CATÓLICA
FACULDADE DE CIÊNCIAS DA SAÚDE E ENFERMAGEM
LISBOA·PORTO

Ispa
Instituto Universitário

MACHINE LEARNING APPROACHES WITH COGNITIVE DATA FOR EARLY DETECTION OF DEMENTIA: A SCOPING REVIEW

Dissertação apresentada à Universidade Católica Portuguesa para obtenção do grau de mestre em Neuropsicologia

Por

Raquel Gonçalves Colaço

Lisboa, 2023

CATÓLICA
FACULDADE DE CIÊNCIAS DA SAÚDE E ENFERMAGEM
LISBOA · PORTO

Ispa
Instituto Universitário

MACHINE LEARNING APPROACHES WITH COGNITIVE DATA FOR EARLY DETECTION OF DEMENTIA: A SCOPING REVIEW

Dissertação apresentada à Universidade Católica Portuguesa para obtenção do grau de mestre em Neuropsicologia

Por
Raquel Gonçalves Colaço

Sob a orientação de Profª. Doutora Maria Vânia Silva Nunes e Prof. Doutor Spyros Angelopoulos

Lisboa, 2023

Para o meu Pai.

"Aproveitem a vida e ajudem-se uns aos outros. Apreciem cada momento. Agradeçam e não deixem nada por dizer, nada por fazer."

- António Feio

Acknowledgments

I would like to, first, thank my supervisor Prof. Dra. Maria Vânia Nunes for trusting me with this work and for guiding me through the process, always being available to answer my questions and helping me organize my time to reach my goals. I would also like to thank Dr. Spyros Angelopoulos for accepting to be my external supervisor in this dissertation without even knowing me, trusting that I was able to develop this work. For always being available and making the time to respond to my questions in record time and never leaving me alone with doubts, Spyros surely taught me a lot more than just AI technology, but also about being kind and helping others through our own knowledge; "*with great power comes great responsibility*" has a whole new meaning for me now.

I would also like to thank my mom that as the strength of more than just one women, a strength that is enough for her and also for me when mine is lacking, that was by my side during this process and without her support none of this would have been possible. I would also like to thank my sister for having the patience to hear me talk about this work even though it is nothing related to her area and for caring about my progression. I would also like to thank my brother-in-law for always being willing to help and giving me answers for technical questions that otherwise I would not be able to find.

Last but not least, I want to thank my partner, the person who was with me since day one always caring and always listening, believing in me especially in moments where I struggled believing in myself.

All of the above were crucial parts of this dissertation, because without them I would not have been able to complete this work, that was without a doubt a major challenge of my life so far.

Table of Contents

Abbreviations

AD – Alzheimer’s Disease

ADAS-Cog – Alzheimer’s Disease Assessment Scale–Cognitive Subscale

AI – Artificial Intelligence

aMCI – Amnestic Mild Cognitive Impairment

ANN – Artificial Neural Networks

CAD – Computer-Aided-Diagnosis

CAIDE – Cardiovascular Risk Factors, Aging, and Incidence of Dementia

CN – Cognitively Normal

CDT – Clock Drawing Test

CNN – Convolutional Neural Networks

CT – Computerized Tomography

DASH – Dietary Approach to Stop Hypertension

DL – Deep Learning

DNN – Deep Neural Network

DSI – Disease State Index

EEG – Electroencephalogram

EHRs – Electronic Health Records

HATICE – Healthy Ageing Through Internet Counselling in the Elderly

HDL-cholesterol – High-Density Lipoprotein Cholesterol

HMM – Hidden Markov Model

KK – Kirtan Kriya

LIBRA – Lifestyle for BRAin Health

MCI – Mild Cognitive Impairment

MIND - Mediterranean-DASH Diet Intervention for Neurodegeneration Delay

ML – Machine Learning

MMSE – Mini Mental State Examination

MoCA – Montreal Cognitive Assessment

MRI – Magnetic Resonance Image

NC – Normal Control

NLP – Natural Language Processing

pMCI – Progressive Mild Cognitive Impairment

PUFA – Polyunsaturated Fatty Acid

RNN – Recurrent Neural Networks

rs-fMRI – Resting State functional Magnetic Resonance Image

sMCI – Stable Mild Cognitive Impairment

sMRI - Structural Magnetic Resonance Image

SVM – Support Vector Machine

WNN – Wide Neural Network

Abstract

Objective: The objective of this scoping review is to map the literature on the types of data commonly used in AI models for the early detection of Alzheimer's Disease (AD). Particularly, we will look for information about the extent to which cognitive data is considered in this type of research.

Introduction: Alzheimer's Disease is the most common form of dementia and the seventh leading cause of death worldwide. Dementia has massive individual, societal and economic impact and because there is no approved cure or modifying therapies available yet, the investigation should focus on early detection of AD and dementia prevention.

Inclusion Criteria: Studies from 2017 until March 2023 were included. The main inclusion criteria included studies that focused their investigation specifically on Alzheimer's disease and aimed to develop/train AI models to assist on early-detection of the disease and on the prevention of dementia.

Methods: This scoping review was developed based on the JBI methodology for scoping reviews (Peters et al., 2020) and conducted on Medline Complete via EBSCO, PubMed and Web of Science on March 2023. Articles included in this review were screened by two independent reviewers and after being analysed, the relevant information was retrieved.

Results: All articles have taken into account either cognitively normal (CN) individuals, mild cognitive impairment (MCI) individuals or Alzheimer's disease (AD) individuals, and trained models based on machine learning, the most common type of data used to train models was brain MRI, cognitive data is mainly used to characterize the cognitive status of the participants (to form the experimental and control groups) and the majority of these investigations are published in engineering journals.

Conclusions: Neuroimage is the type of data where investigation in this area is more focused on, and there is a need to include novel types of data that are also relevant to the early detection of Alzheimer's Disease, namely cognitive data. It is also important to make this type of information more easily available for psychologists as well as educate them on this subject since such technology can bring great benefits to their practice.

Keywords: machine learning, early detection of AD, prevention of dementia, cognitive data, neuroimage

Summary

Objetivo: O objetivo desta revisão *scoping* recai sobre mapear a literatura no sentido de encontrar informação acerca do tipo de dados que são mais comumente utilizados na criação dos modelos de inteligência artificial para a deteção precoce da doença de Alzheimer (DA) e prevenção da demência. Particularmente procuraremos informação acerca de em que medida são os dados cognitivos tidos em consideração neste tipo de investigação

Introdução: A doença de Alzheimer é a sétima principal causa de morte em todo o mundo e a forma mais comum de demência. A demência tem um grande impacto tanto individual, como social e económico, e porque atualmente não existem ainda terapias modificáveis ou mesmo uma cura para a mesma, a investigação nesta área deve focar-se no objetivo da deteção precoce da DA e prevenção da demência.

Critérios de Inclusão: Artigos de 2017 a março de 2023 foram incluídos. Os principais critérios de inclusão passaram por incluir estudos que focaram a sua investigação especificamente na doença de Alzheimer e visavam desenvolver/treinar modelos de inteligência artificial (IA) que auxiliassem a deteção precoce da doença de Alzheimer, bem como a prevenção da demência.

Método: Esta revisão *scoping* foi desenvolvida tendo por base a metodologia da JBI para revisões *scoping* (Peter set al., 2020) e levada a cabo na *Medline Complete via EBSCO*, *PubMed* e *Web of Science* em março de 2023. Artigos incluídos nesta revisão foram revistos por dois revisores independentes e posteriormente analisados detalhadamente na sua íntegra, tendo sido retiradas as informações mais relevantes.

Resultados: Foi verificado que todos os artigos tiveram em consideração indivíduos cognitivamente normativos (CN), com défice cognitivo ligeiro (DCL) ou doença de Alzheimer (DA), todos criaram ou treinaram modelos baseados em aprendizagem automática. O tipo de dados mais commummente utilizados para estas técnicas são os dados de neuroimagem, dados cognitivos foram maioritariamente utilizados para caracterizar o estado cognitivo dos participantes (para formar os grupos experimental e de controlo) e a maioria destas investigações foram publicadas em revistas científicas da área das engenharias.

Conclusão: A neuroimagem é o tipo de dados em que a investigação nesta área está mais focada, podendo haver necessidade de incluir novos tipos de dados que também sejam relevantes para a deteção precoce da DA e prevenção da demência, nomeadamente dados cognitivos. Também é importante tornar este tipo de informação mais facilmente disponível para os psicólogos, bem como educá-los sobre o assunto, uma vez que esta tecnologia pode trazer grandes benefícios para a sua prática.

Palavras-chave: aprendizagem automática, deteção precoce da DA, prevenção da demência, dados cognitivos, neuroimagem

Section I

Introduction

According to WHO (2022) dementia can be described as a syndrome with deterioration in cognitive function that is not a direct consequence of aging. Dementia is not defined as one concept, as there are different types that have different aetiology (Small and Jarvik, 1982). The most common forms of dementia include vascular dementia, dementia with Lewy bodies, frontotemporal dementia, and Alzheimer's Disease (AD); AD is the most common type, representing about 60-70% off all cases (WHO, 2022).

As the lifespan of individuals increases, we face new problems associated with old age, like the growth of dementia cases that are associated with various consequences (Dubois et al., 2016). It is estimated that more than 55 million people live with dementia around the world and that 10 million cases are added to this number every year (WHO, 2022). Brookmeyer et al. (2007) explain that by 2050 the prevalence of AD will quadruple, and predict that if interventions can delay its onset and progression for just 1 year, the number of cases by 2050 would decrease to 9.2 million.

Dementia has just as much individual impact, for the person itself and the caregiver, as it has for the society. The estimated values on the cost of dementia were expected to go up to 1 trillion US$ worldwide in 2018 (DeTure & Dickson, 2019), and AD, being the seventh leading cause of death (WHO, 2022), also taking into account the loss of function that it leads to in its more advanced stage (Jack et al., 2010), becomes crucial to continuing on developing investigation about it.

As we have limited treatment options (Danso et al., 2019), it becomes very relevant to direct the research to early diagnosis and prevention, and minimize the level of risk that a person is in by developing early interventions to slow down its progression. Therefore, the main goal is to detect this disease early and start an intervention to help slow down its progression. To support this, some authors have pointed out the fact that when a person is "finally" diagnosed with AD it is already too late (Alzheimer Europe, 2021), as in the early stages it is possible to see the neuropathological changes with no cognitive impairment, and when the patient starts noticing the cognitive impairment this is due to the neuropathological changes development.

Early diagnosis is important for patients to increase their quality of life, as well as for clinicians to make treatment decisions (Khan and Usman, 2015), but it goes further on importance, taking into account the massive costs that dementia has on society, achieving an early diagnosis can also represent an important step for health economists (*ibid*). The prediction and prevention of this disease with early interventions would allow the slowing down of the progression and also increase the quality of life of the patients as well as decrease the costs that dementia has on society (Brookmeyer et al., 2007).

To achieve this goal, we regard artificial intelligence (AI) techniques as a supporting tool rather than replacement for humans (Dasta et al., 1992). Aiming to be an aid to the diagnosis, helping clinicians to make earlier and more effective decisions (James et al., 2021) and also giving clinicians the opportunity to develop early interventions strategies personalized for each patient, taking into account not just the stage of the disease they are in but also at what level of risk they are in terms of developing a dementia.

This work is organized in the following way: section I presents the introduction to this project and a summarized explanation of the main topics. Section II is the theoretical framework, where these main topics are extensively developed and research questions presented. Section III refers to the methodology this dissertation followed. Section IV presents the results found following this scoping review. Sections V and VI are in reference to the discussion, where the main conclusions are presented and discussed in the context of the research questions and conclusion where the summarized main conclusions and limitations to this dissertation are presented and future research in this topic is proposed.

Section II

Theoretical Framework

2.1 Description of Alzheimer's disease

AD is a progressive neurodegenerative disease (Dubois et al., 2016) that has both clinical and pathological characteristics. Its diagnosis is verified by the presence of amyloid plaques that consist of beta amyloid deposits, microglial cells, dystrophic neurites, and bundles of astrocytic processes (Nagele et al., 2004), and neurofibrillary tangles, that are accumulations of abnormal tau filaments (Nagele et al., 2004). This disease can have a variety of aetiologies, and although there is the guidelines of NINCDS-ADRDA (Mckann et al., 1984) to make a diagnosis, the evidence found in literature about its characteristics, both clinical and neuroanatomical, can be very heterogeneous.

The clinical AD stages can be divided into three phases (Jack et al., 2010). Firstly there is the preclinical (DeTure & Dickson, 2019) or pre-symptomatic (Jack et al., 2010) phase of the disease, where the patient is believed to experience an asymptomatic phase, that is, the individual is cognitively normal (DeTure & Dickson, 2019) at the same time that some pathological changes are already occurring (Dubois et al., 2016), taking into account that the assumption is that an asymptomatic individual with pathological changes that make clinicians believe being indicative of AD, will become symptomatic eventually (Jack et al., 2010). The second phase is the prodromal phase, that commonly can be referred as mild cognitive impairment (MCI), where the main characteristic is the progressive cognitive decline (typically deficits in episodic memory) that does not meet the criteria for dementia yet, before the onset of functional impairment (Dubois et al., 2016; Jack et al., 2010). The severity of cognitive impairment in the prodromal phase can vary from the earliest appearance of memory dysfunction to more generalized dysfunction in other cognitive functions (Jack et al., 2010). As the disease is progressing, it is likely to observe some of the cognitive dysfunction in the form of clinical symptoms. When in an early stage, the clinical symptoms are, typically, memory impairment and some cognitive decline (DeTure & Dickson, 2019); in more advanced stages it can also affect behaviour, speech, visuospatial orientation and the motor system (DeTure & Dickson, 2019); and in the latest phase, the evolution of AD is dementia, characterized by deficits in multiple domains that produce enough impairment to lead to loss of function (Jack et al., 2010). Some studies further emphasize specific changes in the medial temporal cortex, that is involved in episodic memory impairment (that is representative of an early-stage clinical symptom), while some AD models show evidence of amyloid plaque deposition specific in hippocampal regions (Shin et al., 2008).

Currently, only around 5% of AD cases allow an early detection (i.e., early-onset AD) as they occur in patients under 65 years old, and from these, less than half of the patients carry a causal mutation that manifests as an inheritance pattern (known as early-onset familial AD). Late-onset AD (i.e., cases that occur in patients over 65 years old) represents about 95% of all cases. Hereupon, AD can be divided into familial and sporadic cases, where the sporadic (normally late-onset AD) represents more than 90% of all cases (Mishra & Li, 2020). With such different forms of developing AD, the process of diagnosing it needs to be very structured and detailed, and rich in information about the patient. An assessment aiming a diagnose (in both sporadic and familial cases), should be done according to strict guidelines, so it can provide all the information that is necessary to make a decision.

2.1.1 Clinical and pathological diagnostic of Alzheimer's disease

The diagnosis of AD can be done in different ways, as there is not yet a single diagnostic test (Alzheimer's Association, 2022). To make it possible, the National Institute of Neurological and Communicative Disorders and Stroke and the AD and Related Disorders Association (NINCDS-ADRDA) developed a report on clinical diagnosis of AD (McKhann et al., 1984), and to this day, the clinical diagnosis takes into account relevant information like medical history, clinical examination, neuropsychological testing and laboratory assessments. One of the most important criteria established by the Mini Mental State Examination (MMSE), which allows to estimate the severity of cognitive loss (Rodrigues et al., 2021), although it is just a screening test and does not have a diagnostic purpose. There is, thus, still a long way until the diagnosis of AD is established, while the diagnosis is not definite until a post-mortem neuropathologic evaluation (DeTure & Dickson, 2019) and confirmation of specific neuropathologic changes (Dubois et al., 2016; Shin et al., 2008).

The criteria for clinical diagnosis of AD developed by NINCDS-ADRDA (McKhann et al., 1984) is divided in three different forms of diagnosis: Probable AD , Possible AD , and Definite AD (McKhann et al., 1984). To confirm *probable AD,* there is a need to have clinical onset supported by neuropsychological tests, observe deficits in two or more areas of cognition, worsening of memory and other cognitive functions, to perform an onset inside the limitations of age and absence of other systemic disorders or brain diseases that could be responsible for the deficits in memory and cognition (*ibid*). This criterion should be supported by progressive deterioration of specific cognitive functions, impaired activities of daily life and alteration of patterns of behaviour, as well as laboratory tests like lumbar punction (normal), electroencephalogram (EEG) (normal pattern or nonspecific changes like slowed-wave activity) and computerized tomography (CT) (with observable cerebral atrophy with progression). A *possible AD* diagnostic should be made on the basis of dementia syndrome, if other neurologic, psychiatric or systemic disorders are absent in the sense of being insufficient to cause dementia, taking into account that can be other comorbidities (i.e., systemic or brain disorder), insufficient to produce dementia and not considered to be the actual cause of the dementia. Also, it is important that cases with identifiable gradually progressive severe cognitive deficit in the absence of other identifiable cause are used in research (McKhann et al., 1984). Lastly, the criteria to diagnose *definite AD* are the same ones described for probable AD plus histopathological evidence obtain rather from biopsy or autopsy (McKhann et al., 1984).

For the diagnosis of AD , thus, we may need to check medical history, resort on brain imaging, do a physical exam and diagnostic test, test the cerebrospinal fluid, do a neurological exam, do some blood tests, and also perform some mental cognitive status tests to the patient (Alzheimer's Association, 2022). All these tests and exams are likely to be time consuming until we have a final and trustworthy diagnosis. Hereupon, the diagnosis can come "too late" and when it is final, the disease is already too advanced in terms of progression (Alzheimer Europe, 2021).

2.1.2 Genetic component of Alzheimer's disease

AD can also have a genetic aetiology. The pathogenesis of AD is still unclear, and several authors have been conducting a lot of research in that topic, explaining that can be several significant factors that contribute to the development of the disease, not just environmental but also genetic (Mishra & Li, 2020). Heritability of AD can be very heterogenous in its individuality, as the [familial] late-onset from of AD represents 70 to 80% of cases and the heritability of [familial] early-onset AD can go up to 92 to 100%

(Ayodele et al., 2021). Even though AD hereditary (i.e., autosomal dominant) cases are relatively rare, representing approximately 2,5% of all cases (Mishra & Li, 2020), genetic factors can contribute up to 70% to its aetiology (Mishra & Li, 2020).

When clinicians are presented with an autosomal dominant case of AD case, the clinical and behavioural symptoms are observable sooner, since the first symptoms are noticeable from as early as in their late 20s to their early 60s, as the prevalence keeps increasing with age (Ayodele et al., 2021), allowing an early diagnosis (Mishra & Li, 2020). This diagnosis is only possible with the help of clinical criteria from NINCDS-ADRDA report (McKhann et al., 1984) and the verification of the presence of the apolipoprotein E gene (*APOE*) alleles that can represent a risk factor for late-onset [familial] AD (Mishra & Li, 2020).

Because of all the time consuming exams necessary to receive a diagnosis, added to the progressive component of AD and also the fact that there are still no disease modifying therapies (DeTure & Dickson, 2019), it is important to focus the investigation on achieving the possibility of an early diagnosis, as this would allow us to start an intervention in an earlier stage of progression of the disease and have the opportunity of improving the quality of life of patients and their informal caregivers, both of whom are often relieved once the patient is diagnosed (Dubois et al., 2016).

AI-based technology, as an aid to clinicians on a decision-making process mid diagnosis, can help to decrease the time consumed in the process, and achieve an early diagnosis, so much needed in this specific neuropathology.

2.2 Prevention of Alzheimer's disease

Preventive intervention gives the best results if introduced before the first symptoms of AD. That is, when the nutritional status, number of synapses, cognition, and neuropathological changes in the nervous system compensate each other, which increases the chances of staying healthy for a longer period (Sliwinska and Jeziorek, 2021). Because of this, understanding the prevention of AD has a major presence in the literature.

Prevention of AD can be separated in two concepts: Primary prevention strategies and Secondary prevention strategies (Crous-Bou et al., 2017). Primary prevention strategies are based on identifying modifiable risk factors and risk reduction (Crous-Bou et al., 2017). According to Zhang et al. (2021) these modifiable risk factors are: type 2 diabetes; large blood pressure variability; dyslipdemia; midlife obesity; cardiovascular diseases; traumatic brain injury; hyperhomocysteinemia; hearing loss; oral diseases; depression and stress; sleep disturbances; smoking; air pollution and anticholinergic medication. Secondary prevention initiatives are based on the early detection of the pathophysiological hallmarks and intervention at the preclinical phase, as there is a long asymptomatic stage in AD and these authors argue that this time window is ideal to study the implementation of these secondary prevention programmes (Crous-Bou et al., 2017). The implementation of these strategies might offer the best opportunity to start an intervention in earlier stages of the disease as well as delaying the cognitive decline (Crous-Bou et al., 2017).

Pre-existing diseases, unhealthy lifestyles and environmental exposures can represent risk factors to the development of AD, but some factors that include psychosocial conditions and healthy lifestyle might be protective regarding AD (Zhang et al., 2021). According to Zhang et al. (2021) protective factors are: high education; cognitive activity; bilingualism; social engagement; marriage; physical activity; moderate alcohol intake; moderate coffee and tea intake; Mediterranean diet; MIND (Mediterranean-DASH diet intervention for neurodegeneration delay) diet and DASH (dietary approach to stop hypertension) diet; Ketogentic diet; vitamins and flayonoid; PUFA (polyunsaturated fatty acid) and fish intake and HDL-cholesterol (high-density lipoprotein cholesterol).

Zhang et al. (2021) propose some prevention strategies that could have a positive impact on the prevention of AD. On the sight of increasing and maintain cognitive reserve the authors propose strategies like receiving more education in early life and learning new languages as well as maintain social engagement. On reducing neuropathological damage, they propose the close management of blood pressure among other things and the maintenance of good conditions of cardiovascular health. The authors also propose other strategies based on the decrease of risk factors like maintaining a good mental health and relieving stress appropriately, to get regular physical exercise, high quality sleep, to stop smoking and limit alcohol use among other things. The authors also argue that it is imperative to increase the cognitive reserve mainly via enhancing education attainment and promoting social contact. Also, good conditions of body health and healthy lifestyles as well as reducing environmental exposures might be favourable to reduce the neuropathological damage for AD prevention (Zhang et al., 2021).

As primary prevention strategies, Crous-Bou et al. (2017) give some examples of studies that were made. The FINGER (Finish Geriatric Intervention Study to Prevent

Cognitive Impairment and Disability) study aimed to investigate if a multidomain intervention could prevent cognitive decline among older adults and also what effect would this multidomain intervention have on disability, quality of life, depressive symptoms, the use of health care services and vascular risk factors. So far, results from this study have demonstrated that a multidomain intervention including diet, exercise, cognitive training and monitoring vascular risk can improve/maintain cognitive functioning in older people (60–77 years old) from the general population at risk of dementia (Crous-Bou et al., 2017). The PreDIVA that is an investigation that tested whether a multicomponent intervention targeting vascular risk factors could prevent new cases of dementia, and results showed that this multidomain intervention focused on vascular care did not result in a reduced incidence of all-cause dementia. However, these results do not rule out the potential benefit of a better management of cardiovascular risk factors on brain health (Crous-Bou et al., 2017). In fact, there are some risk factors score developed to predict a development a dementia, like the CAIDE Risk Score that stands for "Cardiovascular Risk Factors, Aging, and Incidence of Dementia", a measure developed to predict the risk of developing a dementia in 20 years (for a middle-aged population) based on cardiovascular risk factors (James et al., 2021).

Nutrition is also an important factor for the prevention of AD and to the promotion of healthier lifestyles, also protective to the development of dementia and dietary intervention should be introduced as early as possible in order to minimize, as much as possible, the risk of developing dementia. Sliwinska and Jeziorek (2021) explain that has been proven that dietary habits, which lead to the development of cardiovascular and metabolic diseases, significantly increase the risk of dementia. On the other hand, a Mediterranean diet (rich in antioxidants, fiber, and omega-3 polyunsaturated fatty acids) may have a protective effect on the neurodegenerative process. The beneficial effect of many nutrients (like glutathione, polyphenols, curcumin, coenzyme Q10, vitamins B6, B12, folic acid, unsaturated fatty acids, lecithin, UA, caffeine, and some probiotic bacteria) on the course of AD has been demonstrated. So, these authors conclude that the Mediterranean and DASH diets have been documented to protect against AD, the MIND diet is reported to be much more effective in preventing cognitive decline than either the Mediterranean or DASH diets alone.

In the context of relieving stress (a risk factor for dementia) meditation can play a part on cognition and well-being and offer the possibility to reduce of neurodegeneration and prevention of AD (Khalsa, 2015). Kirtan Kriya (KK) is a meditation technique that had successfully employed to improve memory in studies of people with subjective cognitive decline, MCI, and highly stressed caregivers, all of whom are at increased risk for subsequent development of AD (Khalsa, 2015). This technique also improves psycho-spiritual well-being that is important for maintenance of cognitive function and prevention of AD (Khalsa, 2015). Khalsa (2015) proposes that relying on meditation techniques along with dietary modification, physical exercise, mental stimulation, and socialization, may be beneficial as part of AD prevention programs.

2.3 Artificial intelligence description

The "father" of AI, Alan Turing, has defined this discipline as "the science and engineering of making machines, especially intelligent computer programs" (Azzi et al., 2020; Turing, 2009) and the beginning of AI can be associated with philosophy, possibilities, demonstrations, fantasies, and imagination (Azzi et al., 2020; Buchanan, 2005). The AI field emerged as an attempt to respond to several needs that appear simultaneously, as well as opportunities and interests (Azzi et al., 2020). Patel et al. (2021) assert that AI can be helpful to analyse medical data as well as to the diagnosis itself, as there are many other purposes that it can serve, like to monitor patients and to help clinicians to deal with lots of data in a more efficient way (Patel et al., 2020). This becomes very relevant once these novelties are under-used by clinicians, including neuropsychologists, mainly because the fact that this very innovative technology is available, is unknown by the target audience (i.e., clinicians).

AI is a comprehensive term that covers two other subcategories: machine learning (ML) and deep learning (DL). **AI** can be defined as "algorithms capable of demonstrating cognitive functions associated with the human mind, such as "learning", "problem solving", "adaptation", "logic", "if-then" rules, and/or decision trees" (Patel et al., 2021). **ML** represents a variety of tools to make statistical, probabilistic decisions based on previous learning, and can be defined as "a form of intelligence based on compilation of complex algorithms and software that mimics the human mind to decipher critical problems that include visual perception, speech recognition, and decision-making on exposure of more data over time" (Patel et al., 2021). Particularly in healthcare, there is a specific ML technique that aims to achieve the possibility of taking advantage of the most information possible that presents itself as relevant to the patients care, the **natural language processing (NLP)** method, that consists in turning texts (like clinical notes and medical journals) to "machine-readable structured data", to turn the information into data capable of being read and analysed by ML techniques (Jiang et al., 2017). **DL** is a subcategory of ML and is defined as "a class of artificial neural networks that learns in a supervised and unsupervised manner" (Patel et al., 2021). So, the concept of AI is narrowing down to specific concepts, all based on decision making algorithms.

ML can have three types of learning algorithms: **supervised learning,** where lots of data is provided and the program tries to learn it and learns how to draw the input to the required output; **unsupervised learning**, where the unsupervised learning algorithms employ self-learning based on unclassified and unlabelled data and **reinforcement learning**. The most common ways for algorithms to learn in AD are through supervised learning (Khan and Usman, 2015).

2.3.1 Evolution of AI techniques applied to healthcare

Dasta et al. (1992) conducted a literature review on computer technology, emphasizing institutional-based healthcare applications and pharmacoinformatics. These authors explained the benefits of introducing computers on various hospital departments, in the sense of dealing with the mass amount of information, namely in the pharmaceutic department, pointing the need to pharmacists to become more involved with applications of technology once that these systems, when properly implemented, would allow them to have more time to use their cognitive skills into the pharmaceutical care. Over the years healthcare have been considered as one of the most promising application areas for AI. Azzi et al., (2020) developed a systematic literature review focused on exploring key AI and analytics applications. As the investigation of AI techniques in healthcare continues to grow, Jiang et al. (2017) explained that this type of research has been focused on three main disease types that are leading causes of death: cancer, neurology, and cardiology, and in 2013 data sources like diagnostic imaging, genetic and electrodiagnosis were already being processed by this techniques. Specifically in neurology, AI technology have been making relevant progress in the past recent years. For example, Bouton et al. (2016) developed an AI system that restored the control of movement in patients with quadriplegia by showing that that is possible to restore movement in a paralysis through linking intracortically recorded signals in real time and applying ML algorithms to decode the neuronal activity and control activation of forearm muscles through a custom built high-resolution neuromuscular electrical stimulation system. Also, Farina et al. (2017)

tested the power of an offline man/machine interface that uses the discharge timings of spinal motor neurons to control upper-limb protheses.

AI, thus, can assist clinicals by providing updated information and reducing the time that takes to give a diagnosis and bridge inevitable errors that come with clinical human practice as well as it can extract useful information from a mass patient population data and assist on making inferences for health risk and health outcome prediction (Jiang et al., 2017). To make this possible, AI systems need to be "trained" with data that comes from clinical activities to "learn" similar groups of individuals, associations between these individuals features and outcomes of interest. More specifically, when it comes the diagnosis stage, it relies on data from diagnosis imaging, genetic testing and electrodiagnosis; physical examination notes and clinical laboratory results can also be valuable data to be analysed. There can be different forms of analysing these different types of data. For instance, ML techniques analyse structured data such as imaging, genetic and electrophysiological data, attempting to cluster patients traits, or infer the probability of the disease outcomes, and NLP methods extract information from unstructured data such as clinical notes and medical journals to supplement and enrich structured medical data (i.e., imaging, genetic and electrophysiological data).

2.3.2 Relevance of Artificial intelligence techniques to the neuropsychology practice

Research in AI as an aid to clinicians when dealing with neurodegenerative diseases, has also been achieving great deeds. In the late 90s investigation regarding AI techniques applied into the search for better understanding of AD started to appear. Menschik and Finkel (1998) conducted innovative research where they propose to describe a detailed model of the basic neural mechanisms underlying how memories are stored and recalled in the brain, to then applied it to the study of AD pathology. These authors described what they believe is one of the underlying mechanisms of memory loss and cognitive slowing seen in AD and propose to investigate it through the construction and investigation of a (computed-based) network that is based on a biological implementation of the Hopfield formalism for attractor neural networks (Menschik and Finkel, 1998), that simply explained, are attractor networks that are proposed as models of learning and memory in the cerebral cortex (Pereira e Brunel, 2018). These attractor models have provided the basis for several previous hippocampal-based models of memory and spatial navigation, and to a number of studies of AD, where they were used as a useful tool of exploration for the neuropathological consequences of AD (Menschik and Finkel, 1998).

In the first decade of the 2000's the investigation continued, and some other new discovers were made. Yang et al. (2010) enhanced the importance of volumetric and shape features of hippocampus on the diagnosis of AD and conducted research on the changes of these features in patients, when observed through MRI data. To make this possible they propose to train a Back-propagation artificial neural network (ANN) classifier, to learn the differences and understand this classification, and got very positive results on classifying each individual, whether it was probable AD or normal control, based on hippocampus shape and volumetric features. Reaching an accuracy of 88.27% when only using volumetric features and shape features and a correctness up to 92.17% when using volumetric features and shape features with the aid of principle component analysis (PCA) (Yang et al., 2010).

In decade from roughly 2004 until 2014, Khan and Usman (2015) pointed that there was an evident trend that included a rapid growth in the AD detection and prognosis using machine learning methods. These authors also conducted a literature review based on the fact that machine learning methods, in comparison to standard statistical methods, could improve the accuracy of AD prediction, emphasizing the important role that machine learning can play in AD prediction and prognosis, also proposing a model that used machine learning techniques applied to the study of AD, that they describe as having the potential to distinguish AD patients from healthy controls individuals. (Khan and Usman, 2015). This proposed method consists of four steps: Pre-processing, Attribute selection, Classification and Class Threshold. The data is analyzed in a specific type of way and then the model is trained and used for classification to identify the specific parameters. The association rules resulted in unique associations among the attributes and finally, a certain threshold is used over the resultant rules to classify the instances into one of the two classes: Control and AD (Khan and Usman, 2015). The authors argument their conclusions affirming that this proposed method deals with pathologically proven data and overcomes the class imbalance and overtraining issues, also being based on single modality that overcome the increased cost of computing and combining of different modalities (Khan and Usman, 2015).

More recently the investigation specifically in this area have been growing exponentially, and considering the different aetiology that we see in dementia, specifically AD, have its own variety of factors that can contribute to the development of this disease, not just environmental factors but also genetic. This new phase of blooming investigation got to embrace the various characteristics of AD that are relevant to its diagnosis and prognosis. Research regarding genetic contribution to AD is also available on the literature. Mishra & Li (2020) focus their investigation on the application of AI techniques in genetic research for AD and conclude that ML algorithms were efficient for performing a large data analysis, as well as the fact that they're already being used in studies of diagnosis and prognosis of AD based on factors like genetic variation in its different forms.

Recently, it has also been developed a new EEG signal processing tool that serves the purpose of characterize AD electrical activity in the brain, and it can be helpful to assist the diagnosis on its different stages of progression (Rodrigues et al., 2021). With this tool it's possible to perform statistical analysis to lacstral/cepstral distances between conventional EEG subbands, and this data gives us the chance to characterize AD activity in each electrode, as well as it allows to discriminate AD in its different stages of progression, (Rodrigues et al., 2021). Then, the authors propose that machine learning models are applied to lacstral/cepstral distances to develop an automatic method for diagnosing AD (Rodrigues et al., 2021).

In a more computer-based perspective, it's possible to find several literature that covers the relying of ML algorithms as a basis for clinical decision-making aids to assist on clinical practice (James et al., 2021), and also through "CAD" frameworks that stands for AI "Computer-Aided-Diagnosis" that is, an AI technique that can work in high performance computing infrastructures based on large amounts of data (Heising and Angelopoulos, 2021; 2022) to assist on clinical decision during the process of a diagnosis. While the first (i.e., ML algorithms as an aid to the diagnosis) works based on risk factors and try to assist on prognostic, the second (i.e., CAD – Computer-Aided-Diagnosis) works based on analysing structured data (for example, MRI) and try to assist on the diagnosis itself.

Developing a bit more what this information consists in, it's a fact that there are some strategies that were developed to help predict the risk of developing a dementia.

The LIBRA "Lifestyle for BRAin Health" is a score of risk to develop a dementia that is calculated taking into account 12 modifiable risk and protective factors for dementia and MCI, in both midlife and late-life and also in individuals with different levels of genetic risk (based on the presence of APOEε4) (Deckers et al., 2020).

There is also CAIDE "Cardiovascular Risk Factors, Aging, and Incidence of Dementia" risk score, a measure developed to predict the risk of developing a dementia in 20 years (for a middle-aged population) and BDSI "Brief Dementia Screening Indicator" that is directed to an elderly population, and it evaluates the risk of developing a dementia in 6 years (James et al., 2021).

These are important measures that can help guide more personalised interventions (i.e., LIBRA risk score and CAIDE risk score). However, James et al. (2021) point a gap for these measures, the fact that there is currently no tool that helps clinicians on deciding based on a risk score that has been evaluated in a shorter clinically relevant period. In their research, the authors propose a ML technique that allows the prediction of incidence of dementia over a 2 year (~29months) period, and their results were very significant. Their ML models, when compared with existing non-AI based models (i.e., CAIDE and BDSI, measures based on modifiable risk factors) were superior in terms of predicting whether an individual would develop dementia within 2 years (James et al., 2021), and so these authors manage to create a ML model based on six key factors for dementia risk that can be used as decision-making aid to assist clinicians (James et al., 2021).

Especially in AD and MCI, there is not just a lot of research already available in the literature, but also some projects that are being developed in the present.

Because of the great expansion of interest in studying AI techniques and other innovative technological methods, there is a lot of financial support available to conduct new research. For example, the European Union's research and innovation "Horizon 2020" funding programme had a budget of nearly €80 billion (European Commission, 2020) to fund research and innovation projects, two of which are extremely relevant when we're discussing AI techniques applied to the diagnostic of AD.

The "AI-Mind" is one of the projects that received funding from the European Union's research and innovative "Horizon 2020" programme and that is currently being developed. It's coordinated by the investigator Ira Haraldsen and aims to "develop an "intelligent diagnostic toolkit" for brain connectivity screening and dementia risk estimation in people with MCI containing two new AI based digital tools: the AI-Mind Connector and the AI-Mind Predictor" (Alzheimer Europe, 2021). To sum it up, their proposal it's to use AI to help in the interpretation of EEG data and to reduce the time needed to interpret this data manually, also to improve the understanding of the electrical phenomena and its correlation to dementia (Alzheimer Europe, 2021).

Another relevant project that is currently being developed is the “LETHE project” that also received funding from the European Union’s research and innovative “Horizon 2020” programme and it is coordinated by the investigator Sten Hanke. This project aims to develop a model based on AI and distributed machine learning and its goal is to create a model of prediction and intervention for AD, allowing detection and reduction of risk factors causing dementia (Alzheimer Europe, 2021). In other words, it aims to be a personalised prediction and intervention model for early detection and reduction of risk factors (Alzheimer Europe, 2021).

Recently, it was developed a study based on an eHealth intervention. HATICE (Healthy Ageing Through Internet Counselling in the Elderly) aims to investigate whether a multidomain intervention to optimize self- management of lifestyle-related risk factors for cardiovascular disease in older individuals, delivered through a coach-supported interactive platform, can improve the cardiovascular risk profile, and reduce the risk of cardiovascular disease and cognitive decline (Crous-Bou et al., 2017; Richard et al., 2016). Even more recent investigation on this topic states that self-management of cardiovascular risk factors with coach-support using an interactive internet-based intervention is feasible and was associated with improvements of cardiovascular risk profiles without adverse events, thus, implementation at large-scale could produce effects on clinical outcomes, namely cardiovascular disease, dementia and mortality (Richard et al., 2019).

2.4 Relevance of this review and research questions

The investigation with AI applied to AD diagnosis is relatively recent, and mainly used and investigated in high income countries (Danso et al., 2019; Heising and Angelopoulos, 2022), as these explore AI for neuroimaging and neuropsychological tests as well as for other data sources that can possible monitor effectively brain health and detect early onset of dementia (Danso et al., 2019). AI can help collect data regarding risk and protective factors and analyse it in a faster and more efficient way, allowing to create more personalized interventions in a preclinical stage of the disease as AD is multifactorial and, consequently, interventions should be specific to risk profiles (Crous-Bou et al., 2017), helping to decrease the progression of the disease and increase the quality of life of patients and their caregivers. In addition, AI techniques help understand prevention strategies of AD that would also complement the choice of individuals that are a better fit to participate in studies that seek to understand better these risk and protective factors of AD. Clinicians may not be aware of the benefits that AI can bring to their practice as the benefits of achieving better prevention strategies could start with recommendations addressed to the general population on how to manage lifestyle and cardiovascular risk factors. In parallel, a multidomain long-term intervention could be offered to individuals identified as being at increased risk of developing AD (Crous-Bou et al., 2017).

The study of implementation of prevention programmes at the pre-clinical stage would allow the delaying of the onset of AD for only a few years, but this would already have a huge impact on public health. It is also important to direct the research to cognitively healthy individuals at risk of developing AD, allowing to reduce the incidence and prevalence of AD (Crous-Bou et al., 2017). The identification of at risk individuals via biomarkers identification in the pre-clinical phase is the key factor to identify ideal candidates to participate in these secondary prevention studies, also identifying individuals at risk of developing the disease might be the key to success of intervention studies (Crous-Bou et al., 2017).

AI has been developing rapidly in the past recent years. This has provided an opportunity to understand and solve problems that involve massive amounts of data and very complex structures that exceed the processing capabilities of the human brain (Mishra & Li, 2020). A question that remains unanswered is if cognitive data is taken into account when creating AI models to the early detection of AD, since as it been stated before, a proper diagnosis of AD requires several criteria, being the assessment of cognitive functions a crucial one. It is also important to notice that AI can also be a beneficial tool in analysing the variation of cognitive status in several people, as well as to complement the understanding of the relation of different scores when different cognitive domains are evaluated. These are also important factors to the study of dementia.

Hereupon, with this **scoping review** we aim to find, in the literature, information about what type of AI techniques are being used to analyse data, what type of data is used to create these models, and how available is this information on innovative techniques for psychologists. Additionally, we will explore at what extent is cognitive data taken into account in the creation of these models. This review would help academics and professionals alike to find a summary of information that could help them understand how to develop more personalized interventions to delay the progression of AD in each person, that in current times is very important, as it is pointed by Brookmeyer et al. (2007).

Section III

Methodology

The proposed scoping review will be conducted in accordance with the JBI methodology (Peters et al., 2020). This work aims to map the literature to find what type of AI techniques are being used in early detection of AD, what type of data is used to feed these models, and how available is this information for psychologists and neuropsychologists. And additionally, at what extent is cognitive data taken into account in the creation of these models. To find the most relevant articles published between January 2013 and March 2023 a search was performed in Medline Complete via EBSCO, PubMed and Web of Science. The MeSH terms included "Alzheimer's Disease"; "Neural Networks, Computer", "Machine Learning", "Deep Learning" and "Primary Prevention", "Secondary Prevention", "Tertiary Prevention" to find information regarding Alzheimer's disease, artificial intelligence (neural networks, machine learning and deep learning) and prevention of dementia. The detailed search terms are described in Table 1.

Table 1. Detailed search expression for each bibliographic database

Bibliographic Database	***Search Expression***
Medline Complete via EBSCO	(TI ("Alzheimer* disease" OR "Alzheimer* Syndrome") OR AB ("Alzheimer* disease" OR "Alzheimer* Syndrome") OR MH "Alzheimer Disease") AND (TI ("neural networks" OR "machine learning" OR "deep learning") OR AB ("neural networks" OR "machine learning" OR "deep learning") OR MH ("Neural Networks, Computer" OR "Machine Learning" OR "deep learning")) AND (TI prevention OR AB prevention OR MH ("Primary Prevention" OR "Secondary Prevention" OR "Tertiary Prevention"))
PubMed	("Alzheimer Disease"[MeSH Terms] OR ("Alzheimer Disease"[Title/Abstract] OR "alzheimer syndrome"[Title/Abstract])) AND ("neural networks, computer"[MeSH Terms] OR "Machine Learning"[MeSH Terms]

	OR “deep learning”[MeSH Terms] OR (“neural networks”[Title/Abstract] OR “Machine Learning”[Title/Abstract] OR “deep learning”[Title/Abstract])) AND (“Primary Prevention”[MeSH Terms] OR “Secondary Prevention”[MeSH Terms] OR “Tertiary Prevention”[MeSH Terms] OR “prevention”[Title/Abstract])
Web of Science	TS=(“Alzheimer* disease” OR “Alzheimer* Syndrome”) AND TS=(“neural networks” OR “machine learning” OR “deep learning”) AND TS=(prevention)

Following the search, all identified citations were collected and loaded into *Rayyan* (a tool for systematic literature reviews) and duplicates were removed. Following a pilot test, tiles and abstracts were screened by two independent reviewers for assessment against the inclusion criteria for the review. Studies published in Portuguese, English, Spanish or French were considered. More detailed inclusion criteria included: investigation specific for AD, uses AI and its objective is specific to the prevention of dementia. Systematic reviews, opinion papers and chapters of books were not considered, as well as articles published in predatory journals. All articles that did not meet these criteria were excluded. The reviewers disagreed in 24 articles in the screening process, this was resolved through discussion between the two reviewers. The full text of selected citations was then assessed in detail, one article was removed in this phase of the process due to not having access to the complete paper. The results of the search and the study inclusion process is presented in Figure 1, summarized in a Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for scoping review (PRISMA-ScR) flow diagram (Page et al., 2020).

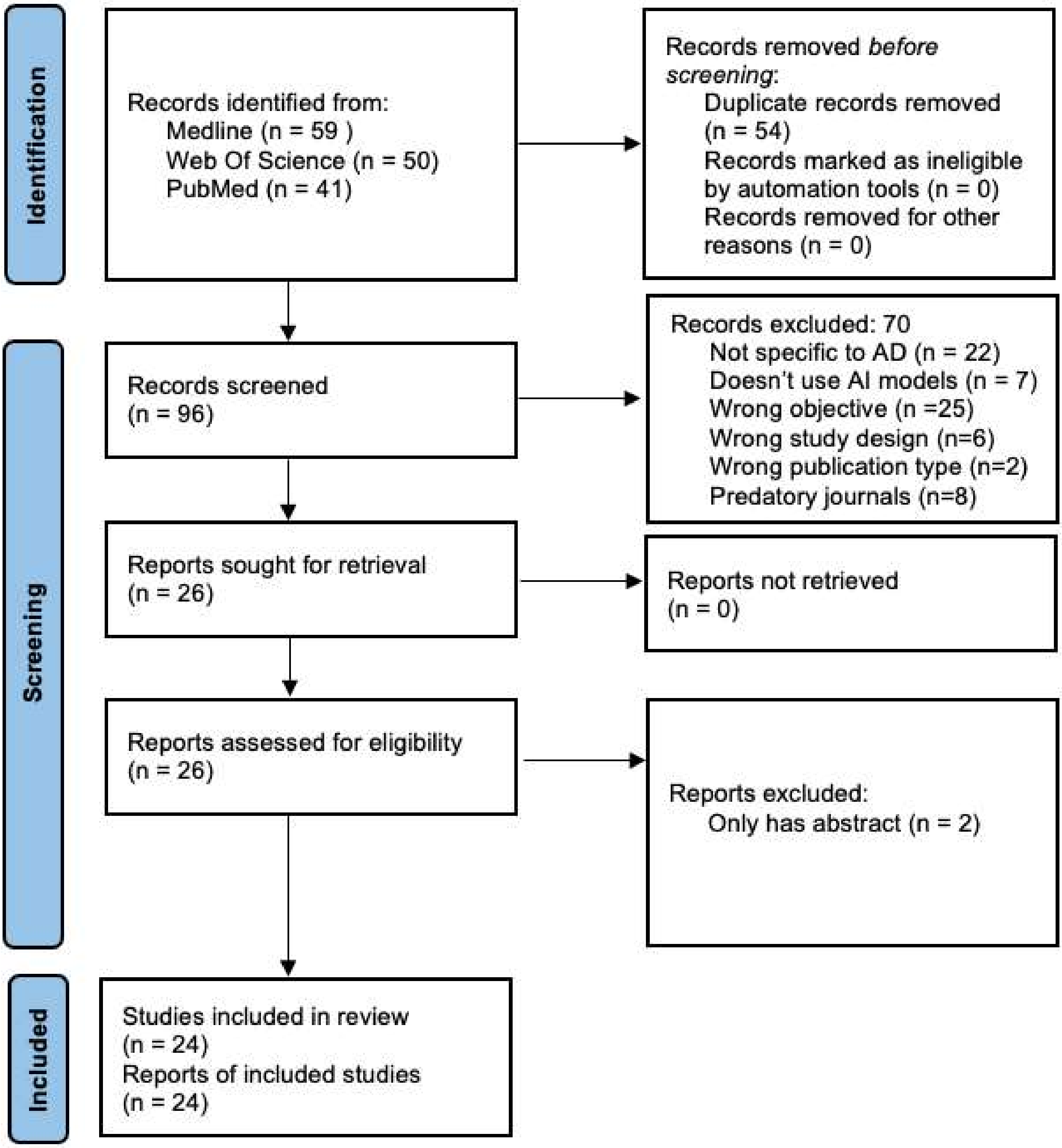

Figure 1. PRISMA-ScR flow diagram of the study selection process (Page et al., 2020)

Section IV

Results

Table 2 contains the descriptive statistics of the included articles, the starting point to these results section.

Table 2. Descriptive statistics of the included articles

Title	***Authors***	***Year***	***Scientifical Journal***	***Knowledge Area of the Journal***
Machine learning based multi-modal prediction of future decline toward Alzheimer's disease: An empirical study	Karaman, B. K., Mormino, E. C., Sabuncu, M. R., & Alzheimer's Disease Neuroimaging Initiative	2022	Plos One	Multidisciplinary
Classifying the lifestyle status of Alzheimer's disease from clinical notes using deep learning with weak supervision	Shen, Z., Schutte, D., Yi, Y., Bompelli, A., Yu, F., Wang, Y., & Zhang, R.	2022	BMC Medical Informatics and Decision Making	Informatics
VGG-TSwinformer: Transformer-based deep learning model for early Alzheimer's disease prediction	Hu, Z., Wang, Z., Jin, Y., & Hou, W.	2023	Computer Methods and Programs in Biomedicine	Informatics
Machine learning approach predicts probability of time to stage-specific conversion of Alzheimer's disease	Wu, X., Peng, C., Nelson, P. T., & Cheng, Q.	2022	Journal of Alzheimer's Disease	Cognitive Area
Comparison of machine learning approaches for enhancing Alzheimer's disease classification	Li, Q., & Yang, M. Q.	2021	PeerJ	Cognitive Area
Alzheimer's disease stage identification using deep learning models	Bringas, S., Salomón, S., Duque, R., Lage, C., & Montaña, J. L.	2020	Journal of Biomedical Informatics	Informatics
Spatio-temporal Tensor Multi-Task learning for predicting Alzheimer's	Zhang, Y., Zhou, M., Liu, T., Lanfranchi, V., & Yang, P	2022	44th Annual International Conference of the IEEE Engineering in	Informatics

disease in a longitudinal study			Medicine & Biology Society (EMBC)	
Dementia risks identified by vocal features via telephone conversations: A novel machine learning prediction model	Shimoda, A., Li, Y., Hayashi, H., & Kondo, N.	2021	Plos One	Multidisciplinary
Sociodemographic data and APOE-ε4 augmentation for MRI-based detection of amnestic mild cognitive impairment using deep learning systems	Pelka, O., Friedrich, C. M., Nensa, F., Mönninghoff, C., Bloch, L., Jöckel, K. H. & Schramm, S., et al.	2020	Plos One	Multidisciplinary
Dual-Task gait assessment and machine learning for early-detection of cognitive decline	Boettcher, L. N., Hssayeni, M., Rosenfeld, A., Tolea, M. I., Galvin, J. E., & Ghoraani, B.	2020	42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)	Informatics
A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease	Liu, M., Li, F., Yan, H., Wang, K., Ma, Y., Alzheimer's Disease Neuroimaging Initiative & Shen, L., et al.	2020	Neuroimage	Cognitive Area
Functional MRI and APOE4 genotype for predicting cognitive decline in amyloid-positive individuals	Zhu, J. D., Huang, C. W., Chang, H. I., Tsai, S. J., Huang, S. H., Hsu, S. W. & Lee, C. C., et al.	2022	Therapeutic Advances in Neurological Disorders	Cognitive Area
An explainable self-attention deep neural network for detecting mild cognitive impairment using multi-input digital drawing	Ruengchaijatuporn, N., Chatnuntawech, I., Teerapittayanon, S., Sriswasdi, S., Itthipuripat, S., Hemrungrojn, S. & Bunyabukkana, P., et al.	2022	Alzheimer's Research & Therapy	Cognitive Area
A transformer-based multi-features fusion model for prediction of conversion in mild cognitive impairment	Zheng, G., Zhang, Y., Zhao, Z., Wang, Y., Liu, X., Shang, Y. & Cong, Z., et al.	2022	Methods	Cognitive Area
Mild cognitive impairment understanding: an empirical study by data-driven approach	Liu, L., Yu, B., Han, M., Yuan, S., & Wang, N.	2019	BMC Bioinformatics	Informatics

Improved prediction of imminent progression to clinically significant memory decline using surface multivariate morphometry statistics and sparse coding	Stonnington, C. M., Wu, J., Zhang, J., Shi, J., Bauer III, R. J., Devadas, V. & Su, Y., et al.	2021	Journal of Alzheimer's Disease	Cognitive Area
Automatic detection of cognitive impairments through acoustic analysis of speech	Nagumo, R., Zhang, Y., Ogawa, Y., Hosokawa, M., Abe, K., Ukeda, T. & Sumi, S., et al.	*2020*	Current Alzheimer Research	Cognitive Area
Protocol for a conversation-based analysis study: PREVENT-ED investigates dialogue features that may help predict dementia onset in later life	de la Fuente Garcia, S., Ritchie, C. W., & Luz, S.	2019	BMJ open	Cognitive Area
Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset	Naz, S., Ashraf, A., & Zaib, A.	2022	Multimedia Systems	Multidisciplinary
TT self-weighted Deep-AD3-Net: An AD stage and risk prediction	Haulath, K., & Mohamed Basheer, K. P.	2023	Journal of Healthcare Management	Multidisciplinary
Predicting Cognitive Impairment and Dementia: A Machine Learning Approach	Aschwanden, D., Aichele, S., Ghisletta, P., Terracciano, A., Kliegel, M., Sutin, A. R. & Brown, J., et al.	2020	Journal of Alzheimer's Disease	Cognitive Area
Development of a two-state gaussian hidden markov model for modelling dementia progression in patients with mild cognitive impairment	Canavan, C., Maguire, L. P., & Bucholc, M.	2021	IEEE 9th International Conference on Healthcare Informatics (ICHI)	Informatics
Development of a Late-Life Dementia Prediction Index with Supervised Machine Learning in the Population-Based CAIDE Study	Pekkala, T., Hall, A., Lötjönen, J., Mattila, J., Soininen, H., Ngandu, T. & Laatikainen, T., et al.	2017	Journal of Alzheimer's Disease	Cognitive Area
Multi-modal discriminative dictionary learning for Alzheimer's disease and mild cognitive impairment	Li, Q., Wu, X., Xu, L., Chen, K., Yao, L., & Li, R.	2017	Computer Methods and Programs in Biomedicine	Informatics

Firstly, we explore in which scientifical journals does this type of investigation gets published on, so we can discuss the ease of access to these papers for psychologists. Figure 2 summarizes the scientifical journals in which the papers included in this review were published on. As we can see in Table 2 and Figure 2, from the seventeen scientific journals that papers included in this review are published on, six of them (including two conferences) are from the informatics area (i.e., BMC medical informatics and decision making, computers methods and programs in biomedicine, journal of biomedical informatics, annual international conference for the IEE Engineering in Medicine & Biology Society EMBC), BMC bioinformatics, International Conference on Healthcare Informatics) resulting on a total of 33,28% of all articles included in this review being published in these journals (frequencies for figure 2 are presented in attachment 1). The other eleven scientific journals are multidisciplinary or related to the cognitive field, being 12,5% of the articles published on Plos One, 16,7% published on Journal of Alzheimer's Disease, 4,2% published on PeerJ, 4,2% published in Neuroimage, 4,2% published on Therapeutic Advances in Neurological Disorders, 4,2% published on Alzheimer's Research & Therapy, 4,2% published in Methods, 4,2% published on Current Alzheimer Research, 4,2% published BMJ open, 4,2% published on Multimedia Systems and 4,2% published in International Journal of Healthcare Management.

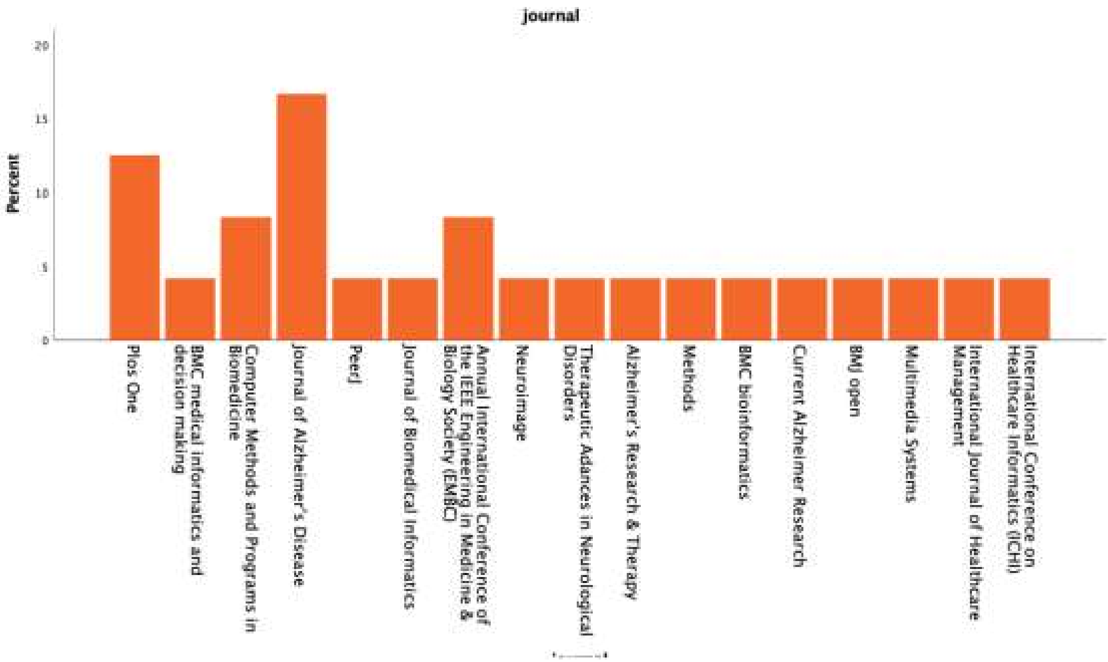

Figure 2. Scientifical Journals

Table 3 contains the more relevant information retrieved from the included studies that have given rise to these results section.

Table 3. More relevant information from the included articles

Study	***Sample Characterization***	***Data***	***AI tech***	***Outcome (results)***
Karaman et al. (2022)	CN and MCI	Demographics, APOE4 genotype, clinical assessment, cognitive assessment, biomarkers	ANN	Capture disease progression dynamics and produce conversion predictions (results show precise levels of prediction of conversion from CN to MCI in a short period of time and from MCI to AD in a longer period of time)
Shen et al. (2022)	Not mentioned	Clinical notes (EHRs)	NLP	Classifying lifestyle status for AD in clinical notes (results show better performance than traditional ML algorithms)
Hu et al. (2023)	MCI	sMRI	CNN	Build an atrophy progression model and improve diagnostic efficiency (results show better diagnostic efficiency compared to other sMRI-based cross-sectional studies)
Wu et al. (2022)	Not mentioned	Clinical and demographic	DNN	Accurately predict how soon disease will convert in a given period of time (results show that this approach is able to predict, specifically, the probability of a subject converting from preclinical AD to early AD)
Li & Yang (2021)	Not mentioned	MRI	ANN	Accurate disease classification and localization of disease-associated regions in the brain (results show better classification performance than a traditional ML algorithm and highlighted the importance of the

				cerebral cortex and the cerebellum as AD associated brain regions)
Bringas et al. (2020)	AD	Mobility	CNN	Accurate classification of all disease stages (results show that this approach improves accuracy in identifying AD stages compared to other common supervised learning models, highlighting the potential of mobility data analysis to the detection of different AD stages)
Zhang et al. (2022)	Not mentioned	Neuropsychological tests and MRI	Not specified ML	Effectively predict disease progression (results show that correlation data (brain biomarkers (i.e., cognitive data) and MRI) can enhance the accuracy and stability of the AD progression model)
Shimoda et al. (2021)	CN and AD	Audio	XGBoost	Accurate classification of audio files from AD and CN (results show a comparable performance to cognitive tests which enhances the possibility of including the assessment of language through vocal features, and regarding on this audio data, to complement the early detection of AD)
Pelka et al. (2020)	CN and MCI	MRI, sociodemographic and genotype (APOE4)	RNN	Accurate classification of MCI vs CN (results show significant accuracy at classifying aMCI vs CN on two different datasets)
Boettcher et al. (2020)	CN and MCI	Mobility and cognitive	SVM	High sensitivity in detecting MCI (results show that this approach has higher sensitivity in detecting MCI subjects when compared to MoCA scores)
Liu et al. (2020)	CN, MCI, AD	sMRI	CNN	Accurate classification of disease status (results show significant accuracy for classifying both AD vs CN

				individuals and MCI vs CN individuals)
Zhu et al. (2022)	CN and MCI	rs-fMRI and clinical (genotype)	WNN	Accuracy for predicting cognitive status (results show accuracy for predicting cognitive status after 1 year in amyloid-positive individuals)
Ruengchaijatuporn et al. (2022)	CN and MCI	Neuropsychological tests/cognitive tasks	Not specified DL	Accurate classification performance and interpretability (results show that adding multiple drawing tasks inputs to the proposed model improves classification performance between MCI and CN individuals)
Zheng et al. (2022)	MCI	sMRI	Not specified DL	Accurate prediction of MCI conversion to AD (results show accuracy in differentiating pMCI from sMCI and highlight brain regions that most contribute to this classification)
Liu et al. (2019)	Not mentioned	Clinical and demographic	Not specified ML	Determine risk factors correlated with diseases (results show that depression, physical health, cigarette usage, education level and sleep time are factors of importance in cognitive decline)
Stonnington et al. (2021)	MCI	sMRI	Not specified ML	Accurate prediction of AD stage progression (results show that this approach can accurately predict imminent progression to MCI even in the absence of other AD biomarkers)
Nagumo et al. (2020)	CN and MCI	Audio	Not specified ML	High discriminatory ability of CN from MCI (results show that individuals acoustic features can be employed in discriminating between CN and MCI (with global cognitive impairment))
de la Fuente et al. (2019)	Not mentioned	Audio	Not specified ML	Novel approach of monitoring early signs of dementia (this research investigates how valuable the

				information obtained from a person's speech in a spontaneous conversation can be to monitor early signs of dementia)
Naz et al. (2022)	CN, MCI, AD	MRI	CNN	Accurate disease stage classification (results show accuracy for classification of AD vs MCI, CN vs MCI and CN vs AD, this approach also outperformed other state-of-the-art AI techniques)
Haulath & Mohamed Basheer (2023)	Not mentioned	Neuropsychological tests and MRI	Not specified DL	Accurate prediction of severity of AD (results show that this approach is better than other traditional techniques in terms of predicting the severity of AD)
Aschwanden et al. (2020)	MCI and AD	Cognitive, psychosocial and physical	Not specified ML	African Americans and high scores of emotional distress represent higher risk for developing cognitive impairment and dementia (results show that the highest risk factors for developing cognitive impairment and dementia were being African American and emotional distress; sociodemographic and health variables were also strong predictors)
Canavan et al. (2021)	CN, MCI, AD	Neuropsychological tests/cognitive tasks	HMM	Accurate prediction of MCI progression to dementia and identify patients with similar disease trajectories (results show that this approach can represent a tool for predicting disease evolution and also to classify patients as MCI converters and non-converters)
Pekkala et al. (2017)	not mentioned	Neuropsychological tests, vascular factors, sociodemographic, APOE4 genotype	DSI	Accurate prediction of dementia development up to 10 years later and identify individuals who are more at risk

				(results show this ML approach demonstrated strong performance in creating detailed profiles that effectively forecast the onset of dementia as far as a decade in advance, with predictors being cognition, vascular factors, age, subjective memory complaints and APOE genotype)
Li et al. (2017)	CN, MCI, AD	sMRI, FDG-PET and florbetapir-PET	SVM	Accurate differentiation between AD and MCI (results show that this approach outperformed other methods in differentiating AD and MCI and that this is a feasible tool for neuroimaging classification tasks)

The majority of the investigations have taken into account data from participants that were diagnosed with MCI or AD and also have taken into account a cognitively normal (CN) group for comparison of results, as we can see in figure 3 (frequencies for figure 3 are presented in attachment 2). Some studies concentrated on MCI, focusing on stable MCI and progressive MCI (Hu et al. 2023; Zheng et al., 2022) emphasizing the struggle to study the differences between the two, as this is a crucial point for early detection and dementia prevention (Hu et al., 2023; Zheng et al. 2022). As we can see in Figure 1, 33,3% of the studies did not mention the characterization of the population used, 25% used CN and MCI individuals, 16,7% used CN, MCI and AD individuals, 12,5% used MCI individuals, 4,2% used only AD individuals, 4,2% used CN and AD individuals and 4,2% used MCI and AD individuals.

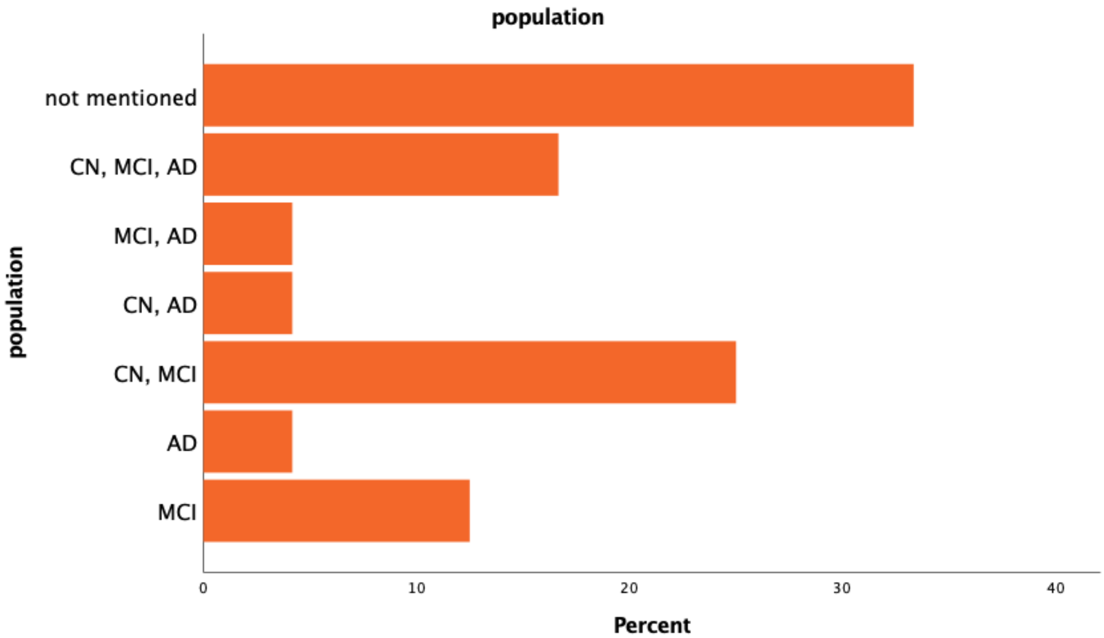

Figure 3. Population characterization

Table 3 summarizes the most relevant information presented in the articles that we included regarding the types of AI technique used. In figure 4 we can observe the AI technique that was on the basis of each of the models developed/tested in the articles included in this review. Some articles created models based on the AI technology that is presented in figure 4, so the authors may have named their own model, but the model is created based on these types of techniques. Some articles did not specify the type of ML or DL techniques they used but other did, as we can see in figure 4 (frequencies for figure 4 are presented in attachment 3). As we can see in figure 4, 25% of the studies adopted a not specified ML technique, 16,7% based their model in Convolutional Neural Networks (CNN), 12,5% adopted a not specified DL technique, 8,3% based their model on Artificial Neural Networks (ANN), 8,3% based their model on Support Vector Machine (SVM), 4,2% based their model on Natural Language Processing (NLP), 4,2% based their model on XGBoost, 4,2% based their model on Recurrent Neural Networks (RNN), 4,2% based their model on Wide Neural Networks (WNN), 4,2% based their model on Hidden Markov Model (HMM), 4,2% based their model on Disease State Index (DSI) and 4,2% based their model on Deep Neural Networks (DNN).

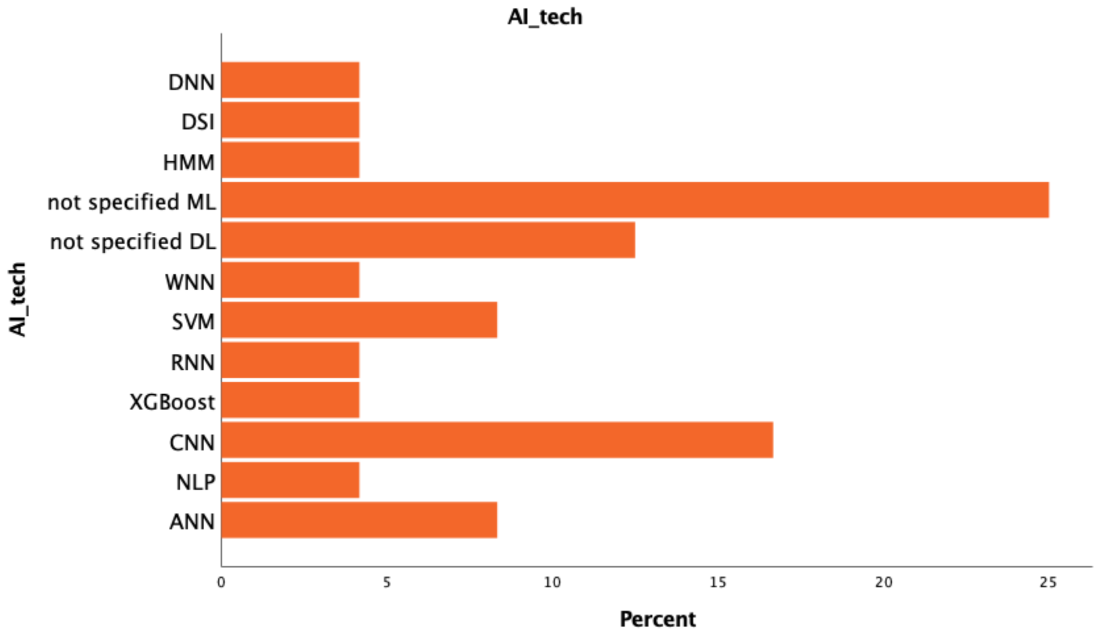

Figure 4. AI technique

Figure 5 presents the type of data used to feed these models. Some authors focused on different types of brain MRI; Haulath & Basheer (2023) for instance focused on rs-fMRI. Even though most investigations uses sMRI, other types of data are also used to feed these innovative techniques. In figure 5 we can observe that 25% of the studies uses only brain MRI to feed the models and other 20,9% of studies uses MRI combined with other type of data. Other innovative types of data to detect and make disease progression predictions are also considered in the studies included in this review, namely clinical notes (EHRs), mobility data and audio data, specifically mobility of the patients captured by an accelerometer incorporated in a smartphone carried in the patients pocket during the day that captured their mobility (Bringas et al. 2020), voice data collected through a telephone conversation with an AI computer programme (Shimoda et al., 2021), acoustic features from short-sentence reading tasks (Naguno et al. 2020) and acoustic and dialogical features retrieved from dialogue data (de la Fuente et al. 2019) (frequencies for figure 5 are presented in attachment 4). In figure 5 we can observe that 25% of studies uses brain MRI to feed their model, 12,5% uses audio data (dialogue and vocal features), 8,3% uses clinical and demographic data, 8,3% uses neuropsychological tests/cognitive tasks data, 8,3% uses neuropsychological tests and MRI data, 4,2% uses demographics, APOE4 genotype, clinical assessment, cognitive assessment and biomarkers data, 4,2% uses clinical notes (EHRs) data, 4,2% uses mobility data, 4,2% uses mobility and cognitive data, namely gait data from a dual task assessment (gait assessment and a cognitive task) (Boettcher et al. 2020), 4,2% uses cognitive, psychosocial and physical data, 4,2% uses neuropsychological tests, vascular factors, sociodemographic and APOE4 data, 4,2% uses MRI and PET data, 4,2% uses MRI, sociodemographic and APOE4 genotype data and 4,2% uses MRI and clinical data.

As it can be seen, even though it's not used exclusively in the majority of the investigations, many researchers rely on cognitive data in addition with other types of data to feed the model developed in their investigation. To obtain these cognitive data many authors apply neuropsychological tests to asses specific cognitive domains.

Karaman et al. (2022) among demographics, APOE4 genotype, biomarkers and clinical assessment also considered cognitive assessment to feed their model. This cognitive assessment included MMSE, ADAS-Cog, MoCA, Rey Auditory Verbal Learning Test Trials 1–6, Logical Memory Delayed Recall, Trail Making Test Part B, Digit Symbol Substitution, Digit and Trails B versions of Preclinical Alzheimer's Cognitive Composite score.

Zhang et al. (2022) in addition to MRI data, the authors also considered cognitive data to feed their model, namely the neuropsychological tests MMSE and ADAS-Cog;

Haulath & Mohamed Basheer (2023) in addition to MRI data, specifically rs-fMRI, the authors considered relevant cognitive variables obtained from the original ADASCog-13 to calculate a risk prediction score.

Boettcher et al. (2020) in addition to a single mobility task to asses gait, in their investigation, the authors considered information from a dual-task that consisted in a mobility task (to assess gait) to be performed at the same time that a cognitive task (subtraction) and so, the dual-task consisted in walking while subtracting.

Aschwanden et al. (2020) considered information from cognitive tasks to feed their model in addition to psychosocial and physical data. These cognitive tasks were immediate and delayed recall of 10, serial 7 subtraction and backward counting.

Ruengchaijatuporn et al. (2022) in their investigation, considered exclusively neuropsychological tests to feed their model, they specifically used the clock drawing test (CDT), cube-copying and trail-making tests.

Canavan et al. (2021) aimed to use only neuropsychological data to train their model, for this they considered a number of neuropsychological factors including neuropsychological tests. These variables were years of education completed, a functional assessment questionnaire, MMSE, Logical memory test, memory span score – digits in order, memory span score – digits in reverse order, Wechsler Adult Intelligence Scale – Digit Symbol and Boston Naming Test.

Pekkala et al. (2017) considered, in their investigation, neuropsychological tests among other types of data like vascular factors, sociodemographic and APOE4 genotype. These neuropsychological tests were used in order to assess five cognitive domains, therefore the authors administrated MMSE to assess global cognition, mean number of recalled words from three 10-word lists to assess episodic memory, one-minute animal naming test to assess verbal expression, mean of normalized scored from Letter Digit Substitution and bimanual Purdue Pegboard tests to assess psychomotor speed, time difference between the color word interference and naming tasks in the Stroop test to assess executive functioning and reminding the investigator to make a phone call at the end of the testing session to assess prospective memory.

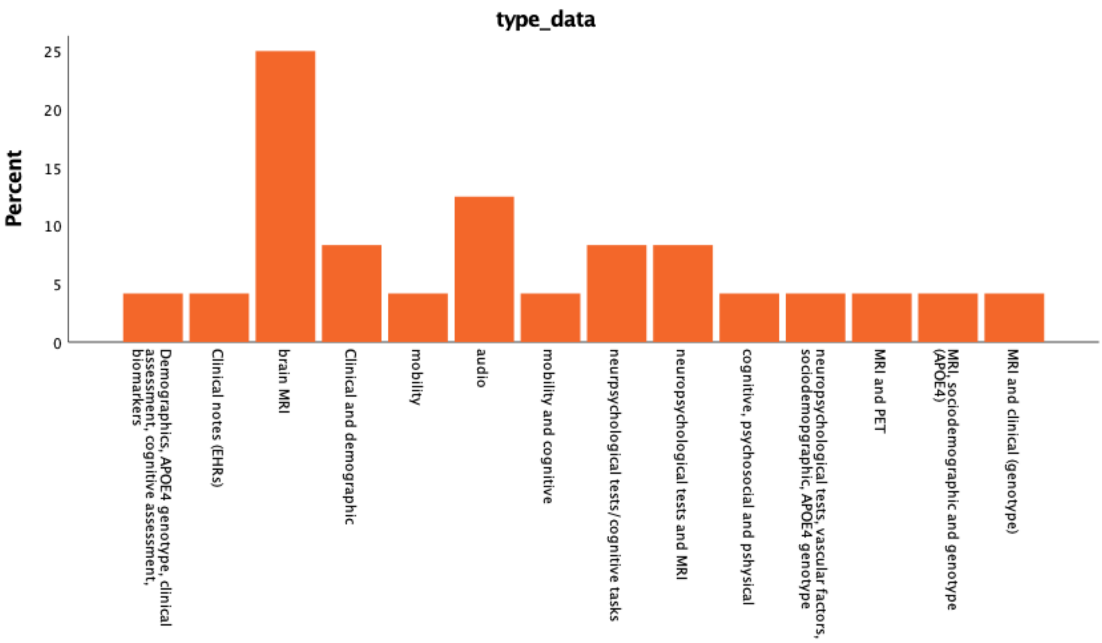

Figure 5. Type of data used to feed the models

Lastly, in table 3 we can observe that cognitive data, that is, neuropsychological tests and/or cognitive tasks, are used as exclusive type of data to feed the model in Ruengchaijatuporn et al. (2022) and in Canavan et al. (2021). This type of data is also used to feed the model combined with other types of data in four other investigations, included in this review, and this leads us to our next question. In figure 6 we can observe that the majority of the studies (41,67%) relies on cognitive data (namely neuropsychological tests) only to characterize the cognitive status of the participants and form the groups (experimental (cognitively impaired) and control (cognitively unimpaired) and does not take this type of data into account in the training of the AI model. Other studies do not mention any cognitive assessment or cognitive data during the investigation (33,33%). The minority of the studies (25%) takes cognitive data into account to feed the model (frequencies for figure 6 are presented in attachment 5).

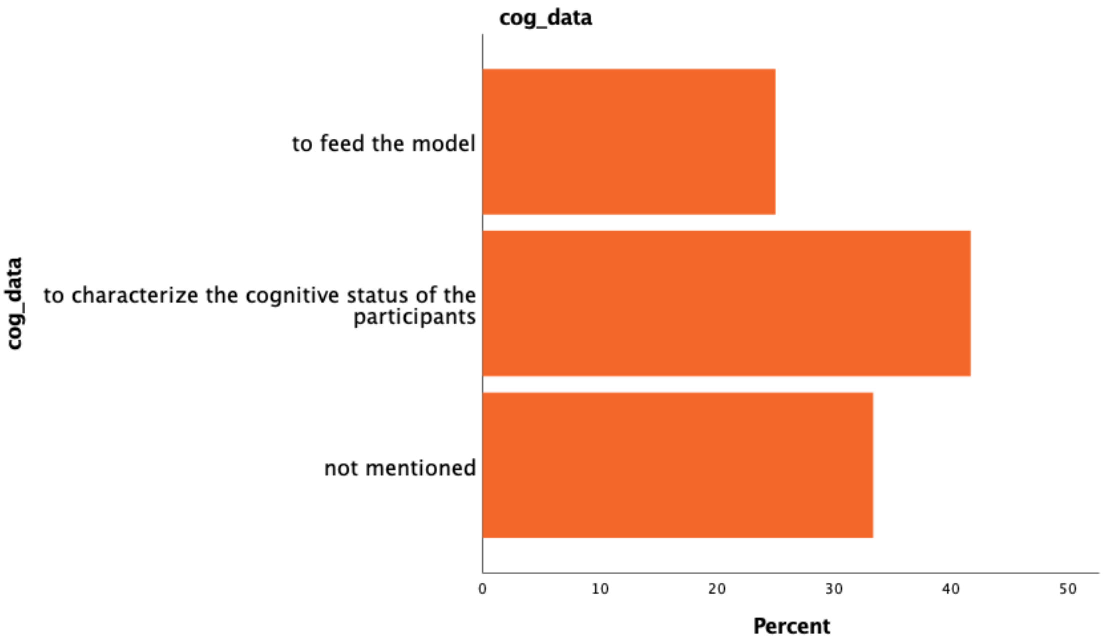

Figure 6. Where's cognitive data considered for

Section V

Discussion

Table 3 contains the outcome from each investigation included in this review and we can conclude that ML demonstrates great achievements in this field, namely the possibility of predict **disease progression** trajectories much before their onset time, which would allow to develop early intervention strategies personalized for each patient and slow down its progression, therefore decrease the onset of dementia. Other outcomes that would also help the achievement of slowing down disease progression and onset of dementia are the possibility of building atrophy progression models that would help improve diagnostic efficiency (Canavan et al., 2021; Haulath & Mohamed Basheer, 2023; Hu et al., 2023; Karaman et al., 2022; Pekkala et al., 2017; Stonnington et al., 2021; Wu et al., 2022; Zhang et al., 2022; Zheng et al., 2022; Zhu et al., 2022).

Other important factor for dementia prevention is the early diagnostic of AD, namely the possibility of MCI detection. Many studies included in this review offer the possibility of accurate **disease classification**, that is, the models presented in these investigations can accurately differentiate CN, MCI patients and AD patients from each other (Boettcher et al., 2020; Bringas et al., 2020; de la Fuente et al., 2019; Li & Yang, 2021; Li et al., 2017; Liu et al., 2020; Nagumo et al., 2020; Naz et al., 2022; Pelka et al., 2020; Ruengchaijatuporn et al., 2022; Shen et al., 2022; Shimoda et al., 2021) and this would also allow the early intervention (i.e., in the prodromic phase of AD) that would help on the slowing down of disease progression. Early detection of AD is considered a crucial factor for the implementation of secondary prevention initiatives (Crous-Bou et al., 2017), therefore the outcomes that allow the early detection of pathophysiological hallmarks before the cognitive impairment represent great advances in this topic. So, regarding on these AI techniques, the early detection of AD and intervention at the preclinical/prodromal phase would allow the implementation of secondary prevention initiatives and it might offer the best opportunity to start an intervention in earlier stages of the disease as well as delaying the cognitive decline (Crous-Bou et al., 2017). It's important to notice that, with this review, we can conclude that this objective is not only possible through neuroimaging data, but also some studies used novel types of data to feed their model, that would help detect the early signs of dementia, namely audio data (dialogue and vocal features) (de la Fuente et al., 2019; Nagumo et al., 2020; Shimoda et al., 2021) and mobility data (mobility and gait data) (Boettcher et al., 2020; Bringas et al., 2020).

On the other hand, there is also primary prevention strategies, which are based on **identifying modifiable risk factors** and work on risk reduction (Crous-Bou et al., 2017). Also, in table 3 some investigations included in this review produced relevant outcomes in this topic. Liu et al. (2019) concluded that relying on clinical and demographic data to feed a machine learning model could accurately determine risk factors correlated with diseases. Also, Aschwanden et al. (2020) relying on cognitive tasks, psychosocial data and physical assessment data to feed a machine learning model concluded that African Americans and individuals that scored higher on emotional distress represented a higher risk for developing cognitive impairment and dementia. Therefore, these innovative techniques enrich the existing knowledge about the modifiable risk factors for cognitive impairment and dementia like the ones pointed out by Zhang et al. (2021), explained in detail in the theoretical framework. The outcomes produced by these studies are very relevant in the sense of allowing the timely detection of AD, by regarding on machine learning to aid the detection and prevention of AD. One of the main goals is that these algorithms achieve immediate conclusions, which saves a lot of time to professionals, that would take much more time to analyse this type of data by themselves.

In the topic of cognitive data specifically, cognitive data is mainly used to characterize the cognitive status of the participants to form the experimental and control groups. Even though some studies produced great achievements relying on neuropsychological tests and/or cognitive tasks exclusively to feed the model (Canavan et al., 2021; Ruengchaijatuporn et al., 2022).

Ruengchaijatuporn et al. (2022) focused their investigation on visuoconstructive and visuospatial tasks (i.e., clock drawing test (CDT) and cube-copying) and in attention and cognitive flexibility tasks (i.e., trail making tests) to feed their model and achieve an accurate disease classification performance and interpretability of the results, this is an important achievement, as the authors point that more drawing tests (i.e., information from all three tasks feeding the model – CDT, cube-copying and trail making tests – rather than just CDT) lead to a better performance of the model, that is better discriminating capability (namely between MCI and CN individuals).

On the other hand, Canavan et al. (2021) focused their investigation in a more global neuropsychological perspective, using MMSE, Logical memory test, memory span score (digits in order and digits in reverse order) Wechsler Adult Intelligence Scale (Digit

Symbol task) and Boston Naming Test to feed their model, and achieve an accurate prediction of MCI progression to dementia and identify patients with similar disease trajectories. Other important conclusion pointed by the authors is the benefits of using neuropsychological information, the fact that these are relatively cheap (in comparison of other methods), their efficiency, the fact that these are non-invasive methods and their easy ability to be introduced in the clinical environment (Canavan et al., 2021).

Other studies have also taken into account cognitive data combined with other types of data to feed the model (Aschwanden et al., 2020; Boettcher et al., 2020; Haulath & Mohamed Basheer, 2023; Karaman et al., 2022; Pekkala et al., 2017; Zhang et al., 2022). It is important to notice that the majority of these investigations also takes into account the assessment of capability to perform daily activities and that all of them achieve great conclusions. Still, as pointed before, the majority of the studies relies on neuroimaging.

Specifically Haulath & Mohamed Basheer (2023) demonstrate in their investigation that their approach performs better than other AI techniques and can produce a risk score for dementia based not just on MRI data but on MRI data correlated with cognitive scores (from the ADAS-Cog13 neuropsychological test).

Also Zhang et al. (2022) demonstrates the importance of cognitive data on the study of prediction, diagnose and progression of AD, as in their investigation the authors found that using correlational data, namely cognitive data (i.e., brain biomarkers, through the application of neuropsychological tests (MMSE and ADAS-Cog)) and MRI data, would boost the accuracy and stability of an AD progression model, demonstrating that by using this correlational information the model could better characterise individual changes in brain structure for AD and MCI patients and also in CN individuals.

From this we can conclude that the investigation on early detection of AD and prevention of dementia using ML should be more multidisciplinary, as it already been referred before, a proper diagnosis of AD requires several criteria, being the assessment of cognitive functions a crucial one. Still in this line of thought, it can also be discussed the ease of access of psychologists to these types of innovative techniques and even to the articles itself. A major part of such articles are published in journals focused on informatics/engineering area or presented in engineering conferences, this may represent a barrier to the access of psychologists to these papers. It is important that not just psychologists but health professionals that work with neurodegenerative diseases to be more educated in this topic, as this could assist their practice. Namely for neuropsychologists, if they were instructed in this subject they would understand the benefits it could bring to their practice, for example, learning how to work with these types of algorithms and having the opportunity to have an immediate outcome produced by one, when in possession of neuroimaging from the patient they are about to perform a cognitive assessment to, this would conduct to more precise cognitive reports and also a more multidisciplinary knowledge, that would only benefit the patient.

It is important to notice that some authors compare the effectiveness of the AI model proposed to the effectiveness of cognitive tests. Shimoda et al. (2021) do this in their investigation, where the authors compare their model to cognitive tests in terms of effectiveness, concluding that their model demonstrated comparable performance to cognitive tests, with almost approached significance. This is an important note once this enhance the need to establish a more strict protocol of neuropsychological assessment for dementia, so professionals from other knowledge areas could take into account that the assessment of cognitive function is a crucial factor in the development of a neurodegenerative disease. Boettcher et al. (2020) also compares the effectiveness of their model to cognitive tests, namely by comparing the effectiveness of detecting MCI subjects to the cut-off points of the MoCA score. This also enhances the need to develop a more strict protocol of neuropsychological assessment for dementia, once if things were more strictly defined, it would be easier for professionals that are not from the area to understand the importance and also the relevance of introducing cognitive data into their investigation.

If there was a strict neuropsychological assessment protocol for dementia it is possible that the importance of cognitive data into the study of dementia was taken into a higher level of consideration in this type of research. This line of thought also has the potential to introduce more innovative types of data into the study of dementia, for example audio (vocal and dialogue features) and mobility data (mobility and gait data), as it been referred before.

There are other innovative forms that ML is contributing to the study of AD. For instance, some articles that appeared in the bibliographic databases search stage of this review and were excluded for not meeting the criteria of inclusion, but these articles also produced very interesting conclusions that should be explored in future research, namely by regarding on machine learning algorithms to enrich AD clinical trials. Lin et al. (2018) explains in their study that by using only non-invasively collected clinical variables they could achieve significant accuracy for identifying subjects converting to MCI within 4 years of initial evaluation and stated that this outcome may be useful in clinical trial enrichment and also assist in creating a shorter battery for screening at-risk subjects (Lin et al., 2018). Also, Ten Kate et al. (2018) developed a classifier, based on ML, that could predict amyloid pathology at a single subject level using a combination of easily obtainable and non-invasive measures and this could be of interest for trial designers who intended to recruit a large number or amyloid positive subjects (Ten Kate et al., 2018). Lastly Casamitjana, et al. (2018) established a ML technique to detect amyloid pathology in cognitively normal individuals using MRI and stated that this method could be used for screening of subjects for secondary prevention trials (Casamitjana, et al., 2018). Cognitive data is of importance in the study of MCI and AD dementia, and this type of data is taken into consideration when research is carried out in this area.

Section VI

Conclusion

This scoping review based on the JBI methodology (Peters et al., 2020) included papers from 2017 until march 2023 and gathered the most relevant information on what type of data is used to train AI models that aim for the early detection of AD. We conclude that neuroimage is the type of data where investigation in this area is more focused on, and argue for the need to include novel types of data that are also relevant to the early detection of AD. Furthermore, we discuss the importance of making this type of information more easily available for psychologists as well as educate them on the subject of these technology, which can bring great benefits to their practice.

The development of this scoping review suffers from some limitations in terms of scientifical rigor.

The first limitation is that this work might suffer from some technical mistakes in the extensive analysis of the specificities of the AI techniques due to limited knowledge in the AI area. Even though this dissertation was supervised by an external supervisor specialized in this area, the extensive analysis of the specificities of the AI techniques may suffer from some technical mistakes due to an academic background in psychology.

Another limitation to this review is the fact that some articles included in it used data from participants from other investigations (for example, the ADNI cohort study) and so it was difficult to characterize the participants in terms of demographic information, also because the participants were evaluated in different life times (as these are longitudinal investigations) and this also contributed to the difficulty in characterize the cognitive status of the participants. So, these difficulties contributed to the limitation of getting an extensive analysis of the demographic information and cognitive status of the participants of each investigation included in this review.

In terms of extensive analysis of the main conclusions, the main focus was the analysis of the conclusions of the included articles that were the most relevant for the main objective of this review, as so not all conclusions of the articles are mentioned, as some of them got more than just one important conclusion. This decision was made based on maintaining this review focused on its main objective, that is to what extent is AI being used in the early-detection of AD and prevention of dementia and to what extent is cognitive data being taken into account.

Lastly, future investigation regarding early detection of AD and prevention of dementia relying on AI models should consider a more multidisciplinary perspective. As it was referred before, the criteria to a final AD diagnostic includes different types of information that should be collected from the patient, therefore to help improve the accuracy of an AI based diagnostic, the algorithms should be trained using not just neuroimaging data, but also clinical, sociodemographic and cognitive.

Hereupon, from this scoping review we retrieve some useful information, namely the fact that AI investigation is growing in the field of early detection of AD and prevention of dementia, namely through the utilization of neuroimaging to the development of these models, so that it can play a part of a complementary tool to the early detection of AD. Through this review is possible to get to know other innovative types of data that are being used to the creation of AI techniques to the early detection of AD and prevention of dementia, and also to understand their relevance to this area. Namely we have the importance of cognitive data, not just in the characterization of the cognitive status of the participants, but also to the creation of the model itself, for both disease classification models and disease progression models. From this scoping review we can conclude that the use of this type of data is lacking in this type of research, and so, the future investigation should take this into account and develop in this way, as it is a more multidisciplinary point of view and offers the scientifical community the opportunity to take into account several of the factors that contribute do the development of a dementia, the first step to achieve its prevention and decrease its prevalence.

References

Alzheimer Europe (2021, June). AI-Mind – New initiative to develop artificial intelligence tools for dementia prevention. *Dementia in Europe – The Alzheimer Europe magazine*, 1(36), 11-13. https://www.alzheimer-europe.org/sites/default/files/dementia_in_europe_magazine_issue_36.pdf

Alzheimer Europe (2021, June). LETHE project looks at digital cognitive biomarkers. *Dementia in Europe – The Alzheimer Europe magazine*, 1(36), 14-15. https://www.alzheimer-europe.org/sites/default/files/dementia_in_europe_magazine_issue_36.pdf

Alzheimer's Association (2022). *How is Alzheimer's Diagnosed?*. Retrieved from: https://www.alz.org/alzheimers-dementia/diagnosis 2022

Alzheimer's Association (2022). *Medical Tests for Diagnosis Alzheimer*. Retrieved from: https://www.alz.org/alzheimers-dementia/diagnosis/medical_tests

American Psychiatric Association. (2014). *DSM-5: Manual diagnóstico e estatística das perturbações mentais*. Artmed Editora.

Aschwanden, D., Aichele, S., Ghisletta, P., Terracciano, A., Kliegel, M., Sutin, A. R. & Brown, J., et al. (2020). Predicting cognitive impairment and dementia: A machine learning approach. *Journal of Alzheimer's Disease*, *75*(3), 717-728. doi: 10.3233/JAD-190967

Ayodele, T., Rogaeva, E., Kurup, J. T., Beecham, G., & Reitz, C. (2021). Early-onset Alzheimer's disease: what is missing in research?. *Current neurology and neuroscience reports*, 21, 1-10. https://doi.org/10.1007/s11910-020-01090-y

Azzi, S., Gagnon, S., Ramirez, A., & Richards, G. (2020). Healthcare applications of artificial intelligence and analytics: a review and proposed framework. *Applied Sciences*, 10(18). https://doi.org/10.3390/app10186553

Boettcher, L. N., Hssayeni, M., Rosenfeld, A., Tolea, M. I., Galvin, J. E., & Ghoraani, B. (2020, July). Dual-task gait assessment and machine learning for early-detection of cognitive decline. In 2020 *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 3204-3207). IEEE. doi: 10.1109/EMBC44109.2020.9175955.

Bouton, C. E., Shaikhouni, A., Annetta, N. V., Bockbrader, M. A., Friedenberg, D. A., Nielson, D. M. & Sharma, G., et al. (2016). Restoring cortical control of functional movement in a human with quadriplegia. *Nature*, 533(7602), 247-250. doi:10.1038/nature17435

Bringas, S., Salomón, S., Duque, R., Lage, C., & Montaña, J. L. (2020). Alzheimer's disease stage identification using deep learning models. *Journal of Biomedical Informatics*, 109. https://doi.org/10.1016/j.jbi.2020.103514

Brookmeyer, R., Johnson, E., Ziegler-Graham, K., & Arrighi, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimer's & dementia*, 3(3), 186-191. https://doi.org/10.1016/j.jalz.2007.04.381

Buchanan, B. G. (2005). A (very) brief history of artificial intelligence. *Ai Magazine*, 26(4), 53-53. https://doi.org/10.1609/aimag.v26i4.1848

Canavan, C., Maguire, L. P., & Bucholc, M. (2021, August). Development of a Two-State Gaussian Hidden Markov Model for Modelling Dementia Progression in Patients with Mild Cognitive Impairment. In 2021 *IEEE 9th International Conference on Healthcare Informatics (ICHI)* (pp. 113-119). IEEE. doi: 10.1109/ICHI52183.2021.00028.

Casamitjana, A., Petrone, P., Tucholka, A., Falcon, C., Skouras, S., Molinuevo, J. L. & Vilaplana, V., et al. (2018). MRI-based screening of preclinical Alzheimer's disease for prevention clinical trials. *Journal of Alzheimer's Disease*, 64(4), 1099-1112. doi: 10.3233/JAD-180299

Crous-Bou, M., Minguillón, C., Gramunt, N., & Molinuevo, J. L. (2017). Alzheimer's disease prevention: from risk factors to early intervention. *Alzheimer's research & therapy*, 9, 1-9. https://doi.org/10.1186/s13195-017-0297-z

Danso, S. O., Muniz-Terrera, G., Luz, S., & Ritchie, C. (2019). Application of Big Data and Artificial Intelligence technologies to dementia prevention research: an opportunity for low-and-middle-income countries. *Journal of Global Health*, *9*(2). doi: 10.7189/jogh.09.020322

Dasta, J. F., Greer, M. L., & Speedie, S. M. (1992). Computers in healthcare: overview and bibliography. Annals of pharmacotherapy, 26(1), 109-117. https://doi.org/10.1177/106002809202600120

de la Fuente Garcia, S., Ritchie, C. W., & Luz, S. (2019). Protocol for a conversation-based analysis study: PREVENT-ED investigates dialogue features that may help predict dementia onset in later life. *BMJ open*, *9*(3). http://dx.doi.org/10.1136/bmjopen-2018-026254

Deckers, K., Barbera, M., Köhler, S., Ngandu, T., van Boxtel, M., Rusanen, M., & Rusanen, M., et al. (2020). Long-term dementia risk prediction by the LIBRA score: a 30-year follow-up of the CAIDE study. *International Journal of Geriatric Psychiatry*, 35(2), 195-203. doi: 10.1002/gps.5235

DeTure, M. A., & Dickson, D. W. (2019). The neuropathological diagnosis of Alzheimer's disease. *Molecular neurodegeneration*, 14(1), 1-18. https://doi.org/10.1186/s13024-019-0333-5

Dubois, B., Padovani, A., Scheltens, P., Rossi, A., & Dell'Agnello, G. (2016). Timely diagnosis for Alzheimer's disease: a literature review on benefits and challenges. *Journal of Alzheime''s disease*, *49*(3), 617-631. doi: 10.3233/JAD-150692

European Commission (2020). Details of the EU funding programme which ended in 2020 and links to further information. *Horizon 2020*. https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-2020_en

Farina, D., Vujaklija, I., Sartori, M., Kapelner, T., Negro, F., Jiang, N. & Bergmeister, K., et al. (2017). Man/machine interface based on the discharge timings of spinal motor neurons after targeted muscle reinnervation. *Nature biomedical engineering,* 1(2). https://doi.org/10.1038/s41551-016-0025

Gruber, T.R (1993). A translation approach to portable ontology specifications. *Knowl. Acquis*. 5, 199–220.

Haulath, K., & Mohamed Basheer, K. P. (2023). TT self-weighted Deep-AD3-Net: An AD stage and risk prediction. *International Journal of Healthcare Management*, 1-13. https://doi.org/10.1080/20479700.2023.2175414

Heising, L., & Angelopoulos, S. (2021, January) (2021). Early diagnosis of mild cognitive impairment with 2-dimensional convolutional neural network classification of magnetic resonance images. In 2021 54th Hawaii International Conference on System Sciences, pp. 3407-3415. URI: http://hdl.handle.net/10125/71030.

Heising, L., & Angelopoulos, S. (2022). Operationalising fairness in medical AI adoption: detection of early Alzheimer's disease with 2D CNN. *BMJ Health & Care Informatics,* 29(1). doi: 10.1136/bmjhci-2021-100485

Hu, Z., Wang, Z., Jin, Y., & Hou, W. (2023). VGG-TSwinformer: Transformer-based deep learning model for early Alzheimer's disease prediction. *Computer Methods and Programs in Biomedicine*, 229. https://doi.org/10.1016/j.cmpb.2022.107291

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W. & Peterson R. C., et al. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. The Lancet Neurology, 9(1), 119-128. https://doi.org/10.1016/S1474-4422(09)70299-6

James, C., Ranson, J. M., Everson, R., & Llewellyn, D. J. (2021). Performance of machine learning algorithms for predicting progression to dementia in memory clinic patients. *JAMA network open*, *4*(12). doi:10.1001/jamanetworkopen.2021.36553

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S. & Wang, Y., et al. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4). http://dx.doi.org/10.1136/svn-2017-000101

Karaman, B. K., Mormino, E. C., Sabuncu, M. R., & Alzheimer's Disease Neuroimaging Initiative. (2022). Machine learning based multi-modal prediction of future decline toward Alzheimer’s disease: An empirical study. *Plos one*, *17*(11). https://doi.org/10.1371/journal.pone.0277322

Khalsa, D. S. (2015). Stress, meditation, and Alzheimer’s disease prevention: where the evidence stands. *Journal of Alzheimer's Disease*, 48(1), 1-12. doi: 10.3233/JAD-142766

Khan, A., & Usman, M. (2015, November). Early diagnosis of Alzheimer's disease using machine learning techniques: A review paper. In 2015 *7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)* (Vol. 1, pp. 380-387). IEEE.

Li, Q., & Yang, M. Q. (2021). Comparison of machine learning approaches for enhancing Alzheimer’s disease classification. *PeerJ*, *9*. doi: 10.7717/peerj.10549

Li, Q., Wu, X., Xu, L., Chen, K., Yao, L., & Li, R. (2017). Multi-modal discriminative dictionary learning for Alzheimer's disease and mild cognitive impairment. *Computer methods and programs in biomedicine*, 150, 1-8. https://doi.org/10.1016/j.cmpb.2017.07.003

Lin, M., Gong, P., Yang, T., Ye, J., Albin, R. L., & Dodge, H. H. (2018). Big data analytical approaches to the NACC dataset: aiding preclinical trial enrichment. *Alzheimer disease and associated disorders*, 32(1), 18. doi: 10.1097/WAD.0000000000000228

Liu, L., Yu, B., Han, M., Yuan, S., & Wang, N. (2019). Mild cognitive impairment understanding: an empirical study by data-driven approach. *BMC bioinformatics*, *20*(15), 1-13. https://doi.org/10.1186/s12859-019-3057-1

Liu, M., Li, F., Yan, H., Wang, K., Ma, Y., Alzheimer’s Disease Neuroimaging Initiative & Shen, L., et al. (2020). A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer’s disease. *Neuroimage*, *208*. https://doi.org/10.1016/j.neuroimage.2019.116459

Lukowicz, P. (2008). Wearable computing and artificial intelligence for healthcare applications [Editorial]. Artificial Intelligence in Medicine, 42(2), 95–98. https://doi.org/10.1016/j.artmed.2007.12.002

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7), 939-944. https://doi.org/10.1212/WNL.34.7.939

Menschik, E. D., & Finkel, L. H. (1998). Neuromodulatory control of hippocampal function: towards a model of Alzheimer's disease. *Artificial intelligence in medicine*, 13(1-2), 99-121. https://doi.org/10.1016/S0933-3657(98)00006-2

Mishra, R., & Li, B. (2020). The application of artificial intelligence in the genetic study of Alzheimer's disease. *Aging and disease*, *11*(6), 1567. doi: 10.14336/AD.2020.0312

Nagele, R. G., Wegiel, J., Venkataraman, V., Imaki, H., Wang, K. C., Wegiel, J. (2004). Contribution of glial cells to the development of amyloid plaques in Alzheimer's disease. *Neurobiology of aging*, 25(5), 663-674. https://doi.org/10.1016/j.neurobiolaging.2004.01.007

Nagumo, R., Zhang, Y., Ogawa, Y., Hosokawa, M., Abe, K., Ukeda, T. & Sumi, S., et al. (2020). Automatic detection of cognitive impairments through acoustic analysis of speech. *Current Alzheimer Research*, *17*(1), 60-68. https://doi.org/10.2174/1567205017666200213094513

Naz, S., Ashraf, A., & Zaib, A. (2022). Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset. *Multimedia Systems*, *28*(1), 85-94. https://doi.org/10.1007/s00530-021-00797-3

Page, MJ., McKenzie, JE., Bossuyt, PM., Boutron, I., Hoffmann, TC., Mulrow, CD., et al. (2020) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021. Retrieved from: http://prisma-statement.org/prismastatement/flowdiagram.aspx

Patel, U. K., Anwar, A., Saleem, S., et al. (2021). Artificial intelligence as an emerging technology in the current care of neurological disorders. *Journal of neurology*, *268*(5), 1623-1642. https://doi.org/10.1007/s00415-019-09518-3

Pekkala, T., Hall, A., Lötjönen, J., Mattila, J., Soininen, H., Ngandu, T. & Laatikainen, T., et al. (2017). Development of a late-life dementia prediction index with supervised machine learning in the population-based CAIDE study. *Journal of Alzheimer's Disease*, *55*(3), 1055-1067. doi: 10.3233/JAD-160560

Pelka, O., Friedrich, C. M., Nensa, F., Mönninghoff, C., Bloch, L., Jöckel, K. H. & Schramm, S., et al. (2020). Sociodemographic data and APOE-ε4 augmentation for MRI-based detection of amnestic mild cognitive impairment using deep learning systems. *Plos one*, 15(9). https://doi.org/10.1371/journal.pone.0236868

Pereira, U., & Brunel, N. (2018). Attractor dynamics in networks with learning rules inferred from in vivo data. *Neuron*, 99(1), 227-238. https://doi.org/10.1016/j.neuron.2018.05.038

Peters, M. D., Marnie, C., Tricco, A. C., et al. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI evidence synthesis,* 18(10), 2119-2126. doi: 10.11124/JBIES-20-00167

Richard, E., Jongstra, S., Soininen, H., Brayne, C., van Charante, E. P. M., Meiller, Y. & van der Groep, B., et al. (2016). Healthy ageing through internet counselling in the elderly: the HATICE randomized controlled trial for the prevention of cardiovascular disease and cognitive impairment. *BMJ* open, 6(6), e010806. http://dx.doi.org/10.1136/bmjopen-2015-010806

Richard, E., van Charante, E. P. M., Hoevenaar-Blom, M. P., Coley, N., Barbera, M., van der Groep, A. & Meiller, Y., et al. (2019). Healthy ageing through internet counselling in the elderly (HATICE): a multinational, randomised controlled trial. *The Lancet Digital Health*, 1(8), e424-e434. https://doi.org/10.1016/S2589-7500(19)30153-0

Rodrigues, P. M., Bispo, B. C., Garrett, C., Alves, D., Teixeira, J. P. & Freitas, D. (2021). Lacsogram: A new EEG tool to diagnose Alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, *25*(9), 3384-3395. doi: 10.1109/JBHI.2021.3069789.

Ruengchaijatuporn, N., Chatnuntawech, I., Teerapittayanon, S., Sriswasdi, S., Itthipuripat, S., Hemrungrojn, S. & Bunyabukkana, P., et al. (2022). An explainable self-attention deep neural network for detecting mild cognitive impairment using multi-input digital drawing tasks. *Alzheimer's Research & Therapy*, *14*(1), 1-11. doi: https://doi.org/10.1186/s13195-022-01043-2

Shen, Z., Schutte, D., Yi, Y., Bompelli, A., Yu, F., Wang, Y., & Zhang, R. (2022). Classifying the lifestyle status for Alzheimer's disease from clinical notes using deep learning with weak supervision. *BMC medical informatics and decision making*, *22*(1), 1-11. https://doi.org/10.1186/s12911-022-01819-4

Shimoda, A., Li, Y., Hayashi, H., & Kondo, N. (2021). Dementia risks identified by vocal features via telephone conversations: A novel machine learning prediction model. *PloS one*, 16(7). https://doi.org/10.1371/journal.pone.0253988

Shin, J., Lee, S. Y., Kim, S. H., Kim, Y. B., & Cho, S. J. (2008). Multitracer PET imaging of amyloid plaques and neurofibrillary tangles in Alzheimer's disease. *Neuroimage*, 43(2), 236-244. https://doi.org/10.1016/j.neuroimage.2008.07.022

Sliwinska, S., & Jeziorek, M. (2021). The role of nutrition in Alzheimer's disease. *Roczniki Państwowego Zakładu Higieny*, 72(1). doi: 10.32394/rpzh.2021.0154

Small, G., & Jarvik, L. (1982). The dementia syndrome. *The Lancet*, 320(8313), 1443-1446. https://doi.org/10.1016/S0140-6736(82)91339-3

Stonnington, C. M., Wu, J., Zhang, J., Shi, J., Bauer III, R. J., Devadas, V. & Su, Y., et al. (2021). Improved prediction of imminent progression to clinically significant memory decline using surface multivariate morphometry statistics and sparse coding. *Journal of Alzheimer's Disease*, *81*(1), 209-220. doi: 10.3233/JAD-200821

Ten Kate, M., Redolfi, A., Peira, E., Bos, I., Vos, S. J., Vandenberghe, R. & Gabel, S., et al. (2018). MRI predictors of amyloid pathology: results from the EMIF-AD Multimodal Biomarker Discovery study. *Alzheimer's research & therapy*, 10(1), 1-12. https://doi.org/10.1186/s13195-018-0428-1

Turing, A. M. (2009). *Computing machinery and intelligence* (pp. 23-65). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6710-5_3

WHO (2022, September). *Dementia*. World Health Organization. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/dementia

Wu, X., Peng, C., Nelson, P. T., & Cheng, Q. (2022). Machine Learning Approach Predicts Probability of Time to Stage-Specific Conversion of Alzheimer's Disease. *Journal of Alzheimer's Disease*, (Preprint), 1-13. doi: 10.3233/JAD-220590

Yang, S. T., Lee, J. D., Huang, C. H., Wang, J. J., Hsu, W. C., & Wai, Y. Y. (2010, August). Computer-aided diagnosis of Alzheimer's disease using multiple features with artificial neural network. In *PRICAI 2010: Trends in Artificial Intelligence: 11th Pacific Rim International Conference on Artificial Intelligence, Daegu, Korea, August 30–September 2, 2010. Proceedings 11* (pp. 699-705). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15246-7_72

Zhang, X. X., Tian, Y., Wang, Z. T., Ma, Y. H., Tan, L., & Yu, J. T. (2021). The epidemiology of Alzheimer's disease modifiable risk factors and prevention. *The journal of prevention of Alzheimer's disease*, 8, 313-321. https://doi.org/10.14283/jpad.2021.15

Zhang, Y., Zhou, M., Liu, T., Lanfranchi, V., & Yang, P. (2022, July). Spatio-temporal Tensor Multi-Task Learning for Predicting Alzheimer's Disease in a Longitudinal study. In 2022 *44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 979-985). IEEE. doi: 10.1109/EMBC48229.2022.9870882.

Zheng, G., Zhang, Y., Zhao, Z., Wang, Y., Liu, X., Shang, Y. & Cong, Z., et al. (2022). A transformer-based multi-features fusion model for prediction of conversion in mild cognitive impairment. *Methods*, *204*, 241-248. https://doi.org/10.1016/j.ymeth.2022.04.015

Zhu, J. D., Huang, C. W., Chang, H. I., Tsai, S. J., Huang, S. H., Hsu, S. W. & Lee, C. C., et al. (2022). Functional MRI and ApoE4 genotype for predicting cognitive decline in amyloid-positive individuals. *Therapeutic Advances in Neurological Disorders*, 15. https://doi.org/10.1177/17562864221138154

Attachments

Attachment 1. Frequencies for scientific journals where each paper was published

journal

		Frequency	Percent
Valid	Plos One	3	12,5
	BMC medical informatics and decision making	1	4,2
	Computer Methods and Programs in Biomedicine	2	8,3
	Journal of Alzheimer's Disease	4	16,7
	PeerJ	1	4,2
	Journal of Biomedical Informatics	1	4,2
	Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)	2	8,3
	Neuroimage	1	4,2
	Therapeutic Adances in Neurological Disorders	1	4,2
	Alzheimer's Research & Therapy	1	4,2
	Methods	1	4,2
	BMC bioinformatics	1	4,2
	Current Alzheimer Research	1	4,2
	BMJ open	1	4,2
	Multimedia Systems	1	4,2
	International Journal of Healthcare Management	1	4,2
	International Conference on Healthcare Informatics (ICHI)	1	4,2
	Total	24	100,0

Attachment 2. Frequencies for population used in each investigation

population

		Frequency	Percent
Valid	MCI	3	12,5
	AD	1	4,2
	CN, MCI	6	25,0
	CN, AD	1	4,2
	MCI, AD	1	4,2
	CN, MCI, AD	4	16,7
	not mentioned	8	33,3
	Total	24	100,0

Attachment 3. Frequencies for AI technology used in each investigation

AI_tech

		Frequency	Percent
Valid	ANN	2	8,3
	NLP	1	4,2
	CNN	4	16,7
	XGBoost	1	4,2
	RNN	1	4,2
	SVM	2	8,3
	WNN	1	4,2
	not specified DL	3	12,5
	not specified ML	6	25,0
	HMM	1	4,2
	DSI	1	4,2
	DNN	1	4,2
	Total	24	100,0

Attachment 4. Frequencies for types of data used in each investigation

type_data

		Frequency	Percent
Valid	Demographics, APOE4 genotype, clinical assessment, cognitive assessment, biomarkers	1	4,2
	Clinical notes (EHRs)	1	4,2
	brain MRI	6	25,0
	Clinical and demographic	2	8,3
	mobility	1	4,2
	audio	3	12,5
	mobility and cognitive	1	4,2
	neurpsychological tests/cognitive tasks	2	8,3
	neuropsychological tests and MRI	2	8,3
	cognitive, psychosocial and pshysical	1	4,2
	neuropsychological tests, vascular factors, sociodemopgraphic, APOE4 genotype	1	4,2
	MRI and PET	1	4,2
	MRI, sociodemographic and genotype (APOE4)	1	4,2
	MRI and clinical	1	4,2
	Total	24	100,0

Attachment 5. Frequencies for where is cognitive data taken into account in each investigation

cog_data

		Frequency	Percent
Valid	not mentioned	8	33,3
	to characterize the cognitive status of the participants	10	41,7
	to feed the model	6	25,0
	Total	24	100,0