if different values of TR are considered. When TR is long, a poor contrast is present because M_z recovers completely in both tissues and their signals are not so different. Instead a short TR accentuate "T_1 - weighting" and provides a good contrast. Finally, when TR is long and TE is short, both T_1 and T_2 effects are minimized. The only remaining factor is the proton density, which is always present and becomes the dominant weighting in this case.

3.2.3 Image signal processing

The generation of the image is related to the MR signal generated by the tissues. Both frequency and phase date are collected in the so-called k-space, to which a two-dimensional Fourier transform is performed in order to produce a grey-scale image.

The k-space is a matrix in which all the raw data acquired are inserted in each row of the matrix (figure 33). The time acquisition of MRI signals ends when the fraction of the matrix is completely full. This matrix is in the frequency domain and through the use of the Fourier transform, data are translated into special domain obtaining the image. The target is split into a number of equally sized volume elements or voxels. Each voxel is coloured by different shade of grey depending on the concentration of Hydrogen. Higher the number of voxel and smaller is the size of the volumes, increasing the resolution of the image.

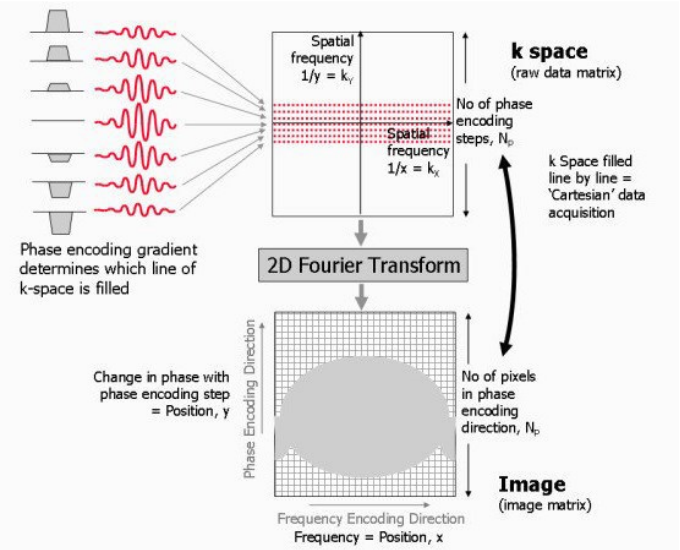


Figure 33: K-space functioning for the generation of the image.

3.2.4 Structural and functional magnetic resonance

Magnetic resonance imaging, as previously explained, is based on the acquisition of signals coming from the tissues of the body. Based on the relaxation time in consideration different kind of images can be obtained, enhancing characteristics of interest. In this way is possible to detect through MRI both structural and functional information.

Regarding the structural MRI, the acquisition of the data is based on the T_1 relaxation, in which is possible to distinguish between tissues thanks to characteristic T_1 values typical of particular tissue composition. This modality is used in the clinical practise in order to observe anatomical changes or tumour lesions. In particular, for the AD analysis

structural MRI is an important tool for the detection of brain tissue atrophy, necessary or the study of the progression of the disease.

Another imaging modality is the functional magnetic resonance (fMRI), which is based on the T_2^* relaxation and it depends on the blood flux inside the body. T_2^* relaxation takes into account both molecular interaction and inhomogeneity of the magnetic field. In the blood is possible to distinguish two kinds of haemoglobin: oxy-haemoglobin, which carries oxygen and deoxy-haemoglobin, which is oxygen-depleted. The first one is diamagnetic and do not affect the magnetic field, while the second one is paramagnetic, so changes of the amount of deoxy-haemoglobin will affect the signal. These blood oxygenated level dependent (BOLD) changes allow the MRI scan to detect neuronal activity. In fact, when brain areas activate they need a larger amount of oxygen, so in these areas is appreciated a decrease in the presence of oxy-haemoglobin and an increase of deoxy-haemoglobin [19].

IV. Machine learning in medicine

4.1 Machine learning

Artificial intelligence (AI) has conquered an important role in recent years in different fields such as economy, computer vison, language and audio recognition, bioinformatics and especially in medicine. In general, AI refers to a series of algorithms able to perform complicated tasks, which require a certain level of intelligence from the human. The first applications of AI in medicine are registered around the 1980s and in the last few years have experienced an exponential increase, since the large amount of clinical data available (constantly increasing) and the necessity to manage them. Machine learning (ML) is a branch of the AI characterized by the use of algorithms able to learn from the provided data without the necessity of predefined rules. ML algorithms can be classified into supervised and unsupervised. The supervised methods are based on a training process where some previous rules and labels are provided to the machine. In the unsupervised algorithms the computer has to recognise itself the classes in the provided data [20].

A basic example of ML model are the Artificial Neural Networks (ANN), which are inspired by the human neuronal communication. In fact, in nature the neuron is able to process an input signal in the cell body and deliver the output from the axon. The synaptic strength of the neuron can be considered as a weight applied to the input. A typical artificial neuron can be represented as in figure 34 a) where all the inputs are multiplied by a certain weight and then summed up together with a bias [20]. To this weighted sum is then applied a function called activation function, modelling the firing frequency of the axon. There are two commonly used activation function:

- Sigmoid: $f(x) = \frac{1}{1+e^{-x}}$ (6)
- Rectified linear unit (ReLU): $f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$ (7)

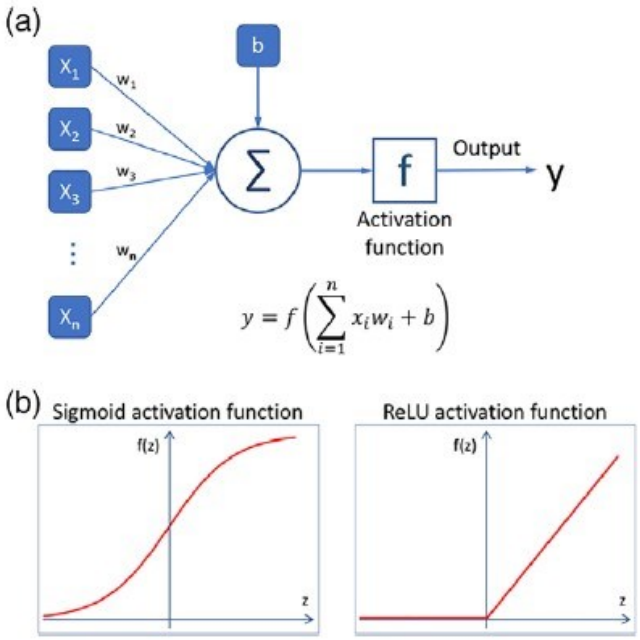


Figure 34: (a) representation of an artificial neuron and (b) two different activation functions [20].

The ANN is then the connection between multiple neurons in which different layers can be recognized. The ANN commonly used are made of two or three layers and an example with two hidden layers of neurons can be appreciated in figure 35. In particular, a layer in which all neurons are connected to all the neurons of the next layer is referred as fully connected layer. The final layer is the output where is obtained the desired outcome or the labelled state.

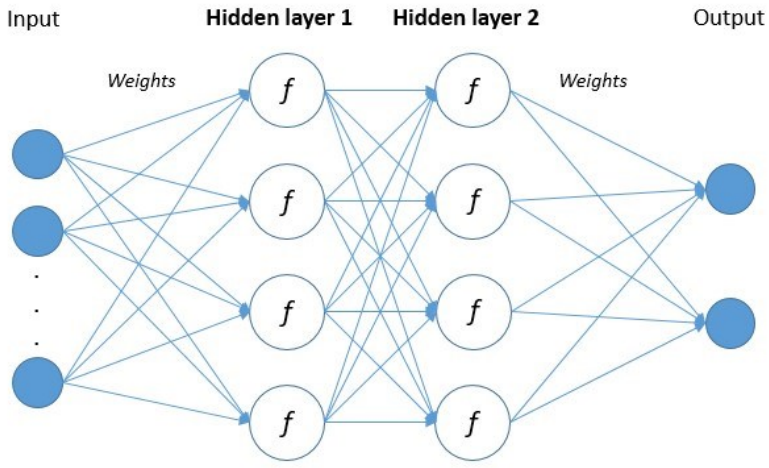


Figure 35: Schematic representation of an artificial neural network.

In order to classify information, for example images, ML algorithms have to accomplish two main phases: a training phase and a test phase. In the training phase the algorithms is trained to associate to certain images a give label, modifying the weights associated to the various connections [21,22]. To do this an output is calculated using the initial weights and is compared with expected label, calculating the error. Then, according to the obtained error, the weights are adjusted until the error is minimized. This mechanism is called error backpropagation [20]. In this phase an important role is given to the feature extraction, which means the definition of interesting patterns inside the images. In the test phase, the trained algorithm obtained is used in order to classify and label new images, considering the same feature extraction of the previous phase.

Deep learning (DL) is a particular application of the ML, widely studied in the last few years for the clinical use, which is more efficient when facing a large dataset. The main difference between DL and traditional ML is the depth; in fact, a higher number of layers, usually between 10 and 30, characterizes DL algorithms. Moreover, the feature extraction phase is automatically made by the algorithm and has not to be done manually, providing a higher accuracy.

The most used among the DL algorithms are the Convolutional Neural Networks (CNN) inspired by the visual cortex of the brain. The main problem of the fully connected layer networks is that when dealing with a big matrix, for example with images (256x256x256 voxels), too many weights have to be evaluated. In the CNN this issue is minimized by the presence of another kind of layers, called convolutional layers, where some neurons are characterized by the same weight [20,23]. This approach is similar to the mathematical convolution and the other advantage is that it allows the features identification independently of their location. Another type of layer that can be present in a CNN is the pooling layer, where the maximum value of a set of outputs is transmitted to the next layer. It is useful for the special size reduction of the output and can control overfitting. A combination of the three different kinds of layers forms the CNN (figure 36).

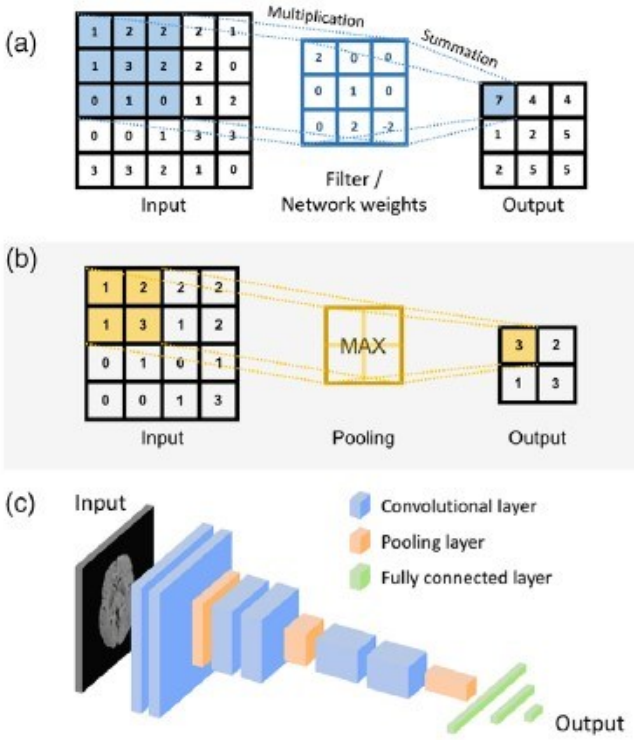


Figure 36: Graphic representation of the convolutional layer (a), of the pooling layer (b) and an architecture of a simple convolutional neural network [20].

In the medical field, an increasing interest is addressed to bioimaging analysis and processing though the use of DL. The areas in which these techniques can be applied are various: oncology, cardiology, ophthalmology, radiology, dentistry, neurology and more [24]. The use of DL in the computer-aided diagnoses (CAD), may increase the accuracy of the diagnoses and treatment calibration, supporting the clinician's decisions[Deep learning in medical image analysis]. Along the pipeline of image acquisition and processing, several steps could appreciate the application of DL:

- Tomographic image reconstruction: where externally measured data are related to internal structures.
- Image segmentation: identifying regions of interest in the image, such as organs, physiological structures and diseases.
- Image registration: the process of obtaining different images related to the same coordinate system. In this way is possible to match the contents of the different images.
- Classification: the process in which images are assigned to a particular class, in order to define the presence of abnormalities, lesion or diseases [24].

4.2 Classification algorithms

Classification algorithms aims to assign to a particular class the image in consideration. To do so, the image voxels are represented as a vector x containing the n voxels. This vector is then assigned to one of the m classes of interest; usually for the clinical purpose the classification is binary, for example AD positive or negative. Therefore, the algorithm is a sort of function f(x) which has as solution the class to which has to be assigned. The main classification algorithms used for the detection of AD from MRI images are listed in the following sections.

Support Vector Machine (SVM) is one of the most used classification algorithm. It is based on the construction of a hyperplane that maximizes the margin, which is the distance between the closest points of the different classes. These points are called support vectors and defines the optimal decision boundary. The classification is then made considering the side of the hyperplane in which the data fall. The decision boundary can be either linear or non-linear depending on the data with which the algorithm has to deal (if they are linearly separable or not). An example of linear and nonlinear SVM can be appreciated in figure 37 [25, 26].

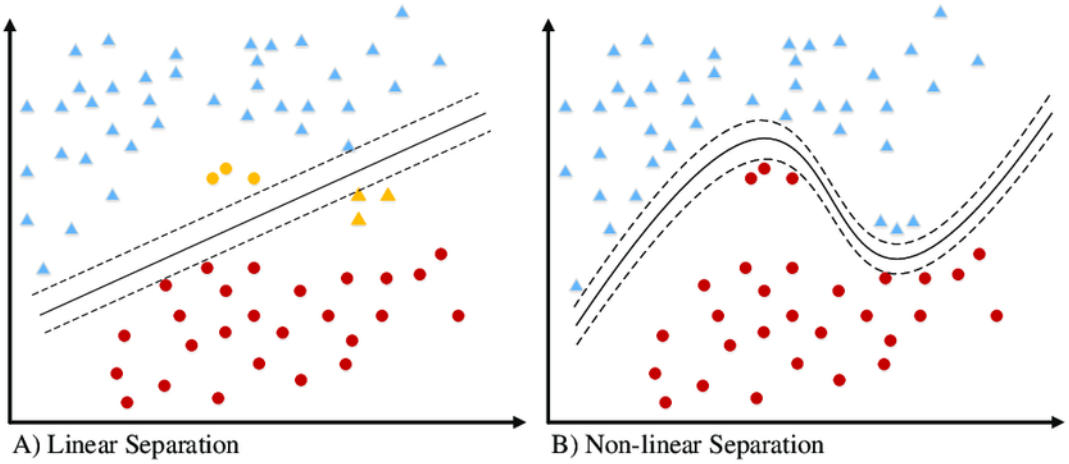


Figure 37: Difference between linear and non-linear SVM [27].

Random Forest (RF) algorithm is a classifier proposed by Breiman in 2001 [28]. It is characterized by a series of decision trees, which have as input a subset of the available

data sampled randomly, known as bootstrap sampling. At each node of each tree some features are selected, and the node is then subdivided in the best possible binary combination, forming two child nodes. The criterion with which the child nodes are defined is the impurity criterion, according to which the homogeneity of the child nodes with the parent one is maximized. The index related to this criterion is defined as Gini index and for two classes C_1 and C_2 is expressed as equation 8.

$$I_G(n) = 1 - \sum_{K=1}^{2} p_K^2 \tag{8}$$

In the Gini index, p_K represents the proportion of samples in the node assigned to C_1 or C_2. at the end, each tree estimate a target based on the selected features and the final predicted class is obtained through the average ore majority vote of all the trees outputs. The structure of the described algorithm is shown in figure 38.

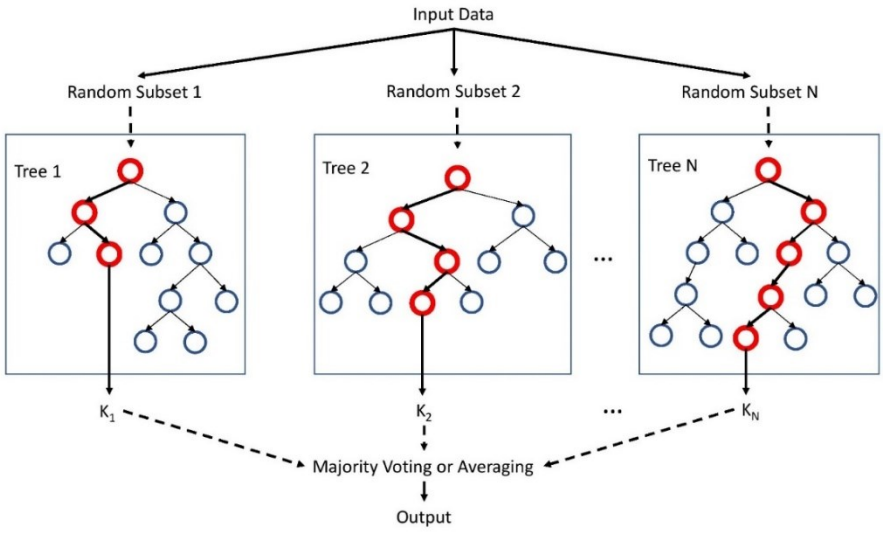


Figure 38: Sketch of the structure of a random forest [26].

4.3 Deep learning for Alzheimer's disease identification

DL in the clinical practise has been widely used in the past few years especially for the clinician support in the diagnoses. The automated classification of the cases based on the image analysis has been a frequent target in the researcher's studies, because it could provide a quicker and accurate identification. Several studies have been directed

on the early identification of AD through classification algorithms able to distinguish between different stages of the disease. In the present study, the focus is addressed on the use of MRI images for the diagnoses and the deep learning algorithms usually taken into account for this analysis.

4.3.1 Non-convolutional neural network

Several studies for the AD identification have been carried out with the use of non-convolutional artificial neural network (NC-ANN) even if they are not the most used. NC-ANN with the implementation of fully connected layers are employed as described previously. The main activation function used in the AD identification are the sigmoid and the ReLU functions. They are often used when dealing with large amount of data, for example the use of Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset which is a big collection of AD cases in different stages of the disease.

4.3.2 Convolutional neural networks

The most widely used for the AD identification are the convolutional neural network. Various combinations of fully connected, convolutional and pooling layers have led to different kind of algorithms. The implementation of convolutional and pooling layers in the algorithm allows a decreasing of the number of model parameters, providing a reduction of the system complexity. They are able to filter data and detect the features of interest in order to maximize the performances. They are often implied in the identification of AD thanks to their efficiency and performances in image recognition.

V. Literature review

A large amount of studies of the last few years have been focused on the detection and identification of AD, in the different stages on which this disease can appear, with the use of machine learning algorithms. In this chapter some of them are summarized, characterized by the use of different kind of ML and DL algorithms and their application on different kind of MRI scans, including structural MRI, functional MRI (fMRI) and diffusion MRI (dMRI).

This literature review has been conducted on PubMed and Scopus, two scientific literature databases based on the use of queries and keywords able to restrict the searching field. The keywords of interest included in this review were: “Alzheimer’s Disease", “AD”, “Dementia”, "Mild Cognitive Impairment", “MCI”, “Deep learning”, “DL”, “Machine learning”, “ML”, “Artificial intelligence”, “AI”, “Magnetic resonance”, “MRI”, “Structural MRI”, “Functional MRI”, “fMRI”. Then the articles have been selected considering the kind of algorithm used, the suitability for this study and the year of publication, choosing the most recent ones. These articles are listed in the following, ordered chronologically and focusing at first on structural and hybrid MRI and then on fMRI. In table 1 is possible to appreciate the selected studies with the main information that characterize the research allowing the quick comparison between them.

Mendoza-Leon R. et al. [29] proposed in 2020 an algorithm for the CAD of AD based on the analysis of a single slice of structural MRI scan. They presented a model based on the application of Supervised Switching Autoencoders (SSAs). The SSAs are a particular kind of supervised autoencoders characterized by the ability to perform unsupervised representation and supervised classification. The work done in this paper considers two applications of SSAs: the classification of AD patients with a prediction of two classes (healthy or AD-demented) and the identification of AD’s significant regions using a patch-based image analysis. The data are selected from the Open Access Series of Imaging Studies (OASIS) dataset, in particular 134 subjects for the training and 40 for testing. They extracted 2D-slices in the axial, coronal and sagittal planes. For the first part of the work, the one interesting for this study, the SSAs training was performed

with local patches taken from the previously described dataset. The classification is then performed on the test data based on the output of the prediction maps. The accuracy of the model depending on the axial, coronal and sagittal plane was respectively 87.5%, 90.0% and 90.0%. Regarding the sensitivity, the results obtained were 90.0%, 95.0% and 95.0% respectively. Finally, the resulted specificity were 85.0%, 85.0% and 85.0%. The proposed algorithm showed performances comparable to the studies carried on volumetric images. The high sensitivity, in particular in the coronal and sagittal plane are an interesting result since it is important in the clinical setting in order to avoid false negatives.

Zhang et al. [30] in a study published in 2019 proposed an extreme learning machine (ELM) model based on voxel-based morphometry (VBM) in order to distinguish between AD patient and normal controls. The ELM technique is characterized by a simple training and allows the optimization of the weights of the neural network also providing a higher speed of calculation. In the clinical field, this method was also used for the classification of other diseases. VBM parameters, on which the ELM model is based, are able to quantify the changes in the structure of cerebral tissue that characterizes the early stages of AD, such as atrophy of the hippocampus. The database from which the images were taken is the ADNI, in particular 58 AD patients and 94 normal control. The classification features taken in consideration were the texture features, the VBM parameters and the clinical background. Other classification algorithm were used on the same dataset in order to compare the performances of them with the one proposed, in particular SVM, partial least square (PLS) and Gaussian process regression (GPR) were considered. The best accuracy obtained was 96.0% with the ELM model compared with the 75.0% of the SVM, the 79.0% of the GPR and the 82.0% of the PLS. According to the authors, the higher performances of ELM can be explained by the fact that the model has to solve a minimum norm least squares problem and the solution is direct and fast, allowing a good generalization. Moreover, all the patient information were used for the creation of the model.

Wang Y. et al. [31] tested the use of machine learning on multimodal MRI (T1-weighted and diffusion MRI). The use of diffusion MRI for the AD identification is related to the ability of assessing white matter abnormalities with the use of tractography, which is an

estimation of white matter tracts. They used these data and the structural data to extract brain phenotypes used as features of the algorithms. The work was carried both on local cohort to test the algorithm and then on the Alzheimer's Disease Neuroimaging Initiative (ADNI) in order to test if could be generalized on a bigger dataset. They distinguished between AD patients, Mild cognitive impairment (MCI), subjects with memory complaints (SMC) and healthy control (HC). The algorithms used for classification were the random forest (RF), the logistic regression (LR) and the SVM with linear kernels, in order to compare the classifiers trained on the different brain characteristics. The accuracy in the classification obtained using the ADNI-2 dataset was: for AD/HC 96%, for MCI/HC 70%, for AD/MCI 75%. The performances in the prediction of MCI changing in AD using the logistic regression as the best approach, related to the ADNI-2 dataset and the combination of morphometry and connectomes are the following: accuracy of 59.0%, sensitivity of 60.0% and specificity of 68.0%. According to the authors, the results show the utility of considering the white matter structural connectomes as biomarker able to classify AD and MCI. In fact, considering the local dataset used, morphometry alone fails in classifying between MCI and SMC, suggesting a higher sensitivity of the white matter connectivity rather than the grey matter morphometry. On the contrary, in the case of the ADNI-2 dataset, the performances decrease of the 10% with respect to the local cohort.

Tam A. et al. in 2019 [32] tried to identify patient with a high risk of AD appearance by the use of cognitive conditions and structural MRI scans as training for machine learning algorithms in a high-specificity regime. T1-weighted MRI scans were taken from the ADNI-1 and ADNI-2 datasets, in particular 228 and 218 normal patients respectively, 397 and 354 MCI patients respectively, 192 and 103 AD patients respectively. Structural features were extracted and images segmented into grey matter, white matter and cerebrospinal fluid (CSF). Then a SVM algorithm have been trained in order to classify AD patients and cognitive normal patients. After that, a 2-step method has been applied in order to identify the patients with highly robust prediction outcomes and then the performances have been evaluated. The prediction of the progression of AD was characterized by a sensitivity of 55.1%, specificity of 95.6% with an accuracy of 69.3% and a quite high predictive value of 80.4%. What they found is a signature of cognitive

deficit and brain damage that highlights patients at high risk in the progression of MCI to AD.

S. Sarraf and G. Tofighi in a study published in 2016 [33] proposed a CNN in order to classify AD patients from fMRI scans. Functional scans of AD patients and normal participants were collected from the ADNI dataset. They were pre-processed applying motion correction, skull stripping and special smoothing. The functional scans were then aligned to the structural ones. The proposed architecture is the LeNet-5, one of the most used CNN, obtaining an accuracy of 96.85%. The authors assessed that this CNN algorithm has a double advantage in the AD classification: it is able to extract high level features thanks to the convolutional layers and is an efficient classifier. Moreover, according to the authors, the advanced pre-processing applied on the images before the use of the CNN has increased the performances of the classification.

Hojjati S. et al. [34] proposed a method in order to distinguish two cases of MCI: the once that will convert into AD, defined as MCI converter (MCI-C), and the once that do not convert into AD, the MCI non converter (MCI-NC). The goal of this study was to combine the use of machine learning and the graph theory, which models the brain as a network of nodes (anatomical parts) and edges (connectivity). The study was conducted on resting-state fMRI images collected from the ADNI dataset. After pre-processing of data, regions of interest (ROI) are identified and graph measures calculation has been performed. Feature extraction is then obtained using five different methods. They proposed a new sequential features collection (SFC) algorithm in order to obtain the best features subset. A SVM algorithm, with linear kernel, was trained and tested for the classification of MCI-C and MCI-NC. They tested also a SVM with non-linear kernel but according to the authors the performances were consistently poorer. Among the five different feature extraction algorithms, the Multivariate Minimal Redundancy Maximal Relevance (MRMR) was the one with the better performances, with accuracy, sensitivity and specificity equal to 91.4%, 83.24% and 90.1% respectively. The given performances suggest that the combination of these two methods can be a valid option in the early diagnoses of AD.

Gao et Al. [35] aim was to analyse changes in the white matter functional connectivity (FC) in order to understand their relationship with the progression of AD. Resting state functional MRI (rs-fMRI) from the ADNI database were used choosing different stages of dementia. Correlations between particular regions of interest (ROI) were calculated, composing the functional correlation matrix (FCM). In order to classify the normal condition and the different combinations of stages of AD, a SVM algorithm with a radial-bias function kernel was used. The initial features considered in the classification algorithm were the elements of the generated FCM and the more efficient features among them were selected through the use or a RF classifier of 200 trees. Data were divided into 10 subset, the SVM algorithm were trained on 9 of them and then tested on the remaining subset. The performances of this algorithm were higher in the identification of the later stages of AD, in fact were registered a sensitivity of 83.0% and a specificity of 81.0% in the recognition of AD demented and cognitive normal (CN) patients. Instead, adding the early stages of the disease, sensitivity and specificity were lower reaching values of 64.0% and 60.0% respectively. White matter FC is related to cognitive scores, so this study highlighted the potential of this feature in the AD analysis and could be used as machine learning feature.

Amini M. et al. [36] in a study compare several machine learning algorithms for the classification of fMRI scans of AD patients, which are: K-nearest neighbour (KNN), SVM, decision tree (DT), linear discriminant analysis (LDA) and RF. They also proposed a new CNN algorithm for the identification of AD severity. The aim is to classify the patients into the different stages of the disease based on the different features of each of them. The work is based on the use of fMRI scans taken from the ADNI dataset that were initially pre-processed applying the Quantum Matched-Filter Technique (QMFT) for the noise reduction. Then, feature extraction were performed with the robust multitasking learning function algorithm. The algorithms were therefore trained on these features in conjunction with the MMSE categories: low, mild, moderate and severe. The results highlight that the proposed CNN has an accuracy of 96%, the highest among the other machine learning algorithms. Considering the different stages, the accuracy for low, mild, moderate and severe AD were respectively 98.1%, 95.2%, 89% and 87.5%.

Janghel R.R. and Rathore Y.K. [37] included a pre-processing method before the use of a CNN algorithm in order to detect AD in both fMRI and PET images. The database on which they worked is the ADNI, choosing patients with a MMSE score of 32 on average. They proposed the conversion of 3D images into 2D before the feature extraction with a CNN. The one used in this study is the VGG-16 architecture, which stands for Visual Geometry Group and 16 is the number of layers. Then, in order to compare the performances of the proposed CNN and other classifiers, the classification is performed also with the use of SVM, linear discriminant, K-means clustering and decision tree. The obtained performances of the classification with VGG-16 of the fMRI images are an accuracy of 99.95%, which is better than the other methods, but according to the authors, the work can be improved in order to reduce the execution time of the classification.

Table 1: List of articles included in this review, from the newest to the oldest, in which are summarized the key information of each.

Authors	Year	Acquisition technique	Classification algorithm	Dataset	Performance *Accuracy*	*Sensitivity*	*Specificity*
Amini et al.	2021	fMRI	3D CNN	ADNI	96%	-	-
Janghel et al.	2021	fMRI	2D VGG-16	ADNI	99.9%	-	-
Gao et al.	2020	fMRI	3D SVM	ADNI	-	83%	81%
Mendoza et al.	2019	MRI	2D SSAs	OASIS	*Axial* 87.5% *Coronal* 90% *Sagittal* 90%	*Axial* 90% *Coronal* 95% *Sagittal* 95%	*Axial* 85% *Coronal* 85% *Sagittal* 85%
Wang et al.	2019	MRI+dMRI	3D LR+RF+SVM	Local+ADNI	59%	60%	68%
Zhang et al.	2019	MRI	3D ELM	ADNI	96%	98%	89%
Tam et al.	2019	MRI	3D SVM	ADNI	69.3%	55.1%	95.6%
Hojjati et al.	2017	fMRI	3D SVM	ADNI	91.4%	83.2%	90.1%
Sarraf et al.	2016	fMRI	3D CNN	ADNI	96.9%	-	-

VI. Materials and Methodology

Starting from the research done by S. Tomassini, A. Sbrollini et al. [38] in a recently-published study where they propose a 3D framework called Brain-on-Cloud for the identification of AD from structural MRI (sMRI) scans, the main purpose of this thesis is to verify the effectiveness of Brain-on-Cloud in pursuing the same classification task by taking into account a different data type from sMRI, precisely fMRI. To this aim, the model hyper-parameters are investigated, the best hyper-parameter combination is selected and used, and the results of the first experiment of Brain-on-Cloud by using both sMRI and fMRI scans are compared and discussed.

6.1 Dataset

The dataset on which this work is carried out is selected from a subset of 3D fMRI of the ADNI dataset, the ADNI2. Two subgroups of patients are considered: the AD patient and the cognitive normal (CN) control. Regarding the AD subgroup, 58 patients are included, in particular 33 males and 25 females with the age ranging between 56 and 88 years-old. Among the three scans available for each patient, “MoCoSeries”, “Perfusion_Weighted” and “relCBF”, only the last two are considered in the study for a consistency issue in the dimensions of the scans. Moreover, two male subjects have a double set of scans and two patients are missing the “Perfusion_Weighted”, leading to a total amount of 118 scans. Whereas, among the CN patients, 30 males and 38 females are chosen with the age between 56 and 84 years-old for a total equal to 68 participants, which provide three scans each and only two of them are taken into account, as for AD subgroup,. Therefore, a female gives a contribution of a double set of scans and a male lacks of a “Perfusion_Weighted” scan, leading to a total amount of 137 scans.

All the scans are obtained from Siemens scanners with a filed strength of 3 T and a slice thickness equal to 4 mm, stored in Digital Imaging and Communications in Medicine (DICOM) format.

6.2 Proposed methodology

In order to obtain a better outcome from the DL architecture that will classify the data, an important role is played by the optimization of the images in the pre-processing phase. Four serial experiments have been carried out in the reference study [38], starting from a basic pre-processing in the first experiment and implementing additional steps in the other ones till reaching the best-performing experiment in identifying AD from sMRI scans.

As for this thesis, the first experiment is the one considered. It includes intensity normalization and cropping of the images, explained in detail in the following.

All the implementations have been made in Python language using Google Colab, a Google notebook allowing the execution of Python code through the browser by exploiting cloud computing. The main steps characterizing the workflow of this thesis are summarized in the block scheme in figure 39.

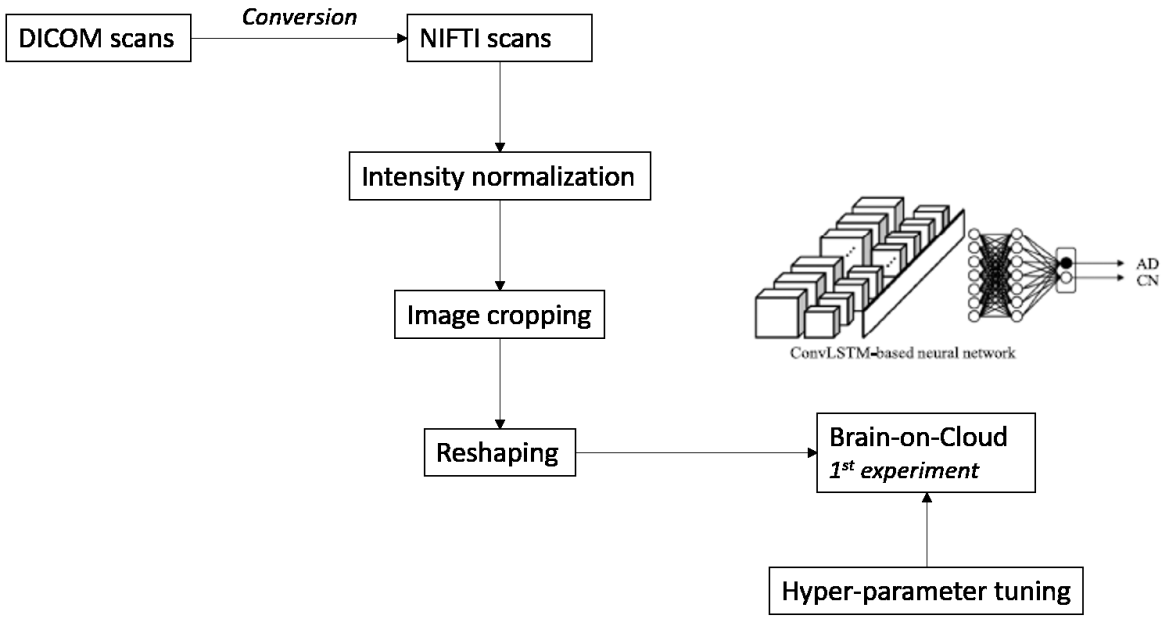


Figure 39: The pipeline of the pre-processing of the fMRI scans available. Both AD and CN scans underwent to the same steps before being used to feed the Brain-on-Cloud neural network.

6.2.1 Pre-processing

The first step of this thesis is the conversion of the images from the DICOM format to the compressed Neuroimaging Informatics Technology Initiative (NIFTI) one, using the Python library “dicom2nifti”. The NIFTI files are then stored maintaining the subdivision in patients in different folders containing the related scans.

Looking at the obtained files, the “MoCoSeries” presented a higher dimension with respect to the other kind of scans, hence are discarded from the considered scans, providing a total amount of 118 AD files and 137 CN files, of shape equal to 64 x 64 pixels for 24 slices.

Given the non-uniformity of intensity among the scans, an intensity normalization is necessary and performed with a designed function “normalize” subtracting the mean intensity and dividing the standard deviation. From the obtained data, the un-informative black voxels around the brain are deleted through a suitable automated cropping program. Due to the poor resolution of the data itself, smoothing is considered a unnecessary step for the improving of the images.

The resulted cropped images present a different shape among them because of the removal of black voxels so a reshaping is performed, obtaining a uniform dimension of 75 x 75 pixels for 24 slices.

The obtained pre-processed images are then used to feed the 3D framework Brain-on-Cloud, after proper adjustments for the right fitting of the data.

6.2.2 Brain-on-Cloud architecture and training

Brain-on-Cloud is a sequential neural network made by 6 layers and 10’917’186 total parameters built to digest sMRI scans. In [38], the first layer is a ConvLSTM layer, which is a variant of a recurrent neural network characterized by 8 convolution filters with a kernel of 3 x 3 allowing the features extraction. Unlike a CNN framework, ConvLSTM performs convolutional mechanisms by which to a given feature is able to assign a specific filter; moreover, the LSTM hidden unit, shown in figure 40, stores inside a memory cell information considered valuable through the opening and closing of three gates. In particular, the input gate collects new input data, the forget gate removes useless information and the output gate provides the information to be stored.

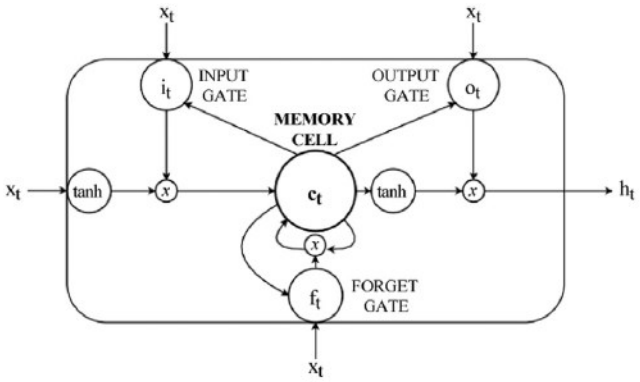


Figure 40: The LSTM hidden unit composed by three gates which decides the information to be stored in the memory cell. [38]

The second layer is a Dropout layer, with an initial dropout rate of 0.5, which allows the removal of random units in order to reduce the complexity of the network. The third layer is a Flatten layer able to level the features into a mono-dimensional tensor. The fourth layer is a Dense layer with 256 neurons characterized by the ReLu as activation function. The fifth layer is another Dropout layer with the initial dropout rate equal to 0.7. The sixth layer is another Dense layer made of 2 neurons and a Softmax activation function.

Always in [38], for the training of the model is used Keras library on TensorFlow backend, for 50 epochs with batch size of 10. In order to minimize the loss function, the optimizer used is the stochastic gradient descent and the binary cross entropy is set as loss function. For the prevention of overfitting early stopping callback with a patience of 5 is used.

In this thesis, instead, Brain-on-Cloud is extended to the analysis of fMRI scans. To do so, the neural network architecture is maintained unaltered, but the hyper-parameters are tuned until reaching the best performing configuration.

6.2.3 Hyper-parameter tuning

Hyper-parameters are the parameters of the model that cannot be estimated from data learning but have to be carefully chosen and applied before training the model. Therefore, in order to obtain an optimal network, it is necessary to explore the different combinations of them through the hyper-parameters tuning. Several techniques are available for this process that could be manual or automated. The manual testing limit is mainly that it is time consuming, especially when dealing with a large number of hyper-parameters. The automatic optimization, called hyper-parameter optimization (HPO) solves this issue,

reducing the human effort required, also improving the model performances and making it more reproducible [39].

Among the HPO methods available, the one selected and applied to the proposed model is the Grid search (GS), which evaluates the optimal hyper-parameters configuration from a set of values specified by the user. GS is a very common method also applied in the reference study [38], therefore it is selected for this thesis in order to obtain a fair comparison of the two applications.

The set of hyper-parameters evaluated are the batch size, the learning rate and the dropout rate of the two dropout layers, which are the ones influencing the model the most. The batch size is the number of images used for the training of the model at each epoch, the complete cycle of training that in this case are 50. This hyper-parameter is highly influencing since with a too large batch size the model will take too long to reach the convergence, while if it is too small the high variance does not provide good performances [40]. The learning rate is the variable determining how much the model should change in relation to the estimated error every time the weights of the model are updated. With a small learning rate, small changes in the weight will be recorded leading to a slow training process, while high learning rate will provide too large weight changes and a not proper optimization of the model [41]. Regarding the dropout rate, for dropout is intended the temporal removal of random units of the network and their connections, in order to prevent overfitting. The probability with which a unit is retained is the dropout rate [42].

Several combinations of the aforementioned hyper-parameters are investigated with the GS algorithm and are listed in Table 2. Between the selected values, are present the initial ones considered in the reference study (0.5 for the dropout rates, 0.001 as learning rate and 10 for batch size) and the ones obtained with the tuning of the best of the four experiments performed (0.6, 0.005 and 10).

Table 2: The combinations of the hyper-parameters considered for the tuning: dropout rate of the two dropout layers is set equal to [0.5, 0.6]; the learning rate is considered equal to [0.001, 0.005]; the batch size equal to [6, 10].

Dropout-rate 1	Dropout-rate 2	Learning rate	Batch size
0,5	0,5	0,001	6
0,5	0,5	0,001	10
0,5	0,5	0,005	6

0,5	0,5	0,005	10
0,5	0,6	0,001	6
0,5	0,6	0,001	10
0,5	0,6	0,005	6
0,5	0,6	0,005	10
0,6	0,5	0,001	6
0,6	0,5	0,001	10
0,6	0,5	0,005	6
0,6	0,5	0,005	10
0,6	0,6	0,001	6
0,6	0,6	0,001	10
0,6	0,6	0,005	6
0,6	0,6	0,005	10

GS is performed by “scikit-learn” Python library which allows the tuning of hyper-parameters of Keras models. The model is firstly wrapped with the “KerasClassifier” class from the “SciKeras” module, to be used in scikit-learn. Then, the “GridSearchCV” class is applied for the tuning. Given the set of values to evaluate, the default score optimized is the accuracy, so it is necessary to design a scorer in order to obtain the Area Under the Curve (AUC). At first each hyper-parameter was tested alone but given the dependency between them, especially learning rate and batch size, they are all evaluated together.

Given the stochastic nature of the algorithm, the procedure is run several times and considered the most recurrent outcome and its relative AUC. The best combination, so the one with higher AUC is then implemented in the model.

6.3 Statistics

The evaluation of the performances of the proposed model is made by the computation of the following parameters: accuracy (ACC), specificity (SP), sensitivity (SE), F1-Score (F1-S), Receiver Operating Characteristic (ROC) curve and AUC. The first four indexes are obtained from the equations (9), (10), (11) and (12).

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$SP = \frac{TN}{TN+FP} \tag{10}$$

$$SE = \frac{TP}{TP+FN} \tag{11}$$

$$F1 - S = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \tag{12}$$

From these equations, True Positives (TP) indicates the number of AD patients classified as affected by AD, True Negatives (TN) are the CN subjects classified as healthy, while False Positive (FP) and False Negative (FN) are the subjects classified in the wrong category.

VII. Results

In Figure 41 are displayed four consecutive slices of a pre-processed and reshaped scan in order to show the images that are used as input of the Brain-on-Cloud framework.

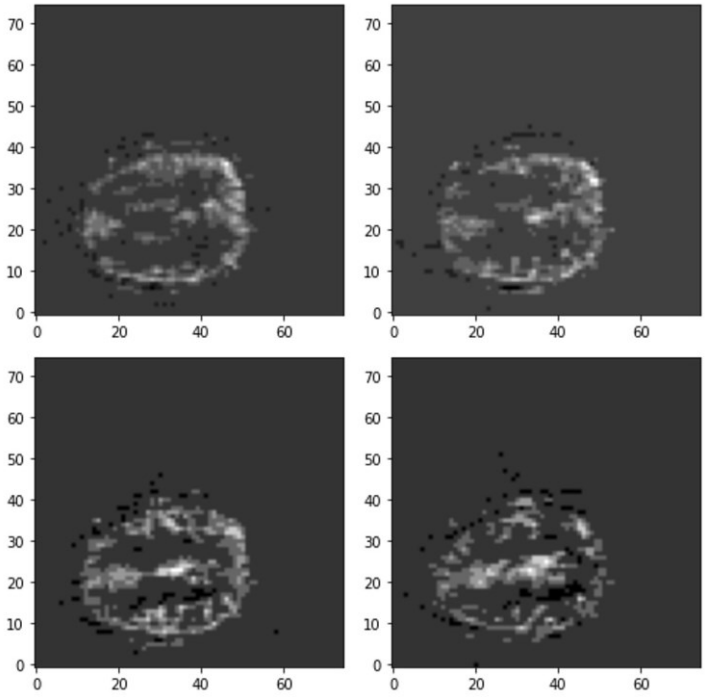


Figure 41: Four consecutive slices of the 24 composing a pre-processed and reshaped scan with dimension 75 x 75.

In Table 3 are listed the different combinations of the values considered in the hyper-parameter tuning performed with the GS algorithm according to the AUC metric, specified in the last column.

Table 3: The combinations of the hyper-parameters considered for the tuning: dropout rate of the two dropout layers equal to [0.5, 0.6]; the learning rate equal to [0.001, 0.005]; the batch size equal to [6, 10]. The AUC is specified as output of the GS algorithm.

Dropout-rate 1	Dropout-rate 2	Learning rate	Batch size	AUC
0,5	0,5	0,001	6	0.510
0,5	0,5	0,001	10	0.539
0,5	0,5	0,005	6	0.507
0,5	0,5	0,005	10	0.490
0,5	0,6	0,001	6	0.555
0,5	0,6	0,001	10	0.515

0,5	0,6	0,005	6	0.550
0,5	**0,6**	**0,005**	**10**	**0.558**
0,6	0,5	0,001	6	0.510
0,6	0,5	0,001	10	0.437
0,6	0,5	0,005	6	0.467
0,6	0,5	0,005	10	0.496
0,6	0,6	0,001	6	0.556
0,6	0,6	0,001	10	0.465
0,6	0,6	0,005	6	0.509
0,6	0,6	0,005	10	0.495

Implementing the best hyper-parameters combination as model parameters, the output obtained is displayed in figure 42. From this plot it is possible to appreciate the ROC curve for each split of the stratified shuffle-split cross validation scheme of five splits, and their average in different colours; moreover, in brackets are specified the relative AUC values. In order to allow the comparison between this experiment and the first one of [38], which includes the same basic pre-processing steps but a different data type, a similar plot is shown in figure 43.

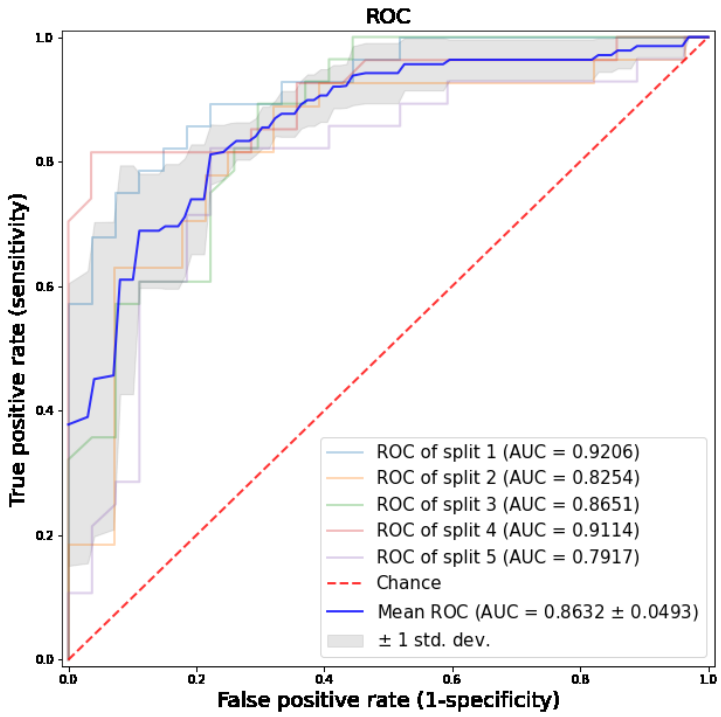


Figure 42: ROC of the five different splits and relative AUC. The average of the splits is displayed in blue.

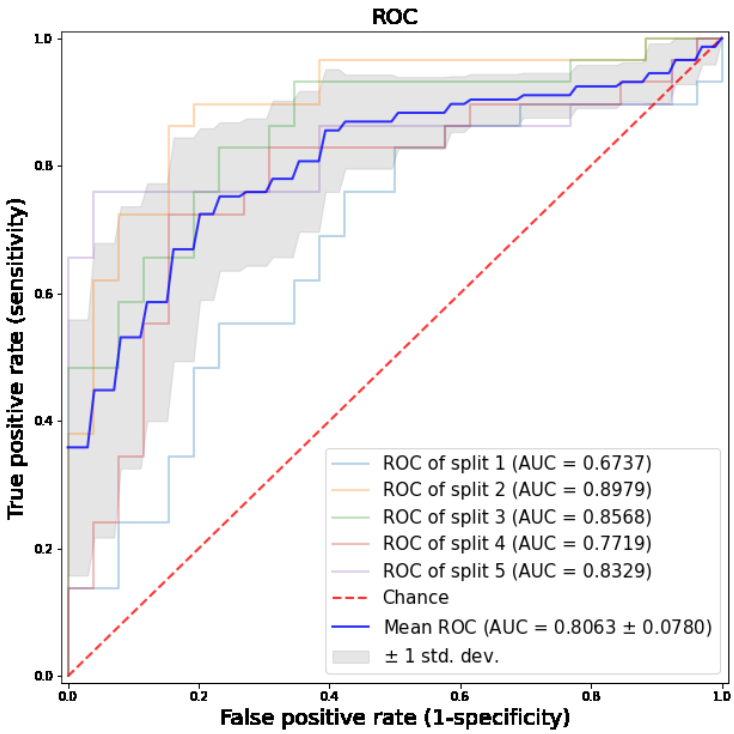


Figure 43: ROC of the five different splits and relative AUC of the first experiment performed on Brain-on-Cloud in [38].

In Table 4 are listed the main parameters considered for the evaluation of the model: ACC, SE, SP and F1-score; all of them expressed in percentage. The first column is relative to the tuned model with fMRI scans as input; while the second column is the evaluation of the first experiment of Brain-on-Cloud in [38].

Table 4: Comparison between the results obtained from the fMRI and sMRI scans. Accuracy, Sensitivity, Specificity and F1-score of the proposed model are displayed in percentage.

	fMRI scans	*sMRI scans*
Accuracy %	77.45 ± 3.17	75.27 ± 6.15
Sensitivity %	81.08 ± 10.59	68.46 ± 11.77
Specificity %	73.73 ± 6.13	81.38 ± 8.04
F1-score %	77.95 ± 4.93	77.64 ± 5.07

VIII. Discussion

This study explored an innovative 3D framework recently proposed by Tommasini S. and Sbrollini A. [38] called Brain-on-Cloud. Starting from their work on structural input images, the application of this neural network on functional brain scans was analysed. In [38] four experiments were performed on the data, considering different degrees of image pre-processing, improving it adding a step at each experiment.

The present study took in consideration the first experiment implementing intensity normalization and image cropping, the basic pre-processing steps necessary to handle the data. After the conversion from DICOM to NIFTI format and the pre-processing, ADNI images appears to be as shown in figure 41. From these plots, characterized by the same shape given by the reshaping process performed, is appreciable the not so high resolution of the images, which may affect the performances of the model.

A crucial role is played by the hyper-parameters governing the model: batch size, learning rate and dropout rate. The values included in the tuning are chosen in order to include the values used in the reference study in order to make a fair comparison among the models and see their behaviour when dealing with different datasets. The resulted AUC obtained for each combination of values (Table 3), highlights as optimal set of hyper-parameters the following: batch size equal to 10, learning rate equal to 0.005, dropout rate of the first dropout layer equal to 0.5 and of the second one equal to 0.6. These results are comparable with the ones of the tuned model in [38] except for the second dropout rate, which is the same of the first one in the reference study. It is observable that generally for an higher value of learning rate is preferable a larger number of batch size and, similarly, for a smaller learning rate is better a lower batch size; so a direct proportionality characterizes the dependency among the two hyper-parameters.

Implementing the aforementioned combination of hyper-parameters, the output of Brain-on-cloud is surprisingly better than before the tuning process and is displayed in figure 42. The mean AUC obtained is about 0.86, comparable with the one in figure 43, which is relative to the first experiment in [38] on structural brain scans. Moreover, the evaluation made on the model through the calculation of the statistical parameter explained in the previous chapter, highlights the comparable or even higher performances of the model

with respect to the reference one (Table 4): ACC=77.45 $\pm$ 3.17 %, SE=81.08 $\pm$ 10.59 %, SP=73.73 $\pm$ 6.13 % and F1-S =77.95 $\pm$ 4.93 %. In particular, the sensitivity appears to be higher than the one with sMRI scans and this is a crucial parameter for a diagnosis: lower the sensitivity higher will be the number of false negatives labelled by the model and false negatives lead to a delayed diagnosis of AD.

Therefore, an observation has to be made: the values relative to sMRI scans considered for the comparison are not from a tuned model; the tuning process was performed on the fourth experiment, reaching even higher performances. Despite this, the obtained performances are more than acceptable if considering the low resolution of the input fMRI images of the used dataset. Even if this preliminary evaluation on fMRI data seams to provide better results than sMRI scans, further improvements on the pre-processing phase, maybe implementing all the four experiments in [38], could lead to a different outcome.

This preliminary evaluation on fMRI scans shows how versatile Brain-on-cloud is, providing good results with different types of image, and the importance of the hyper-parameter tuning, fundamental for the improving of the performances.

Future developments of this study could be the implementation of a more complete pre-processing of the fMRI images that may improve the model performances. Moreover, the promising results obtained may open the perspective of a hybrid approach, using Brain-on-Cloud on both sMRI and fMRI images in order to obtain even better performances.

Conclusion

AD is the most common kind of dementia affecting millions of people worldwide. Its severity and social impact is given by the consequences related to this pathology, affecting clearly patients' lives but also the ones of family and friends. Physical and cognitive impairments are often reasons of shame and social isolation for the subjects and this condition worsen in time, given the degenerative nature of the pathology. The large amount of research in this field in the last decades has opened the perspective of the use of ML for the computer-aided diagnoses of AD, in order to accelerate and improve this phase.

The purpose of this study is the analysis of this pathology, not well understood yet, and the possible diagnoses, proposing the implementation of Brain-on-Cloud framework on function magnetic resonance images (fMRI) in order to explore different applications of this neural network. Deep learning algorithms, in fact, have demonstrated to be the more accurate and precise ones for this purpose, given the automated feature extraction which also lightens the human computational effort.

In this thesis only a preliminary evaluation is made on fMRI scans, implementing the first experiment performed in [38], which includes the basic pre-processing steps necessary to handle data. The promising results obtained opens the perspective of the use of this kind of images as input data of Brain-on-Cloud for AD diagnosis, obtaining comparable or even better performances with respect to the already tested structural magnetic resonance images (sMRI). The outcome of this research also highlights the importance of hyper-parameter tuning as necessary tool for the improvement of the model, calibrated for the different input data.

Future improvements on this study can be adding the implementation of the other experiments proposed in [38], incrementing the quality of the image with a more complete pre-processing. Moreover, the handling of both fMRI and sMRI images could open the perspective of a hybrid approach for the diagnosis of AD.

Bibliography

[1] Dale Purves, George j. Augustine et al. 2004. Neuroscience: Third Edition. Sinauer Associates, Inc.

[2] Lauren Thau, Vamsi Reddy, Paramvir Singh, 2021. Anatomy, Central Nervous System. National Library of Medicine.

[3] Ackerman S. 1992. Discovering the brain. Major structures and functions of the brain.

[4] Dheaa Zageer, Sundus F Hantoosh 2019. Amino Acids Deficits in Brain. Book Publisher International, India, United Kingdom.

[5] Kenia A. Maldonado; Khalid Alsayouri 2021. Physiology, Brain. National Library of Medicine.

[6] Christian Meisel. 2008 Self-organized criticality in adaptive neural networks

[7] Kevin Ashley; Forshing Lui 2022. Physiology, Nerve. National Library of Medicine.

[8] Jianshi Tang, Fang Yuan, 2019. Bridging Biological and Artificial Neural Networks with Emerging Neuromorphic Devices: Fundamentals, Progress, and Challenges.

[9] Jia-Qin Yang, Ruopeng Wang et al. 2020. Neuromorphic Engineering: From Biological to Spike-Based Hardware Nervous Systems.

[10] Alzheimer's association. 2017 Alzheimer's disease facts and figures. Special Report.

[11] National Institute on aging. Alzheimer's Disease Fact Sheet, https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet.

[12] Mark W. Bondi, Emily C. Edmonds et al. 2017. Alzheimer's Disease: Past, Present, and Future.

[13] D. Dawbarn and S. J. Allen (editors). Neurobiology of Alzheimer's disease. Oxford University Press, New York, 3rd edition, 2007.

[14] Wikimedia Commons Contributors. File:Alzheimer’s disease brain comparison.jpg, https://commons.wikimedia.org/wiki /File:Alzheimer%27sdiseasebraincomparison.jpg.

[15] Sneham Tiwari, Venkata Atluri, 2019. Alzheimer’s disease: pathogenesis, diagnostics, and therapeutics.

[16] D. Scheuner, C. Eckman, M. Jensen et al. Secreted amyloid beta protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. Nature Medicine, 2(8):864{870, 1996.

[17] Katherine Rachel Gray. Machine learning for image-based classification of Alzheimer's disease. 2012

[18] Keith A. Johnson, Nick C. Fox et al. 2012 Brain Imaging in Alzheimer Disease.

[19]https://home.csulb.edu/~cwallis/482/fmri/fmri.html#:~:text=What's%20the%20Difference%20Between%20MRI,dimensional%20pictures%20of%20anatomic%20structure

[20] Maciej A. Mazurowski, Mateusz Buda, MS et al. 2019. Deep Learning in Radiology: An Overview of the Concepts and a Survey of the State of the Art With Focus on MRI

[21] Nicolas Brunel, Vincent Hakim, andMagnus JE Richardson. Single neuron dynamics and computation. Current Opinion in Neurobiology, 25:149–155, 4 2014.

[22] Michael London and Michael Häusser. Dendritic Computation. Annual review of neuroscience, 58, 2005.

[23] G. Zaharchuk, E. Gong et al. 2018. Deep Learning in Neuroradiology.

[24] Karen Drukker, Pingkun Yan et al. 2021. Biomedical imaging and analysis through deep learning.

[25] V. N. Vapnik and A. Lerner. Pattern recognition using generalized portrait method. Automation and Remote Control, 24(6):774{780, 1963.

[26] Tory O. Frizzell, Margit Glashutter at al. 2021. Artificial intelligence in brain MRI analysis of Alzheimer's disease over the past 12 years: A systematic review.

[27] Kamran Kowsari, Kiana Jafari Meimandi et al. 2019. Text Classification Algorithms: A Survey.

[28] Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

[29] Mendoza-Leon, R., Puentes, J., Uriza, L.F., Hernandez Hoyos, M., 2020. Single-slice Alzheimer's disease classification and disease regional analysis with supervised switching autoencoders. Comput. Biol. Med. 116, 103527

[30] Zhang, F., Tian, S., Chen, S., et al., 2019. Voxel-based morphometry: improving the diagnosis of Alzheimer's disease based on an extreme learning machine method from the ADNI cohort. Neuroscience 414, 273–279

[31] Wang, Y., Xu, C., Park, J.H., et al., 2019. Diagnosis and prognosis of Alzheimer's disease using brain morphometry and white matter connectomes. NeuroImage Clin. 23, 101859

[32] Tam, A., Dansereau, C., Iturria-Medina, Y., et al., 2019. A highly predictive signature of cognition and brain atrophy for progression to Alzheimer's dementia. Gigascience 8 (5), 1–16.

[33] Sarraf S., Tofighi G. 2016. Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks. arXiv. 20161603.08631

[34] Hojjati, S.H., Ebrahimzadeh, A., Khazaee, A., Babajani-Feremi, A., 2017. Predicting conversion from MCI to AD Using resting-state FMRI, graph theoretical approach and SVM. J. Neurosci. Methods 282, 69–80.

[35] Gao Id, Y., Sengupta, A., Li, M., et al., 2020. Functional connectivity of white matter as a biomarker of cognitive decline in Alzheimer's disease. PLoS One 15 (10), e0240513.

[36] Amini M., Pedram M., Moradi A., Ouchani M. 2021. Diagnosis of Alzheimer's Disease Severity with fMRI Images Using Robust Multitask Feature Extraction Method and Convolutional Neural Network (CNN) Comput. Math. Methods Med.

[37] Janghel R.R., Rathore Y.K. 2021. Deep Convolution Neural Network Based System for Early Diagnosis of Alzheimer's Disease.

[38] Tomassini S., Sbrollini A. et al. 2022. Brain-on-Cloud for automatic diagnosis of Alzheimer's disease from 3D structural magnetic resonance whole-brain scans.

[39] Yang L. and Shami A. 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice.

[40] Kandel I. and Castelli M. 2020. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset.

[41] Attoh-Okine N. 1999. Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance.

[42] Srivastava N., Hinton G. et al. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15.

Sitography

[45] https://jnjinstitute.com/sites/default/files/2019-03/092892-181219-Anatomy-Physiology-of-the-Brain.pdf

[46] http://what-when-how.com/neuroscience/the-cerebellum-motor-systems-part-1/

[47] https://content.byui.edu/file/a236934c-3c60-4fe9-90aad343b3e3a640/1/module6/readings/neuron_structure.html

[48] https://www.blendspace.com/lessons/s2RuMIVmOkfTig/nervous-system

[49]https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/Book%3A_General_Biology_%28Boundless%29/35%3A_The_Nervous_System/35.05%3A_Ho

w Neurons Communicate - Nerve Impulse Transmission within a Neuron-Action Potential

[50] https://www.airalzh.it/i-numeri/

[51] https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf