UNIVERSITÀ
POLITECNICA
DELLE MARCHE

FACULTY OF ENGINEERING

MASTER'S DEGREE IN BIOMEDICAL ENGINEERING

CLASSIFICATION OF ALZHEIMER'S DISEASE FROM MAGNETIC RESONANCE AND PATIENTS' METADATA

Candidate:

Muaaz Alshalak

Supervisor:

Prof. Laura Burattini

Co-supervisor:

Prof. Aldo Franco Dragoni

Dr. Agnese Sbrollini

Academic Year 2021-2022

ABSTRACT

Alzheimer's disease (AD) is the most frequent type of dementia, that usually affects elderly people. It's a progressive disease that begins with modest memory loss and advances to the inability to talk and respond to the environment. Parts of the brain that control thought, memory, and language are involved in the injury that AD causes. It can have a significant impact on a person's capacity to carry out daily tasks. Since there is no completely efficient treatment for AD, it is useful to be detected in its early stage. This increases the chance of benefiting from treatment and increases its efficiency. The common way to do so is the use of Magnetic resonance imaging (MRI). MRI is being used in computer-aided diagnosis (CAD) to help the medical doctor in taking medical decisions. To increase the efficiency of AD detection using artificial intelligence-based CAD systems, the usage of patients' metadata is being used combined with the MRI scans.

The aim of this work is to study how this usage of metadata could affect CAD systems for the early detection of AD.

This was done by investigating literature that combined the metadata with MRI scans. Those literature were selected by applying a two-steps criterion for selection; first step selection of 35 papers that contains the usage of the deep learning (DL) for AD diagnosis, while the second one resulted in the 7 papers that focus on the utilize of metadata for AD classification. The first papers showed that a novel DL framework was developed, that connects a Fully convolution network (FCN) to Multilayer perception (MLP)to classify AD from multimodal inputs of imaging data (MRI) and non-imaging data age, gender, and Mini-Mental State Examination (MMSE) score. This paper shows that inserting the non-imaging variables into the imaging input has a significant impact on the performance of the DL model in classifying the AD.

The second one indicated developing a convolution neural network-based AD classification algorithms using a coronal slice of T1-weighted MRI scans from AD patients which cover the medial temporal lope and age/gender-matched CN control from two populations that differ in education level and ethnicity. The results obtained show the generalizability of our models to different patients with different education levels and ethnicities.

The third one represented a deep multi-task multi-channel learning (DM^2L) framework for simultaneous clinical score regression and brain disease classification. Using the MRI scans and demographic data (age, gender, and education) helps improve the learning performance of the proposed method in term of AD classification and clinical score regression.

The fourth one showed an End-to-End 3D Convolution Neural Network (3D-CNN) model was created to diagnose AD and its prodromal stage, MCI. Moreover, this model contributes to identifying significant disease biomarkers. The result obtained shows that the proposed end-to-end classification model can classify the subjects more accurately when the metadata is integrated.

The fifth one demonstrated a DL model to diagnose AD by using sagittal MRI images with considering age and sex as metadata. The result shows that there are minor improvements when including the metadata,

especially for the first two classes (Cognitive Normal and Very Mild Dementia). The main reason for that is due to the more minor imbalance between the stages and the larger amount of data.

The sixth one illustrated that combining structural information contained in the MRI images with the cognitive function assessed (The Wechsler Memory Scale Logical memory (LM) test and MMSE) helps to improve the performance of the fusion model to predict the Mild Cognitive Impairment (MCI).

The final paper presented a significantly inferior results when the metadata was used in comparison with the case of not using it. The possible reason for that is because the metadata are highly biased by the characteristics of the population included in the datasets

The input of the final classification layer is almost the same in all the studies. This input is a features vector that contains two types of features. The first type is the hidden features extracted from MRI scans using DL networks. The other features are the metadata contained in the description of the scans data. This combination shows a slight increase in the overall performance of the classification problem. Some papers showed no significant change. On the other hand, introducing those metadata features results in problems such as bias of the classification model. This could be explained, because of the low number of studies. Therefore, it is advised to conduct more experiments in order to obtain uniform and generalized insights.

ACKNOWLEDGMENTS

First and foremost, I wish to express my sincere appreciation to my supervisor, *Prof. Laura Burattini*, who has the substance of genius for giving me the chance to work with her. I wish to express my deepest gratitude to my co-supervisors *Dr. Agnese Sbrollini*, and *Dr. Selene Tomassini* for their valuable guidance throughout my internship period, you provided me with the precise information that I needed to choose the right direction and successfully complete my dissertation. In addition, I would like to thank *my parents* for their wise counsel and sympathetic ear. My warm and heartfelt thanks go to *my brothers,* you are always there for me, wherever and whenever that was. They kept me going on and this work would not have been possible without their input. I would like to thank my beloved, my soul, *Ola*, which was the cause of all kinds of support on the most difficult days and nights. I could not have completed this dissertation without the shoulders of one of my best friends *Nibras*, the partner, the brother. I would like to thank my friends and colleagues that I have met in this home far away from my home, which is called UNIVPM. Especially *Bilel*, the prince of optimism.

INDEX

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

AD is an irreversible progressive neurological disorder characterized by progressive memory loss and a decline in activities of daily life, the retardation of thinking, and changes in personality and behaviors, starting with mild memory impairment in the early stages and progressing to a total loss of the physical and mental faculties. AD is an age-related but not an age-dependent disease. Additionally, Accurate diagnosis and early detection of AD are essential in improving the quality of life of affected patients and their families, which require precise medical assessment, including patient history in addition to physical and neurological examinations. MRI provides a non-invasive solution to potentially identify abnormal structural brain changes more sensitively. MRI is widely used to capture the structural changes in the brain caused by AD in its early stages.

The DL methods have shown promising results, garnered significant attention between the researchers, for computer-aided diagnosis. DL is a part of machine learning in artificial intelligence that permits the mode to learn classification tasks from raw data. A branch of DL neural network is CNN. CNN is used to extract high-level features from images automatically to analyze and classify the images. The convolutional layers' nodes exchange connection weights to help find the same local feature across all input channels. The following pooling layers between convolution layers minimize computing complexity and generate a hierarchical collection of visuals. Modern CNN systems replace the pooling layer with a convolution layer with a stride higher than one to improve computational efficiency.

Increasing the performance of the DL model is important to enhance the efficiency of CAD systems. To enable the usefulness and interoperability of neuroimaging data, several needs and limits must be satisfied. Data without description is pointless. As a result, the picture dataset, or any dataset in general, should contain metadata along with the image. To do so, additional useful features could be combined with the deep features, which were automatically extracted from CNN. Perfect candidates for the added features are the metadata information of the patents. There are numerous metadata that can be used clinically. Medical, demographic, and experimental data can all be categorized. Examples of medical data include blood type, the apolipoprotein E (APOE) genotype, Body Mass Index (BMI), and organ volumes. Demographic factors include age, gender, race, and educational level. The experimental data also includes information about the tools used to acquire the data. If you're using MRI imaging, it's possible you've been injected with a chemical, which affects the magnetic field strength, sampling frequency, and the number of slices.

Metadata is undeniably useful when employing data collection. It might help both the operator and the statistical model. The operator may acquire significant insights when using the analysis. However, the statistical model gets the information together with the original data and utilizes it to create predictions. The researchers claim that classification tasks based on imaging and genetic data beat clinical expertise in digital pathology. When patient demographics are integrated with imaging data, high accuracy scores in binary classification tasks have been attained.

The aim of this work is to examine the effect of adding the metadata information to the deep features vectors, produced using CNN. Knowing how such features impact the process is critical for designing more accurate and resilient systems.

CHAPTER 1

ANATOMY AND PHYSIOLOGY OF THE BRAIN

The brain is a complex organ, weighing about three pounds (1 - 1.5 Kg) in the average adult, that regulates and controls all functions of the body, and interprets information from the external environment. Judgment, intelligence, creativity, emotion, and memory are a few of the many things controlled by the brain. it also regulates aspects of the body including body temperature, blood pressure, and the activity of internal organs to remains the human body's safety and response to its environments. Protected within the skull, surrounded by three membranes, meninges. The order of three from outside to inside is Dura matter, though and fibrous, Arachnoid, an intermediate thin layer, Pia matter. The main parts of the brain are the cerebrum, cerebellum, and brainstem.[1]
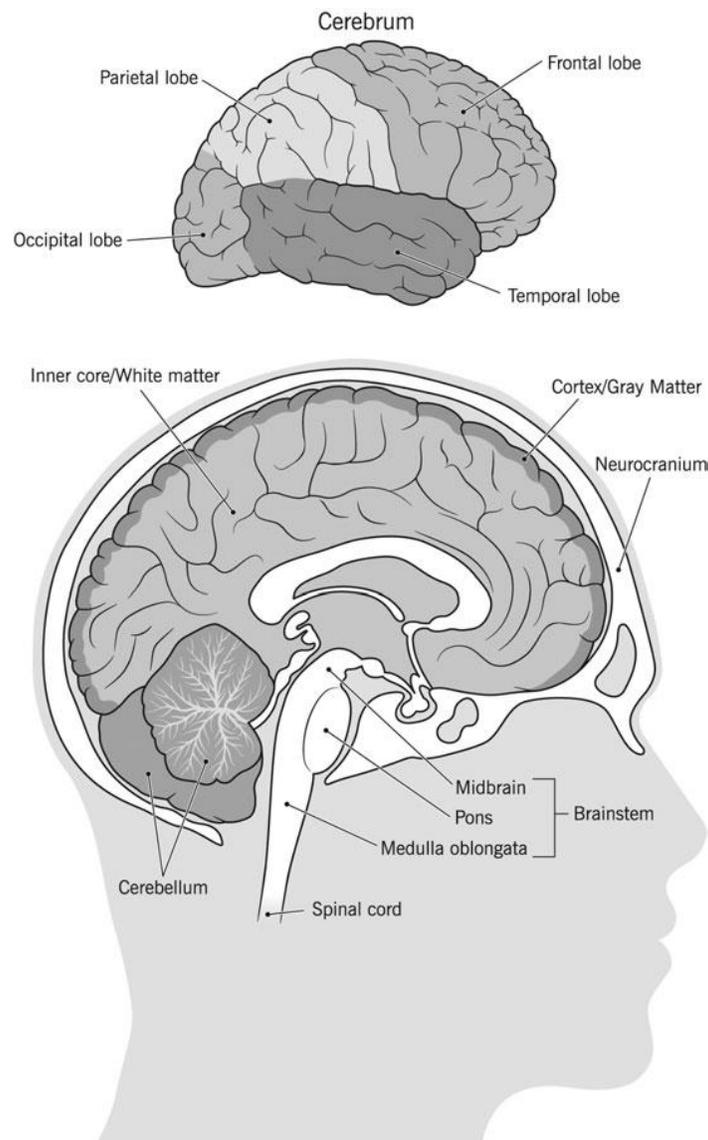
FIGURE 1 BASIC DIAGRAM OF THE BRAIN [1].

1.1 CEREBRUM
Presents the largest part of the brain and is composed of two hemispheres (right and left).

Responsible for the performance of higher functions: interpreting touch, vision, and hearing, emotions, learning, and fine control of movement. The extrinsic cortex of the cerebrum is composed of gray matter, billions of neuron cell bodies and unmyelinated axons arranged in six discrete layers. The internal region is composed of white matter, tracts of myelinated axons. A group of subcortical gray matter called basal nuclei located deep within the cerebral white matter is a third basic region of the cerebrum. These nuclei are important regulators of skeletal muscle movement. Elevated ridges of tissue called gyri, separated by shallow groves called sulci mark nearly the entire surface of the cerebral hemispheres. Fissures, which are deeper groves, separate large regions of the brain. Each cerebral hemisphere is responsible for a motor function to the opposite side of the body. They are not functionally equal, in each individual, there is one hemisphere that is always dominant. The dominant one controls language, mathematical and analytical functions. The non-dominate hemisphere controls simple spatial concepts, recognition of faces, some auditory aspects, and emotion.[1]
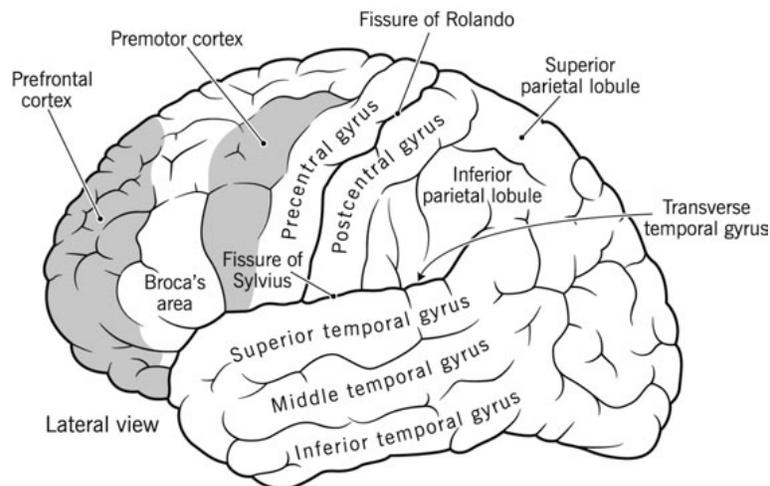
FIGURE 2 CEREBRUM [1].

1.1.1 CEREBRAL LOBES
The surface of each cerebral hemisphere is commonly divided into four lobes depending on the formation of a constant pattern over the cortex by sulci and gyri. The four lobes are:

- Frontal lobe: The largest of the cerebral lobes, located under the frontal bone. the precentral gyrus is the most important structural, which constitutes the primary motor region of the brain. Paralysis on the opposite half of the body occurs by injury to parts of the precentral gyrus. The premotor cortex is located in front of this region, where complex movements are orchestrated. Forward to this region, there is the prefrontal cortex, which exerts inhibitory control over actions, like distinctly human abilities as expecting the consequences of an action depend on the normal functioning of the prefrontal cortex. Other parts of the frontal lobe constitute Broca's area, a region involved with speech.

- Parietal lobe: Existing behind the central sulcus, this lobe is composed of three parts: the postcentral gyrus, the superior and inferior parietal lobules. The postcentral gyrus receives sensory signals from the other half of the body. The location of the superior parietal lobule is below and behind the postcentral gyrus. It is regarded as the association cortex. The integration of multiple sensory signals is involved by the inferior parietal lobule. The parietal lobe is responsible for the Interpretation of language, words, sense of touch (pain, temperature), and spatial and visual perception. which assist the people to perceive and interpret the world around them.

- Temporal lobe: located along the side of each hemisphere, below the lateral sulcus. The outer surface of the temporal lobe is an association area, where movement and sensory functions are integrated. Two transverse gyri, near the margin of the lateral sulcus, form the primary auditory area of the brain. The secondary auditory area which surrounded the gyri is a less finely tuned area. The temporal lobe is responsible for processing the information associated with hearing and equilibrium, Involved in memory formation and retention through connection to the hippocampus.

- Occipital lobe: The occipital lobes are the most posterior of the cerebral lobes, lying behind the parietal and temporal lobes. These lobes are concerned generally with vision. The occipital lobes house the primary visual cortex which receives visual information from the eyes. The primary visual cortex exists on the most posterior part of the occipital lobe. The primary visual cortex divisions (right and left) receive and process information from the contralateral visual field. The visual association cortex is located anterior to the primary visual cortex, which processes and interprets visual information received from the primary visual cortex allowing for convenient visual perception. This lobe is responsible for processing visual information (color, light, movements) by interpreting signals sent by the eyes and is related to our understanding of the written word [1].

1.1.2 CEREBRAL VENTRICLES

The brain has hollow fluid-filled cavities called ventricles. The choroid plexus is located inside the ventricles, a ribbon-like structure, which makes clear colorless cerebrospinal fluid (CSF). CSF flows inside and around the brain and spinal cord to help it from injury. This fluid is constantly absorbed and replenished. There are two ventricles deep within the cerebral hemispheres called the lateral ventricles, which connect with the third ventricle, interventricular foramina are responsible for the draining of the fluid between the first two ventricles and the third one. A long narrow tube called the aqueduct of Sylvius serves as a connection between the third ventricle with the fourth ventricle. CSF flows from the fourth ventricle into the subarachnoid space where it bathes and cushions the brain. Arachnoid villi are special structures in the superior sagittal sinus, which are responsible for recycling or absorbing the CSF. A balance is maintained between the absorbed amount of CSF and the produced amount. The buildup of CSF, which can cause enlargement of the ventricles (hydrocephalus) or collection of fluid in the spinal cord (syringomyelia) can be caused by a disruption or blockage in the system [1].
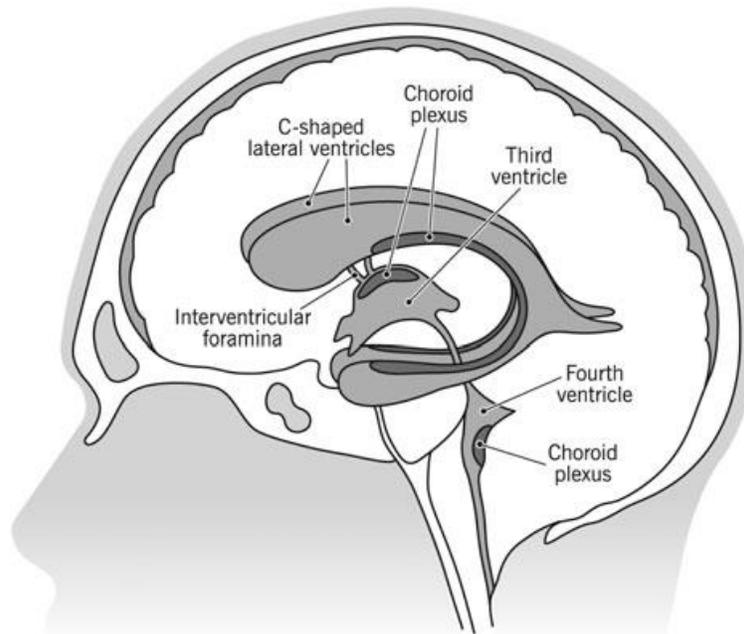
1.2 LIMBIC SYSTEM

Is a set of brain structures that are centrally located in the brain. The limbic system is underneath the cerebral cortex and above the brainstem. The hippocampus and the amygdala are two of the major structures. They are necessary for the brain's processing of emotion and motivation. There is a pair of the hippocampus, one in each hemisphere of the brain, and looks like the shape of a curvy seahorse. It is involved in recognizing new experiences, learning, and memory. The hippocampus is particularly involved in the creation of short-term memories and associating memories with various senses. The amygdala, on the other hand, is located right next to the hippocampus and plays a central role in emotional responses, including feelings like pleasure, fear, anxiety, and anger.

- Thalamus: Located close to the center of the brain; egg-shaped. considered as a relay station of the brain, receives input from all the senses, except smell (olfaction), performs preliminary analyses, and directs messages to various parts of the brain. Responsible for processing the information from the cerebellum and other brain areas involved in the movement.
- Hypothalamus: Located below the thalamus, the most basic function is in homeostasis, maintaining a steady internal state, regulating hormone release. Most of the autonomic functions are controlled by this region: hunger, blood pressure, heart rate, and sexual activity. The hypothalamus also acts as an interface between the nervous system and the endocrine system, and in the regulation of sexual motivation and behavior.

- Basal ganglia: The basal ganglia are a group of structures, situated at the base of the forebrain and top of the midbrain. Includes the caudate, putamen and globus pallidus. These nucleus work with the cerebellum to coordinate fine motions like fingertip movements. The main function is to regulate voluntary movements, including eye movements, help with balance [1].
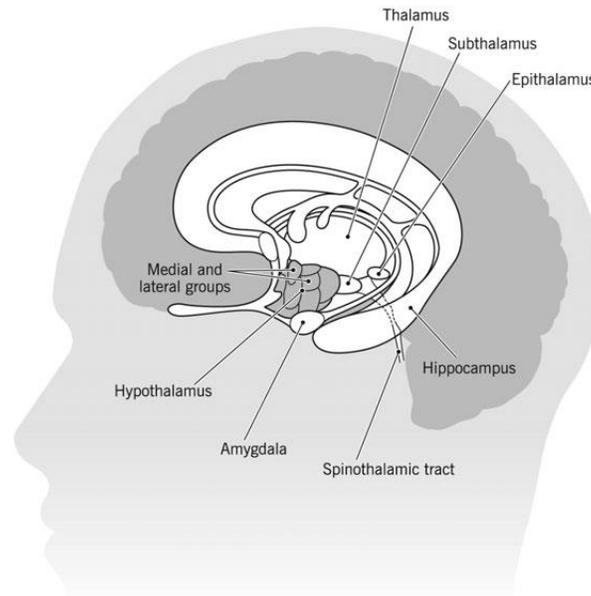
FIGURE 4 LIMBIC SYSTEM [1].

1.3 BRAINSTEM

Is the structure that connects the spinal cord to the brain, composed of three main divisions of the brain stem. From top to bottom these are the medulla, the pons, and the midbrain. The superior most section of the spinal cord is the Medulla, which connects the spinal cord to the pons. The medulla is the place where many of the motor fibers that transmit impulses of motor movement cross over or decussate to the other side of the body and control the fundamental mechanism of life: heartbeat, blood pressure, and breathing. Above the medulla is the pons, a broad, horseshoe-shaped mass of nerve fibers, provides an attachment between the cerebellum and the rest of the central nervous system. Which is also the connection between the medulla below to the midbrain above. The pons is correlating with sensation and movement of the face. Above the pons is the midbrain, which contains a mass of cranial nerves that stimulate the muscles that move the eye and control the shape of the lens and the diameter of the pupil [1].
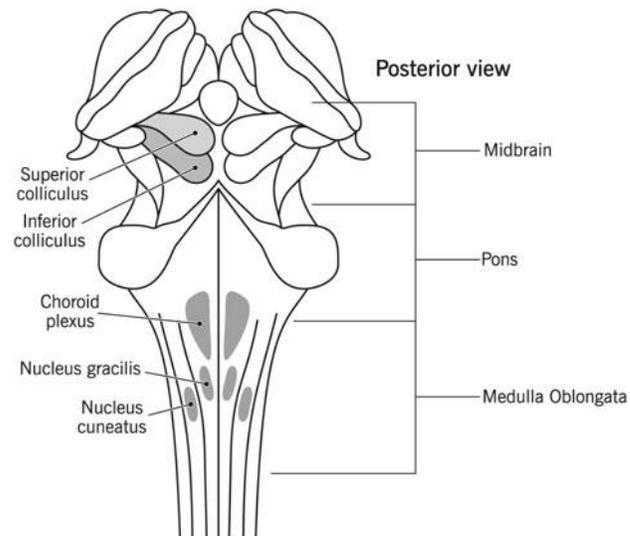
FIGURE 5 BRAINSTEM [1].

1.4 CEREBELLUM

Located on the lower dorsal aspect of the brain, behind the pons and medulla, and fills the greater part of the base of the skull. The cerebellum is composed of two major hemispheres, the outer cortex made up of gray matter with an inner region of white matter containing four paired nuclei: the dentate, globose, emboliform, and fastigial. The cerebellum provides the appropriate patterns and accurate timing of skeletal muscle contraction for smooth and coordinated movements needed for daily lives. Cerebellar activity occurs subconsciously. It exerts profound influences upon equilibrium, muscle tone, and the coordination of voluntary motor function [1].

CHAPTER 2

ALZHEIMER'S DISEASE

2.1 DEMENTIA

The expression dementia, or what is known as a major neurocognitive disorder, is not one specific disease, in fact, is a group of symptoms that happen because of a disease. It affects memory, behavior, thinking, calculation, learning capacity, language, and social abilities. AD is the most common cause of dementia in people more than 65 years old, it is not the only one. Most people more than 80 years have multiple reasons to account for their dementia, such as small strokes or Parkinson's disease [2]. According to the most recent data of the World health organization, there are around 55 million people who have dementia worldwide. This number is expected to rise to 78 million in 2030 and 139 million in 2050 [3]. There are many types of dementia such as AD, Vascular dementia, Dementia with Lewy bodies, and Frontotemporal dementia. For each type and each person affected, the symptoms can present in different ways [2]. Different types of dementia are correlated with specific types of brain cell damage in particular regions of the brain. For example, in AD, the hippocampus is the center of learning and memory, and the brain cells in this region are often the first to be damaged [4]. According to the report by Lancet commissions 2020, the risk factors in early life such as education, midlife like hypertension, obesity, and alcohol misuse, and in later life for example smoking, depression, physical inactivity, and diabetes can contribute to increasing the risk of dementia [5]. In Alzheimer's Association International Conference (AAIC) 2019, five research studies suggest that adopting multiple healthy lifestyle choices reduces the risk of AD by 60% compared with nonadopting. These healthy lifestyles including a healthy diet, not smoking, not drinking alcohol, regular exercise, and cognitive stimulation, may decrease the risk of cognitive decline and dementia [6].

2.2 ALZHEIMER'S DISEASE

AD is an irreversible progressive neurological disorder characterized by progressive memory loss and a decline in activities of daily life, the retardation of thinking, and changes in personality and behaviors[7]. AD is the most common form of dementia that is affecting elder people. AD causes nerve cells to die frequently, which leads to the loss of tissue in the brain that reduces the brain volume dramatically [8]. The characteristics of AD are intracellular neurofibrillary tangles and extracellular β-amyloid plaques in addition the extensive synaptic loss and neuronal death within the brain[9]. The proportion of deaths in the United States related to AD is going up, rising 89% between 2000 and 2014. Around $500 billion annually are the direct and indirect costs for the health care related to AD [10]. Alois Alzheimer presented his first signature case and the pathological features of the disease at the 37th convention in Germany in 1906 [11]. AD typically progress in three stages: early, middle, and late. Due to Alzheimer's affecting patients in different ways, various symptoms appear in each stage, and the following symptoms should only be used as a general guide. Early-stage Alzheimer's: the patients in this stage could do their activity independently, still driving, working. Despite the feeling of having memory lapses, such as forgetting familiar words. Middle-stage Alzheimer's: during the middle stage of Alzheimer's, the dementia symptoms are more pronounced. the patients may confuse words, get angry, and act in unexpected ways. Damage to the brain's nerve cells can also make it difficult to express thoughts and perform routine tasks without help. Late-stage Alzheimer's: in this stage the dementia symptoms are severe. Individuals lose the ability to respond to their environment, to carry on a conversation, and, eventually, to control movement [12].

2.3 MILD COGNITIVE IMPAIRMENT

MCI causes serious cognitive changes that could be noticeable by the patient, in spite of that, the patient has the ability to perform everyday activities. People 60 years of age or older live with an MCI of 12-18%. Experts classify MCI depending on the thinking skills affected:
- Amnestic MCI: MCI that primarily affects memory. which leads to starting to forget important information such as appointments, or conversations.

- Non-amnestic MCI: MCI affects thinking skills other than memory, ability to make the right decisions. In some individuals, MCI returns to a normal state or remains stable. on the other hand, the MCI could develop for various reasons, resulting in the MCI individuals going on to develop dementia. MCI can be an early stage of AD if the hallmark changes in the brain are present [13]. A meta-analysis of 41 studies showed that the conversion rate from MCI to dementia when the individuals were tracked for 5 years or more, averaged at 38% [14].

2.4 RISK FACTOR

Researchers suppose that there is no one reason for AD. It probably develops depending on several factors, like lifestyle, genetics, and environment. some risk factors cannot be changed like age, family history, and heredity. while there is other factors, we can effect on it.

2.4.1 AGE

Increasing age is very clearly associated and the most important factor for AD and other dementias, but these disorders are not a normal part of aging. The probability to develop these diseases increases with age (65 or older). After this age, the risk of AD doubles every 5 years [15].

2.4.2 FAMILY HISTORY

Another strong risk factor is family history. Those who have relatives with AD are more likely to develop the disease. The risk of developing AD will increase if one or more family member has the illness. Maybe genetics and environmental factors play a role when AD is speared in families [15].

2.4.3 GENETICS (HEREDITY)

Two categories of genes influence whether a person develops a disease: risk genes and deterministic genes, which are the direct cause of diseases, it is estimated that less than 1% of AD cases are caused by deterministic genes. AD genes have been found in both categories. The researchers defined head injury, Heart-head connection, and overall healthy aging as risk factors that we may be able to influence. While age, family history, and heredity are all risk factors we can't change [15]

2.5 BIOMARKERS

A biomarker is a substance, measurement, or indicator of a biological state that can be measured to indicate the presence of disease accurately and reliably. Even before the clinical symptoms arise, the biomarker could exist. As AD is a progressive disease over time, biomarker magnitudes reach abnormal levels in a predictable order (see Figure 6). Alzheimer's Disease Neuroimaging Initiative (ADNI) uses several biomarkers to help predict the beginning of AD. This figure shows biomarkers as indicators of dementia, the curves indicate changes caused by five studied biomarkers from normal cognition to dementia:
1) β-Amyloid imaging detected in CSF and Positron emission tomography (PET) amyloid imaging.

2) Neurodegeneration detected by rising of CSF tau species and synaptic dysfunction, measured via fluorodeoxyglucose-   Positron   emission   tomography   (FDG-PET).
3) Brain atrophy and neuron loss measured with MRI (especially in the hippocampus, caudate nucleus, and medial temporal lobe).

4) Memory loss is measured by cognitive assessment.

5) Clinical function, indicated by general cognitive decline measured by cognitive assessment. The first three biomarkers can be observed prior to diagnosis of dementia, while the last two are "the classic indicators of dementia diagnosis" [16].
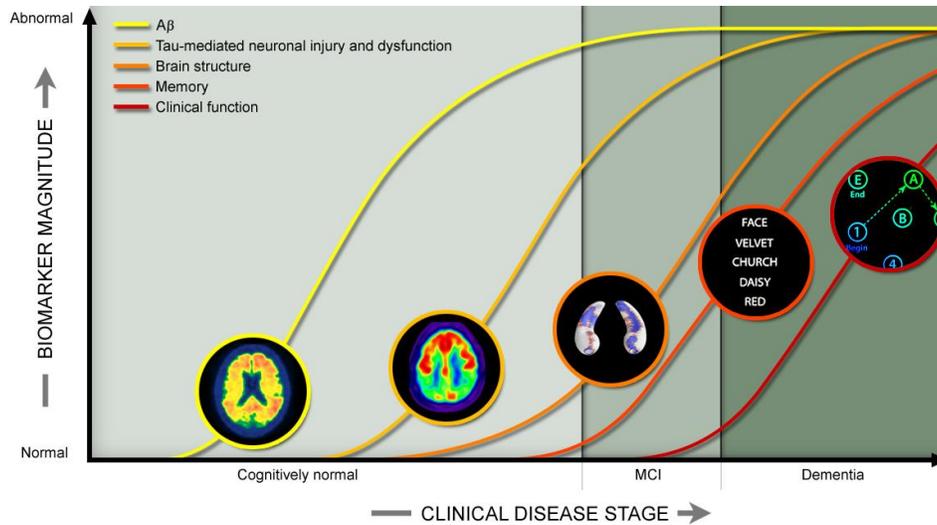
FIGURE 6 ALZHEIMER'S BIOMARKERS OVER THE COURSE OF THE DISEASE [16].

2.6 SYMPTOMS

Difficulty remembering newly learned information is the most famous early symptom of AD. The advancing situations of AD, causing progressively severe symptoms appearance such as disorientation, mood, and behavior changes; making more confusion about time, events, and place; more serious memory loss and behavior changes; and difficulty speaking, swallowing, and walking. Symptoms may in some cases be more obvious to family or friends than to the patient. The patient may have trouble following or joining a conversation. The destructive pairing of plaques and tangles starts in the hippocampus. Usually, the first symptom of AD is short-term memory loss. Then the other parts of the brain are invaded progressively by the proteins, creating unique changes, indicate to various stages of the disease. Finally, control centers which responsible for governing heart rate and breathing are overpowered, resulting in death [17].

2.7 DIAGNOSIS

In fact, that there are no accurate diagnostic methods for AD except the autopsy after death. However, the early diagnosis is very important because at this stage the pharmacological therapy can be successful to treat reversible diseases and delay the patient to enter the hospital as long as possible. The patients with MCI have a high risk of development to dementia 12 % per year, while the normal patients develop

dementia at a rate of 1-2 % annually. According to these numbers, most patients with MCI would develop dementia 3 to 4 years after the diagnosis [18]. A study with 404 healthy people and patients with MCI, observed the evolution over 9.5 years. The result was 100 % of patients with MCI developed AD. The current view is that these patients may have the best performances if they are given early treatment. MCI patients have neurofibrillary degeneration in the hippocampus. Atrophy of the hippocampus has been observed by neuroimaging [19]. The primary step for dementia diagnosis depends on the doctor's studies based on the patient's clinical history and reports from the individual, family members, and friends. Neuropsychological tests help with diagnoses, such as Mini-Mental State, Alzheimer's Disease Assessment Scale (ADAS), and Dementia Questionnaire. Brain computed axial tomography (CAT) gives information about the brain cortical and subcortical atrophy, ventricular enlargement, vascular injuries, tumors, and hydrocephalus. Functional studies such as the Single Photon Emission Computed Tomography (SPECT), and the PET allow us to detect alterations in the neuronal metabolism. In AD, the hypoflow or the hypometabolism appears in the temporal and parietal cortex, making it global with the evolution of the disease [19].

2.8 PATHOGENESIS AND TREATMENT

AD is characterized by a specific brain regions atrophy with a significant loss of neurons. However, the exact process behind this disease is not completely understood yet. At the microscopic level, the principal hallmarks are the existence of β-amyloid plaques (Aβ) and neurofibrillary tangles, along with neuronal degeneration [20]. β-amyloid plaques are generated by a split of a larger protein called Amyloid-Precursor Protein (APP) which is involved in cell membrane function. There is a large amount of Aβ presence throughout the body, the most concentrated in the brain. In normal conditions, the Aβ present in the brain is degraded and cleared. The main hypothesis for AD is the failure of this technique leading to the accumulation of toxic concentrations of Aβ which aggregates into plaques thus causing neuronal degeneration and leading to dementia - the amyloid cascade hypothesis (Fig 7) [21].
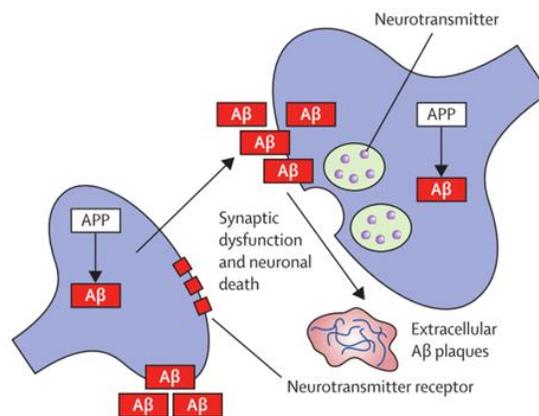
FIGURE 7 THE AMYLOID CASCADE HYPOTHESIS, Aβ IS GENERATED BY APP, WHICH IS ACCUMULATED INTO PLAQUES INSIDE AND OUTSIDE OF THE CELLULAR, LEADING TO SYNAPTIC DYSFUNCTION AND NEURONAL DEATH [21].

The APOE genotype was found to be the gene responsible for the clearance of Aβ [22], hence being associated with a major risk of AD development. Tangles are composed of Tau proteins have an important role in the nerve cell structure stabilizing the microtubules. Tau proteins are the fundamental proteins for

the normal function of the cells because they allow the passage of nutrients. The formation of twisted strands of tau proteins due to the abnormally hyperphosphorylated of the microtubule-associated tau protein in neurons, which are accumulated into neurofibrillary tangles. Tangles cause disassembly of the microtubules, disrupting the function and structure of the neuron, so no guarantees of nutrient supply. leading to abnormal neuronal and synaptic functions. The main hypothesis to confirm AD in the autopsy is the presence of these neurofibrillary tangles [23], [24]. The tau protein malfunction starts in the earlier stage of the AD affecting the transentorhinal region, spreading in more advanced stages to the hippocampus and amygdala, and in the high-order association of the frontal, temporal, and parietal regions [25].
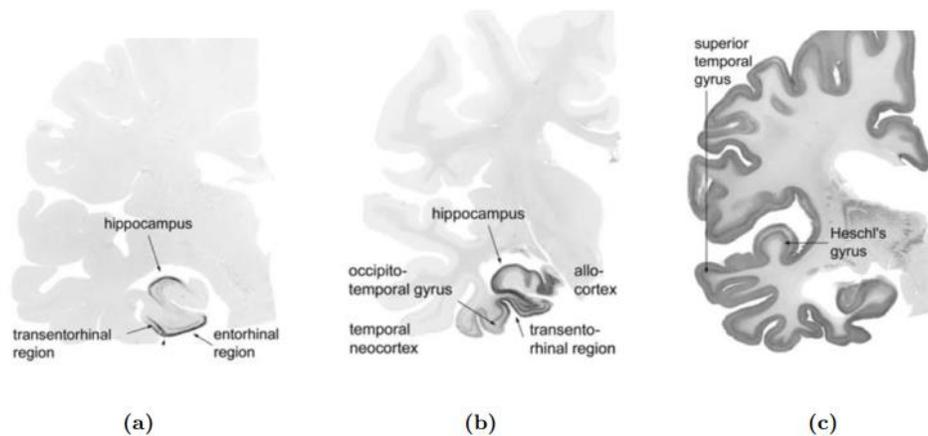
FIGURE 8 MOST AFFECTED BRAIN REGIONS IN AD ALONG DIFFERENT STAGES. (A) EARLIER STAGE: IN THE EARLY PROCESS OF AD, THE ALTERATIONS APPEAR TRANSENTORHINAL REGION. (B) MIDDLE STAGES: THE LESIONS EXTEND INTO THE FUSIFORM AND LINGUAL GYRI AND BECOME MORE SEVERE. (C) STAGE V-VI: THE PATHOLOGY SPREADS TO THE FRONTAL, TEMPORAL, AND OCCIPITAL REGIONS [25].

There is currently no specialized cure for Alzheimer's. The current goal of treatments is to make the progression of the disease slow and manage its symptoms. More concern is given to pharmacological treatments that aim to improve cognition and functional deficits. This treatments approach consists only of a small group of approved drugs, namely the cholinesterase inhibitors Rivastigmine, Galantamine, Donepezil, and the NMDA receptor antagonist Memantine. The treatment of the symptoms of AD consists of a wide range of unspecific pharmacological and non-pharmacological interventions. However, it can only soften the burden of symptoms for AD and caregivers, facing behavioral and psychological symptoms of this disease, and help to maintain the daily living activities as long as possible. Most drugs aim to prevent Aβ aggregation, reduce Aβ production, or promote Aβ clearance. In addition to strategies against Aβ, much effort is produced in drugs that inhibit tau-phosphorylation or prevent tau aggregation [26]–[28].

CHAPTER 3

MAGNETIC RESONANCE IMAGING MODALITY

The first description for the physics of MRI was in 1946, and Nuclear Magnetic Resonance (NMR) images have been used since 1973. The NMR phenomenon was primarily described independently at Stanford University by Felix Bloch and at Harvard by Edward Purcell, in 1946, they received the Nobel Prize in Physics in 1952. There are several advantages of using MRI as a non-invasive diagnostic modality and imaging technique compared to the other commonly used alternative diagnostic techniques, conventional X-ray, and Computed Tomography (CT) scans. There is no use of ionizing radiation which may have biological hazards, sometimes MRI patients are injected with a contrast agent to make abnormalities appear clearer. MRI generates 3D images with high spatial resolution. This imaging technique works by basically mapping the position of water molecules, which spread in different densities in different types of tissue, therefore exhibiting excellent soft-tissue contrast [29].

3.1 HARDWARE OF MAGNETIC RESONANCE IMAGING SYSTEM

MRI technique development requires an associated performance improvement in the system hardware combine of three components, which mainly includes: the main magnet which is constant in time and is used to align the spins within the patient, three gradient coils that produce a gradient field, radiofrequency (RF) coil which is pulsed on and off to produce repetitive excitation of the spins within the patient and the patient table. Figure 9 shows the schematic representation of a typical clinical scanner system [33].
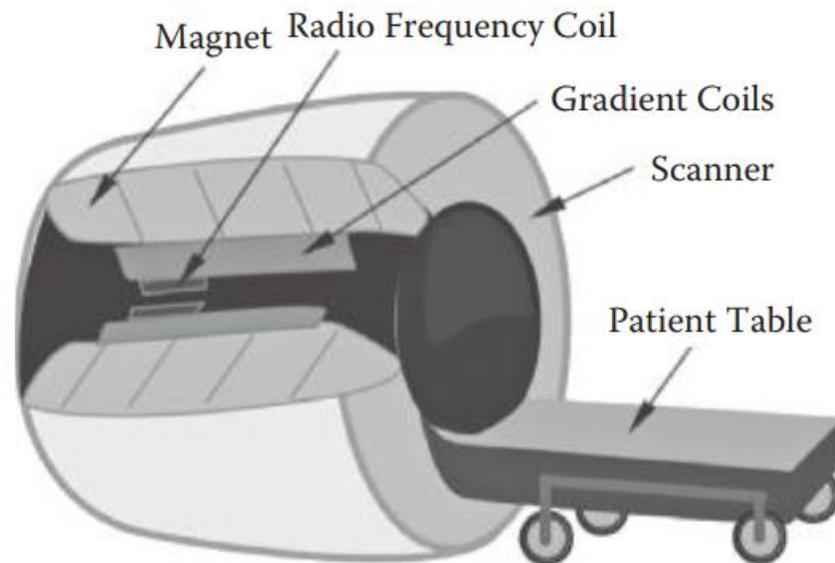
FIGURE 9 SCHEMATIC REPRESENTATION OF A TYPICAL CLINICAL SCANNER SYSTEM [33].

- The main magnetic field needs to be uniform regarding its strength and direction. It defines a z-direction in space, the z component of the main field at the center of the magnet is designated Bo, which usually runs along the cylindrical axis of the main magnet (in the direction from the patient's head to the feet), which is constant in time and space. Superconducting magnets are operated at a constant field and near the maximum limit of their capabilities. Depending on the superconducting properties of the magnet coils, the power supply can be disconnected when the desired field strength is reached, and the magnet will operate in the persistent mode and maintain the full field strength for months or years. Most superconducting whole-body magnets run between (1.5T_3T), recently, whole-body systems operating at 7 T have been placed in several research centers. A cross-sectional drawing of a generic cylindrical design for a whole-body superconducting magnet is shown in figure 10. This is a multi-coil design suitable for both 1.5 and 3 T. There are 6 main field coils, shield coils, and additional superconducting shim, the temperature is maintained close to absolute zero. A bath of liquid helium provides this low temperature and an associated cryocooler/recondenser that recycles helium as it boils and reduces the process of refilling the helium bath. The cryostat is a structure that contains cryogenically cooled coils. The patient table moves into the center of the magnet by computer control. The field strength and homogeneity in the region, labeled field of view (FOV), near the center of the cylindrical bore of the machine is satisfactory for MRI [30].
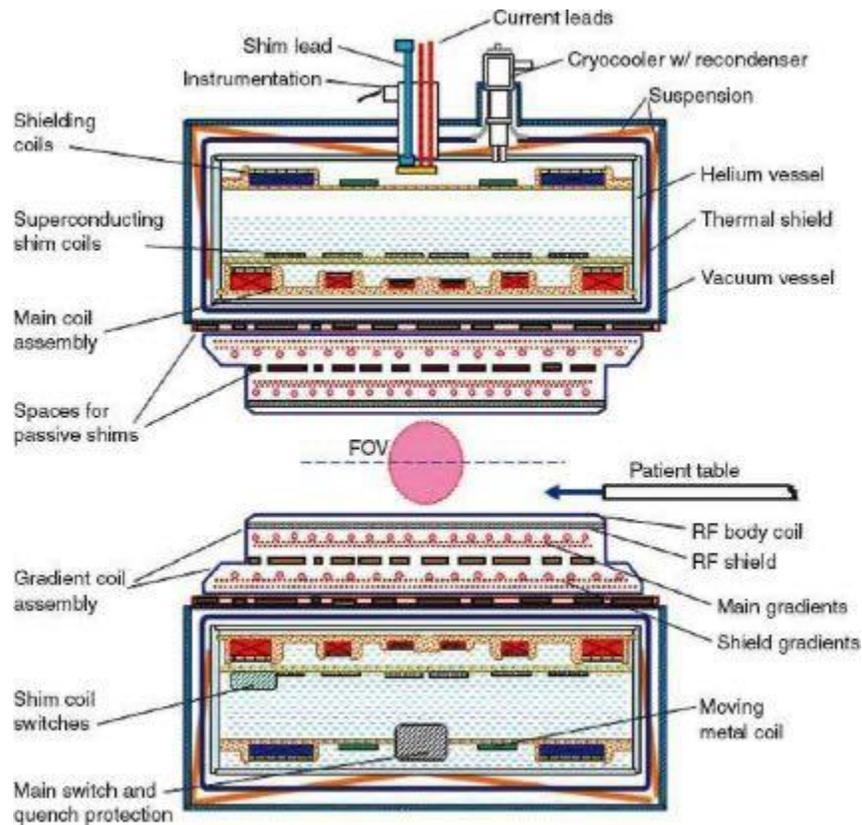
FIGURE 10 CROSS-SECTIONAL DRAWING OF A GENERIC CYLINDRICAL DESIGN FOR A WHOLE-BODY SUPERCONDUCTING MAGNET [30].

- Three separate gradient coils are used to produce controlled variations in Bo, to permit slice selection and to encode position-dependent information in the MR signal. Activating two or all the gradient coils generates a gradient field in any arbitrary direction in space. That facilitates the selection of imaging planes in any one of the principal orientations [30].

- Transmitter coil/coils carry currents at radio frequencies equal to or near the Larmor frequency, in the direction perpendicular to z that is completely uniform in space, which is used to excite the spins, which lead to rotating in the main magnetic field and make the signal that used to generate the image. This signal is detected by a receiver coil, it could be the same transmitter coil, or it may be a separate structure. Data collection can be from one or more receiver coils, maybe built-up in the analog domain, but it is more desired in the digital domain because it has a correction and spatial match-up. Data processing in the digital domain allows for some correction of the inhomogeneous sensitivity of the receiver coils and improves the ability to provide an accepted image over the desired FOV. Using more number coils (smaller and closer to the area of interest) rather one large coil is better for receiving less unwanted signal and the total collected signal energy is also reduced [30]. There are several kinds of RF coils. The difference between coils depends on different parts of the human body and different field strengths. According to the imaging part of the human body, it can be classified for example into head coils, body coils, knee coils, or foot coils. all the coils can be basically divided into two kinds of coils, surface coil, and volume coil. Regarding the surface coils, usually, it takes a circle shape, which facilitates the manufacturing of the coil. It is usually used as a receiver because it produces an inhomogeneous field, which is unuseful to the imaging process. However, the surface coils have a signal-to-noise ratio (SNR) which is higher than the SNR of the volume coils, partly because they can be located close to the imaging area. For good performance in receiving and transmitting surface coils, a group of surfaces could be used for that. An illustration of the surface coil is shown in Figure 11 [31]. The second type of RF coil is the volume coil, and the most popular volume coil is the birdcage coil, which is shown in Figure 12 [30].
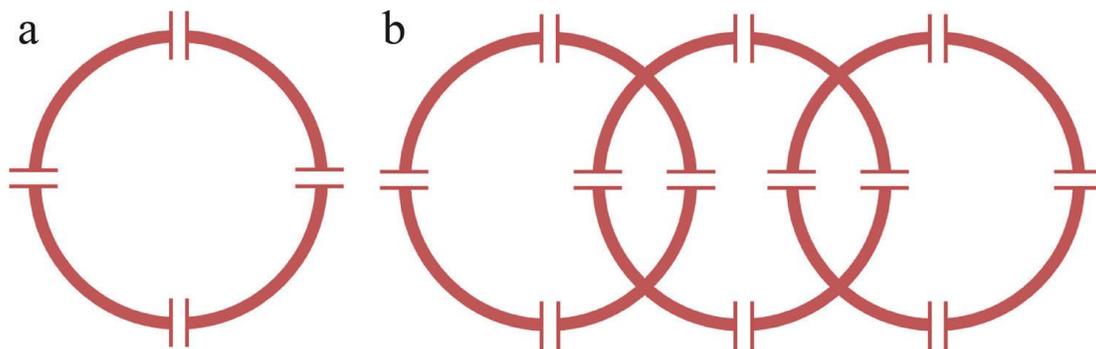
FIGURE 11 DIAGRAM OF SURFACE COILS: (A) SINGLE COIL AND (B) ARRAY COIL [31].
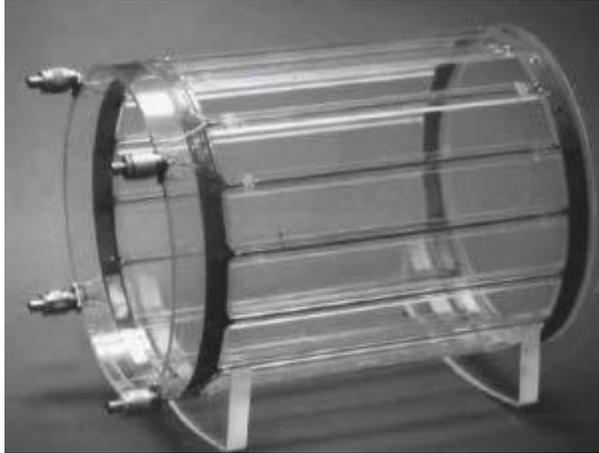
FIGURE 12 BIRDCAGE RESONATOR [30].

- MR scanners contain in addition to these main components, a set of several additional static field coils called shim coils, responsible for correcting the inhomogeneities that may be present in the Bo field that results from manufacturing tolerances or from other noise sources lead to the presence of error fields which reduce the uniformity of the field within the FOV [30].

3.2 PHYSICS OF MAGNETIC RESONANCE IMAGING

All atomic nuclei consist of protons and neutrons, with a net positive charge. Specific atomic nuclei, like the Hydrogen nucleus or the Phosphorus nucleus, have a property known as spin, dependent on the number of protons [32]. MRI uses the magnetic properties of the human tissue and the ability of contrast agents to alter them, MRI depends on the detection of the magnetic moments created by single protons in the hydrogen atoms. Movement of electrical charge produces a magnetic field, a spinning proton produces a small magnetic field and can be considered as a little magnet, when a patient is placed in the cavity of a large magnet, MR scanner, hydrogen protons align with the externally applied, static magnetic field (B_0) to create a net magnetization vector. The protons' axes realign with B_0, around half of them face in the direction of the field, around the other half in the opposite direction. Spinning protons wobble about the axis of B0, the angular frequency around the field is the Larmor frequency, which is proportional to the file strength, it is given by the equation (1), which is known as Larmor equation:

$$\omega_0 = \gamma B_0 \qquad (1)$$

ω_0: angular frequency of the protons, γ is the gyromagnetic ratio, a constant fixed for a specific nucleus (For hydrogen γ=42.6MHz/Tesla), and B_0 is the field strength. With the application of a radio frequency (RF) pulse at the resonance frequency of the wobble, the protons will absorb the energy and jump to a higher energy state. This RF pulse deflects the protons away from the main axis of the B0. Thereby manipulating the net magnetization vector in a direction away from the direction of the externally applied magnetic field, from longitudinal to the transverse plane (transverse magnetization). In the case of applying stronger RF, the angle of deflection for the magnetization will be greater [29].

The protons will be flipped by the RF pulse and the net magnetization vector is deflected by a flip angle. After the RF pulse takes the spins out of alignment with the main magnetic field, protons start to realign with the main magnetic field at a rate determined by the T1 relaxation time. Energy is released as the spins move from the high energy states to lower states. The absorbed RF energy is retransmitted at the resonance frequency, which is collected by RF coils placed around the patient. These signals form the MR images. The process of excitation of protons with an externally applied RF field is repeated at short intervals to form images. This MR parameter is referred to as the time of repetition or TR. For conventional anatomic MRI, TR is typically 0.5 to 2 seconds. All spins are processing in a phase when they are tipped to the transverse plane. The speed of wobbling depends on the strength of the magnetic field that the proton experiences. Some protons spin faster while the others spin slower, and they quickly get out of phase relative to each other, during this process the MR signal fades away and decays. This loss of phase is termed T2 transverse relaxation or relaxation time. T1 and T2 are unique for every type of tissue and provide an additional basis for contrast on MR images. inhomogeneity of B_0 results in quick loss of transverse magnetization. T2* is the relaxation time that reflects the sum of these random defects with tissue T2. To obtain an MRI signal, must bring back these spins in phase. The time of occurring is referred to as echo time (TE) [29].

3.3 SIGNAL WEIGHTING

Transverse and longitudinal magnetizations happen at the same time, but they are two different processes that reflect the properties of different tissues in the body. T1 is a measure of signal recovery, tissues with long T1 are dark, whilst tissues with short T1 are bright. Fat has a very short T1. T1-weighted images are especially well suited for brain scans and show important differences between grey matter and white matter. T2 is a measure of the loss of signal, tissues with short T2 are dark whilst tissues with long T2 are bright. CSF has the longest T2. To distinguish between the tissues based on these relaxation times, MR images can be designed to be T1-weighted, T2-weighted, or proton density-weighted. Exogenous contrast agents can be used to improve the sensitivity if there is a small intrinsic difference in contrast between structures [29].

3.4 IMAGING

To compose an effective MR image, localization of the signals obtained from tissues in space is required. External, time-varying extra magnetic fields are applied to spatially encode the MR signal and form cross-sectional images. Magnetic fields gradients are required to completely define the location of the MR signal in space. Which are on the x, y, and z axis. The magnitude of the magnetic field changes linearly in the direction of the gradient. Each gradient coil is controlled independently and can be pulsed with current to generate a small local field gradient, and selectively excite particular anatomical slices of interest that impose a corresponding spatially and linearly varying angular frequency change on the spins [29]. The term "resonance" in MRI is related to the fact that the hydrogen nuclei in water molecules have a strong response to electromagnetic energy when the frequency of that energy matches the magnetic field strength. which is called the Larmor rotation frequency. If the field strength varies in a controlled way, then the frequency content of the signal will vary too, and an image of the spin distribution can be created. This observation led to the award of the Nobel Prize in Physiology or Medicine to Peter Mansfield and Paul Lauterbur. This controlled variation is created by the gradient coils by adding a small shift to the main

magnetic field which is proportional to the position. This shift is generally used in three ways in a pulse sequence selective excitation, phase encoding, and frequency encoding. Each gradient coil produces a field whose Z component along the axis of the magnet varies linearly with either X (right-left direction), Y (anterior-posterior), or Z (superior-inferior). Combinations allow linear encoding in any spatial direction which is a unique feature of MRI among other imaging modalities. During the RF pulse, if the gradient field is applied, the spins will only absorb the energy from this pulse if the frequency of the pulse content is the Larmor frequency of the spins in question. Usually, the amplitude of the RF is modulated, a sine wave at the Larmor frequency of the center of the slice of interest is multiplied by a narrow-band pulse shape or kernel. This modulation spreads the spectrum of the sine wave over the bandwidth of the shape, and at the same time as the pulse if a gradient field is applied, only spins in a certain slice through the object will absorb energy from the pulse. The slice is selectively excited. After this pulse of excitation, the spins will rotate at the Larmor frequency, producing a free induction decay (FID). By applying another gradient during the FID, the Larmor frequencies will depend on the position of the rotating spins. A receiver coil collects the signal from the rotating spins, the density of spins along the axis of the gradient will be encoded by the frequency content of the signal. The density can be recovered mathematically by a Fourier transform, this process is called frequency encoding. By applying a short pulse over a third spatial direction, the spin density along this axis will be encoded as a phase shift in the rotation. The Larmor frequency is altered by this third gradient for a short period, at different locations of spins, and the difference in rotation frequency along this direction will lead to a phase shift proportional to the duration of the pulse and its amplitude. Repeating the experiment many times with slightly different amplitudes for the phase encoding gradient allows full encoding of the spin density at the desired resolution. The gradient fields must be switched on and off many times within a pulse sequence [33].

3.5 IMAGE ACQUISITION

Three major concepts are essential to generating an image by MRI: Firstly, the raw signal acquired in an MRI experiment corresponds to the 2D or 3D Fourier transform (FT) of the subject being imaged; Secondly, the use of magnetic field gradients define a trajectory through the acquired FT (also called "k-space"); and Finally, the object image is obtained by calculating the inverse Fourier transform (iFT) on the obtained k-space data. In the nomenclature of MRI, the temporal or spatial domain of data points is indicated as "image space," and their corresponding frequency domains as "k-space." The most popular way to acquire 2D k-space data is line-by-line, in a raster pattern. The position in k-space is correlating to the strength of the linear field gradient, and the length of time it is applied. The gradients act to move the position through k-space, with phase encoding used to select the ky position and frequency encoding used to select the kx position. Through the application of the positive frequency-encoding gradient, the signal is sampled at regular timing increments, acquiring each kx point along the selected ky line. This process is then repeated N times, with uniformly changes in amplitude of Gy to acquire N ky lines (Figure 13, A) and fill in the complete k-space. After the desired k-space has been obtained, the 2D iFT is calculated to recover the image (Figure 13, B). The contrast of the MR image generated depends on the intrinsic MR properties of tissue, such as the longitudinal and transverse relaxation times (T1, T2) and proton density (PD), the acquisition strategy (i.e., acquisition parameters and pulse sequence), imaging hardware (e.g., field strength and radiofrequency [RF] coil geometry), and image reconstruction methodology [30].
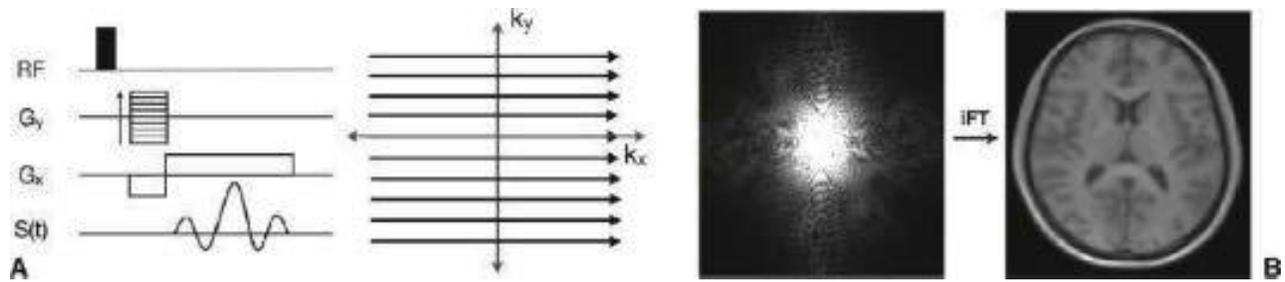
FIGURE 13 A: SHOWS THE REPETITION OF THE SEQUENCE WITH G**Y** UNIFORMLY INCREMENTED TO ACQUIRE KY LINES. **B:** THE CALCULATION OF THE INVERSE **FT** TO RECOVER THE **2D** IMAGE **[30].**

3D images acquisition can be accomplished in one of two ways: multi-slice 2D imaging, or true 3D imaging. Multi-slice imaging involves the acquisition of multiple 2D slices ordered in a stack to cover a 3D volume. Along the z-axis of the object, a linear gradient is applied, Gz, causing dispersion of rotational frequencies in this direction. Like as the duration and strength of Gx and Gy defined the position (kx,ky) in the 2D FT, the strength and duration of Gx, Gy, and Gz define a point (kx,ky,kz) in the 3D FT. The 3D FT is acquired as with the 2D case, a line in K-space is defined by using phase-encoding gradients (Gy and Gz), and the frequency-encoding gradient (Gx) is used to sample this line. This process is then repeated, with Gy and Gz uniformly incremented (Figure 14, A). Finally, to reconstruct the object's image a 3D inverse FT is calculated (Figure 14, B) [30].
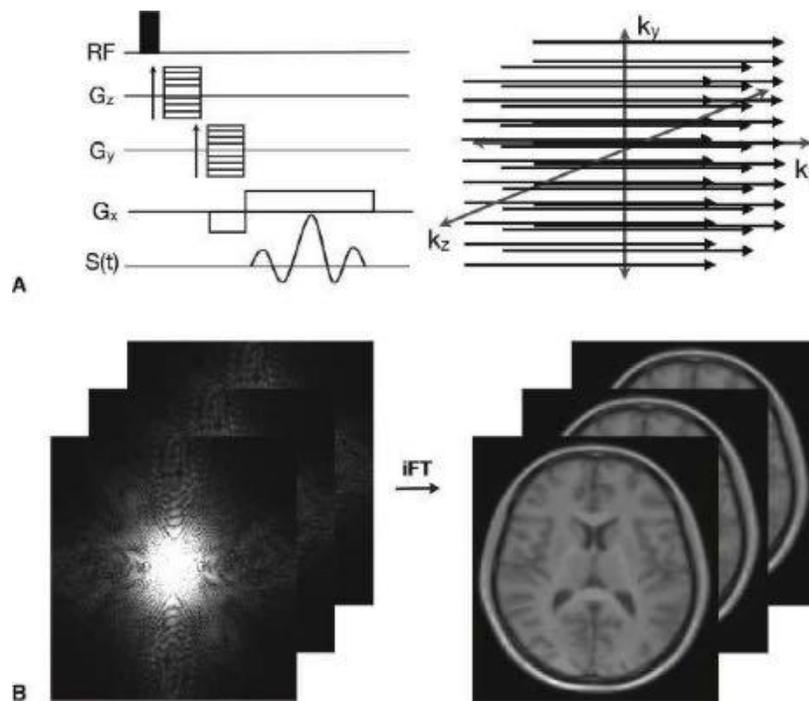
FIGURE 14 A: THE ADDITIONAL PHASE-ENCODING GRADIENT G**Z,** FOR THE THIRD DIMENSION OF K-SPACE. **B:** A **3D** INVERSE **FT** IS THEN PERFORMED TO RECOVER THE **3D** IMAGE VOLUME. **[30].**

CHAPTER 4

ARTIFICIAL INTELLIGENCE

In the 1950s the Artificial Intelligence (AI) was born when humans were asking: Could machines learn patterns as humans do? We are still trying to answer this question today. AI did not get popular until the 1980s when the computation power was capable of transforming equations from papers to ones and zeros efficiently [34].

A new term was coined when the question: "what we know how to order it to perform" was raised in [35], which is Machine Learning (ML) At that time, traditional programming was to build roles, based on human knowledge, and feed the inputs to it in order to obtain answers. The game was changed when the ML tries to construct the rules by looking at the inputs and the answers. The two processes are illustrated in figure 15.[34]
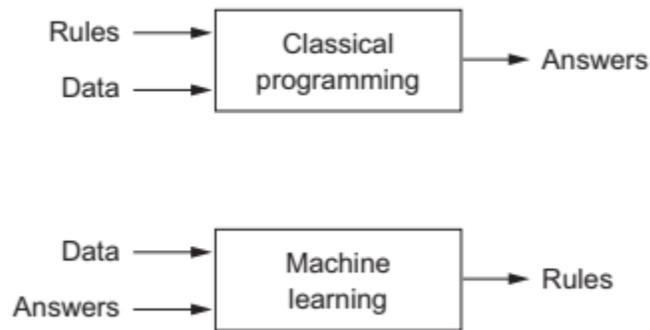
FIGURE 15 COMPARISON BETWEEN CLASSICAL PROGRAMMING AND MACHINE LEARNING PARADIGMS [34].

4.1 NEURAL NETWORKS

The neural network is a model of assembling artificial neurons model in a hierarchical way. A neuron is a mathematical model of biological neurons. They share the same structure and the principle of multiple input, summing, activation, and output [36]. figure 16 illustrates both neurons.
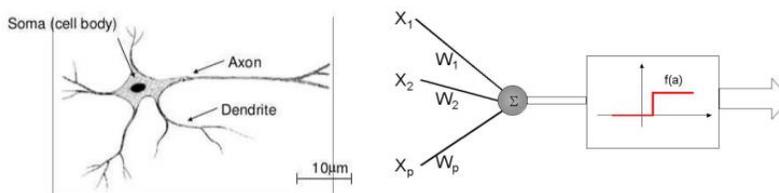
FIGURE 16 COMPARISON BETWEEN BIOLOGICAL AND ARTIFICIAL NEURON. [36]

The inputs x_p are summed after being multiplied by their associated weights W_p. The output of the neuron is subject to the activation function f [36]. The neuron model is given in the following equation:

$$output = f(\textstyle\sum^p (x_p.W_p) + b) \qquad (2)$$

Where b is the bias of the neuron.

Training of the neuron means assigning values the weights and the bias. The most common architecture of neural networks is the MLP. The MLP consists of an input layer of neurons, connected to several hidden layers and an output layer [34].

4.2 DEEP LEARNING

By increasing the number of successive layers, the networks come deeper, and that's what the concept of DL stands for. The deep neural networks, usually, involve tens or hundreds of layers. The depth, number of layers, increases the complexity of the model, which in turn is useful for complex problems. This suitability for solving complex problems makes the pipeline even easier by automating the feature extraction, in many DL architectures. The autonomy of feature extraction, aka feature engineering. simplified machine-learning workflows [34], [37].

One of the most known DL architectures is the CNN. CNN can be thought of as a condensed version of the Neocognitron model, which was created in 1980 to replicate the human visual system [38]. CNNs first developed in the early 1990s [39]. However, due to restricted computational capabilities, they did not gain widespread appeal at the time. However, with the introduction of fast graphics processing unit (GPU) computers and availability of labeled training data, CNNs, as illustrated in fig (17), have re-emerged as potent feature extraction and classification method, achieving record-breaking performance in a variety of significant computer vision issues. The success of CNNs in computer vision has motivated a large number of investigators in the medical imaging field, resulting in a flurry of publications in a short amount of time demonstrating the usefulness of CNNs for a range of medical imaging tasks [40].

CNNs are so-called because their architectures incorporate convolutional layers. Convolutional layers are in charge of recognizing specific local features across all of their input images' locations. Each node in a convolutional layer is connected to only a limited subset of spatially related neurons in the input picture channels in order to detect local structures [40] .

To facilitate the search for the same local feature across all input channels, the convolutional layers' nodes share their connection weights. A kernel or convolution kernel is a collection of shared weights. Thus, a convolutional layer with n kernels learns to detect n local features whose intensity is observable throughout the input pictures in the output n feature maps. The convolutional is demonstrated in fig (18).

Each sequence of convolution layers is followed by a pooling layer to reduce computational complexity and produce a hierarchical set of image features. The max-pooling layer compresses feature maps by picking the maximum feature response in overlapping or non-overlapping local areas and eliminating the precise location of such maximum responses. As a result, max pooling can enhance translation invariance even further. CNNs are typically composed of many pairs of convolutional and pooling layers, followed by a series of sequential convolutional layers (a.k.a., fully connected layers (FCL)), and eventually a SoftMax layer or regression layer to obtain the desired outputs. In more contemporary CNN systems, the pooling layer is replaced by a convolution layer with a stride greater than one in order to increase computational efficiency [41]. A CNN often contains a significant number of convolutional and FCLs; hence, it is not uncommon for the architecture of a CNN to contain millions or billions of weights. The pooling operation is presented in fig (19).

CNNs are mostly used for automated feature extraction from images. Those networks are known for their special two characteristics:

- The patterns they learn are translation invariant. This means once a pattern is learned in a location in the image, it could be detected wherever it appears in any future image. That can explain the need for fewer training examples to generate a higher generalization capability.
- Spatial hierarchies of patterns could be learned. This is to say that each convolutional layer is able to learn a different level of detail. For example, the first layer is able to learn simple patterns like edges and lines while deeper layers can learn larger patterns of the features.
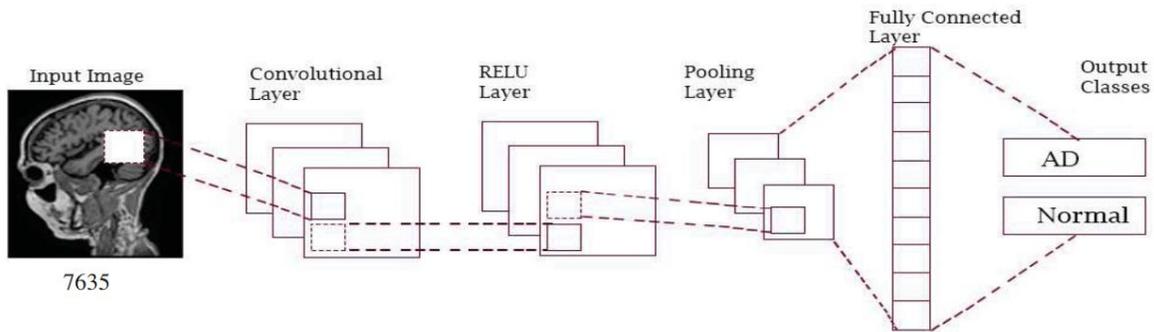
FIGURE 17 GENERAL STRUCTURE OF CNN [42]
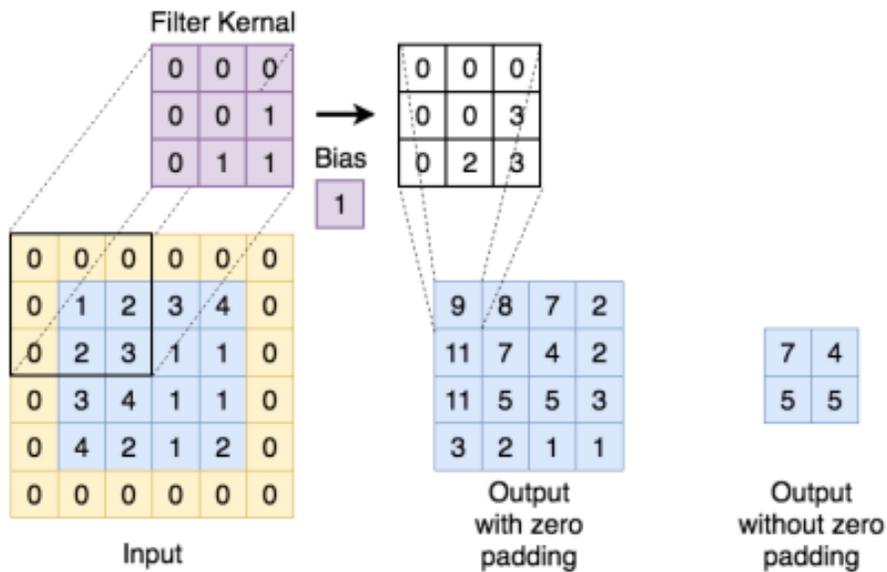
FIGURE 18 CONVOLUTIONAL LAYER [43].
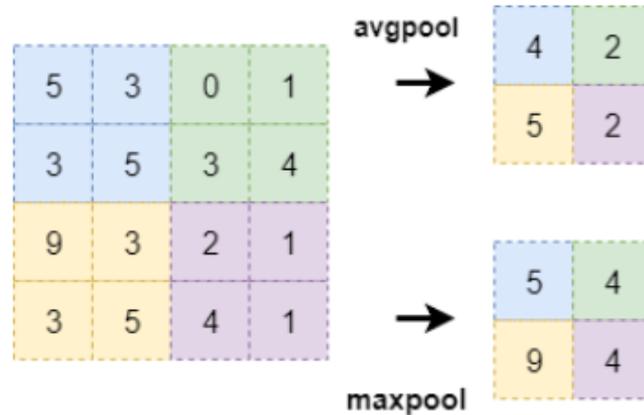
FIGURE 19 THE POOLING OPERATION [43].

4.3 INTEGRATING CONVOLUTIONAL NEURAL NETWORK WITH METADATA

When it comes to neuroimaging data, some specific requirements and constraints must be met to ensure usability and interoperability. With no description, digital image data becomes useless. Consequently, the image dataset, or any dataset in general, should include additional information known as metadata in addition to the images.

Metadata is a collection of descriptors for a data set that makes it easier to find, understand, and use. Metadata can include information about the reason for the data set's creation, the experimental setup, and the data formats used to store the data. Typically, it assists the user of the data, rather than the data's creator, in reproducing the experiment or making decisions about any potential bias in the dataset. Furthermore, it assists in determining the generalizability of the findings. Moreover, depending on the nature of the investigation, metadata can be considered data in and of itself because it contains information about the population sample and common characteristics between data points from different sources [44].

There is no doubt that metadata is beneficial for utilizing the data set. Both the operator and the statistical model could benefit from it. While putting the analysis into practice, the operator can gain valuable insights. However, the statistical model, which is a classification model in our case, receives the metadata as an additional input along with the original data and uses it to make predictions. In particular, the researchers have shown classification tasks based on a combination of imaging and genomics data to outperform clinical experts in digital pathology. When relevant information about the sample, such as patient demographics, is combined with imaging data, high accuracy scores in binary classification tasks have been achieved, for example [45].

Data can be expressed in many different ways. A categorical variable such as gender (male or female) could be used, or a multi categorical variable such as blood type could be used. The metadata can also be identified using a continuous variable such as age, weight, or BMI. If necessary, how metadata is provided to the user could be modified. As an example, BMI as a continuous variable can be expressed as a categorical variable by grouping the numbers into relevant categories [45].

There is plenty of metadata that can be used for clinical purposes. It is possible to classify the metadata into medical, demographic, and experimental information. Some examples of medical information include the blood type, APOE genotype, BMI, and the volumes of various organs. On the other hand, the demographic variables include things like age, gender, race, and educational level. In addition, the

experimental data contains information about the tools used to collect the data. In the case of MRI imaging, it could be about the magnetic field strength, the type of MRI, the sampling frequency, the number of slices, or whether or not a substance has been injected into the patient [46].

Integrating metadata with scan features is done to every MRI image (M) input for the standard neural networks. This integration may require a mapping function G (M), which converts the data to the desired format before continuing. This mapping could be accomplished by referring to the gender with a numerical expression, such as males with a 0 and females with a 1. Alternatively, padding the data to the same length and converting it to an expression in an ASCII table, as described in [47]. In order to provide the most significant amount of flexibility, the conversion should be performed element-wise for an input string. As shown in Fig.20, the encoded metadata inputs, G(M), are concatenated with the image data by concatenating the image features obtained by the CNN at its deepest layer prior to classification, F(X), (blue vector in Fig. 20), with the encoded metadata inputs, G(M), (red vector in Fig. 20), to form F(X) and G(M).The characteristic vector of a sample is represented by the following equation (3) [47].

$$H = (f(x) \quad G(x)) \qquad (3)$$

By Generalizing this problem to include N datapoint, we obtain equation (4):

$$H = \begin{pmatrix} F_{1,1} & \cdots & F_{1,d_k} & G_{1,1} & \cdots & G_{1,d_{k'}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ F_{N,1} & \cdots & F_{N,d_k} & G_{N,1} & \cdots & G_{N,d_{k'}} \end{pmatrix} \qquad (4)$$

Where d_k is the number of the features extracted by CNN while $d_{k'}$ is the number of metadata features.
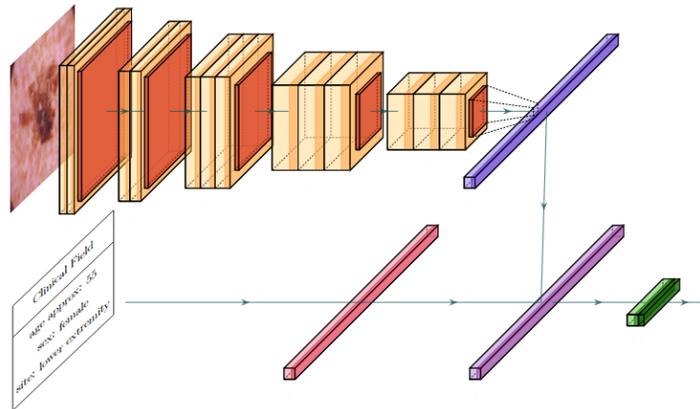
FIGURE 20 INTEGRATING CNN WITH METADATA [47].

4.4 PERFORMANCE METRICS

Performance metrics are used to evaluate the performance of neural networks. Basically, the main tool used for evaluation is called the confusion matrix. For a binary classification problem (2 classes) or a multi-class classification problem (more than 2 classes), a Confusion Matrix is a summary of anticipated outcomes presented in a specific table structure that permits visualization of the performance measure of the machine learning model (more than 2 classes). The confusion matrix is a way to compare the

predicted value of the classifier and the true values [34]. Figure 21 shows the main structure of the confusion matrix [43].

FIGURE 21 THE CONFUSION MATRIX [43].

From the confusion matrix, there are four main variables to be considered:

- True Positive is abbreviated as TP. It means that the model anticipated a positive class, and it is correct.
- False Positive is abbreviated as FP. It is possible to read this as the model predicting a positive class, but this is incorrect.
- False Negative is abbreviated as FN. It is possible to read this as the model predicting a negative class, but this is incorrect.
- True Negative is abbreviated as TN. It is True and can be read as the model's expected negative class.

Performance metrics could be calculated as follows:

Accuracy

Accuracy is defined as the simple ratio between the number of correctly classified points and the total number of points in a set of questions. It is given as in equation (5):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad (5)$$

If the data set is very imbalanced and the model classifies all of the data points as the majority class data points, the accuracy of the model will be very high in this situation. As a result, accuracy is no longer a reliable performance criterion when dealing with imbalanced data. The likelihood of the model's predictions can be calculated based on the accuracy of the model's predictions. As a result, we can't tell how accurate the model's predictions are based on their accuracy [34].

Precision

Precision is defined as the proportion of correctly classified occurrences in relation to the total number of classified instances. Precision assists us in determining the usefulness of the results [34]. It is given as in equation (6):

$$Precision = \frac{TP}{TP+FP} \qquad (6)$$

Recall

The recall rate is the percentage of correctly classified cases among the total number of classified instances. The recall function assists us in determining how complete the results are [34]. It is given as in equation (7):

$$Recall = \frac{TP}{TP+FN} = Sensitivity \qquad (7)$$

F1 Score

The F1 score, on the other hand, is utilized to limit the number of times pockets are checked. Precision and recall are combined to provide the F1 score, which is the harmonic mean of precision and recall [34]. It is given as in equation (8):

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall+\ Precision} \qquad (8)$$

For example, in the field of information retrieval, the F-score is frequently employed as a means of assessing search, document classification, and query classification performance. For example, the F-score has been frequently utilized in the natural language processing literature to assess named entity recognition and word segmentation [34].

Specificity

Specificity refers to the likelihood of receiving a negative result if one does not have the condition in question. If someone does not have the condition yet tests negative for it, this is referred to as the specificity of the testing procedure [34]. It is given as in equation (9):

$$Specificity = \frac{TN}{TN+FP} \quad (9)$$

Balanced Accuracy

When analyzing the performance of a binary classifier, it is possible to utilize a metric called balanced accuracy. Particularly effective in cases where the classes are unequally distributed; for example, when one of the two classes appears significantly more frequently than the other This occurs frequently in a variety of circumstances, including anomaly detection and the presence of a disease [34]. It is given as in equation (10):

$$Balanced\ Accuracy = \frac{Sensitivity+Specificity}{2} \quad (10)$$

CHAPTER 5

LITERATURE REVIEWS

This part will discuss the literature and what was done previously in the field of AD diagnosis. Reviewing literature was done by searching the popular academic search engines like google scholar, PubMed, Scopus, and Elsevier. To do the search, the following keywords were used:

- Alzheimer's disease
- Magnetic resonance imaging / MRI
- Metadata
- demographic data
- Non-imaging data
- Deep learning
- Classification
- Convolutional Neural Network / CNN

Those keywords were combined to make several queries. These queries are conducted using "AND", "OR" quotation marks statements. Examples of those queries are: "Metadata AND MRI AND Alzheimer's disease". Another example is "Deep learning OR Convolutional Neural Network AND MRI AND Metadata OR Non-imaging data". The keywords were utilized with different queries on the previously stated search engines resulting in many papers publications.

The selection criteria start with manual investigation to know whether the paper is relevant to the topic or not. This step was conducted by reading the abstract, conclusion, and the last part of the introduction. This step resulted in 35 papers after all searching queries. The next step is to read further details in the paper and to know whether the metadata was for training and creation of DL model the diagnosis of AD. This step resulted in 7 papers. Figure (22) shows a block diagram of the selection strategy of the literature review.
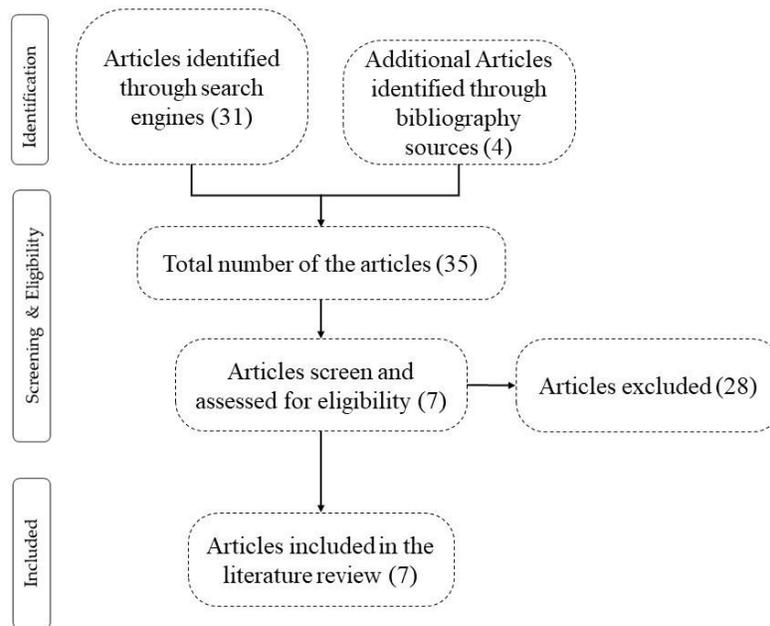

FIGURE 22 A BLOCK DIAGRAM OF THE SELECTION STRATEGY OF THE LITERATURE REVIEW.

QIU, S. ET AL. 2020

Many studies interested in the classification of AD that Integrate patient history, neuropsychological testing, and MRI to identify likely cases have it lacking in sensitivity and specificity. To address these limitations, the authors in [48] developed a novel DL framework that connects a FCN to MLP to classify AD from multimodal inputs of imaging data (MRI) and non-imaging data like age, gender, and MMSE score. Data from four data sets were used: ADNI dataset (NC=229, AD=188), Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing (AIBL) (NC=320, AD=62), the Framingham Heart Study (FHS) (NC=73, AD=29), and the National Alzheimer's Coordinating Center (NACC) (NC=356, AD=209) for training, validation and testing the model.

The model training, internal validation, and testing were performed on ADNI data, and the validation was performed on the other three datasets. Individuals over 55 years old with 1.5 T, T1-weighted MRI scans collected within 6 months from the date of a clinically confirmed diagnosis of AD or normal cognition were included in the study. This study uses a FCN and MLP.

This study followed three steps as shown in figure 23; an FCN model was developed using a patch-based strategy that randomly selected 3000 volumetric patches of size 47x47x47 voxels of T1-weighted MRI to train the model, the output of this model is a binary classification of the AD and NC, Adam optimizer with a 0.0001 learning rate and a mini-batch size of 10 was used. In the second step, After FCN training, a complete array of disease probabilities can be obtained from a single volumetric MRI scan, referred to as disease probability maps. After generating the disease probability maps for all subjects, the authors selected disease probability maps voxel from 200 fixed locations that were indicated to have high Matthew's correlation coefficient values.

The third step used MLP models; The first model used the feature extracted from these locations were served as input for the classification of AD and NC. The second MLP model used the non-imaging clinical variable (age, gender, MMSE) to classify AD and NC. The third MLP model used the 200 extracted features from the disease probability maps in addition to age, gender, and MMSE to predict the AD status. To assess the significant differences between NC and AD groups, a two-sample t-test and the Chi-squared test were used for continuous and categorical variables, respectively.

The ability of the FCN model to recognize the regions of high AD risk was evaluated by overlapping the disease probability maps with autopsy results, in order to achieve that, a group of 11 individuals from the FHS study had undergone brain autopsy. The locations and frequencies of β-amyloid and tau pathologies were related to AD risk regions. In order to determine the strength and direction of the relationship between the regional AD probabilities and pathology score, Spearman's rank correlation coefficient test was used.

Accuracy, sensitivity, specificity, and F1-score for these three models were obtained; for the MRI model predicted AD, the performance metrics on the ADNI dataset were the Accuracy= 0.834, sensitivity= 0.767, specificity= 0.889, and F1-score= 0.806. For the non-imaging model built by MLP, the performance metrics on the ADNI dataset were Accuracy=0.957, sensitivity=0.924, specificity=0.983, and F1-score=0.951. The fusion model, MLP model that used the multimodal (imaging/ non-imaging) input, the performance metrics which increased significantly were Accuracy=0.968, sensitivity=0.957, specificity=0.977, and F1-score=0.965. Regarding the AIBL dataset, the accuracies on the three models were (0.870, 0.915, and 0.932), respectively. For the FHS dataset, the accuracies on the three models were (0.766, 0.760, and 0.792), respectively. Finally, for the NACC, dataset the accuracies were (0.818, 0.854, and 0.852), respectively. And the sensitivities for the proposed model were (0.764, 0.881, and 0.942), respectively.

Based on these results, it's clear that adding the non-imaging variable to the imaging input has a significant impact on the performance of the DL model in classifying the AD, in addition to facilitating the control for the natural progression of cerebral morphological changes over the time. Moreover, the Model-predicted regions of high AD risk overlapped with the regions that indicated to have a high localized deposition of β-amyloid and tau, which is useful for tracking conspicuous brain regions implicated in AD during diagnosis.
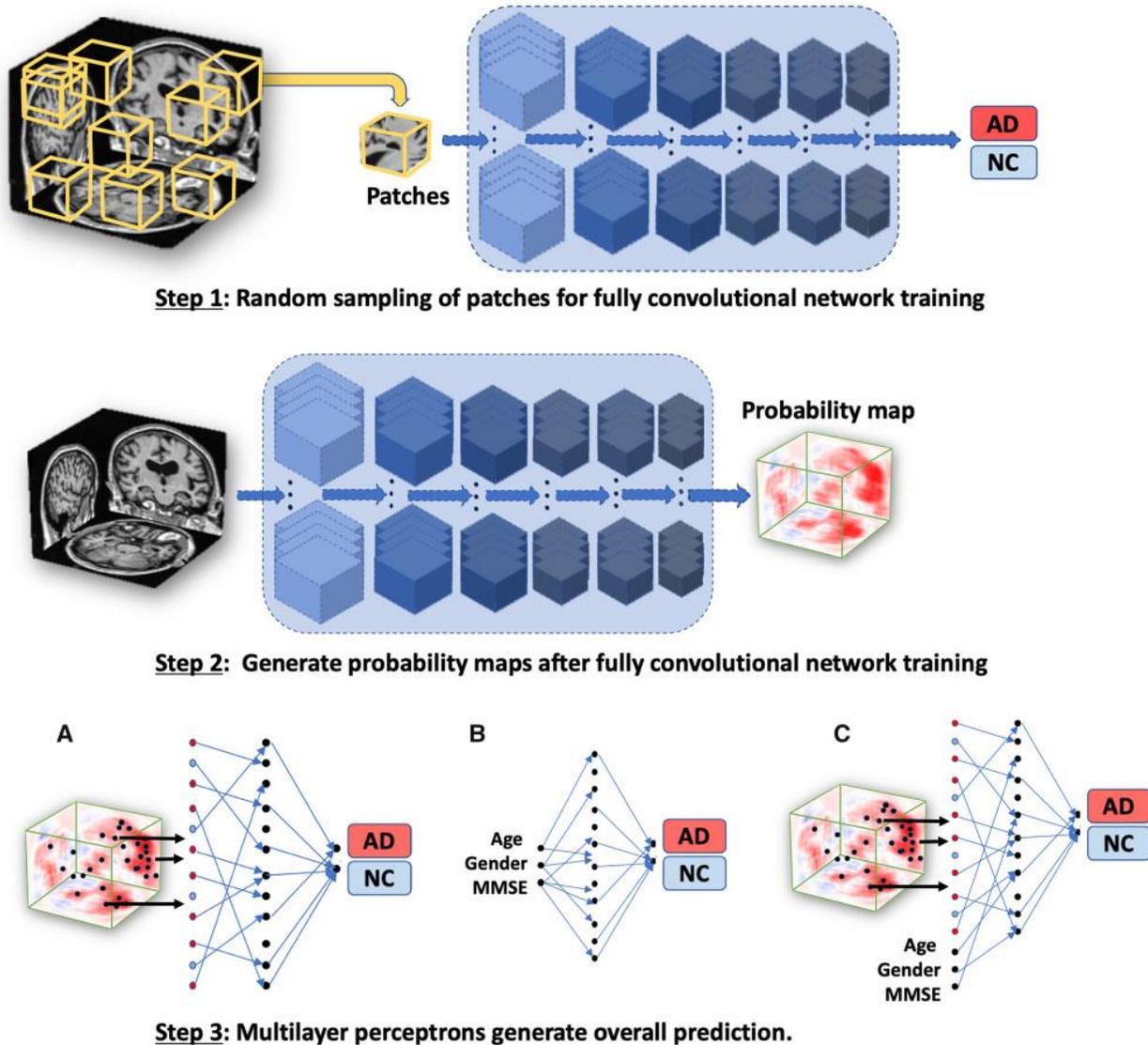

FIGURE 23 GRAPHICAL REPRESENTATION OF THE DEEP LEARNING FRAMEWORK USED BY QIU, S. ET AL. 2020 [48].

BAE, J. B. ET AL. 2020

The classification of AD using DL methods has shown promising results, garnered significant interest between the researchers. However, the successful application in clinical settings requires a combination

of short processing time, high accuracy, and generalization of the model's performance to several populations. In [49] study, the authors aim to develop convolution neural network-based AD classification algorithms using a coronal slice of T1-weighted MRI scans from AD patients which cover the medial temporal lope and age/gender-matched cognitively normal (CN) control from two populations that differ in education level and ethnicity.

The two data sets used in this study are ADNI and the Seoul National University Bundang Hospital (SNUBH); from ADNI the author included the subject who had 3.0 T T1-weighted images and had CDR 0.5 or 1 i.e., the subjects were diagnosed as CN or Mild AD. From the SNUBH, CN control and AD with T1-weighted MRI images whose sex, age, and CDR were matched to the patients in the first dataset, were selected. Each data set contains 390 participants (195 AD and 195 NC), the participant's characteristics like age, sex, education (in a year), MMSE, and clinical dementia rating scale (CDR) are available. The participants from the ADNI performed better on the MMSE and were more educated than those from the SNUBH the authors were not able to match the education and cognitive level.

In addition to selecting only one MRI scan based on diagnosis at the time of assessment and the participant's age. To a more demographically balanced dataset, the authors selected the scan whose demographic factors would contribute to this balance. The 3D T1-weighted brain image was resampled into a grid of 256x256x256 voxels (256 slices/ of 256 × 256 pixels size) using the mri_convert routine in FreeSurfer.

Then, two rounds of rigid transformation were applied to extract coronal slices around the medial temporal lobe (MTL); the first rigid transformation is to match the position of the input images with a template constructed from a CN elderly population, then extract brain parenchyma using Brain Extractor tools. The second rigid transformation is registering the images from the first transformation to a skull-stripped version of the template from the first step. This two-step transformation process was used to increase the accuracy of the registration of each subject's brain parenchyma to the template. To avoid changing the morphological structure of the brain parenchyma, Rigid transformation was utilized. Finally, the 2D coronal slice was extracted. Among the 256 coronal slices, 30 slices were extracted starting from the corpus of the hippocampus.

The Inception-v4 architecture was used in this study with a few modifications; this architecture shows outstanding performances in a 2D image classification with low computational cost. It was pre-trained weights that were obtained from a group of ImageNet used for the classification tasks. For consistency, the authors replicated grayscale coronal slices into three channels due to the architecture of Inception-v4 to take 2D images with three RGB channels as inputs. One of 30 coronal slices was fed into the model independently. Feature vector contains 1024 values representing the result of the convolution procedure. After that, the authors add three values to the end of this vector which is related to the demographic data of the subjects, were age, gender, and the number of the slices being evaluated, the reason for adding the subject age is to observe mild medial temporal atrophy in CN elderly individuals, therefore the magnitude of atrophy should be evaluated with reference to the object's age.

The CNN model contains 497 layers, the architecture of the model used is illustrated the figure 24. The 1027 values of the feature vector are fed to the FCL with two output nodes. The output of the FCL was fed into a SoftMax output layer to predict the probability of the presence of AD. Classification the AD is considered a binary classification problem.

The results for all slices are averaged after labeling the slice as AD or CN. For ADNI and SNUBH datasets, the MRI scans from 390 participants were randomly divided and assigned 80% for development and 20% for testing. During validation and testing procedures, the authors used the average of the predicted

probabilities for each of the 30 extracted slices. The estimation of the final prediction values for the test sets was from the average ensemble values of the five runs of fivefold cross-validation for the development set.

The authors evaluated the proposed models on validation subsets from the other populations (between-dataset validation) and both the same population (within-dataset validation). The models developed from the ADNI dataset presented a mean within-dataset accuracy, sensitivity, AUC, and specificity of 0.89, 0.88, 0.94, and 0.91, respectively, and a between-dataset accuracy, sensitivity, AUC, and specificity of 0.83, 0.76, 0.88, and 0.89, respectively. The models developed from the SNUBH dataset presented a mean within-dataset accuracy, sensitivity, AUC, and specificity of 0.88, 0.85, 0.91, and 0.91, respectively, and a between-dataset accuracy, sensitivity, AUC, and specificity of 0.82, 0.79, 0.89, and 0.85, respectively. The authors compared the AUCs of the algorithms developed from the two datasets using the DeLong test and compared accuracy, sensitivity, and specificity using the student's t-test. Then they compared continuous variables based on independent samples using a paired t-test or Student's t-test as appropriate and compared categorical variables using the chi-squared test. Two-sided p-values can be considered statistically significant if they are less than 0.05.

In conclusion, this study proposed a CNN-based algorithm that uses MRI coronal slices covering the MTL to classify AD. Two independent populations with different ethnicities and education levels were used to train and validate these algorithms. Using 2D images has multiple practical advantages. Such as, 2D images are more widely applicable and available in clinical settings, while 3D MRI scans may not. Moreover, it can reduce processing time (the mean processing time per person was 23s–24s) and computational resources significantly.
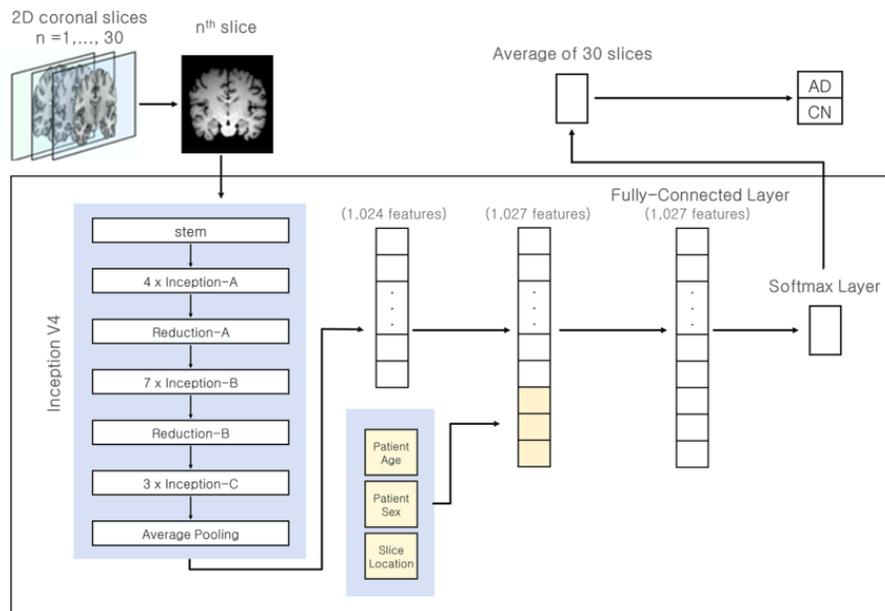
**FIGURE 24 GRAPHICAL REPRESENTATION OF THE NETWORK ARCHITECTURE, WHICH CONSISTS OF INCEPTION V4 NETWORK, AND THE ADDITION OF METADATA, USED BY BAE, J. B. ET AL. 2020 [49].**

LIU, M. ET AL. 2020

In the area of computer-aided AD diagnosis, a combination of identifying brain diseases and predicting clinical scores using MRI has attracted increasing interest since these two tasks are highly correlated in

diseases classification. By automatically learning discriminative features for MR images and the contribution of the demographic information of subjects into the learning process, the authors in [50] aimed to propose a DM^{2}L framework for simultaneous clinical score regression and brain disease classification, using MRI data and non-imaging information of individuals.

Four public datasets containing 1,984 subjects are used in this study, including ADNI-1, ADNI-2, Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD), and AIBL. The same subjects in both ADNI-1 and ADNI-2 were removed from ADNI-2 for independent testing. The ADNI-1 contains (NC=226, AD181, 165 progressive MCI (pMCI), 225 stable MCI (sMCI)). In the ADNI-2 dataset, there are (143 AD, 185 NC, 37 pMCI, and 234 sMCI) subjects. Four types of clinical scores are utilized for subjects in both ADNI-1 and ADNI-2, including Clinical Dementia Rating Sum of Boxes (CDRSB), classic Alzheimer's Disease Assessment Scale Cognitive subscale (ADAS-Cog) with 11 items (ADAS11), modified ADAS-Cog with 13 items (ADAS13), and MMSE. MIRIAD dataset contains (46 AD and 23 NC) subjects. But in the MIRIAD dataset, only MMSE score, gender, and age information are available for all 69 subjects, while other scores are not available for all subjects. For this reason, the authors use only this information in the experiments. AIBL dataset,519 subjects including (72 AD and 447 NC) subjects. Like the MIRIAD dataset age, and gender, as well as the MMSE score, are available for all subjects in AIBL.

To achieve the goal, firstly, identifying anatomical landmarks via a data-driven algorithm from the pre-processing MR images, then based on these identified landmarks, image patches from MRI were extracted. By utilizing image patches and demographic information (age, gender, and education) to input a deep multi-task multi-channel CNN to jointly execute classification and regression tasks. For the pre-processing procedure, firstly, the authors perform the correction, using the MIPAV software and re-sample each image in a resolution of 256 × 256 × 256, (256 slices/ of 256 × 256 pixels size). Then correcting the intensity inhomogeneity by using the N3 algorithm. Finally, they remove the other structures like the skull, dura, and cerebellum.

Few studies report anatomical landmarks which can model both global (whole-image-level) and local (voxel-level) information of brain MRI. Anatomical landmark identification from MRI via a data-driven landmark detection algorithm, which aims to identify the landmarks that have statistically significant group differences between AD and NC in local brain structures. linear and non-linear registration processes are performed for all training MRI in the ADNI-1 dataset using the Colin27 template. For each voxel in the template, the morphological features (local energy pattern) are extracted from its corresponding voxels in all linearly aligned training images that include both AD and NC subjects in ADNI-1. Then, to perform a voxel-wise group comparison between AD and NC groups, the authors use a multivariate statistical test and thus can obtain a p-value for each voxel in the template space. In the end, the local minima in the obtained p-value map in the template space are defined as locations of discriminative anatomical landmarks. A higher discriminative capability of the corresponding landmark in distinguishing AD from NC is denoted by a smaller p-value and vice versa.

Due to the high number of anatomical landmarks identified (around 1700), a spatial Euclidean distance threshold was defined to eliminate the overlapping among the image patches and control the distance between the landmarks. The authors extract a 3D image patch centered at a specific landmark location. the author selected the top L = 50 landmarks, which are located in the areas which supposed to be related to AD/MCI depending on the literature, these areas are bilateral parahippocampal, bilateral hippocampal, and bilateral fusiform. Further, the authors randomly sample different patches centered at each landmark location with the displacements within a cubic with volume 5 × 5 × 5, in order to suppress the impact of registration error and to augment the training set. That is, 25 patches centered at each landmark location

can be generated. Finally, the authors treat a combination of L patches as a training sample, with each patch extracted from a particular landmark location. Each patch fed to the input of the model, in addition to the three demographic information (age, gender, and education), as seen in figure 25.

To evaluate the performance of this model, two experiments were performed. The first one, the model trained on the ADNI-1 dataset and tested on ADNI-2, MRI, and three demographic information (age, gender, and education) used as input of the model. In the model trained for multi-class classification (AD vs. pMCI vs. sMCI vs. NC), the overall classification accuracy (ACC) for four categories is 0.518, and the accuracy for each category (AD =0.490, pMCI = 0.243, sMCI = 0.513, and NC= 0.600). The performance of the four clinical score regression (CDRSB, ADAS11, ADAS13, and MMSE) is calculated by the root mean square error (RMSE) (1.666, 6.200, 8.537, and 2.373), respectively. and correlation coefficient (CC) (0.533, 0.565, 0.590, and 0.567), respectively. While the performance metrics without considering the metadata, the overall classification accuracy (ACC) for four categories is 0.487, and the accuracy for each category (AD =0.427, pMCI = 0.297, sMCI = 0.415, and NC= 0.665).

The second one, the model trained on the ADNI-1 and tested on MIRIAD, the model train for binary classification (AD, NC) and for MMSE score regression. The binary classification performance is evaluated by the ACC= 0.937, SEN=0.946, SPE= 0.932, AUC=0.986. The regression performance is calculated by RMSE= 4.136 and CC=0.736. While the performance metrics without considering the metadata were: ACC= 0.920, SEN=0.963, SPE= 0.898, AUC=0.969., and the regression performance which calculated by RMSE= 4.295 and CC=0.683.

In addition to this comparison, the authors performed comparisons with four state-of-the-Art approaches. The first approach is called Landmark-based Morphometrical Feature (LMF), in this method the tasks of regression and classification are performed separately. The second one is called Multi-Modal Multi-task (M^3T) learning method. The third one is Manifold Regularized Multi-Task Feature (M^2TF) learning method. Finally, the Matrix-Similarity based Joint Learning (MSJL) method, the last three methods are joint classification and regression. Note that LMF, MSJL, and M^3T rely on support vector machine (SVM) and linear support vector regressors (SVR) for separate classification and regression, while the proposed model DM^2L and M^2TF can jointly perform classification and regression, The result of these comparisons is shown in the figure (26).

In general, using the demographic data helps improve the learning performance of the proposed method. The DM^2L model was proposed for simultaneous AD classification and clinical score regression, utilizing subjects' MRI and demographic information (age, gender, and education). The proposed model performs classification and regression at a speed of close to real-time, which is particularly important in real-world applications, needs around 15 seconds.
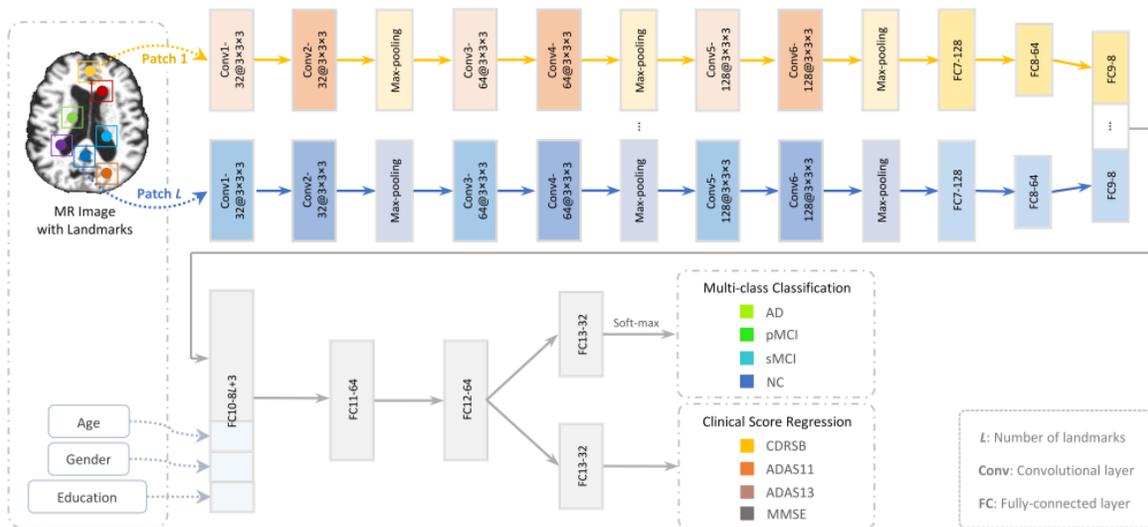
FIGURE 25 GRAPHICAL REPRESENTATION OF THE PROPOSED FRAMEWORK USED BY LIU, M. ET AL. 2020 [50].
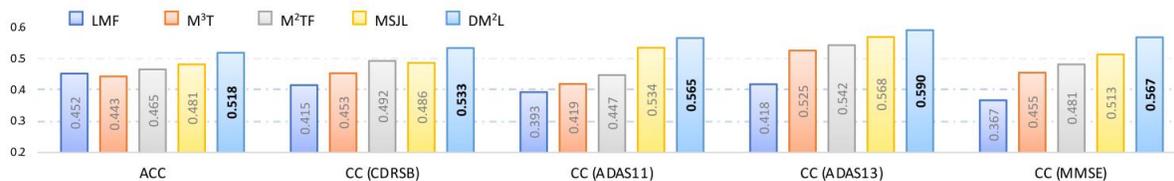
FIGURE 26 COMPARISON BETWEEN THE PROPOSED MODEL AND FOUR STAT-OF-THE-ART APPROACHES [50].

ESMAEILZADEH, S. ET AL. 2018

End-to-end classification AD and identifying the AD biomarkers were the target of [51]study; the authors reached this goal by building a 3D-CNN, that beginning on classifying two classes then ends into three categories.

The data used in this study is T1-weighted structural MRI scans from the ADNI dataset for 841 subjects (200AD, 230NC, and 411MCI). Age and gender for the subjects are available. The clinical criteria to diagnose the subjects with AD or MCI were MMSE scores and CDR.

A simple preprocessing step was performed on the MRI images to remove the non-cerebral tissue using the Brain Extraction Technique. A 3D-CNN was built using the TensorFlow framework for the end-to-end classification task. The authors proposed two architectures of CNN, a complex one (shown in figure 27) and a simple architecture with fewer filters and one less FCL. The number of trainable parameters in the complex architecture is greater than the number of the simple ones. The fewer parameters avoid the overfitting of the network on a limited number of subjects. The input MRI images were re-seized to 116x130x83 voxel. The demographic data (Age and Gender) were added to the first FCL. A RELU activation

function was used, and the loss by the cross-entropy cost function, which is minimized by utilizing the Adam optimizer. A drop-out (D/O) and l2-regularization (Reg) were added to optimize the parameters and improve the model. The authors augment the data for the model training by flipping all subjects by swapping the left and the right hemispheres. To improve the classification task and identify the related biomarkers, firstly, the model trains on two-class (AD and NC), then use the cross-entropy in the last layer to extend for the multi-class cases, which means adding the third output class (MCI).

Training the three-class model from scratch provided results worse than this transfer learning strategy's results as the MCI is the middle stage between the cognitive decline of aging and the more pronounced decline of dementia, therefore, first learning to identify the differences between the AD and NC classes by separating AD from NC. After that, by adding the third class and fine-tuning the network transfers the learned knowledge to classify the third condition.

To identify AD biomarkers, the authors perform an image occlusion analysis on the proposed model by sliding a box of 1×1×1 zero-valued voxels along with the whole MR image of AD patients labeled as AD. Hence, the importance of each voxel can be characterized as the relative confidence of the subjects that are classified as AD. The resulting heat map is shown in Fig 28. The color map refers to the relative importance of each voxel. The regions in the heat map coincide with the hippocampus, amygdala, thalamus, and ventricles of the brain which are mentioned in the previous literature to be responsible for short-term memory and early stages of AD.

To evaluate the performance of the model in binary classification, classification accuracy (ACC), F2-score, precision (Pre), and recall (Rec) were used. The best model was that it used a data augmentation and drop-out (D/O) and l_2-regularization (Reg). The simple model achieved Acc 94.1%, F2= 0.93, Pre= 0.92, and Rec=0.94. While the complex model achieves Acc 88.3%, F2= 0.89, Pre= 0.88, and Rec=0.91. Due to the simple network being less prone to overfitting, it outperforms the complex one. The authors used a Fisher exact test to test the significance of the classification results, and both models led to a p-value of less than 0.001.

In comparison with two prior works that did not use the metadata, the ACC were 90.8% and 91.1% which are lower than the results of the proposed model, the results show that the proposed end-to-end classification method has a better result. The three-class classification models were tested in both (with transfer learning/ without transfer learning). The results show that the model with transfer learning perform better classification than the other model. the simple model achieves Acc 61.1%, F2= 0.62, Pre= 0.59, and Rec=0.63, while the complex model Acc 57.2%, F2= 0.59, Pre= 0.55, and Rec=0.61. while the simple model without transfer learning had Acc 0.54%, F2= 53.4, Pre= 0.49, and Rec=0.55. However, if the model is trained from scratch with AD, MCI, and NC classes, the model results in worse accuracies, Because of the difficulty to distinguish the MCI class from AD or NC. On the other hand, training the model based on all three classes at once, the model gets stuck in local optima easier and overfit to the training data.

In conclusion, Using MR images, a 3D-CNN model was created to diagnose AD and its prodromal stage, MCI. This end-to-end approach not only outperformed other methods in terms of classification but also contributed to identifying significant disease biomarkers. The hippocampus region of the brain was found critical in the diagnosis of AD. This research discovered that the simple architecture performed better in testing than the more sophisticated architecture because it is less prone to overfitting the training data.
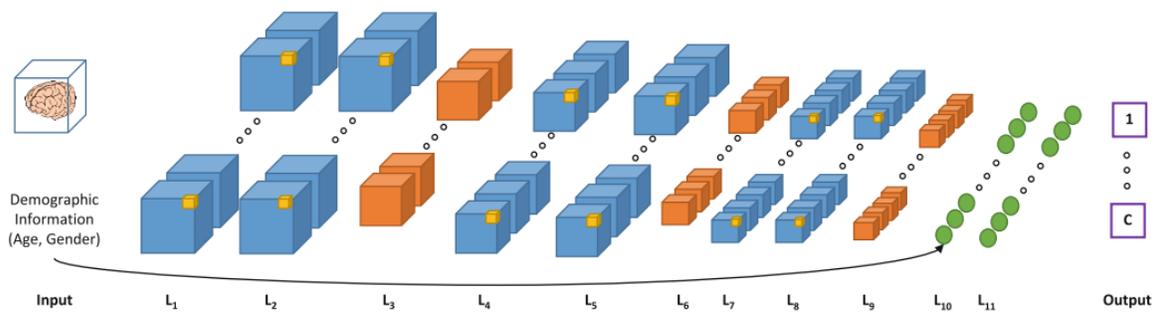
FIGURE 27 GRAPHICAL REPRESENTATION OF 3D-CNN USED BY ESMAEILZADEH, S. ET AL. 2018 [51].
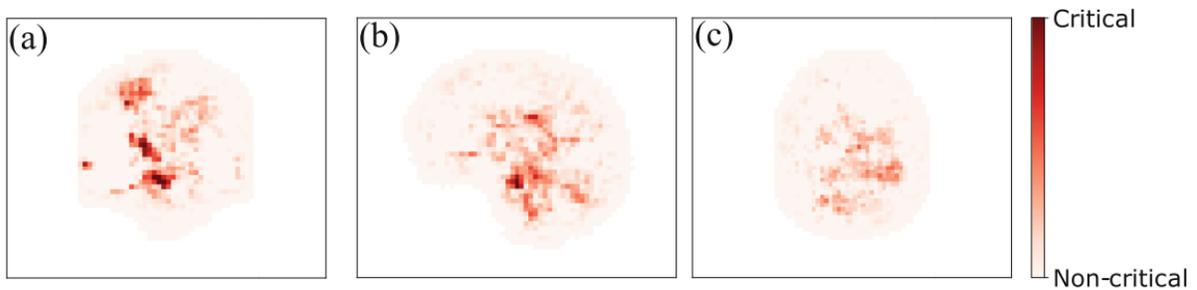
FIGURE 28 RELATIVE IMPORTANCE OF DIFFERENT VOXELS ASSOCIATED WITH AD DIAGNOSIS [51].

PUENTE-CASTRO, A. ET AL. 2020

Most of the studies interested in AD diagnosis have used horizontal MRI images or volumes. The [52] study proposes a DL model to diagnose AD by using sagittal MRI images without the assistance of any other kind of acquisition plan and this study proves that the transfer learning strategy (TL) is crucial for training models from a few available data. Using the sagittal plane which has information at different vertical levels of the brain, makes it possible to extract information from other regions of the brain, with lower computational cost, without the need to process a huge amount of images.

The datasets used in this paper were: 436 sagittal images from the Open Access Series of Imaging Studies (OASIS) dataset. Out of these, 2 were moderate AD cases, 70 very mild dementia, 28 were mild dementia, and 316 were cognitively normal. In addition to 1743 sagittal images from the ADNI dataset, divided into 297AD, 525 CN cases, and 921 mild cognitive dementia.

To prevent overfitting and overcome the imbalances of the dataset, an initial parameterization was defined in the weights of the classes. Therefore, the weight adjustment was balanced. The highest represented class will be given the lowest weight or priority. When multiplying this weight by the number of elements in each class, the same value will always be given to have a value proportional to the number of elements in each class.

MRI scans were fed to a ResNet artificial neural network (ANN) in order to extract feature vectors, then the age and sex of the patients are concatenated to the vectors (shown in figure 29). 100,352 numerical

values extracted from each MRI scan were contained in the obtained vectors. Training and testing data were divided from the total set of vectors. The training data were used for training the SVM model with Radial Basis Function (RBF) kernel. The testing data were utilized for evaluating the effectiveness of the trained SVM model in order to improve it.

Due to the small size of the MRI datasets, the authors used a transfer learning strategy, which was done by training the model for another task with a huge number of images. Therefore, fewer weights were required to be tuned, which means that fewer new data were needed to adapt the model to the current problem because of the small number of weights that were required to be tuned. So, using an ANN trained required less amount of data and makes the experimentation faster. The output of the last layer of the ResNet was concatenated to the age and sex of the patients, and then it was fed to a SVM for the classification task.

To measure the performance of the proposed model, the Accuracy, Specificity, Precision, Recall, and F1 scores were chosen. For the OASIS dataset, Because of the few cases contained in the data set for the advanced stages of AD (28 for mild dementia and 2 for moderate AD) out of the 416 available cases, the results are shown in Table 1, with and without considering age and sex, for these stages of AD were not detected (recall= 0%). It could be solved by performing data augmentation, increasing the number of cases, and solving the imbalance. However, the proposed model showed a remarkable ability to diagnose AD at the earliest stage. Compared to another study from the literature that was not good for diagnosing AD in the earlier stages, which detected fewer positive cases (recall= 50%), it is important to find positive cases as soon as possible to start preventive treatment.

The result obtained from only MRI images were compared to the result when the sex and age of the patient were considered. The latter shows that there are minor improvements, especially for the first two classes (CN and Very Mild Dementia). The result was better than OASIS for the ADNI dataset due to the more minor imbalance between the stages and the larger amount of data, i.e., more data, representing a greater phenotypic variety, for each stage. However, because of the few examples of AD cases, the AD stage was the lowest stage was correctly detected (recall=30.62%). Despite this, the proposed model is satisfactory in addressing the diagnosis at the earliest stages (Mild Cognitive Dementia) compared to a previous literature study, as shown in table 2.

Considering the demographic variables gives results worse than not considering them in all cases except for the earliest stage of AD. This model was designed to assist in the early diagnosis of AD using MRI sagittal images and the DL technique. Transfer learning technique, ResNet ANN used as feature extractor, with SVM. The proposed model detects better CN cases and early stages of the disease in both data sets which are the most important and difficult stages to detected. The detection on this stage led to greater efficacy of the therapy. This model presents satisfactory results compared to those found in the state-of-the-art that used horizontal plane MRI. Using pre-trained models that require less data accelerates the experiment design and makes processing a task faster.
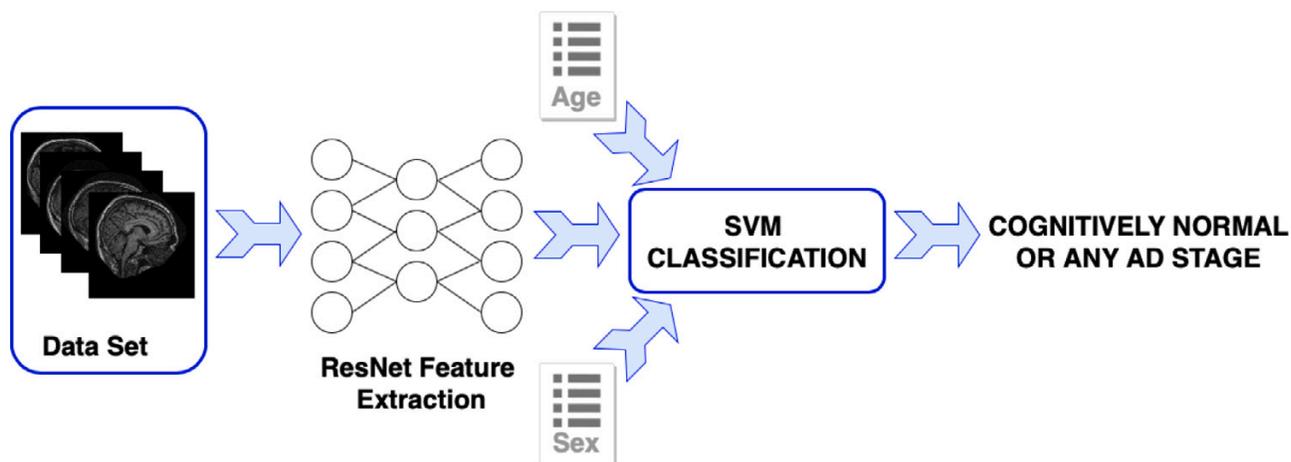
FIGURE 29 GRAPHICAL REPRESENTATION OF THE WORKFLOW DIAGRAM USED BY PUENTE-CASTRO, A. ET AL. 2020 [52].

TABLE 1 BEST RESULTS WITH THE OASIS DATASET [52].

	Class	Accuracy	Precision	Recall	Specificity	F1
Horizontal MRI	Cognitively Normal	N/A	99.00%	99.00%	N/A	99.00%
	Very Mild Dementia	N/A	75.00%	50.00%	N/A	60.00%
	Mild Dementia	N/A	63.00%	71.00%	N/A	67.00%
	Moderate AD	N/A	33.00%	50.00%	N/A	40.00%
	Average	N/A	67.50%	67.50%	N/A	66.50%
Proposed model without considering sex and age (sagittal MRI)	Cognitively Normal	79.36%	89.94%	82.44%	69.00%	86.02%
	Very Mild Dementia	74.31%	33.06%	58.57%	77.32%	42.27%
	Mild Dementia	92.66%	0.00%	0.00%	99.02%	0.00%
	Moderate AD	99.54%	0.00%	0.00%	100.0%	0.00%
	Average	86.47%	30.75%	35.25%	86.34%	32.07%
Proposed model (sagittal MRI)	Cognitively Normal	80.05%	92.54%	81.25%	87.00%	86.53%
	Very Mild Dementia	75.00	35.77	70.00	75.96	47.34
	Mild Dementia	92.66%	0.00%	0.00%	99.02%	0.00%
	Moderate AD	99.45%	0.00%	0.00%	100.0%	0.00%
	Average	86.81%	32.08%	37.81%	88.25%	33.47%

TABLE 2 BEST RESULTS WITH THE ADNI DATASET [52].

	Class	Accuracy	Precision	Recall	Specificity	F1
Horizontal MRI	Cognitively Normal	N/A	100.00%	100.00%	N/A	100.00%
	Mild Cognitive Dementia	N/A	60.00%	80.00%	N/A	69.00%
	AD	N/A	70.00%	47.00%	N/A	56.00%
	Average	N/A	76.67%	75.67%	N/A	75.00%
Proposed model without considering sex and age (sagittal MRI)	Cognitively Normal	78.25%	64.44%	59.80%	86.05%	62.03%
	Mild Cognitive Dementia	71.51%	69.02%	84.32%	56.95%	75.91%
	AD	86.40%	73.42%	31.88%	97.62%	44.45%
	Average	78.72%	68.96%	58.66%	80.21%	60.79%
Proposed model (sagittal MRI)	Cognitively Normal	78.36%	63.69%	59.60%	85.97%	61.58%
	Mild Cognitive Dementia	71.50%	69.00%	84.61%	56.50%	76.01%
	AD	86.05%	73.93%	30.62%	97.72%	43.31%
	Average	78.64%	68.87%	58.28%	80.06%	60.30%

QIU, S. ET AL. 2018

Accurate diagnosis and early detection of AD need careful medical assessment, including patient history and physical and neurological examinations. The MMSE is a cognitive assessment tool utilized to screen for dementia, which is a 30-point questionnaire. The LM test is considered a sensitive AD test and is commonly used to assess verbal memory. The authors in [53] aim to develop a predictive model of MCI by considering detailed structural information contained in the MRI images and a cognitive function assessed using the LM tests and MMSE.

The authors selected a cross-sectional collection of individuals from the NACC data set, the following criteria were utilized to select one MRI scan from each subject: subjects were NC or MCI; subjects have LM and MMSE test score; MRI scans with at least 20 slices in the axial plane; type of MRI was either T1 weighted or FLAIR; the dimension of each cross-sectional slice of the MRI was about 256 pixels. The dataset used in this study is T1 weighted MRI images for 386 subjects from the NACC, containing 303 Normal Cognitive NC and 82 MCI. For each of 386 subjects, three MRI slices were selected manually that covered the regions of interest which correlated to AD and MCI, these regions are inferior temporal, lateral ventricles, and middle temporal cortices. The main data were split in 2:1:1 for training, validation, and testing.

In this work, the Oxford University's Visual Geometry Group's (VGG-11) model was adapted. That pre-trained on millions of images with 1000 object classes. This adaption includes inserting a batch normalization layer after every convolutional layer and a dropout layer after every max-pooling layer. To perform binary classification on the output of VGG-11 that formed the deep neural network, two fully connected layers after the output layer of VGG-11 were added. For the MMSE and LM models, the authors developed two MLP models, that can be used to train non-imaging data sets, to predict the MCI. In the MMSE model, there are 3 input nodes, four input nodes for the LM model, one hidden layer with sigmoid function, and one output layer with SoftMax non-linear function with two outputs nodes. (As shown in

figure 30) three VGG-11 models were independently trained on the three slices. Max, mean, and majority voting were performed on the outputs of these three models.

The final prediction in the mean voting takes the mean probabilities predicted from each modified VGG-11 model trained on the three slices. The final prediction is defined as:

$$P_{NC}^{mean} = (P_{NC}^1 + P_{NC}^2 + P_{NC}^3)/3 \qquad (11)$$

$$P_{MCI}^{mean} = (P_{MCI}^1 + P_{MCI}^2 + P_{MCI}^3)/3 \qquad (12)$$

P_{NC}^i and P_{MCI}^i are the probabilities predicted from the ith model that one specific scan belongs to NC or MCI.

In max voting the final prediction depends on the prediction from the model with the largest confidence. where confidence is defined as:

$$\max{(P_{NC}^i, P_{MCI}^i)} \qquad (13)$$

The final prediction in the majority voting takes the label NC or MCI, which was predicted for equal or more than 2 times.

In order to generate a model based on MMSE results alone, three features were selected: MMSEORDA (orientation subscale score—time), NACCMMSE (total MMSE score), and MMSEORLO (orientation subscale score—place). For generating an LM test-based model the authors select all features that are considered as part of the LM test: LOGIPREV (total score from the previous test administration), MEMUNITS (total number of story units recalled), MEMTIME (time elapsed since first recall to delayed recall), and LOGIMEM (total number of story units identified from this current test administration). LM and MMSE models were trained using backpropagation with a gradient descent optimizer, 0.9 was the momentum with learning rate 0.05.

To generate a fused MRI model, predictions from the three voting methods were combined by majority voting. Two MLP models were independently developed on LM test and MMSE results. To produce a final prediction of the multimodal fusion model, the projections from the three principal models were combined by applying majority voting.

Model's performance was evaluated by accuracy, precision, F1 score, AUC, and Matthews's correlation coefficient (MCC). The result of the MRI model shows that the VGG11 architecture is able to capture the small details of the 2D image regions. Regarding the LM and MMSE models, in terms of accuracy, precision, F1, and MCC the authors observed that the LM model has better performance than the MRI and MMSE models. In terms of Accuracy, F1 score, recall, and MCC, the performance of the fused multimodal model outperformed the other three base models, as seen in Table 3. The precision of the LM model is slightly higher than that of the fused multimodal model.

In this study, a DL model automatically distinguishes the NC and MCI from the data set. MRI models can further push the diagnostic accuracy from MMSE and LM models to a superior level. The fused MRI model outperformed the performances of each other model, depending on the performance metrics. Regarding the fusion model results show that multiple neural network models can be combined and trained to

generate a fused model to predict MCI. In terms of predicting MCI, the majority voting on the three base models contributes to enhancing the accuracy of the multimodal fusion model.
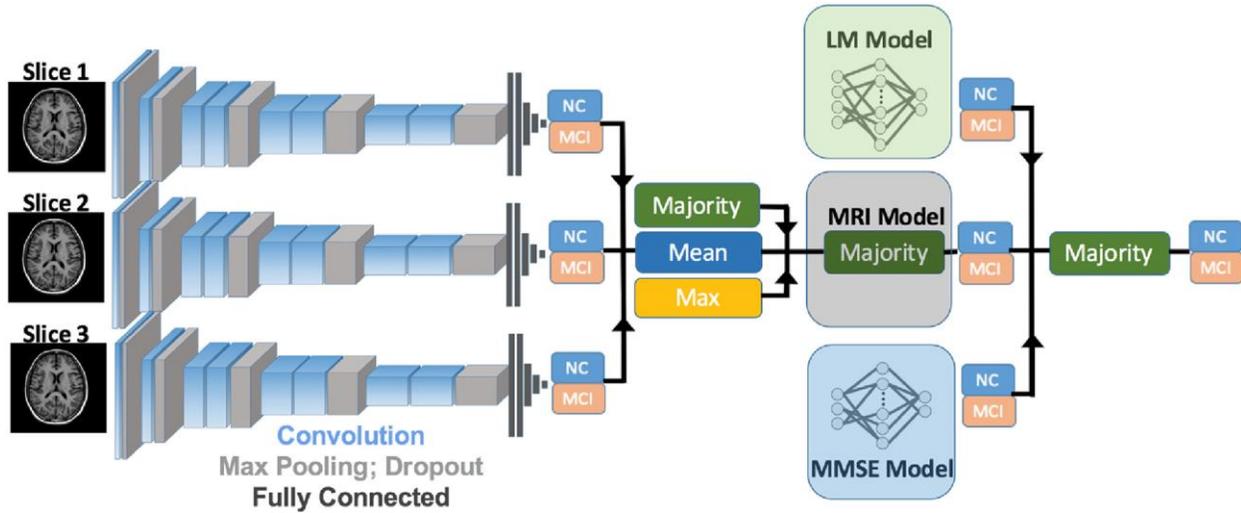
FIGURE 30 GRAPHICAL REPRESENTATION OF THE MODELING ARCHITECTURE USED BY QIU, S. ET AL. 2018 [53].

TABLE 3 PERFORMANCE METRICS FOR THE 3 BASE MODELS ALONG WITH THE MULTIMODAL FUSION MODEL [53].

Model Type	Accuracy	Precision	Recall	F1	MCC
MRI model	83.1±4.2%	0.878±0.031	0.913±0.050	0.894±0.27	0.481±0.132
MMSE model	84.3±2.3%	0.888±0.044	0.921±0.049	0.902±0.14	0.518±0.111
LM model	89.1±1.9%	**0.951**±0.024	0.908±0.001	0.929±0.011	0.698±0.067
Majority voting	**90.9**±2.7%	0.926±0.037	**0.963**±0.015	**0.944**±0.015	**0.719**±0.101

SARATXAGA, C. L. ET AL. 2021

The [54] study aims to propose and develop models to estimate the presence of AD by brain MRI analysis and compare the results with previous literature works, by calculating the accuracy and balanced accuracy (BAC) which is a more realistic metric to evaluate the accuracy of the approach in the presence of the high data imbalance of this dataset, in addition to including metadata such as age and sex. In the OASIS collection, the severity of AD can be labeled through the CDR indicator.

The data used in this study was obtained from the OASIS dataset, considering the CDR sample was categorized. From OASIS-1, the total number of subjects is 436 (336NC, 70very-mild dementia, 28 mild dementia, and two moderate dementia). Regarding the OASIS-2, the number of subjects is 305 (177 NC, 98 very-mild dementia, 27 mild dementia, and three moderate dementia). A CDR was used as an indicator of the severity of AD, the value could be 0 (healthy), 0.5 (very mild), 1 (mild), or 2 (severe). To avoid overfitting and increase variability, a data augmentation strategy was performed, where images are randomly horizontally or vertically flipped or rotated.

2D and 3D input images were used depending on the neural network's architecture. The architecture utilized to complete the comparison in this paper is BrainNet2D (2D Conv), BrainNet3D, and ResNet.

The input of the BrainNet2D is considered as a 3D subject level and one output. The dimension of the input data is (M, N, K), the size of every slice is M and N, the number of slices is K. The information is

treated slice by slice. BrainNet2D includes four convolution blocks, each one contains a 2D convolution layer and 2D max-pooling layer in order to reduce spatial resolution and extract a representative number of high-level descriptors. The activation function is 'SoftMax' for the FCLs for exclusive classes. BrainNet3D is similar to the previous one but with a real 3D approach. The dimension of the Input data is M, N, K, and 1, the size of every slice is M and N, and K is the number of slices. Five convolutional blocks are proposed, each one of them contains the 3D convolutional layer and 3D max-pooling layers. This approach appropriates better with biology and human understanding. The activation function is 'SoftMax' for exclusive classes.

Cyclical Learning Rate (CLR) and Batch Normalization were also introduced in this study, which helps stabilize the learning process and contributes to a better training process and performance. Due to the small amount of the dataset used, Batch normalization can help reduce overfitting. The dimensions of MRI volume are 176 × 208 × 176, (176 slices/ of 176 × 208 pixels size). The training, validation, and test sets are automatically randomly generated for the used datasets. In order to eliminate possible bias, each set contains a balanced number of samples per sex (Male/Female). Just the horizontal MRI scans were used for the training and testing of the proposed model.

To include the MRI information two different strategies have been considered in the experiments. One approach uses a limited number of volume slices, which has been used with models built with 2D convolutional layers considering the center slice to be 88 i.e., the half of the 176, with an extracted number of slices from each side, the authors suggested that 10 slice is the perfect number of slices were extracted which contain the most important and relative information for diagnosing the AD. The other strategy has been to use the whole 3D volume for models constructed with 3D convolutional layers. The patient's sex and age have both been implicated in the diagnosis. As a result, these two variables have been integrated into the network to help with the classification output. The metadata has been considered as two classes for age (<60 and >60), and two classes for sex (male, female). The inclusion in the network has been done before the final activation layer and after the embeddings.

Depending on the accuracy level, the results were compared by a two-class model (NC vs AD) and a three-class model (NC vs very-mild dementia vs mild and moderate dementia). The comparison results between the different models were performed, taking into account the contribution of the metadata (age/sex) and the normalization methods on the performance.

The results for the two-class classification, over the OASIS-1 dataset, were the following: regarding the BrainNet2D with CLR triangular learning rate, the accuracy (ACC) was 0.81and the balanced accuracy (BAC) was 0.82. When metadata are included, the ACC is 0.79, and BAC is 0.84. For the BrainNet3D using Batch Normalization, the ACC is 0.79, and BAC is 0.83. In addition to CLR triangular learning rate and metadata, the ACC is 0.80, and BAC is 0.84. The results over the OASIS-2 were the following: for BrainNet2D model using CLR triangular learning rate, the ACC was 0.92 and BAC was 0.92, with metadata the ACC is 0.82 and BAC is 0.83. For BrainNet3D with Batch Normalization, the ACC is 0.84 and BAC is 0.84. by adding metadata, the ACC was 0.77, and BAC was 0.78. OASIS-2 dataset seems to demonstrate that the sex/age metadata are not a stable source of information to feed the model during the classification process. The two experiments that contained the metadata achieved BAC values significantly inferior to the other experiments with the same model. The possible reason for that is because the metadata are highly biased by the characteristics of the population included in the datasets, but are not representative of the real clinical incidence of AD.

From the results above, the BrainNet2D with CLR triangular learning rate and the inclusion metadata show slightly better results. 2D approaches have an improvement rate higher than the 3D approaches. The

possible reason for the poorer improvement may be the good skull removal and pre-processing applied to all the slices in the sequences is required in the input.

OVERVIEW COMPARISON

The seven literature studies used in this thesis represent the effectiveness of including the metadata in the DL models in order to AD classification. The most noticed metadata used was the age and sex. it was noticed in all studies except the [52] study, which used another type of metadata, which was cognitive assessment tests (MMSE and LM). While the [50] used education level in addition to age, sex. In [48], the authors mixed demographic information (age, sex) and cognitive assessment test (MMSE).

The problem of AD diagnosis is simplified as a classification problem. Some of the studies used two classes (NC and AD) [48] or (NC and MCI) [53] while other papers used multiple stages of AD (AD, NC, and MCI) [51]. Performing a multiclass classification problem is more complex and has fewer chances to succeed than the case of binary classification. Therefore, the comparison between those studies might not be completely fair.

The input images of the studies used different resolutions. Those resolutions represent the amount of information stored in the data. By comparing the performances of each study with its resolution, we could not notice any effect. This means the lowest used resolution (116x130x83) [51], which achieved 94% accuracy, does not affect the performance in a negative way. Consequently, it is possible to downsample the high resolution of the scan if the computational power was insufficient for the task.

Among the literature studies, the transfer learning strategy was utilized obviously, due to the significant effect in increasing the performance of the model and accelerating the training procedure. Transfer learning strategy was done by training the model for another task with a huge number of images, it is useful when the dataset used has a few samples. The [48, 50] studies train the model from the scratch, while the other studies pre-trained the model on the ImageNet subset. In order to make the comparison in a legit way, the pre-trained model of transfer learning should be taken into account.

Finally, the comparison might be hard to be conducted due to the usage of different strategies in terms of input, output, and processing. However, the most constant variable among the studies was the used metadata, in particular age and sex.

TABLE 4 SUMMARY OF THE LITERATURE.

Ref.	Year	DB	Cardinality of classes	Resolution	Classifier and architecture	Metadata	Input	Optimization	Output classes	Acc (%)	Sp (%)	Se (%)	F1 (%)	Recall (%)	AUC (%)
[48]	2020	ADNI	NC=229, AD=188	181x 217x 181	FCN + MLP	Age, Sex, MMSE	T1-weighted full MRI	Adam	AD, NC	97	96	98	97	-	100
		AIBL	NC=320, AD=62							93	88	94	81		97
		FHS	NC=73, AD=29							79	74	81	63		88
		NACC	NC=356, AD=209							85	92	81	82		95
[49]	2020	ADNI	NC=195, AD=195	256x 256x 256	Inception-v4 + MLP	Age, Sex	T1-weighted full MRI	mini-batch SGD	AD, NC	89	88	91	-	-	94
		SNUBH	NC=195, AD=195							88	91	85	-		94
[50]	2019	ADNI-1	NC=226, AD=181 pMCI=165 sMCI= 225	256x256x256	CNN + MLP	Age, Sex, Education	MRI	SGD	-	-	-	-	-	-	-
		ADNI-2	AD=143 NC=185 pMCI=37 sMCI=234						AD vs. pMCI vs. sMCI vs. NC	52	-	-	-	-	-
									AD	49					
									pMCI	24					
									sMCI	51					
									NC	60					

		MIRIAD	AD=46 NC=23						AD vs.NC	94	95	93	-	-	99
		AIBL	AD=72 NC=447						N/A	-	-	-	-	-	-
[51]	2018	ADNI-1	AD=200 NC=230 MCI=411	116x 130x 83	3D-CNN + MLP	Age, Sex	T1-weighted structural MRI	Adam	AD, NC	94	-	-	-	94	-
									AD, NC, MCI	61	-	-	-	63	-
[52]	2020	OASIS	AD=2 VMD=70 MiD=28 NC=316	256x 256	ResNet +SVM	Age, Sex	Sagittal MRI	NA	NC	80	78	-	87	81	-
									VMD	75	76	-	47	70	-
									MiD	93	99	-	0	0	-
									AD	99	100	-	0	0	-
		ADNI	AD=297 NC=525 MiD=921	256x256					NC	78	86	-	62	60	-
									MiD	72	56	-	76	84	-
									AD	86	98	-	43	31	-
[53]	2018	NACC	NC=303 MCI=82	224x 224	VGG-11 +MLP	MMSE, LM	MRI		NC, MCI	91	-	-	94	96	-
[54]	2021	OASIS-1	NC=336 VMD=70 MiD=28 MoD=2	176 x 176	BrainNet +MLP	Age, Sex	MRI	NA	NC vs AD	79	-	-	-	-	-
										80	-	-	-	-	-
		OASIS-2	NC=177 VMD=98 MiD=27 MoD=3	224 x 224						82	-	-	-	-	-
										77	-	-	-	-	-

ADNI: Alzheimer's Disease Neuroimaging Initiative; AIBL: Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing; FHS: Framingham Heart Study; NACC: National Alzheimer's Coordinating Center ; SNUBH: Seoul National University Bundang Hospital ; MIRIAD: Minimal Interval Resonance Imaging in Alzheimer's Disease ; OASIS: Open Access Series of Imaging Studies; FCN: Fully convolution network; MLP: Multilayer perception; CNN: Convolution Neural Network; SGD: Stochastic gradient descent; AD: Alzheimer's disease; NC: Normal Control; MCI: mild cognitive impairment; VMD: very mild dementia; MiD: mild dementia; MoD: Moderate dementia; pMCI: progressive MCI; sMCI: stable MCI; ACC: accuracy; Se: sensitivity; Sp: specificity; AUC: area under the curve; MMSE: Mini-Mental State Examination LM: The Wechsler Memory Scale Logical memory test; MRI: Magnetic resonance imaging;

Discussion

This research is a theoretical investigation of the literature. It aims to identify the efficiency of combining the metadata of AD patients with the MRI scans in order to enhance the performance of CAD.

The research included 7 papers as final candidates for the investigation. The selection criteria of are consist of two steps. Firstly, utilizing the number of keywords to create several search queries. Those search queries were applied to the number of search engines. A manual skimming of the publications was performed to identify whether or not the paper is a matter of interest to this work. The first step resulted in 35 articles. The second step was performed on the resulting 35 papers in order to know whether they should be included or not in the study. The main question to be addressed was, "Does the methodology combine the metadata with the DL model in the training and creation step?". Only 7 papers could successfully pass the second step and those are the ones included in the research.

Previous literature used five types of metadata: gender, age, MMSE, LM, Education. The first and second studies show that combining MRI images with the adjacent metadata, age, gender, and MMSE, shows a slight improvement in performance metrics. The third and fourth studies showed similar conclusions considering age, gender, and education metadata. Interestingly, the fifth study shows a slight increase in performance only in the case of early AD detecting. In particular, the usage of metadata increases recalls scores in early diagnosis of AD, which is helpful for early treatment. In the 6th one, using MMSE and LM combined with MRI images, the accuracy, and recall scores increase. On the other hand, the 7th study shows that age and gender suffer from information reaches for diagnosis purposes. This could be explained because the metadata is highly biased to the population. Despite the fact that metadata could be a useful improvement of the DL model, it suffers from several drawbacks.

Previous literature used different sample sizes from the main dataset. This leads to a lack of a uniform pattern in the pipeline, which means the possibility of making a fair comparison is low. Noticeably, the main problem faced by most of the researchers was the imbalance in the dataset. This could lead to rapidly getting stuck in the local minimum. An additional limitation is the bias of the model. A model with a strong bias won't match the data set closely, while a model with low bias will match the data set very well. Another problem that causes bias is that the data was sampled from a specific population with specific characteristics. This means the ability to generalize the results on all humankind is low. Another limitation is that there is no standardized way to select the biomarkers and the metadata that truly reflects the existence of AD. Therefore, more studies should compare the effect of the population and the generalization possibility.

CONCLUSION

This work aims to determine whether or not the metadata is contributing to enhancing the performance of CAD systems for AD.

The work reviewed seven papers from the literature. It was found that there is very little research on this topic even though it has potential. The results of this work showed that a slight increase in performance could be detected. However, it still suffers from numerous problems due to the lack of balance in the dataset. Moreover, there is no noticeable uniform approach in the literature.

It is recommended to conduct further experiments to be able to successfully answer the question. The experiments should have a common approach and aim to increase the generalizability of the methods. Moreover, new datasets could be added in order to have a uniform sample from the society and to distinguish between local society and global society research.

The contribution of this work is summarizing the literature that aimed to combine metadata with MRI scans. It gives an overview of the research field and states the potential of such methods.

REFERENCES

[1]   Fine, C. (2008). The Britannica Guide to The Brain: A guided tour of the brain-mind, memory, and intelligence. Encyclopaedia Britannica.

[2]   Gauthier, S., Rosa-Neto, P., Morais, J. A., & Webster, C. (2021). World Alzheimer Report 2021: Journey through the diagnosis of dementia. Alzheimer's Disease International.

[3]   World Health Organization. (2021). Dementia. Retrieved December 14, 2021, from https://www.who.int/en/news-room/fact-sheets/detail/dementia.

[4]   Alzheimer's association. (2021) What is dementia? Alzheimer's Disease and Dementia. Retrieved December 14, 2021, from https://www.alz.org/alzheimers-dementia/what-is-dementia#symptoms

[5]   Livingstone, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., & Banejee, S. (2020). The Lancet Commissions, Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet, 396, 413-446.

[6]   Alzheimer's association. (2019). Lifestyle Interventions may offset elevated Alzheimer's risk due to genetics, environment. Retrieved December 15, 2021, from https://www.alz.org/aaic/releases_2019/sunLIFESTYLE-jul14.asp

[7]   Liu, P. P., Xie, Y., Meng, X. Y., & Kang, J. S. (2019). History and progress of hypotheses and clinical trials for Alzheimer's disease. Signal transduction and targeted therapy, 4(1), 1-22.

[8]   Hosseini-Asl, E., Keynton, R., & El-Baz, A. (2016, September). Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In 2016 IEEE international conference on image processing (ICIP) (pp. 126-130). IEEE.

[9]   Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., & Alzheimer's Disease Neuroimaging Initiative. (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage, 104, 398-412.

[10]   Weller, J., & Budson, A. (2018). Current understanding of Alzheimer's disease diagnosis and treatment. F1000Research, 7.

[11]   Lane, Christopher A.; Hardy, John; Schott, Jonathan M. (2017). Alzheimer's disease. European Journal of Neurology, –. doi:10.1111/ene.13439.

[12]   Alzheimer's association. (2021). Stages of Alzheimer's. Alzheimer's Disease and Dementia. Retrieved December 20, 2021, from https://www.alz.org/alzheimers-dementia/stages

[13]   Alzheimer's association. (2021). Mild cognitive impairment (MCI). Alzheimer's Disease and Dementia. Retrieved December 20, 2021, from https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment

[14]   Mitchell, A. J., & Shiri-Feshki, M. (2009). Rate of progression of mild cognitive impairment to dementia–meta-analysis of 41 robust inception cohort studies. Acta psychiatrica scandinavica, 119(4), 252-265.

[15]   Alzheimer's association. (2021). Causes and risk factors for Alzheimer's disease. Alzheimer's Disease and Dementia. Retrieved December 20, 2021, from https://www.alz.org/alzheimers-dementia/what-is-alzheimers/causes-and-risk-factors

[16]   Alzheimer's Disease Neuroimaging Initiative. (2017). Study design. Retrieved February 9, 2022, from http://adni.loni.usc.edu/study-design/#background-container

[17]   Alzheimer's association. (2021). What is Alzheimer's? Alzheimer's Disease and Dementia. Retrieved February 9, 2022, from https://www.alz.org/alzheimers-dementia/what-is-alzheimers

[18]   Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., ... & Winblad, B. (2001). Current concepts in mild cognitive impairment. Archives of neurology, 58(12), 1985-1992.

[19]   Grossberg, G. T. (2003). Diagnosis and treatment of Alzheimer's disease. Journal of Clinical Psychiatry, 64, 3-6.

[20]   Blennow K, de Leon MJ, Zetterberg H. Alzheimer's disease. Lancet. 2006 Jul 29;368(9533):387-403. doi: 10.1016/S0140-6736(06)69113-7. PMID: 16876668.

[21]   Hardy, J., & Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. science, 297(5580), 353-356.

[22]   Raber, J., Huang, Y., & Ashford, J. W. (2004). ApoE genotype accounts for the vast majority of AD risk and AD pathology. Neurobiology of aging, 25(5), 641-650.

[23]   Grundke-Iqbal, I., Iqbal, K., Tung, Y. C., Quinlan, M., Wisniewski, H. M., & Binder, L. I. (1986). Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. Proceedings of the National Academy of Sciences, 83(13), 4913-4917.

[24]   Iqbal, K., Alonso, A. D. C., Chen, S., Chohan, M. O., El-Akkad, E., Gong, C. X., ... & Grundke-Iqbal, I. (2005). Tau pathology in Alzheimer disease and other tauopathies. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1739(2-3), 198-210.

[25]   Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., & Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta neuropathologica, 112(4), 389-404.

[26]   Zhang, F., Zhong, R. J., Cheng, C., Li, S., & Le, W. D. (2021). New therapeutics beyond amyloid-β and tau for the treatment of Alzheimer's disease. Acta Pharmacologica Sinica, 42(9), 1382-1389.

[27]   Ballard, C., Khan, Z., Clack, H., & Corbett, A. (2011). Nonpharmacological treatment of Alzheimer disease. The Canadian Journal of Psychiatry, 56(10), 589-595.

[28]   Hort, J. O. B. J., O'brien, J. T., Gainotti, G., Pirttila, T., Popescu, B. O., Rektorová, I., ... & EFNS Scientist Panel on Dementia. (2010). EFNS guidelines for the diagnosis and management of Alzheimer's disease. European Journal of Neurology, 17(10), 1236-1248.

[29]   Creager, M., Loscalzo, J., & Beckman, J. A. (2012). Vascular medicine E-book: A companion to Braunwald's heart disease. Elsevier Health Sciences.

[30]   Atlas, S. W. (2016). Magnetic Resonance Imaging of the Brain and Spine. United States: Wolters Kluwer Health.

[31]   Wang, Q. (2019). Hardware of MRI System. In Nuclear Magnetic Resonance. IntechOpen.

[32]   Grover, V. P., Tognarelli, J. M., Crossey, M. M., Cox, I. J., Taylor-Robinson, S. D., & McPhail, M. J. (2015). Magnetic resonance imaging: principles and techniques: lessons for clinicians. Journal of clinical and experimental hepatology, 5(3), 246-255.

[33]   Constantinides, C. (2016). Magnetic resonance imaging: the basics. CRC press.

[34]   Chollet, F. (2018). Keras: The python deep learning library. Astrophysics source code library, ascl-1806.

[35]   Turing, I. B. A. (2007). Computing machinery and intelligence-AM Turing. Mind, 59(236), 433.

[36]   Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford university press.

[37]   Stork, D. G., Duda, R. O., Hart, P. E., & Stork, D. (2001). Pattern classification. A Wiley-Interscience Publication.

[38]   Deutsch, S. (1981). A simplified version of Kunihiko Fukushima's neocognitron. Biological Cybernetics, 42(1), 17-21.

[39]   LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4), 541-551.

[40]   Tajbakhsh, N., Gotway, M. B., & Liang, J. (2015, October). Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 62-69). Springer, Cham.

[41]   Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., & Shen, D. (2015). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage, 108, 214-224.

[42]   Salehi, A. W., Baglat, P., Sharma, B. B., Gupta, G., & Upadhya, A. (2020, September). A CNN model: earlier diagnosis and classification of Alzheimer disease using MRI. In 2020 International Conference on Smart Electronics and Communication (ICOSEC) (pp. 156-161). IEEE.

[43]   Moen, U. (2018). Early detection of Alzheimer's Disease using 3D texture features and 3D convolutional neural networks from structural MRI (Master's thesis, University of Stavanger, Norway).

[44]   Smith, N. A., Sinden, D., Thomas, S. A., Romanchikova, M., Talbott, J. E., & Adeogun, M. (2020). Building confidence in digital health through metrology. The British Journal of Radiology, 93(1109), 20190574.

[45]   Wolz, R., Aljabar, P., Hajnal, J. V., Lötjönen, J., & Rueckert, D. (2011, March). Manifold learning combining imaging with non-imaging information. In 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (pp. 1637-1640). IEEE.

[46] Neu, S., Crawford, K., & Toga, A. W. (2012). Practical management of heterogeneous neuroimaging metadata by global neuroimaging data repositories. Frontiers in neuroinformatics, 6, 8.

[47] Thomas, S. A. (2021, November). Combining Image Features and Patient Metadata to Enhance Transfer Learning. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 2660-2663). IEEE.

[48] Qiu, S., Joshi, P. S., Miller, M. I., Xue, C., Zhou, X., Karjadi, C., ... & Kolachalama, V. B. (2020). Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain, 143(6), 1920-1933.

[49] Bae, J. B., Lee, S., Jung, W., Park, S., Kim, W., Oh, H., ... & Kim, K. W. (2020). Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. Scientific Reports, 10(1), 1-10.

[50] Liu, M., Zhang, J., Adeli, E., & Shen, D. (2018). Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. IEEE Transactions on Biomedical Engineering, 66(5), 1195-1206.

[51] Esmaeilzadeh, S., Belivanis, D. I., Pohl, K. M., & Adeli, E. (2018, September). End-to-end Alzheimer's disease diagnosis and biomarker identification. In International Workshop on Machine Learning in Medical Imaging (pp. 337-345). Springer, Cham.

[52] Puente-Castro, A., Fernandez-Blanco, E., Pazos, A., & Munteanu, C. R. (2020). Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques. Computers in Biology and Medicine, 120, 103764.

[53] Qiu, S., Chang, G. H., Panagia, M., Gopal, D. M., Au, R., & Kolachalama, V. B. (2018). Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 10, 737-749.

[54] Saratxaga, C. L., Moya, I., Picón, A., Acosta, M., Moreno-Fernandez-de-Leceta, A., Garrote, E., & Bereciartua-Perez, A. (2021). MRI Deep Learning-Based Solution for Alzheimer's Disease Prediction. Journal of personalized medicine, 11(9), 902.