A Machine Learning Architecture to detect Alzheimer's Disease Progression based on the Evaluation of Cognitive and Functional Attributes of Neuropsychological Assessments

Fadi Abdeljaber

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)

Health and Human Sciences
Department of Psychology
University of Huddersfield
Huddersfield, UK

February 2022

DECLARATION

I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other university for a degree.

Copyright Statement

i. The author of this thesis (including any appendices and/ or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching.

ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

ACKNOWLEDGMENTS

I would like to thank God almighty for the strength, patience, health, and knowledge given to me to pursue my studies. I would like also to thank my supervisor Professor David Peebles for the ongoing help and support during my research studies. Without his guidance and support this work would not have been possible. My thanks also to Dr Jenny Retzler for her feedback in the early stages of my PhD.

Special thanks go to my mother, Tina, Timmy, Zo, and the rest of my family for their spiritual support.

DEDICATION

This thesis is dedicated to my father who always believed in me and my capabilities and was a giver all his life; love you dad and I will always pray for you.

PUBLICATIONS

Thabtah F., Peebles D., Retzler J, Hathurusingha C. (2020) Dementia medical screening using mobile applications: A systematic review with a new mapping model. Journal of Biomedical Informatics, Volume 111, 2020,103573, ISSN 1532-0464,https://doi.org/10.1016/j.jbi.2020.103573.

Thabtah F., Peebles D., Retzler J, Hathurusingha C. (2020). A review of dementia screening tools based on Mobile application. Health and Technology, 10(5), 1011-1022. https://doi.org/10.1007/s12553-020-00426-5.

Thabtah F. Peebles D. (2022). A Machine Learning Algorithm for Detecting Alzheimer's Disease Progression. To be submitted to the journal of Journal of Biomedical Informatics.

Thabtah F., Spencer R., Peebles D. (2022). Common Dementia Screening Methods: DSM-5 Fulfilment and Mapping to Cognitive Domains. Behavioural and Healthcare Research Journal. Vol. 8, No. 1-2, pp. 104-120.

Thabtah F. Swan Y., Peebles D. (2022) Detection of Dementia Progression from Functional Activities Data using Machine Learning Techniques. To Appear at the Journal of Intelligent Decision Technologies.

Contribution Statement:

I was the main author for the published work in this thesis; the contributions of Prof. Peebles, and Dr Retzler were mainly guidance and feedback on the quality of the work; they provided useful feedbacks and suggestions that improved the quality of the publications. The rest of the authors contributed to validating the mapping between the mobile apps and the medical tests besides validating the correlation between cognitive tests and cognitive domains, and normalizing the ranking figures.

ABSTRACT

With an aging population and the increased burden and economic cost of dementia on the community, coupled with an ongoing pandemic, it is becoming critical to develop a faster, cheaper, reliable way of diagnosing and screening for dementia and its progression. Dementia is a condition associated with memory decline, cognitive impairment, and difficulties in language, problem-solving, and sometimes functional impairment. Pre-diagnosis of common dementia conditions such as Alzheimer's disease (AD) in the initial stages is crucial to help in early intervention, treatment plan design, disease management, and for providing quicker healthcare access. Current pathological assessments are often physically invasive, psychologically stressful, and their availability in poor countries and rural areas is low. In addition, many neuropsychological assessments are time-consuming, rarely cover all cognitive domains involved in AD diagnosis, and they do not measure the individual's cognitive and functional abilities together. Therefore, the design and implementation of an intelligent method for AD progression from few cognitive and functional items in a manner that is accessible, easy, affordable, quick to perform, and does not require special and expensive resources, is desirable. This thesis investigates the issue of AD progression based on cognitive and functional items using machine learning to offer good performance (accuracy, time, etc.) and accessibility besides providing valuable knowledge for clinicians during the clinical assessment.

The thesis proposes a Machine Learning Architecture for Alzheimer's Disease Progression (MLA-ADP), which contains a novel classification algorithm called Alzheimer's Disease Class Rules (AD-CR). The proposed AD-CR algorithm learns models from the distinctive feature subsets that contain rules with low overlapping among their cognitive and functional items yet are easy to be interpreted by the clinicians during clinical assessment. More importantly, our research investigated cognitive elements, functional abilities and their correlations during dementia progression using computational intelligence, and it was able to identify sets of key cognitive and functional items within neuropsychological assessments. The cognitive items mainly covering the cognitive domains of learning and memory, and language, when processed by machine learning techniques, produced models that performed well and showed superior models using the AD-CR algorithm. As for functional items, demographic features must be included with functional items for AD progression detection, at least when using machine learning techniques. Overall, cognitive items appear to be more influential, the AD-CR algorithm model was still able to generate results across the dissimilar evaluation metrics that are within the standard medical research.

To measure the performance of the dementia progression models, extensive experiments have been conducted using classification methods on the Disease Neuroimaging Initiative data repository (ADNI) datasets. Results obtained by ten-fold cross validation showed that fewer cognitive and functional items can be processed by the AD-CR algorithm generating models that maintain adequate performance in terms of accuracy, sensitivity, and specificity. The models derived by the AD-CR algorithm are competitive in terms of accuracy, specificity, and sensitivity rates. For the cognitive subsets, processing three items by the AD-CR algorithm with the addition of demographic information derives models with 91.25% accuracy, 89.50% sensitivity, and 92.90% specificity. Whereas for

functional items, when four items are processed by the AD-CR algorithm, the models derived show 87.57% accuracy, 86.70% sensitivity, and 88.30% specificity.

TABLE OF CONTENTS

Title ..... *i*
Declaration and Copyright ..... *ii*
Acknowledgments ..... *iii*
Dedication ..... *iv*
Publication ..... *v*

ABSTRACT ..... **VI**

CHAPTER ONE **5**

INTRODUCTION **5**

1.1 INTRODUCTION 5
1.2 THE PROBLEM 8
1.3 SCOPE, AIMS, AND RESEARCH QUESTIONS 9
1.4 DEMENTIA AND CAUSES 111
1.5 RESEARCH GAPS AND CONTRIBUTIONS 122
1.5.1 Detecting Disease Progression *122*
1.5.2 Interpretable Models *133*
1.5.3 Classification Performance *13*
1.5.4 Time and Accessibility *144*
1.5.5 Composite Elements *155*
1.5.6 Cognitive and IADL Correlations *166*
1.5.7 Non-Overlapping Models *177*
1.5.8 Mapping to DSM-5's Domains and Cognitive Digital Tools Review *177*
1.6 MACHINE LEARNING: AN INTRODUCTION 188
1.7 THESIS OUTLINE 200

CHAPTER TWO **22**

LITERATURE REVIEW **22**

2.1 INTRODUCTION 222
2.2 DEMENTIA DIAGNOSIS PROCESS, AND REQUIREMENTS 233
2.3 COMMON DEMENTIA COGNITIVE ASSESSMENT METHODS 255
2.3.1 Alzheimer's Disease Assessment Scale - Cognitive Subscale (ADAS Cog) *255*
2.3.2 Clinical Dementia Rating Scale Sum of Boxes (CDR-SB) *266*
2.3.3 Everyday Cognition (ECog) *266*
2.3.4 Functional Activities Questionnaire (FAQ) *277*
2.3.5 Mini Mental State Examination (MMSE) *28*
2.3.6 Montreal Cognitive Assessment (MoCA) *28*
2.3.7 Rey's Auditory Verbal Learning Test (RAVLT) *29*
2.4 DEMENTIA ASSESSMENT METHODS COMPARISON 29
2.4.1 DSM-5 Fulfilment *300*
2.4.2 Performance *322*
2.4.3 Accessibility and Administration *322*
2.4.4 Cognitive-Functional Diagnostic Procedures *333*
2.5 RECENT LITERATURE ON COGNITIVE AND FUNCTIONAL ABILITIES IN DEMENTIA DIAGNOSIS AND SCREENING 377
2.5.1Discussion *432*
2.6 REVIEW ON MACHINE LEARNING STUDIES IN COGNITIVE DEMENTIA ASSESSMENT METHODS 433
2.6.1 Decision Trees and Rule-based Classification *43*
2.6.2 Support Vector Machines *466*

2.6.3 Probabilistic Classification *477*
2.6.4 Artificial Neural Networks *49*
2.6.5 Regression and Statistical-based Models *511*
2.6.6 Discussion *515*
2.6.6.1 Advanced Learning Techniques for Diagnostic Decisions *515*
2.6.6.2 Dementia Data *516*
2.7 MOBILE DEMENTIA ASSESSMENT METHODS 599
2.7.1 Single Assessment Methods *599*
2.7.2 Multiple Assessment Methods *60*
2.7.3 Non-conventional Assessment Methods *62*
2.7.4 Discussion on Digital Assessment Methods *655*
2.7.4.1 DSM-5 Neurocognitive Domain Coverage *655*
2.7.4.2 Validity and Reporting *659*
2.7.4.3 Performance *73*
2.7.4 The role of intelligent Methods *76*
2.8 CHAPTER SUMMARY 777

CHAPTER THREE 799

A MACHINE LEARNING ARCHITECTURE FOR ALZHEIMER'S DISEASE PROGRESSION 799

3.1 INTRODUCTION 799
3.2 PROPOSED MACHINE LEARNING ARCHITECTURE FOR ALZHEIMER'S DISEASE PROGRESSION: MLA-ADP 80
3.3 DATA UNDERSTANDING, ACCESS, PREPARATION, AND MODELLING 833
3.3.1 ADNI Data Access *833*
3.3.2 Cognitive and Functional Data Understanding *844*
3.3.3 Data Integration *866*
3.3.4 Data Modelling and Data Balancing *888*
3.4 FEATURE SELECTION 90
3.5 THE PROPOSED CLASSIFICATION ALGORITHM 933
3.5.1 Class Association Rule *963*
3.5.2 Class Association Terms *966*
3.5.3 Learning Phase *977*
3.5.4 Classification Phase *999*
3.6 MOBILE APPLICATION DESIGN AND CONTENT 100
3.7 CHAPTER SUMMARY 104

CHAPTER FOUR 1066

DATA, METHODS USED, AND EXPERIMENTAL SETTINGS 1066

4.1 INTRODUCTION 1066
4.2 DATASETS 1077
4.2.1 Unprocessed Datasets *1077*
4.2.2 Processed Datasets *11515*
4.3 PLATFORMS, EXPERIMENTAL SETTING, AND METHODS USED 1188
4.3.1 Platforms Used and Settings *1188*
4.3.2 Classification Methods Used *1199*
4.4 CHAPTER SUMMARY 12222

CHAPTER FIVE 12323

EXPERIMENTS AND RESULTS ANALYSIS 12323

5.1 INTRODUCTION 12323
5.2 FEATURE SELECTION EXPERIMENTS 12323
5.3 COGNITIVE FEATURE SELECTION RESULTS & ANALYSIS 12525

5.4 Functional Feature Selection Results Analysis 13030
5.5 Combined Cognitive-Functional Feature Selection Results Analysis 13535
5.6 Classification Results Analysis 1366
5.6.1 Cognitive Classification Results Analysis 13737
5.6.2 Functional Classification Results Analysis 14545
5.7 Chapter Summary 1522

CHAPTER SIX 15353

CONCLUSIONS AND IMPLICATIONS 15353

REFERENCES 16060

LIST OF FIGURES

Figure 1.1: Supervised Learning Algorithm Lifecycle 20
Figure 3.1: The Proposed Machine Learning Architecture for Alzheimer's Disease Progression (MLA-ADP) . 81
Figure 3.2: Data Preparation Processes of the MLA-ADP 833
Figure 3.3: Modelling Process of the Data 89
Figure 3.4: The AD-CR Algorithm 98
Figure 3.5: The Proposed Dementia Pre-diagnosis Application Navigation 102
Figure 3.6a: Landing Screen 103
Figure 3.6b: Information Screen 1 103
Figure 3.6c: Information Screen 2 103
Figure 3.6d: Sample Question A 104
Figure 3.6e: Sample Question B 104
Figure 3.6f: Output screen 104
Figure 4.1: Age Distribution of the Participants 109
Figure 4.2: Frequency of the Patients' Visits 109
Figure 4.3: Class Labels' Distribution at Baseline and at Last Examination Visit 110
Figure 4.4: Gender Distribution at Baseline Diagnosis and at Last Examination Visit Diagnosis 111
Figure 4.5: Age Distribution at Baseline Diagnosis and at Last Examination Visit Diagnosis 11111
Figure 4.6: Gender and Age Distribution at Baseline Diagnosis and at Last Examination Visit Diagnosis Respectively 11111
Figure 4.7a: Number of Data Subjects Progressed from CN to MCI or MCI to Dementia 11212
Figure 4.7b: Gender Distribution of the Data Subjects Progressed from CN to MCI or MCI to Dementia 11212
Figure 4.8: Age Distribution of the Data Subjects Progressed from CN to MCI or MCI to Dementia 11313
Figure 5.1: Correlation Coefficient Matrix of Cognitive Items 1266
Figure 5.2a: Feature-Feature Correlation Matrix of Cognitive Items for CN to MCI Group 1299
Figure 5.2b: Feature-Feature Correlation Matrix of Cognitive Items for MCI to AD Group 1299
Figure 5.3: Correlation Coefficient Matrix of Functional Items 13130
Figure 5.4a: Feature-Feature Correlation Matrix of Functional Items for CN to MCI Group 13333
Figure 5.4b: Feature-Feature Correlation Matrix of Functional Items for MCI to AD Group 134
Figure 5.5: Feature-Feature Correlation Matrix of Combined Cognitive and Functional Items for all Dementia Stages 13636
Figure 5.6a: Performance of the Classification Methods against MCI-AD Cognitive Groups with and without Demographics Respectively ..... **Error! Bookmark not defined.**42
Figure 5.6b: Performance of the Classification Methods against CN-MCI Cognitive Groups with and without Demographics, Respectively ..... **Error! Bookmark not defined.**42
Figure 5.7a: Performance of the Classification Methods against CN-MCI Functional Groups with and without Demographics, Respectively 15050

FIGURE 5.7B: PERFORMANCE OF THE CLASSIFICATION METHODS AGAINST MCI-AD FUNCTIONAL GROUPS WITH AND WITHOUT DEMOGRAPHICS, RESPECTIVELY 15050

LIST OF TABLES

TABLE 2.1: MAJOR ND AND MINOR ND DESCRIPTION 244
TABLE 2.2: AD DIAGNOSTIC CRITERIA BASED ON DSM-5 244
TABLE 2.3: COGNITIVE DIAGNOSTIC PROCEDURES DETAILS .....**ERROR! BOOKMARK NOT DEFINED.**4
TABLE 2.4: COGNITIVE DOMAIN COVERAGE 355
TABLE 2.5: MAPPING OF PROCEDURE SECTIONS TO COGNITIVE DOMAINS 366
TABLE 2.6: SUMMARY OF ALL CONSIDERED DIGITAL DEMENTIA SCREENING METHODS AVAILABLE ON A MOBILE PLATFOR 633
TABLE 2.7: NUMBER OF DOMAINS COVERED BY DIGITAL DEMENTIA METHODS 67
TABLE 2.8: VALIDITY DETAILS OF THE CONSIDERED DEMENTIA DIGITAL SCREENING METHODS 71
TABLE 2.9: REPORTED SENSITIVITIES AND SPECIFICITIES OF MEDICAL EXAMINATIONS USED BY THE APP 744
TABLE 3.1: ADAS-COG ITEMS AND THEIR ASSOCIATED DSM-5 COGNITIVE DOMAINS 866
TABLE 3.2: GENERAL STATISTICS OF THE DATASETS BEFORE PRE-PROCESSING 877
TABLE 3.3: MAPPING SCORES IN THE FAQ DATASET AND FAQ SCALE VALUES 877
TABLE 3.4: FAQ ITEMS AND THEIR ASSOCIATED DSM-5 COGNITIVE DOMAINS 101
TABLE 4.1: SAMPLE OF THREE DATA SUBJECTS OF ADNI-MERGE WITH DIAGNOSTIC CLASS (DX) 1088
TABLE 4.2: SAMPLE OF 3 DATA SUBJECTS FROM THE ADAS-COG-SHEET DATASET WITH 16 ATTRIBUTES 11414
TABLE 4.3: SAMPLE OF 3 DATA SUBJECTS FROM THE FAQ-SHEET DATASET WITH 19 ATTRIBUTES..... 11414
TABLE 4.4: GENERAL STATISTICS AFTER DATA PRE-PROCESSING AND DATA BALANCING..... 11515
TABLE 4.5: GENERAL STATISTICS FOR THE GROUPS OF PARTICIPANTS WITHIN 36 MONTHS FROM THE BASELINE AND 1177
TABLE 4.6: SAMPLE OF 5 DATA SUBJECTS FROM THE PROCESSED AND INTEGRATED DATASET WITH 16 ATTRIBUTES 1177
TABLE 5.1: SUMMARY OF THE METHODS USED TO DERIVE EACH NEUROPSYCHOLOGICAL SUBSET 12525
TABLE 5.2: SUMMARY OF THE COGNITIVE ITEMS FOR EACH DATA SUBSET 12525
TABLE 5.3A: COGNITIVE ITEMS WITH COMPUTED SCORES AND NORMALISED SCORES DERIVED BY THE FEATURE SELECTION METHODS (ADAS-SUBSET3 AND ADAS-SUBSET4)..... 1277
TABLE 5.3B: COGNITIVE CLUSTERS IDENTIFIED BASED ON THE % DROP BETWEEN THE RANKED ITEMS ACCORDING TO THEIR NORMALISED AVERAGE SCORES (ADAS-SUBSET3 AND ADAS-SUBSET4) 1288
TABLE 5.4: SUMMARY OF THE DERIVED FUNCTIONAL ITEMS FOR EACH SUBSET OF FAQ..... 13131
TABLE 5.5A: FUNCTIONAL ITEMS WITH COMPUTED SCORES AND NORMALISED SCORES DERIVED BY THE FEATURE SELECTION METHODS (FUNC-SUBSET3 AND FUNC-SUBSET4) 13232
TABLE 5.5B: CLUSTER ITEMS IDENTIFIED BASED ON THE % DROP BETWEEN THE FUNCTIONAL ITEMS ACCORDING TO THEIR NORMALISED AVERAGE SCORES (FUNC-SUBSET3 AND FUNC-SUBSET4)..... 13232
TABLE 5.6: PERFORMANCE OF THE CLASSIFICATION METHODS FROM DIFFERENT SUBSETS OF THE COGNITIVE ITEMS..... 1388
TABLE 5.7: SAMPLE COGNITIVE RULES DERIVED BY THE AD-CR ALGORITHM FROM THE COGNITIVE DATA SUBSETS..... 14545
TABLE 5.8: PERFORMANCE OF THE CLASSIFICATION METHODS FROM DIFFERENT SUBSETS OF THE FUNCTIONAL ITEMS ... 1468
TABLE 5.9: SAMPLE COGNITIVE RULES DERIVED BY THE AD-CR ALGORITHM FROM THE COGNITIVE DATA SUBSETS..... 15151

Chapter One

Introduction

This chapter introduces dementia, the thesis aims, objectives, and research questions, the problems considered, and thesis contributions. The chapter also introduces basic concepts related to machine learning. Most of this chapter's content has been disseminated in the Journal of Biomedical Informatics, the journal of Health and Technology, and the Journal of Behavioural and Healthcare Research.

1.1 Introduction

Dementia is a neurodegenerative disorder characterised by difficulties in memory, disturbance in language, psychological and psychiatric changes, and impairments in activities of daily living, and mainly occurring in the elderly (Burns & Iliffe, 2009; Kim et al., 2020). There are 50 million people worldwide living with dementia, a condition that has a devastating impact on a person's physical and psychological status, as well as damaging effects on the society and the economy (World Health Organization, 2020). This number has been predicted to rise considerably in the future if there is no improvement in preventive interventions (Prince et al., 2013; Wimo et al., 2003). In 2007, Wimo et al. (2007) estimated the total worldwide cost of dementia in 2005 to be US$315 billion. In 2013, they updated this figure to US$604 billion for 2010, and in 2017 assessed it as being to US$818 billion for 2015 (Wimo et al., 2013; Wimo et al., 2017). According to Dementia Statistics (2021), the cost of dementia in the UK is £25 billion annually, and globally US$817 billion.

In 2019 in the United Kingdom, there was an estimated 885,000 people with dementia, and this is projected to surge by 80% to approximately 1.6 million people by 2040 (Wittenberg et al., 2019). With this growing prevalence of dementia and the associated costs, there has also been a call for an increase in research initiatives (Pickett et al., 2018). Topics for research include preventing future cases of dementia, bettering the quality of

life for people with dementia, enhancing and enabling dementia health and social care systems and the dementia workforce, and the focus of this thesis: improving the process of dementia pre-diagnosis using intelligent data-driven approaches. To date, there is no cure or an effective treatment for the disease. Particularly during a pandemic where focus and resources are diverted, it is crucial that we devise a fast, affordable, and reliable way for screening and prognosis using innovative technologies such as machine learning and artificial intelligence.

Recently, machine techniques have been explored in the neurological and behavioural science arenas, such as for the detection of dementia, to enhance the accuracy/efficiency of conventional medical assessments (Clemmensen et al., 2020; Chua et al., 2019; Jammeh et al., 2018; Bang et al., 2017; Balsis et al., 2015). The reason for adopting machine learning technology in medical applications is because it efficiently explores data to derive classification systems and predictive models; these offer clinicians and medical professionals rich information and knowledge for making better quality decisions. Machine learning models can explain influential cognitive and neuropathological features related to the diagnosis of dementia, which can help in early detection, therefore rapid intervention and medical resource accessibility for the patients and caregivers (Xefteris et al., 2020; Kemp et al., 2020; Gupta & Katarya, 2020; Youn et al., 2018; Weakley et al., 2015).

According to Wessels et al. (2015), identifying a few cognitive items that cause progression of the disease can assist in early intervention. While pathological assessments of Alzheimer's disease (AD) diagnosis such as using biological markers (biomarkers), cerebrospinal fluid (CSF), and magnetic resonance imaging (MRI) can be used to predict the disease, they are also time and cost intensive, stressful, and with results requiring laboratory study and professional personnel who may not be available (Zhu et al., 2020; Battista et al., 2017). Cognitive psychological assessments such as the Functional Activity Questionnaire (FAQ), the Alzheimer's Disease Assessment Scale-Cognitive 13 (ADAS-Cog-13), Mini Mental State Examination (MMSE), and others (Pfeffer et al., 1982; Rosen et al., 1984; Folstein et al., 1975), are useful methods that can screen for signs of early impairment. These assessments are usually easier to carry out than pathological procedures and show acceptable performance with reference to validity, sensitivity, and specificity (Zhu et al., 2020; Pereira et al., 2018; Teng et al., 2010). However, few

research studies have measured the progression of AD using both functional and cognitive features, i.e. Jutten et al. (2020), Battista et al. (2017), and Shahbaz et al. (2019).

The above studies acknowledged that early detection and distinguishing the stages of AD were important for appropriate treatment plan purposes, but they did not consider each cognitive item to identify disease progression to assist early intervention. While Battista et al. (2017) used machine learning to assess cognitive and behavioural measures, their study focused on diagnosing AD and not the progression of the disease, which is more challenging, and it did not consider the neuropsychological criteria defined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) framework (American Psychiatric Association, 2013). Mapping between the assessed cognitive items and DSM-5's criteria would have been helpful for the clinicians' understanding of linking the machine learning results and actual diagnosis of the disease.

To fill the gap, the aims of this research are:

- To improve neuropsychological methods' performance in terms of dementia progression detection rate using machine learning
- To understand cognitive and functional elements and their mapping to neurocognitive conditions like dementia according to the DSM-5 framework
- To assess the associations among cognitive items as well as functional items within neuropsychological assessments using computational intelligence
- To identify key neuropsychological items that can primarily trigger the advancement of AD experimentally using real data observations from the Alzheimer's Disease Neuroimaging Initiative data repository (ADNI) (ADNI, 2021)
- To propose a new classification algorithm with the ability to not only improve predictive performance of the disease progression detection, but also provide clinicians with a useful and easy-to-understand knowledge base derived from real data related to the dementia stages.

The scope of our research is limited to neuropsychological assessments—results related to other pathological procedures, neuroimaging, or biomarkers are excluded. To achieve the aim, we propose a Machine Learning Architecture for Alzheimer's Disease Progression (MLA-ADP) that incorporates a novel classification method and models

cognitive-functional activities from the ADNI data repository. In this thesis, we attempt to contribute the following benefits to clinicians:

1. The ability to predict the progression of AD objectively using influential sets of features and machine learning to model and analyse the data subjects.
2. To disseminate a rule-based classification model comprised of easy-to-interpret rules on the association of the cognitive and functional items to the DSM-5 framework as a knowledge base toolkit that can be used during early screening.
3. To improve the classification accuracy of AD progression when using models derived objectively from cases and controls by machine learning methods.
4. To incorporate the learnt classification models into digital platforms such as mobile for accessibility and fast and accurate screening for quick referrals.
5. To discover a few impactful cognitive items that are crucial to disease progression so clinicians can use this knowledge in early screening.
6. To present functional items related to the instrumental activities of daily living (IADL) that are key to detect the disease progression so that clinicians can use this knowledge to improve the design of future functional assessments.

1.2 The Problem

Early dementia diagnosis, also known as dementia screening, can be defined as the process by which an individual, who might be in the prodromal stage of dementia, is verified as a case of dementia using neuropsychological assessment (Panegyres et al., 2016). The prognosis of AD progression has been recognised as a challenging problem due to the massive number of cognitive and pathological features recorded for patients and controls. While many studies have investigated the diagnosis of dementia using these characteristics, predicting the advancement of the disease has not been heavily studied, particularly using technologies such as artificial intelligence and machine learning.

Two intercorrelated areas that are normally assessed by clinicians for determining dementia are cognitive and functional abilities. For example, the MMSE is one of the screening methods for cognitive impairments in dementia, while the FAQ is an established method used for measuring functional performance (Folstein et al., 1975; Pfeffer et al., 1982). Cognition involves remembering, thinking, problem solving,

decision making, judgment, and knowing, among others—all mental processes that are associated with activities that an individual performs routinely (Jutten et al., 2019).

Functional abilities involve basic functions (ADL) such as walking, dressing, and bathing as well as more complex functions (IADL) such as preparing a multi-recipe meal, managing finances, and shopping, besides others (Pfeffer et al., 1982). According to Barberger-Gateau et al. (1992), functional ability or status can be defined as the ability of the individual to perform physical tasks, self-manage, and self-care. Functional ability, one of the criteria used to assess quality of life, enables individuals to maintain functional independence (Hoffmann et al., 2013). When patients with dementia start to exhibit functional loss in their daily activities, they experience negative consequences such as hospitalisation, the need for more medical care and services, depression, and hardship in personal care, among others (Brown et al., 2014).

Since the problem is to detect the progression of AD using specific neuropsychological and IADL activities, we consider it as a supervised learning problem in data science in which a classification algorithm is used to construct classification models. These models in turn are exploited by the clinicians to detect any possible advancement of the AD using a few, yet effective, items. The models are tested automatically to evaluate their effectiveness in terms of multiple performance metrics such as accuracy, specificity, and sensitivity. In this research, we assess sets of functional and cognitive items separately, and then consider integrating these sets to show fewer items that, when processed by a classification technique, can detect AD advancement, if any.

The proposed data driven architecture will be implemented and evaluated in a digital mobile platform to increase accessibility and reduce costs associated with resources. Part of the problem under consideration is the process of mapping between the key cognitive items to the DSM-5 framework, which is used to describe the standard criteria to diagnose neurocognitive disorders.

1.3 Scope, Aims, and Research Questions

One of the aims of this research project is to improve the performance of screening methods for early dementia and Mild Cognitive Impairment (MCI) advancements using machine learning technology. A further aim is to determine influential neuropsychological features based on computational intelligence methods from real data

observations. We intend to design and implement an architecture that consists of a classification algorithm plus feature selection that can distinguish between dementia stages using models extracted from real ADNI data subjects.

To narrow the scope of work, we focus on two primary domains in the neuropsychological assessments of AD: cognitive and functional. We use two accepted psychological methods that emphasise cognition and functional activities: the ADAS-Cog-13, and the FAQ. The selection of ADAS-Cog-13 is primarily because this method has been used widely to capture in cognitive-related research cognitive impairment for MCI and dementia conditions such as AD for example by Tabatabaei-Jafari, et al. (2020), Nogueira et al. (2018), Kueper et al. (2018), and Liu-Seifert et al. (2015). ADAS-Cog-13 is the most common of the 31 versions of the ADAS cognitive assessment. More importantly, it has more features than other available cognitive methods in ADNI such as MMSE, potentially covering larger cognitive domains. Lastly, ADAS-Cog-13 is associated with enough real data observations in ADNI thus can be easily evaluated by a data-driven study such as ours. However, the FAQ method is one of the popular assessments to capture an individual's independence in daily living including functional activities, and it is the only functional method available in the ADNI data repository with a sufficient number of data observations.

Being able to identify the items that have the highest impact on the progression of AD will assist the clinician, not only in time and cost-savings, but in early detection and provide a quick-to-deploy information sheet on AD diagnosis. This will give a better understanding of how neuropsychological items can trigger the progression of AD.

The research questions that this research study aims to answer are:

- How can the neuropsychological items, at each dementia stage, if any, be determined using machine learning?
- Which is more influential in AD progression: functional or cognitive activities?
- Do functional or cognitive activities/items vary when the dementia stage changes?
- How can the performance of screening methods of AD progression be improved when providing a new classification method based on rules?

It should be noted that the Clinical Dementia Rating-Sum of Boxes (CDR-SOB) assessment method (Hughes et al., 1982) has not been used in the experimental analysis

of this research since it was primarily used to assign the class labels in ADNI project besides MMSE method; we therefore avoided any biased results, especially in feature-class correlations, by excluding the CDR-SB scores.

1.4 Dementia and Causes

As defined by the Mayo Clinic (2021), "dementia is a term used to describe a group of symptoms affecting memory, thinking, and social abilities, severely enough to interfere with daily life". Dementia does not refer to any one disease—there are many diseases which cause dementia, the most common of which is AD. This disease attributes between 60–80% of dementia cases to AD, 5–10% of cases to vascular dementia, 5–10% of cases to Lewy body dementia, and 5–10% of cases to frontotemporal dementia. It is possible to suffer from more than one of these types simultaneously, resulting in what is known as mixed dementia.

Nall (2017) further explains these common types of dementia. AD is a progressive form of dementia caused by brain cell death. Typical symptoms begin with a loss of short-term memory which progresses to confusion, mood changes, and difficulty in speaking and walking. Vascular dementia can be either progressive or present rapidly. It is caused by a lack of blood flow to the brain following a stroke or otherwise, and tends to result in confusion, disorientation, and affects the ability to concentrate and complete tasks. Lewy body dementia is caused by protein deposits in nerve cells. It shares similar symptoms with AD including a loss of memory and difficulty in walking. People with Lewy body dementia may also have difficulty falling asleep at night and often become disoriented. Frontotemporal dementia refers to several types of dementia which specifically affect the frontal and temporal lobes of the brain. This form of dementia usually affects people's motivation, behaviour, and speaking ability.

Many forms of dementia tend to progress in severity over a person's life, so it has been a focus of research to detect or even predict the development of dementia as early as possible (Knopman et al., 2003; Xefteris et al., 2020). MCI, which has been explored as a possible diagnostic precursor to dementia, is defined by Gauthier et al. (2006) as "cognitive decline greater than expected for an individual's age and education level but does not interfere notably with their activities of daily life". While cognitive decline is also consistent with dementia, MCI differs in that the cognitive decline is not yet severe

enough to interfere with the individual's daily life. Over half of the people with MCI deteriorate to the point of having dementia within five years, so MCI is considered as an early marker for possible dementia; some authors have introduced minor functional disability among the criteria for diagnosing MCI (Winblad et al., 2004).

With the arrival of the DSM-5, there has been a redefining of dementia and MCI to the new categories of Major Neurocognitive Disorder (Major ND) and Minor Neurocognitive disorder (Minor ND). This change in terminology reflects a distancing from the stigma associated with dementia and a shift towards placing more emphasis on other cognitive areas besides memory (Crisis Prevention Institute, n.d.). As with MCI, Minor ND differs from Major ND in that people with Minor ND can retain independence with daily activities. Details on the diagnostic criteria of dementia based on the DSM-5 are discussed in Chapter 2.

1.5 Research Gaps and Contributions

1.5.1 Detecting Disease Progression

In the early stages of the condition (mild dementia) or before establishing a prognosis of the condition (MCI), most of the cognitive methods have difficulty in detecting any progression due to floor or ceiling effects in the scores (Mura et al., 2014; Podhorna et al., 2016). Subsequently, patients whose cognitive test scores are close to the cut-off that differentiates dementia stages are hard to classify, resulting in high false positives and false negatives. Therefore, the composite methods used by Harrison et al. (2013) and Jutten et al. (2017) can be considered promising directions. However, such methods are limited and do not measure the progression of the disease as they only deal with detecting dementia traits by conventional neuropsychological methods.

To deal with the above issue, cognitive elements and functional features can be combined in a single data repository with real cases and controls. A computational approach that considers feature-feature and feature-class assessments can then be conducted to distinguish types of dementia. In this thesis, we use a combined approach of feature selection that examines multiple feature selection methods besides clustering to pinpoint limited, yet non-overlapping, cognitive and functional subsets related to the dementia

subcategories. These subsets can assist the clinician during a clinical setting to differentiate possible progression points.

1.5.2 Interpretable Models

Limited data-driven research studies related to dementia have considered building a knowledge base with simple-to-understand rules for clinicians and other stakeholders to exploit (Das et al., 2019). Having interpretable classification models that can be converted into rules will offer the patient and their families answers as to which components have triggered dementia progression. This is aligned with the patient's right outlined in the General Data Protection Regulation (GDPR), particularly the section on decision-making using automated algorithmic methods and the 'right for an explanation' (Goodman & Flaxman, 2016; GDPR, n.d.).

Furthermore, the GDRP requires that for any data collected from a subject (the individual undergoing dementia screening) in an automated decision process (the screening process using the machine learning), that the subject should have the right to be given the rationale behind the decision-making process. Consequently, having an intelligent and easy-to-understand classification system that provides information to the patients and their family members, besides clinicians, is advantageous. The system can articulate to the different stakeholders useful information to answer many questions.

We propose within MLA-ADP a rule-based classifier called Alzheimer's Disease Class Rules (AD-CR) to derive understandable If-Then rules, and more importantly, to learn models for predicting the progression status of anyone undergoing pre-diagnosis. The rules derived by the AD-CR can be exploited by clinicians to understand correlations among features and AD progression, while its classification method can predict changing points of disease advancement.

1.5.3 Classification Performance

Healthcare professionals, in particular clinicians, use neuropsychological assessments to assess the presence of dementia conditions such as AD. These methods, including the ADAS-Cog, MMSE, and CDR-SOB (Rosen et al., 1984; Folstein et al., 1975; Hughes et

al., 1982) produce indicative scores based on the patient's learning, memory, executive function, comprehension, social cognition, etc. Generally, these methods have shown acceptable diagnostic performance in terms of sensitivity and specificity (Liu-Seifert et al., 2015; Carrion et al., 2018). However, the quality of these cognitive methods for detecting gradual changes in the condition remains a challenge (Bossers et al., 2012; Jutten et al., 2017).

We propose within MLA-ADP architecture a classification algorithm that learns models objectively from real data for predicting the disease advancement of any participants undergoing the dementia screening process. The algorithm contains a novel rule discovery method that not only reduces the number of rules significantly, but also performs better in terms of predictive accuracy, sensitivity, and specificity rates. Empirical results reveal that the proposed AD-CR algorithm within the MLA-ADP architecture outperforms conventional methods besides common classification algorithms including SVM, ANN, statistical, rule-based classifiers, and probabilities. Chapter 4 provides details of the classification algorithms' details.

1.5.4 Time and Accessibility

A deficiency of neuropsychological assessments is that the assessment time can be lengthy, which may be a burden for the patients and possibly result in a loss of concentration and fatigue (Jutten et al., 2017; Ong, 2017)—some patients may feel tired and stop before completing the assessment. In addition, many cognitive activities in neuropsychological tests require direct interaction with the patient in a clinical setting which is difficult in the era of pandemics such as Covid 19.

Previously, a study by the European Task Force concentrated on a few cognitive domains that could potentially trigger the progression of AD in the early stages (Vellas et al., 2008). Our research fits within the scope of the European Task Force by not only identifying a few, albeit influential, cognitive elements and daily functions that may be key performance indicators for the progression of the condition, but by providing these elements within an interactive digital platform to expedite the screening process and improve accessibility for patients, caregivers, and clinicians. The proposed MLA-ADP can be implemented in a cloud-based environment to provide clinicians with an accessible

and cost-effective platform since he/she can access the items of the assessment via mobile or tablet. The AD-CR models can be invoked whenever a prediction is needed.

1.5.5 Composite Elements

Some research studies have reported that a decline in cognitive skills can be used to red flag the possibility of developing functional impairment (Cipriani et al., 2020; Clemmensen et al., 2020; Lim et al., 2018; Liu-Seifert et al., 2015; Razani et al., 2009). Two of the requirements for diagnosing dementia/major ND are a substantial decline in one of the six cognitive domains defined in the DSM-5 framework, and that the cognitive deficits experienced by the individual interfere with their independence during everyday activities (American Psychiatric Association, 2013). Winblad et al. (2004) showed that not only does minor cognitive decline not interfere with everyday life, but that this decline is associated with minor functional impairment. These findings suggest that functional impairment should be used as a specific criterion to diagnose patients with AD.

Unfortunately, most of the existing neuropsychological methods used for dementia pre-diagnosis rarely measure cognitive and functional activities in a single assessment or establish DSM-5 criteria coverage. Both the functional and cognitive perspectives should be considered during the screening process for dementia in ageing individuals so that healthcare professionals can capture any concealed correlations. It is also imperative to distinguish between influential functional and cognitive items for the subgroups of dementias, especially prior to dementia conditions like MCI and mild dementia as these are more challenging than the later stages.

In the proposed architecture, we have not only assessed cognitive and functional elements separately, but also in combination to comprehensively cover larger sets of features that might directly prompt dementia advancement. Thus, unlike typical neuropsychological assessment, the proposed architecture takes into account both the individual's cognitive and functional areas so that clinicians are able to have a wider perspective on which elements may prompt any changing points in the disease, and for which dementia sub-groups.

1.5.6 Cognitive and IADL Correlations

Cognitive elements within existing neuropsychological methods, as suggested by some previous research studies, may be used to predict some daily functions (Barberger-Gateau et al., 1992; Cahn-Weiner et al., 2000; Sperling et al., 2011; Zahodne et al., 2013; Yu et al., 2018). A study by De Paula and Malloy-Diniz (2013) showed that memory and executive functions correlate with some functional deficiencies. However, the scores from cognitive methods partly capture the functional status which limits their relevance in clinical settings (Robin et al., 2012).

Additionally, many patients who exhibit cognitive impairment, particularly for MCI or mild dementia, may still have normal ability to function in their daily life. Moreover, Clemmensen et al. (2020) and Boyle et al. (2003) revealed that the executive function scores of some neuropsychological assessments can be used as indicators for functional performance of IADLs. Krall et al. (2014) in a longitudinal study, researched the possibility of a relationship between memory and functional performance. However, a functional and cognitive performance relationship in early-stage dementia is not well characterised as described by Zucchella et al. (2017) and Jutten et al. (2019).

The correlation between cognitive and functional areas at the pre-dementia and mild dementia stage is complex, as a particular function is typically assessed by multiple cognitive elements. For example, the ‘assembling tax records, business affairs, or other papers’ function necessitates cognitive processes related to complex attention, executive function, learning, and memory. Playing a game of skill, and working on a hobby also requires these as well as perceptual motor function. As considered in this research, it is then essential to capture the relationship between cognitive and functional performance defined in IADLs so that clinicians can predict a decline in functional abilities or cognition before it reflects in the patient’s quality of life. Specifically, we investigate whether a positive correlation exists between functional activities and cognitive tasks using computational intelligence. We not only evaluate cognitive elements during a dementia progression assessment experimentally, but also their impact on functions related to IADL, and for dissimilar sub-groups of participants in different dementia stages.

1.5.7 Non-Overlapping Models

Another major challenge in dementia pre-diagnosis methods is the overlapping of the cognitive elements as defined in the DSM-5 framework. More precisely, some activities that are captured using functional methods, such as the IADL, rely on a combination of cognitive functions. For instance, within the DSM-5 in the FAQ, managing finances such as writing cheques, paying bills, and balancing a cheque book covers multiple cognitive domains including executive functions, memory, and complex attention. Currently, and despite such functional methods being a key part of screening and diagnosing AD, there is no 'gold standard' for measuring the functional activities defined in IADL (Loewenstein et al.,1989). More importantly, it is difficult to reduce the overlapping between daily functions defined in IADL and cognitive elements in neuropsychological assessments in terms of the cognitive domains defined in the DSM-5—only a small set is needed for screening assessment. Crucial for clinicians is a screening method that minimises overlapping between functional and cognitive activities retaining the non-overlapping items only.

The classification models of the proposed AD-CR algorithm are derived from distinctive subsets of cognitive and functional activities. One principle used to select these subsets is minimising the items' overlapping in terms of the DSM-5's criteria for neurocognitive disorders. Therefore, models derived by the AD-CR algorithm contain rules with antecedents that may reduce the similarity among cognitive and functional items.

1.5.8 Mapping to DSM-5's Domains and Cognitive Digital Tools Review

There is a variety of cognitive diagnostic procedures that aim to measure a patient's cognitive ability in different areas. While these procedures have been used successfully to demonstrate the presence of dementia, there is no clear consensus as to where these procedures fit into the process of dementia diagnosis (Sachdev et al., 2014; Baldwin & Farias, 2009). Therefore, one of this thesis's aims is to critically analyse an array of commonly used cognitive diagnostic procedures to determine how they can be used specifically within the DSM-5 criteria for dementia diagnosis. We seek to identify the procedures that cover a wider range of diagnostic criteria based on the DSM-5 neurological areas related to dementia. This involves mapping tasks or sub-sections of

cognitive diagnostic procedures to their corresponding neurological areas. This can be a difficult task since neurological domains are not strongly defined, are hard to quantify, and diagnostic tasks or subsections can often overlap when measuring aspects of multiple neurological domains.

Additionally, since systematic reviews on digital tools for dementia cognitive methods are rare, we review digital tools that have utilised dementia diagnostic procedures with a focus on neuropsychological assessments built on a mobile platform. Currently, little research has been systematically conducted to evaluate the available tools with respect to inclusion criteria such as accessibility, fulfilment of DSM-5, and coverage of cognitive areas. The few available mobile-based review studies on dementia screening tools are not comprehensive, leaving an area which this research aims to address (e.g. Groppell et al., 2019; Berauk et al., 2017; Yamagata & Kowto, 2013). More importantly, we introduce a mapping between digital assessment tools and the neurocognitive domains to help clinicians identify the tools that cover multiple cognitive domains which will be more appropriate for dementia screening.

1.6 Machine Learning: An Introduction

With the advancements of computer networks, mobile applications, and web-based systems, access to massive data stored in various internal systems and cloud data storage necessitates smart methods of data exploration, for instance, innovative technologies such as Artificial Intelligence. Machine learning, the core of Artificial Intelligence, has emerged as a field of research which can discover new knowledge or organise information to not only make the computer intelligent to study human behaviour, but also to improve its own performance (Xue & Zhu, 2009). Consider, for example, image and text recognition applications where the system intelligently recognises a face or handwriting in an image or within text. In a stock trading application, predicting the volatility of a stock and the price affected by short-term movement, or the effectiveness of a treatment in healthcare can be assessed using machine learning.

Using such techniques, models from historical datasets are built as decision-making systems, many of which facilitate predicting an outcome based on the available features of the problem (Choi et al., 2020). Machine learning technology is applied widely to

forecast probable results by unveiling the hidden trends and smartly deciphering from the historical data, whether structured or unstructured, to make data-driven decisions (Thabtah, 2018). For instance, a retail store can use this technique to derive models which help managers to organise stock and find correlations among purchased items. In the medical field, machine learning can be used to investigate the effectiveness of the treatment for any illness, predict the disease progression—such as for cases of AD, screen illness early, and enhance medical service accessibility in mobile-health and digital-health systems. Usually, the process of training in a machine learning algorithm lifecycle requires minimal human intervention to produce desirable results from the historical datasets (Maglogiannis & Kotsiants, 2007).

Machine learning techniques are generally categorised into two types of learning: supervised learning and unsupervised learning. The former involves predicting a special feature called the target class based on a classification model learnt from an historical dataset that comprises a number of characteristic features (Janiesch et al., 2021). The problem under consideration in supervised learning is called 'classification'; loan approval, credit card scoring, and weather forecasting are examples of supervised learning problems. Unsupervised learning, which is not restricted to the target class, involves deriving useful information for decision-making from input datasets that have no target class. Tasks that are relevant to unsupervised learning include categorising customers based on demographics (cluster analysis), and discovering the correlations among items in transactional databases (association rule mining) (Gheware et al., 2014). Since this research focuses on predicting the progression of dementia, we limit the scope to supervised learning and in particular, classification.

The lifecycle of the supervised learning algorithms consists of three phases (see Figure 1.1):

1. Data Preparation: In this phase, the dataset is pre-processed by applying necessary data preparation methods such as data cleansing, data normalisation, data balancing, data modelling, feature selection, etc. In addition, when enough data are available, the historical dataset can be divided into training and testing subsets with an option to include a validation data subset.

2. Training: The training data subset is processed by using a learning algorithm to find and then produce a classification model. Usually, the learning algorithm discovers the relationships between the characteristic independent features and the target class in the dataset to build the classification model. During the training phase, the learning algorithm always adjusts the model to reach the one that can be generalised in predicting test data instances when the model is utilised later in the classification phase. Thus, adjusting the model during learning is a critical procedure to ensure that, a) the model is effective in prediction and can be generalised, and b) the model does not overfit the training data subset (Janiesch et al., 2021).
3. Testing (classification): The learnt model is assessed on an independent test data subset to check its performance using dissimilar evaluation metrics such as predictive accuracy, specificity, sensitivity, error date, precision, recall, etc. (Powers, 2011). When the performance of the learnt model shows good predictive power, then the model can be used by the end-user to assist in the decision-making process.

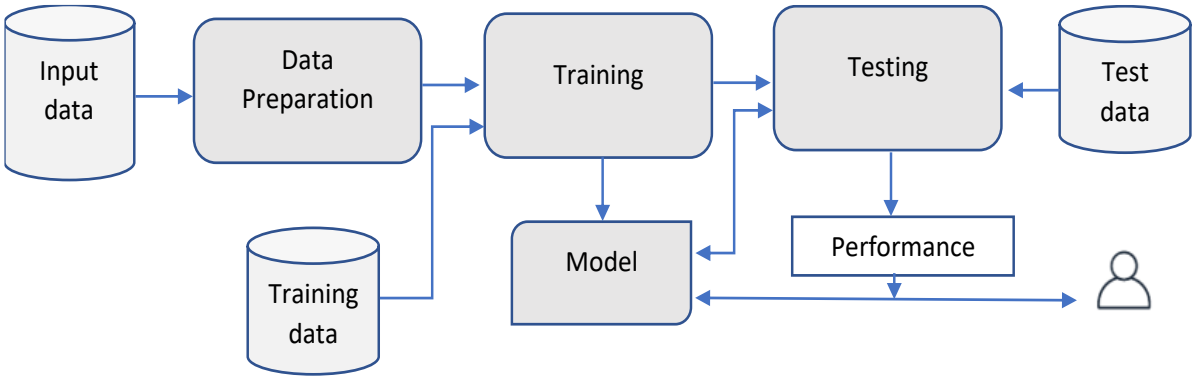

Figure 1.1: Supervised Learning Algorithm Lifecycle

1.6 Thesis Outline

This thesis consists of six chapters; Chapter 2 introduces common dementia neuropsychological methods, critically analyses the literature, and reviews machine learning approaches related to dementia prediction. In Chapter 3, the proposed MLA-ADP architecture is discussed and each of its components explained in detail. Chapter 4

demonstrates the data, the experimental settings, and the methods used. The experiments, results, and results analysis are discussed thoroughly in Chapter 5. Lastly, we conclude and highlight the future work and limitations of the thesis in Chapter 6.

Chapter Two

Literature Review

This chapter reviews dementia screening methods from a neuropsychological perspective, critically analyses related works on cognitive and functional elements of dementia besides their correlations, and considers mapping these elements into the DSM-5 framework. Digital screening methods of dementia are critically analysed, and machine learning studies on dementia prediction based on data-driven methodologies are reviewed. Most of this chapter's content has been disseminated in the Journal of Biomedical Informatics, the Journal of Health and Technology, and the Journal of Behavioural and Healthcare Research.

2.1 Introduction

In this chapter, we study functional abilities and cognitive domains and their correlations from a psychological perspective and provide relevant recent literature on cognitive and functional abilities defined in neuropsychological assessment theories. More importantly, we critically analyse common cognitive diagnostic procedures, show their advantages and disadvantages, and their mapping to the DSM-5 framework. Another aim of this chapter is to review and critically analyse recent literature on the cognitive and functional areas of dementia using intelligent techniques, mainly machine learning. The focus is on classification methods used in AD progression such as decision trees, rule induction, artificial neural networks (ANNs), statistical approaches such as logistic and linear regression, and probabilistic approaches. Machine learning studies that cover dementia classification or diagnosis aside from cognitive assessments are excluded since they are out of the scope of this research. Section 2.6 provides more details.

We have organised this chapter as follows: Section 2.2 sheds light on dementia and its diagnostic criteria. Sections 2.3 and 2.4 compare and critically analyse common dementia diagnostic methods in terms of the different criteria, and Section 2.5 reviews related

works on cognitive and functional elements within neuropsychological assessments, and the associations between cognitive domains and functional abilities. Section 2.6 explains recent research works of classification algorithms in dementia diagnosis with a narrow scope on cognitive assessments used in dementia diagnosis and screening. Section 2.7 systematically reviews digital dementia assessment methods that are based on a mobile platform and presents a new comparison based on dissimilar criteria. Lastly, we summarise the chapter in Section 2.8.

2.2 Dementia Diagnosis Process, and Requirements

The American Psychiatric Association (2013) in the DSM-5 describes the criteria for diagnosing possible and probable AD. First, some level of dementia or neurocognitive disorder must be established. Major ND requires that the patient experiences a significant decline over time in at least one of the following six cognitive domains: complex attention, executive function, learning and memory, language, perceptual-motor, or social cognition. This can be reported either by the patient, an informant, or the clinician, and then corroborated by performing a cognitive procedure that is associated with the domain. Secondly, the cognitive deficits experienced by the patient must interfere with their independence during everyday activities. Alternatively, minor ND requires a moderate decline in cognitive domains over time and that the patient's independence during everyday activities is not affected. Both Major ND and Minor ND also require that the cognitive defects are not only observed when the patient is delirious, and that the cognitive defects are not better explained by another mental disorder.

Once Major ND or Minor ND has been demonstrated, the DSM-5 criteria can be used to determine if the observed disorder is being caused by AD. In addition to the above criteria, there must be a gradual progression of impairment in one or more of the six cognitive domains mentioned earlier.

For a patient with Major ND, AD is probable if either there is evidence of AD from family history or genetic testing, or there is clear evidence of a decline in memory and learning as well as one other cognitive domain, and the cognitive decline must have a steady but gradual progression (see Tables 2.1 and 2.2). If either condition is not met in a Major ND patient, then AD is possible.

For a patient with Minor ND, AD is probable if there is evidence of AD from family history or genetic testing, otherwise AD is possible if there is clear evidence of a decline in memory and learning as well as one other cognitive domain, and the cognitive decline must have a steady but gradual progression.

Typically, when AD diagnosis is needed for research purposes, a more rigorous diagnosis criteria is required. In the ADNI study (ADNI, 2021), AD subjects were diagnosed as having probable AD according to the Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria (McKhann et al., 1984). This involves a MMSE score between 20 and 26 and a CDR-SB of 0.5 or 1.0.

Table 2.1: Major ND and Minor ND Description

Major ND	Significant cognitive decline in one or more cognitive domains.	The cognitive deficits interfere with independence in everyday activities.	The cognitive deficits do not exclusively occur in the context of delirium.	The cognitive deficits are not better explained by another mental disorder.
Minor ND	Modest cognitive decline in one or more cognitive domains.	The cognitive deficits do not interfere with independence in everyday activities.	The cognitive deficits do not exclusively occur in the context of delirium.	The cognitive deficits are not better explained by another mental disorder.

Table 2.*2*: AD Diagnostic Criteria based on DSM-5

			Major ND	**Minor ND**
Probable AD	Major or Mild Neurocognitive Disorders criteria are met.	Evidence of AD from family history or genetic testing Or: Clear evidence of decline in memory and learning and one other cognitive domain; steady gradual decline in cognition, and no evidence of other neurodegenerative disease.	Evidence of AD from family history or genetic testing.	
Possible AD	Major or Mild Neurocognitive Disorders criteria are met.	None of the above.	All three of the following: Clear evidence of decline in memory and learning and one other cognitive domain. Steady gradual decline in cognition. No evidence of other neurodegenerative disease.	

2.3 Common Dementia Cognitive Assessment Methods

2.3.1 Alzheimer's Disease Assessment Scale - Cognitive Subscale (ADAS Cog)

ADAS-Cog is a procedure designed to measure the level of a patient's cognitive dysfunction (Rosen et al., 1984). While its use for monitoring pre-dementia and MCI have been criticised (Kueper et al., 2018), it is generally accepted as one of the commonly-used procedures for assessing dementia. This test is typically administered by a professional and consists of structured and unstructured tasks: word recall, naming objects and fingers, commands, constructional praxis, ideational praxis, orientation, word recognition, language, comprehension of spoken language, word finding, and remembering test instructions. ADAS-Cog usually takes an hour to complete in a clinical setting and produces a score between 0 and 70, with 70 indicating the most severe cognitive dysfunction. In addition to the 11-question version, some variations exist: ADAS-Cog 13 additionally contains a delayed word recall section, a maze or number cancellation section, and is scored between 0 and 85 (Mohs et al., 1997). Monllau et al. (2007) tested the ADAS-Cog's ability to diagnose AD on a sample of 451 subjects (of which there were 254 control subjects with normal cognition, 86 with MCI and 111 with AD). They found that the best cut-off score for describing AD was >=12 which had a sensitivity of 89.19% and a specificity of 88.53%.

Mohs et al. (1997) deduced that the ADAS-Cog-11 did not assess attention and concentration, planning and executive function, verbal memory, nonverbal memory, and praxis. Therefore, the authors recommended to include, in addition to the ADAS-Cog-11 items, a test for delayed word recall and number cancellation which became the ADAS-Cog13.

Kueper et al. (2018) determined that the ability of ADAS-Cog-13 to identify disease progression was better than that of ADAS-Cog-11, subsequently pinpointing the various versions of ADAS-Cog assessment available making it difficult for cross-comparison of validity and reliability. This was useful to us in that it brought to our attention the various versions available so we could understand their differences. For our research, we are interested in the findings of ADAS-Cog-13 (Mohs et al., 1997)—our dataset retrieved from the ADNI uses this version.

A frequently-used assessment in validation studies is the ADAS-Cog13, where for MCI patients the accuracy is 82–83%, sensitivity 58–61% and specificity 91–93%; and for AD patients, the accuracy is 90.5–99.6%, sensitivity 74–94% and specificity 92–98% (Nogueira et al., 2018; Yang et al., 2019).

2.3.2 Clinical Dementia Rating Scale Sum of Boxes (CDR-SB)

After extensive research, Hughes et al. (1982) were unable to produce an appropriate scale to separate the stages of dementia. To address the issue, the authors introduced a new rating scale - the Clinical Dementia Rating Scale (CDR). The CDR consists of two sets of interrelated semi-structured interview guidelines/questionnaires, which incorporate an assessment of the following six cognitive domains: memory, orientation, executive (judging and problem solving), community affairs, home and hobbies, and personal care.

In the CDR method, an informant is interviewed ahead of the subject to form a baseline understanding of the subject's cognitive degradation level. The clinician validates the subject's answers by comparing to the informant's information, then assigns a score of 0 (normal cognitive function), 0.5 (very mild cognitive impairment), 1 (mild cognitive impairment), 2 (moderate cognitive impairment) or 3 (severe cognitive impairment) for each category. The final sum of the scores (SB: Sum of Boxes) ranges from 0 (normal) to 18 (severe), where a score less than 4 will indicate very mild dementia or light cognitive degradation. A mildly-demented subject will likely score between 4.5 and 9, while a moderately-demented subject will score between 9.5 and 15.5; anyone scoring over 16 is considered severely demented (Mennella & Heering, 2015). Within the study, sensitivity and specificity have not been disclosed, however, in another validation study conducted by O'Bryant et al. (2010), 80% sensitivity and 69% specificity have been reported.

2.3.3 Everyday Cognition (ECog)

ECog (Farias et al., 2008) is a questionnaire which was created with the goal of increasing sensitivity, especially for the MCI group, over pre-existing cognitive diagnostic procedures. Based on input from domain experts such as clinicians and neurologists, the ECog initially had 138 potential questions which was reduced to a final total of 39 questions with four possible responses to each question:

1 = better or no change compared to 10 years earlier

2 = questionable/occasionally worse

3 = consistently a little worse

4 = consistently much worse.

There are two versions of ECog: an informant-based version (ECogSP) where the questions are answered on behalf of the patient by their caregiver, and a patient-based version (ECogPT) where the subject answers the questions.

The questions can be divided into six categories: Memory, language, semantic knowledge, visual spatial, planning, organisation and divided attention. All questions also count towards a general category, which is a function of all other categories. Each category is then rated on a scale of 1– 4, with 4 representing the most severe decline in everyday function. The authors reported that when targeting cut off points for a specificity of 80%, sensitivity was 93% in discriminating dementia from Cognitively Normal (CN), 75% in discriminating MCI from dementia, and a 67% in discriminating MCI from CN.

Recently, Thabtah et al. (2020) investigated the difference between ECog versions experimentally using correlation analysis with classification on data subjects from the Alzheimer's Disease Prediction of Longitudinal Evolution (TADPOLE) challenge - ADNI dataset (Marinescu et al., 2019). The results derived by the authors using a few machine learning algorithms revealed that scores of the ECogSP for data subjects with AD at baseline showed an increase over the period when compared to those of the ECogPT. Moreover, classification results showed an increase in the detection rate of the AD progression when ensemble learners were used, and after the input dataset was resampled.

2.3.4 Functional Activities Questionnaire (FAQ)

To access the living independence level of elderly in a community setting, Pfeffer et al. (1982) considered Lawton and Brody's (1969) IADL Scale, which was initially developed for the assessment of independent living for injured patients, to be relevant. Based on the foundation of IADL, the authors developed the FAQ which contains questions that are the grouped daily tasks for independent living. Each question is rated

on a scale of 0–3, where lower scores indicate better autonomy. Scores are then calculated for a total score between 0–30. The cut-off point is 9 for determining if one has impaired functioning. Within the study, FAQ reported 85% in sensitivity while IADL achieved only 57%, however, in regard to specificity, IADL achieved 92% when FAQ reported 81%.

Wessels et al. (2015), Podhorna et al. (2016), Marshall et al. (2015), and Battista et al. (2017) have consistently referenced their findings to cognitive domains, particularly executive function, memory, and attention, to provide the association to the DSM-5 framework.

2.3.5 Mini Mental State Examination (MMSE)

Folstein et al. (1975) created the MMSE with a shorter administration time than other cognitive diagnostic procedures. MMSE consists of 11 questions to address the issue of a shorter period of mental concentration in elderly patients and is administered in a clinical setting. In comparison, other mental state tests usually assess the full mental state and take more than 30 minutes. The authors focused on assessing the cognitive functions of perception, visual, memory, language, and fine motor skills. MMSE has a total score of 30, and scoring less than 24 indicates some level of cognitive impairment.

The authors validated the screening method on a sample of 269. Of these, 63 were healthy subjects and 206 were patients who exhibited various types of abnormal mental symptoms, including dementia. Within the initial publication of MMSE, there is no information on its sensitivity and specificity. However, in 2000, a validation study with 151 subjects in Greece reported 90.8% sensitivity and 90.6% specificity with the cut off score 23/24 (Fountoulakis et al., 2000), as specified by Folstein et al. (1975).

2.3.6 Montreal Cognitive Assessment (MoCA)

MMSE is known for its insensitivity in separating MCI subjects from normal subjects (Ihl et al., 1992; Tombaugh & McIntyre, 1992; Wind et al., 1997), which is why Nasreddine et al. (2005) created MoCA as a screening instrument for detecting MCI and Mild AD. If a subject scores within normal ranges on the MMSE, but is still experiencing memory issues, further testing in MoCA can help determine if they have MCI or Mild AD. The

authors designed 12 structured and unstructured questions to complete, including activities of drawing a cube and trail making, to assess the subjects' attention, abstraction, executive, memory, orientation, and visuospatial capabilities in a clinical setting. Within the study, MoCA was able to separate 90% of MCI subjects whereas MMSE only managed to separate 18% (Nasreddine et al., 2005). The impairment cut off score was set to 26 where the maximum score was 30; the study reported a specificity level of 87%, the sensitivity for discriminating MCI and Mild AD was 90% and 100%, respectively.

2.3.7 Rey's Auditory Verbal Learning Test (RAVLT)

RAVLT (Rey, (1941); Schmidt, (1996)) is a test where the subject is given a set of 15 nouns which they are asked to recall. They are given the nouns again and asked to recall these a total of five times. Next, they are given a second set of 15 different nouns but asked to again recall the first set. The subjects are then given a 30 minute break and asked to recall the first set of nouns. Finally, the subjects are asked to identify the original 15 nouns from a set of 50 nouns including the first and second sets.

The RAVLT test produces a number of different summary scores, of which two are used in the ADNI dataset (Moradi et al., 2017; ADNI, 2017). *RAVLT immediate* is the total number of nouns recalled in the first five repetitions, giving a score that ranges from 0 to 75. *RAVLT percent forgetting* is the proportion of nouns remembered in the 5th trial that are forgotten after 30 minutes (the score of the 5th repetition, minus the score after the 30 minute break, divided by the score of the 5th repetition, and multiplied by 100). *RAVLT percent forgetting* can in theory range -1400 (assuming at least 1 is recalled in the 5th trial since you can't divide by 0) and 100, with negative scores meaning the subject remembered more nouns after the break and 100 meaning the subject forgot all nouns after the break. *RAVLT immediate* is typically associated with learning memory and *RAVLT percent forgetting* with delayed memory.

2.4 Dementia Assessment Methods Comparison

In the cognitive diagnostic procedures selected (see Table 2.3), it is obvious that the intentions of each procedure are different, since the clinical researcher's study focus varied. For instance, the MMSE has been used as a dementia screening tool since its

inception in 1975, and while MoCA is focused on differentiating MCI patients from AD, the MMSE is known for its insensitivity (Ihl et al., 1992; Tombaugh & McIntyre, 1992; Wind et al., 1997; Nasreddine et al., 2005). CDR-SB and ECog are considered as dementia diagnosis procedures, which segregate the subjects into different stages of dementia progression, such as mild MCI, MCI, moderate dementia, and severe dementia. RAVLT is unique in that it is not designed primarily for the screening of dementia however; it is an effective tool for measuring the cognitive domain associated with dementia: memory. As such, its use has been explored as a possible component of robust dementia screening toolkits (Ghafar et al., 2019). In the following subsections we critically analyse the dementia assessment methods in terms of DSM-5 coverage, performance, and accessibility among others.

2.4.1 DSM-5 Fulfilment

The DSM 5 criteria do not require that any specific cognitive diagnostic procedures be used in the process of AD diagnosis, however cognitive tests within these procedures can be used to measure decline in one or more of the six cognitive domains: complex attention, executive function, learning and memory, language, perceptual-motor, or social cognition. Decline in at least one of these domains is required by DSM 5 for a diagnosis of Major or Minor ND, and decline in learning and memory as well as at least one other domain is required for AD. The severity of decline can determine whether the neurocognitive disorder is Major or Minor and if AD is probable or possible. Therefore, it is beneficial to use cognitive tests which are quantified and equitable to measure decline in cognitive domains. Table 2.4 depicts which of the six cognitive domains is covered by each diagnostic procedures, and Table 2.5 depicts which sections/cognitive tests of each diagnostic procedure relate to each cognitive domain.

We have used the definition of each cognitive domain according to the Johns Hopkins Psychiatry Guide (Peters & Rabins, 2017). Complex attention includes sustained, divided, and selective attention, ability to concentrate and to perform mental calculations. Executive function includes planning, decision making, and ability to organise. Learning and memory includes immediate, recent, and long-term memory. Language includes remembering names of objects or people and using correct grammar. Perceptual motor includes visual perception and ability to navigate familiar environments and perform

spatial tasks. Social cognition includes the recognition of emotions, theory of mind, and an awareness of social standards.

RAVLT was the only cognitive test we reviewed that covers one cognitive domain only, being learning and memory, however it can be useful to individually measure different aspects of the learning and memory domain such as immediate recall and delayed recall. All other considered cognitive diagnostic procedures could be applied to more than one cognitive domain, although there were no cognitive diagnostic procedures that covered all six domains. ECog had the largest coverage including all domains apart from social cognition. In fact, CDR-SB was the only procedure found that could be considered to cover social cognition as part of the 'community affairs' test. While this assesses behaviours likely to relate to social cognition, such as reactions in social situations and consideration of others, it does not directly assess the socio-emotional or mentalising skills usually considered central to the social cognition domain. Nor does the CDR-SB cover complex attention or language. To cover all cognitive domains in DSM-5, you would need to combine CDR-SB with another cognitive diagnostic procedure such as ECog or alternatively ADAS-Cog, MMSE, or MoCA which each cover all cognitive domains apart from social cognition and executive function.

It is worth highlighting that while some of the cognitive diagnostic procedures include cognitive tasks to measure functioning in DSM domains directly, others rely on descriptions of routine behaviours that we assume will be impacted by poor cognitive functioning. For example, the ADAS-Cog measures perceptual and motor skills directly by asking patients to copy geometric shapes. In contrast, the CDR-SB 'home and hobbies' test asks about the patient's abilities to conduct chores and routine tasks and to use appliances. Functioning impairment in these tasks could indicate difficulties with perceptual and motor skills, but may also reflect issues in executive functioning.

The cognitive domains required for diagnosing possible or probable AD according to the NINCDS-ADRDA criteria differ slightly: replacing social cognition with orientation and adding constructive abilities and problem solving. Therefore, a different set of cognitive diagnostic procedures is required. All cognitive diagnostic procedures mentioned here aside from RAVLT also cover orientation. ADAS-Cog covers constructive abilities and CDR-SB can evaluate a subject's problem-solving ability.

2.4.2 Performance

Table 2.3 shows the performance of cognitive diagnostic procedures in terms of sensitivity and specificity based on critical reviews. Sensitivity refers to the ability to correctly classify subjects with AD, specificity refers to the ability to correctly classify subjects without AD. In general, sensitivity and specificity are inversely proportional to one another as the cognitive procedure's cut off point is changed. Therefore, cut off points are chosen carefully to maximise both measures as needed.

The results below show that MoCA was the highest performing procedure, with a sensitivity of 90% towards MCI subjects and a sensitivity of near 100% towards AD subjects, as well as a specificity of 87% for both MCI and AD. For some of the original studies, there were no disclosed sensitivity, specificity, or cut off scores. In these cases, we have found validation studies or comparison studies which provide such information (O'Bryant et al., 2010; Hsu et al., 2017; Li et al., 2017). For RAVLT, we were unable to locate any studies specifically linked to dementia diagnosis with an overall sensitivity and specificity. The likely reason is that RAVLT was not comprehensive enough to use for dementia screening or diagnosis due to its narrow scope. For the rest, CDR-SB seems to perform worse with only 80% sensitivity and 69% sensitivity when the others were all above 85% sensitivity and 80% specificity. ECog performed better than CDR-SB in discriminating dementia from the control, with a 93% sensitivity and 80% specificity.

2.4.3 Accessibility and Administration

Most of these cognitive diagnostic procedures require clinicians or trained clinical professionals to administer; FAQ is the only cognitive test which can be performed in a community setting.

Since most of these cognitive diagnostic procedures, except ECog and RAVLT, were constructed to have less than 15 items, the administration duration is reasonably short. The 11 tasks of ADAS Cog are usually administered over a period of 60 minutes—CDR-SB usually takes 40 minutes. MMSE and MoCA are the quickest of the cognitive diagnostic procedures described in this thesis with an administration time of up to 15 minutes.

Since demonstrating cognitive decline in memory and learning and one other cognitive domain is the requirement for possible and probable AD according to the DSM-5 criteria, it would be beneficial to begin testing with a quick cognitive diagnostic procedure such as MMSE or MoCA. This can be used to demonstrate impairment in memory and learning and three of the other cognitive domains and could be followed by use of ADAS-Cog13 or CDR-SB if there is found to be an impairment of memory and learning but no impairment in complex attention, language and perceptual motor skill.

2.4.4 Cognitive-Functional Diagnostic Procedures

Research studies have shown that functional performance has been underrated within neuropsychological dementia tests which do not cover most of the functioning demonstrated in the ADL or IADL (Bossers et al., 2012; Brown et al., 2014; Jutten et al., 2019). ADL regression often occurs in later stages of AD; thus it is less reliant on functions related to cognition. However, some cognitive elements are related to IADL decline as reported by multiple studies such as by Clemmensen et al. (2020), Allen et al. (2009) and De Paula and Malloy-Diniz, (2013). Despite the fact that these dementia diagnostic procedures can show clear cognitive impairment, they do not necessarily reveal deficits in functional activities (Mioshi, 2011). Therefore, there is a need to develop new dementia diagnostic procedures that measure both cognitive and functional sections. The next section reviews studies that investigate functional and cognitive elements together in dementia diagnosis research.

Table 2.3: Cognitive Diagnostic Procedures Details											
Procedure	**Name**	**Type**	**No. of Questions**	**Cut Off Score**	**Min Score**	**App. Time**	**Max Score**	**Sensitivity**	**Specificity**	**Administration Settings**	**Activities**
ADAS	Alzheimer's Disease Assessment Scale	Dementia Screening	11	12	0	30 to 35	70	89.20%	88.50%	Clinical	Drawing
CDR-SB	Clinical Dementia Rating Scale Sum of Boxes	Dementia Severity	6	2*	0	30	18	80%*	69%*	Clinical	N/A
ECog	Everyday Cognition	Dementia Severity	39	1.23* 1.92*	1	Not reported	4	93% Dementia from Normal; 75% MCI from Dementia; 67% MCI from Normal	80%	Clinical	N/A
FAQ	Functional Activities Questionnaire	Dementia Screening	10	9	0	8 to 10	30	85%	81%	Community	N/A
MMSE	Mini-Mental State Exam	Dementia Screening	11	23/24*	0	10	30	90.8%*	90.62%*	Clinical	Drawing
MoCA	Montreal Cognitive Assessment	MCI Detection	12	26	0	30	30	90% MCI; 100% AD	87%	Clinical	Drawing & Trail making
RAVLT	Rey Auditory Verbal Learning Test	Memory	8 Trials	5 (20- to 30-delay)	0	N/ A	75	N/A	N/A	Clinical	N/A

Procedure	Table 2.3: Cognitive Domain Coverage						
	Social Cognition	**Complex Attention**	**Learning and Memory**	**Executive Function**	**Language**	**Perceptual Motor Skill**	**Original Ref**
ADAS	No	Yes	Yes	No	Yes	Yes	Rosen et al., (1984)
CDR-SB	Yes	No	Yes	Yes	No	Yes	Hughes et al., (1982)
ECog	No	Yes	Yes	Yes	Yes	Yes	Farias et al., (2008)
FAQ	No	Yes	Yes	Yes	No	Yes	Pfeffer et al., (1982)
MMSE	No	Yes	Yes	No	Yes	Yes	Folstein et al, (1975)
MoCA	No	Yes	Yes	No	Yes	Yes	Nasreddine et al., (2005)
RAVLT	No	No	Yes	No	No	No	Schmidt, (1996)

Procedure	Table 2.4: Mapping of Procedure Sections to Cognitive Domains						
	Social Cognition	**Complex Attention**	**Learning and Memory**	**Executive Function**	**Language**	**Perceptual Motor Skill**	**Original Ref**
ADAS-Cog	None	Ideational Praxis	Word recall, Orientation, Word recognition, Remembering test instructions	None	Commands, spoken language ability, naming objects/fingers, word finding difficulty, comprehension.	Constructional Praxis	Rosen et al., (1984)
CDR-SB	Community affairs	None	Memory, Orientation	Judgment and problem solving	None	Home and Hobbies	Hughes et al., (1982)
ECog	None	Divided Attention	Memory	Planning, Organisation	Language	Visuospatial	Farias et al., (2008)
FAQ	None	Questions 1,2,4,7,8	Question 9	Questions 3,10	None	Questions 5,6	Pfeffer et al., (1982)
MMSE	None	Attention and Calculation	Orientation, Registration, Recall	None	Language	Copying	Folstein et al., (1975)
MoCA	None	Attention.	Memory/delayed recall, Orientation	None	Naming, Language, Abstraction	Visuospatial/executive	Nasreddine et al., (2005)
RAVLT	None	None	Immediate, Percent forgetting	None	None	None	Schmidt, (1996)

2.5 Recent Literature on Cognitive and Functional Abilities in Dementia Diagnosis and Screening

Cipriani et al. (2020) reviewed research works related to functional elements for dementia patients to show how these elements change in the different stages of dementia. The authors aimed at understanding how functional disabilities vary in the dementia subgroups, and how these disabilities impact the progression of the disease. Using the Cochrane digital library and the PubMed database, 207 research papers were retrieved by the authors using terms such as 'functional activities', 'instrumental activities of daily living', and 'daily functioning' among others; 81 papers were retained and analysed in a qualitative manner. The findings of the review showed a correlation between cognitive domains such as executive function and certain functional activities. The findings of this study support previous conclusions such as the meta-analysis study conducted by Martyr and Clare (2012). Using a meta-analysis approach, their research investigated 49 studies with 3,663 data subjects with AD , using 17 different tests that included executive function cognitive domain to establish a relationship between executive function and ADL. The findings revealed a relationship between ADL including driving and executive function.

The impact of cognitive and functional performance on ADL was studied by Clemmensen et al. (2020) on 185 data subjects with AD using MMSE, Stroop Color and Word test [Stroop] (Golden, 1978), Symbol Digit Modalities Test [SDMT] (Smith, 1982), Timed Up & Go [TUG] (McGough et al., 2011), the Astrand Cycle test, and the Sit to Stand test [STS] (Eggermont et al., 2010) cognitive and functional performance tests. Results derived using Pearson Correlation analysis on the data collected from the patients revealed that the total scores of ADL and IADL were associated with executive functions, but no correlations were observed between other cognitive elements and physical performance with ADL or IADL for patients with mild to moderate AD.

Yu et al. (2018) showed that neuropsychological assessments may be utilised to pinpoint some functional activities of daily living, particularly the ones that rely mainly on cognitive processes involving memory and reasoning. For example, memory and executive functions have shown correlations with some functional performance (De Paula & Malloy-Diniz, 2013; Cipriani et al., 2020). Conversely, Lim et al. (2018) revealed in an applied study using 61 patients that in-depth neurological assessments are needed to use cognitive scores of neurological tests as predictors for functional deficiencies. In addition, De Paula and Malloy-Diniz (2013) showed that people at risk of dementia who exhibit impairment in executive functions are likely to have issues in complex activities such as managing their medications, organising finances, and preparing a meal with multiple recipes. Thus, inferring the level of which cognitive elements within neuropsychological assessments relate to the functional ability of patients is key information for clinicians treating

dementia patients since therapeutic plans can be developed early for the appropriate condition level to optimise resources and to prepare for the later progression.

Jutten et al. (2019) investigated quality issues (clinical relevance, validity) in a previously designed composite battery test for mild dementia called cognitive-functional composite [CFC] (Jutten et al., 2017). CFC comprises multiple cognitive tests and one functional ability questionnaire. A total of 184 data subjects with different dementias and predementia (MCI, AD, Lewy body) along with baseline diagnosis have been analysed using linear regression and confirmatory factor analyses including variables such as CFC score, CDR-SB score, ADL-scale, ADAS-Cog13, age, education, disease severity, and sex, among others. The results showed that the CFC's main components (executive functions, episodic memory, and Amsterdam IADL Questionnaire score (A-IADL)) are more correlated with the disease severity than conventional cognitive tests such as CDR-SB score and ADAS-Cog13 score, at least when using the statistical measures and the sample data considered.

A longitudinal study conducted by Najar et al. (2019) evaluated the impact of functional and cognitive activities on the risk of dementia for women in their midlife. Cognitive, functional, and demographic data using different cognitive tests, as well as informant interviews and medical records of 800 different women with a mean age of 47, were followed between 1968 and 2012, and then archived. The Cox regression model was used to analyse the associations of cognitive-functional activities at risk of dementia. The results showed independent associations of the cognitive and functional activities in midlife to dementia types for the women followed aged over 44 years. To be specific, cognitive activities in midlife may reduce the risk of AD, and functional activities may reduce the risk of other dementia types including cerebrovascular and mixed dementia. Gender and age seem to be factors impacting the progression of dementia. This outcome supports the findings of another recent study by Qin et al. (2020) which showed that demographic features such as gender, age, and education influenced AD progression in a cohort of 2901 subjects with MCI, AD, and CN.

Jutten et al. (2017) combined multiple battery tests for cognition and functions in a composite test they named CFC to measure any possible decline in early dementia and predementia stages (MCI). The authors concentrated on executive functions and memory based on the findings of the previous literature, particularly measuring the sensitivity to changes in the conditions in mild dementia and MCI, i.e., Vellas et al. (2008), Martyr & Clare, (2012), Harrison et al. (2013), which are the domains that show decline in early dementia. To clarify, CFC consists of a method that combines two episodic memory and three executive functions called cognitive composite which was initially reported by a data driven study by Harrison et al. (2013), and an everyday informant functional changes questionnaire called A-IADL (Sikkes et al., 2012). The A-IADL evaluates functional

abilities related to preparing meals, managing finances, and using technological devices, among others.

Zucchella et al. (2017) investigated functional and cognitive relationships for patients diagnosed with mild dementia using the Disability Assessment for Dementia (DAD) functional performance method (Gelinas et al., 1999), and neuropsychological methods like MMSE. A cohort of 158 dyads (patients and informal caregivers) were used in which patients were screened using the MMSE for cognitive impairment and DAD for functional deficits; the clinicians used the CDR-SB method to measure the patients' dementia severity, and the caregivers were interviewed to evaluate behavioural issues. Results obtained using statistical analysis showed that logical and executive functions were associated with functional status.

Lim et al. (2018) examined 61 data subjects with cognitive impairment to determine the relationship between cognitive elements of the MoCA and functional results of the Modified Barthel Index (MBI) functional questionnaire (Nasreddine et al., 2005; Collin et al., 1988). The authors divided the patients into two groups based on their MoCA scores—a score of 11 was used to distinguish between the two groups with those scoring below 11 categorised as moderate to severe and those with MoCA scores above 11 considered mild. The results obtained using linear regression analysis demonstrated that MBI functional outcomes were improved when comparing between initial and final assessments; MBI outcomes showed more improvement for patients within the first group (MoCA scores of 11 and greater). More interestingly, there was a positive correlation between initial MoCA scores, and the functional improvement represented in the MBI outcomes—this correlation was not seen using the MMSE test scores. However, the MMSE's improvement on MBI was established in the patients in the first group. The authors concluded that multiple and detailed neurological assessments are needed to evaluate whether cognitive elements can be used as predictors for functional abilities.

Brown et al. (2014) investigated the association between MMSE scores and functional activities for patients with dementia. Thirty patients were recruited from three hospitals in the Melbourne (Australia) area in acute settings. The score of the MMSE test was adopted to measure cognitive areas and the Functional Independence Measure (FIM), one of the methods for measuring functional activities of self-care (Hall et al., 1993), was used to report the functional performance. The authors analysed the scores obtained for the participants using linear regression and Spearman's correlation and discovered that the MMSE total score had some correlation with the FIM total score with a coefficient of determination of 0.405 for $p < 0.05$; the MMSE items' sub-scores are not correlated with the FIM motor sub-score. Therefore, it is highly recommended that clinicians when assessing

individuals at risk of dementia do not generalise patients' functional performance, especially motor skill abilities, and should include a functional performance questionnaire to measure IADLs.

A number of previous research has shown that for dementia patients, cognitive decline may precede functional decline (Sperling et al., 2011; West et al., 2012; Yu et al., 2018); however, the relationship between the cognitive and functional changes is not well studied especially in the early dementia stage. Liu-Seifert et al. (2014) examined the cognitive and functional progression for patients with mild AD, i.e., MMSE baseline scores between 20 and 26. Cognitive and functional data were collected for different patients diagnosed with mild dementia using the AD Assessment Scale-Cognitive subscale (ADAS-Cog14) and the AD Cooperative Study-Activities of Daily Living (ADCS-ADL) (Rosen et al., 1984; Galasko et al., 1997). The results produced using regression models showed that cognitive impairments precede functional impairments. More importantly, the relationship between the scores obtained from ADAS-Cog14 and ADCS were low at baseline and then consistently increased slightly over a period of 80 weeks. A higher correlation score was observed between cognitive scores (ADAS-Cog14) and IDALs when compared with basic ADL functions.

Mioshi (2011) pinpointed that those diagnostic methods for cognitive impairment of dementia would not give a clear indication of functional deficits because these methods contain scores of the cognitive elements based on the category dementia levels (mild, moderate, and severe). The author expressed a concern that neuropsychological methods that measure cognitive impairment of dementia should not be used in isolation of methods that measure functional activities. This is because cognitive methods are primarily dependent on the skills of language and memory and may not measure the degree of independence level presented in the IADLs.

Allen et al. (2009) studied the correlation between cognitive methods such as MMSE, besides variations of the clock drawing, and learning new executive functions of the elderly and inhaler performance. The authors hypothesized that tests of executive functions such as drawing clocks or drawing the intersecting pentagons item of the MMSE can be better predictors for inhaler performance than MMSE total score. From a population of 83 inpatients, 63 women were recruited excluding patients with severe dementia (MMSE total score less than 11). Clock drawing tests [CLOX1 and CLOX2] (Royall et al., 1988), the MMSE intersecting pentagon item score, and MMSE total score were recorded for the subjects before the inhaler training accordingly. Then, a statistical analysis using Yates' Chi-Square test was performed on the data collected; the results illustrated that the MMSE total score and the MMSE's ideo-motor praxis (drawing pentagon score) were better predictors than CLOX1 and CLOX2 for inhaler performance, at least on the subjects considered in the study. One probable explanation for the results obtained is that learning to inhale

can be less determined by executive control function as the drawing test activities do not reflect the instruction provided within the inhaler training, rather it evaluates the executive function in a more real and uncertain situation. In conclusion, learning new functional skills for patients with dementia seems more dependent on general cognitive elements and ideo-motor praxis activities.

Razani et al. (2009) showed that MMSE's attention and orientation test scores are correlated with some functional activities, at least for the 61 data subjects used in the study and using the Direct Assessment of Functional Status (DAFS) questionnaire (Loewenstein et al., 1989). In addition, the authors determined that the language elements of the MMSE test have some sort of correlation with functional activities other than activities related to time orientation and shopping—this supports previous findings by Lecky and Beatty (2002). The time orientation and shopping activities of the DAFS have shown some correlation with the MMSE recall items. However, most of the items assessed in the MMSE are linked with orientation raising a question about the accuracy of the individual's cognitive screening (Lim et al., 2018).

Payne et al. (1998) studied the interaction between both cognitive and functional elements for dementia patients and their correlation with features related to depression. The authors sought to answer questions such as, 'Are patients with cognitive deficits more likely to be depressed when they exhibit functional impairment?' 'Can the level of cognitive or functional impairment be used to predict depression?' Data related to the Cornell Scale for Depression in Dementia [CSDD] (Alexopoulos et al., 1988), MMSE, and the Psychogeriatric Dependency Rating Scale, physical dependency subscale [PGDRS-P] (Wilkinson & Graham-White, 1980) from 569 patients were collected from two clinical sites in Maryland, USA—the data for just 259 patients were analysed using logistic regression after excluding patients with missing values or patients without a primary diagnosis. The results produced implied that there are no clear associations between the cognitive elements and depressive features for vascular dementia, and mixed dementia patients. In addition, depressive features seem to occur in all levels of vascular dementia, and AD patients may show depressive features in the early stages. More interestingly, early-stage AD patients with functional impairment are more likely to show symptoms of depression. The study had some limitations, such as the sample data subjects considered were healthier than the generation dementia population; additionally, the mixed dementia subgroup consists of multiple dementia types and thus the findings do not necessarily support each subtype in that group.

A more recent study that investigated cognitive leisure activities associated with depression for a larger population was published by Kim et al. (2019). Data related to cognition and functional activities was measured using the NCGG Functional Assessment Tool [FAT] (Makizako et al., 2013). Demographic information (age, gender, employment status, body mass index, education

level, alcohol use, etc) and cognitive activity scale was collected from 9,380 individuals. In addition, the 15-item Geriatric Depression Scale [GDS]'s (Friedman et al., 2005) data were also collected for the same individuals. In conducting the analysis, Pearson correlation and t-test statistical measures were derived from the considered data and the correlations between leisure activities and cognitive impairment were measured using logistic regression. A negative relationship was found between the frequency of leisure activities and cognitive impairments in the older adult population. The same negative correlation was noticed in older adults with depressive features. The results suggest a complex association between cognitive-functional impairments for dementia patients and depression.

DiBenedetti et al. (2020) observed patient and caregiver perspectives on the stages of AD in terms of symptoms, impacts, and outcomes. Interviews of 60 participants reported qualitative results, often summarising that memory impairment was the largest concern in progressing AD. Other areas of concern from MCI caregivers seemed to be communication, language, and emotional changes. A common theme between the participants was the desire for memory restoration and preventing further cognitive decline.

2.5.1 Discussion

Most of the reviewed works that have tackled cognitive and functional assessments of dementia have focused on predicting dementia or MCI rather than detecting points when the disease is progressing (e.g. Weakley et al., 2017; McCombe et al., 2020). The latter is more challenging as some of the dementia levels and even precursors such as MCI can share common cognitive impairment traits, and the cognitive indicators have common cognitive domains as defined in the degenerative conditions of the DSM-5 framework. For example, individuals with MCI or mild dementia can share memory decline and thus could have close final scores when they are assessed using cognitive tests. Therefore, it is imperative not only for the clinicians to differentiate dementia level, but also to determine factors that may trigger advancements in the disease. This can be achieved by detecting cognitive indicators during a clinical assessment in which any change may signal the disease's advancement.

Most of the current memory and neuropsychological assessments related to dementia such as RAVL, ADAS-Cog-13, MoCA, and ECog, are mainly used by clinicians to predict MCI or dementia level using scoring functions—these methods do not detect the progression of dementia. To do so, the sequence of cognitive and medical assessments of the individual needs to be considered and the time elapsed between the assessments to capture any signs of change that may yield to disease

advancement, no change, or even disease regression. Currently, most of the cognitive studies related to dementia use medical assessment datasets that do not capture dementia progression, rather they only capture the diagnostic class using the DSM-5 or NINCDS -ADRDA criteria. These challenges make it difficult for researchers to deal with the issue of dementia progression especially using data-driven approaches.

Most of the research studies that we reviewed had not considered integrating both cognitive and functional areas in a single clinical assessment for dementia despite the fact that these areas can be correlated. For example, studies by Clemmensen et al. (2020) and Liu-Seifert et al. (2015) reported that cognitive impairment may yield to functional impairment over time. Other researchers pinpointed that cognitive decline can be associated with minor functional decline that does not necessarily impact the individual's independence of daily living (Winblad et al., 2004). Hence, integrating cognitive and functional elements in a single clinical assessment session could be advantageous to identify impactful elements and their associations to establish cognitive criteria related to degenerative conditions in the DSM-5's frameworks.

Research studies, for example by Brown et al. (2014), Zucchella et al. (2017), Lim et al. (2018) and Clemmensen et al. (2020), showed that discovering the associations between cognitive and functional elements can be useful to detect functional decline. However, according to Zuchella et al. (2017), measuring these correlations in dementia subgroups and dementia precursors is a difficult task as it is not well characterised, besides the difficulty of having real data observations for individuals that capture both functional and cognitive elements—most cognitive tests scores in neuropsychological assessments can partially detect the functional status only. Thus, it can be useful to study the associations between the cognitive and functional elements of individuals in the ageing community and at different dementia levels. This can assist clinicians in the early detection of a deterioration in functional activities caused by a cognitive decline.

2.6 Review on Machine Learning Studies in Cognitive Dementia Assessment Methods

In this section, we review common conventional machine learning techniques and their applications to cognitive assessments of dementia using data-driven methodologies. The choice of the machine learning techniques is based on their popularity in the medical research literature particularly dementia-related screening and diagnosis research, i.e. Bang et al., (2018), Pereira at al. (2018), McCombe et al. (2020), Thabtah et al., (2020). and others. We focus on data-driven studies that

used statistical techniques including regression analysis, support vector machines, decision trees, probabilistic methods, artificial neural networks, rule-based techniques, among others.

2.6.1 Decision Trees and Rule-based Classification

Decision tree is a divide-and-conquer approach in machine learning in which a tree-based structure is built from the training dataset to represent the classification system (Witten & Frank, 2002). A decision tree algorithm such as C4.5 utilises information theory metrics to construct the tree in a recursive top-down fashion (Quinlan, 1986). The fundamental question is how to grow the tree and which attribute to choose for the data split. Common metrics to grow the tree and split the data quantitatively are information gain and Gini-Index (Quinlan, 1993; Ceriani & Verme, 2012). The former, which is more popular, employs Entropy (Shannon, 1948) to compute the degree of uncertainty of any data split using the attribute under consideration and the target class information as shown in Equations (2.1–2.2) and the latter is computed as per Equation 2.3. Usually, the decision tree algorithm selects the attribute that yields the maximum information gain after considering all independent attributes in the input dataset.

$$\text{Information Gain } (D, y) = \text{Entropy } (D) - \sum ((|D_y| / |D|) * \text{Entropy } (y)) \quad (2.1)$$

where

$$\text{Entropy } (D) = \sum -P_y \, log_2 \, P_y \quad (2.2)$$

Where

D = the dataset

P_y = the probability of a to belong to class y

D_y = subset of D for that $Y = y$

$|D_y|$ = The number of instances in D_y, and $|D|$ = Size of D

$$\text{Gini} = 1 - \sum_{i=1}^{n} (p_i)^2 \quad (2.3)$$

Where p_i is the probability of an instance belonging to a specific target class.

The decision tree algorithm keeps growing the tree until the termination condition is met (the tree cannot grow any further or the complete training data instances are used). When the decision tree is completed, then the algorithm trims unnecessary branches or sub-trees to minimise any chance of overfitting the training dataset. In the final decision tree, a rule denotes a path from the tree's root

to any leaf. Common decision tree algorithms are C4.5, C5, and random forest (Quinlan, 1993; Quinlan, 1998; Breiman, 2001).

Thabtah et al. (2020) investigated the problem of AD progression when using the Everyday Cognition (ECog) test based on real data subjects collected from ADNI (TADPOLE) (Marinescu et al., 2019). In particular, data subjects (diagnosed as CN, MCI, and AD) who participated in ECog's two versions: the patient and study partner, were analysed using machine learning algorithms with a focus on rule-based classifiers. The results obtained using the machine learning algorithms showed that when the input dataset was balanced prior the training phase the classification performance in terms of accuracy was enhanced for the models derived. Moreover, rule-based classifiers such as RIPPER, PART (Frank & Witten, 1998) and C4.5-Rules, besides random forest, derived competitive classifiers for AD progression.

Common rule-based classification algorithms are RIPPER, Multi-class Association Classification (MAC), CN2, Ripple Down Rule Learner (RIDOR), and Fuzzy Unordered Rule Induction Algorithm (FURIA) (Abdelhamid et al., 2012; Clark & Niblett, 1989; Gaines & Compton, 1995; Hühn & Hüllermeier, 2009).

Das et al. (2019) introduced a rule-based machine learning approach called the Sparse High-Order Interaction Model with Rejection Option (SHIMR), to predict dementia. The authors utilised data subjects from the ADNI dataset with some proteomics features as well as cerebrospinal fluid-related features collected from 141 data subjects. A comparative analysis on the data subjects was conducted using the C4.5 algorithm (Quinlan, 1993) and the rule-based approach (SHIMR). Results derived based on classification accuracy revealed that the SHIMR approached was superior to C4.5 by deriving classifiers with 84% accuracy on the data subjects.

Bang et al. (2017) applied machine learning algorithms to clinical data to help physicians diagnose dementia more accurately. The authors applied their model to the CREDOS study consisting of data gathered by 37 universities in Korea from 2005–2013. The proposed model recognises the need for a diagnostic process that is made up of four steps or modules. First the proposer module: the kScale variable selection method is used because it provides flexibility by verifying various results from several methods. Next a classification model is built to determine dementia symptoms. This model uses the variable used by the proposed model as input and with the CDR-SB score as the class label. Machine learning algorithms including ANN, Decision Tree, and SVM are used to build the models. A descriptive analysis was then performed to describe the process of the classification of patients based on the classification model learnt by the classification algorithms. In this descriptive analysis, Decision Tree outcomes are transformed into easy-to-interpret knowledge. Finally, the visualisation model helps to seek useful characteristics of instances which are clustered in a descriptive step. It

uses contrasting colours so that the patients with normal, dementia, and other diagnostic indicators can be distinguished immediately.

Weakley et al. (2015), used classification techniques for two datasets, each consisting of participants who were classified as exhibiting 'normal ageing' to dementia. The first dataset consisted of 310 participants who were diagnosed according to clinical diagnosis criteria; the second dataset contained 272 participants who were diagnosed according to CDR scores: 0 for normal ageing, 0.5 for MCI, and 1 and above for AD. Both datasets initially contained 27 variables related to various cognitive test scores, but this number was reduced using a wrapper feature selection method. To evaluate the generated models, 5-fold cross-validation was used. The main comments from the authors are that the classification techniques used showed no statistically different ability in classifying the data, with the most difficult (least successfully classified) group being the middle group, alternatively MCI, or CDR = 0.5.

2.6.2 Support Vector Machines

SVM is one of the classification approaches that can deal with both classification and regression problems in machine learning. The main idea of the SVM approach is choosing the best hyperplane (a decision boundary) that can segregate two target class values (Platt, 1998). The algorithm initially plots the data instances of the available *n* attributes in the training dataset in *n*-dimensional view for which each attribute denotes a specific coordinate. The SVM algorithm then classifies data instances by selecting the hyperplane that accurately distinguishes between the two target classes. One way to find the best hyperplane to differentiate between the classes is to compute the maximum margin (the maximum distance between the data instances of the classes) (Joachims, 1999).

Stamate et al. (2018) investigated how to distinguish dementia and MCI diagnoses using feature selection with intelligent learning methods. The authors utilised ReliefF and Information Gain (Kira & Rendell; Quinlan, 1986) measures to select influential attributes from the ADNI's Merge dataset (ADNI, 2021). Since a patient could have had multiple visits, this study limited each patient to their first visit to the research centre along with the diagnosis. This resulted in 49 attributes and 1851 observations being included and one response variable (the diagnosis). The response variable has three possible values (CN, MCI, Demented). The dataset consists of attributes related to cognitive, brain areas, magnetic resonance imaging (MRI), and positron emission tomography (PET), among others. To produce classification models, a number of classification algorithms including random forest, SVM, Stochastic Gradient Boosting, eXtreme Gradient Boosting, and Gaussian Processes have been applied to differentiate between the three possible values of the response variable. Results using cross validation and Monte Carlo simulation as testing and validation methods respectively

reported that the eXtreme Gradient-Boosting algorithm was able to outperform the models derived by the remaining algorithms. Models derived by the eXtreme Gradient Boosting achieved 88% accuracy, 93% sensitivity, and 94% specificity in identifying dementia. The same algorithm was able to detect MCI with 86% sensitivity and 90% specificity for MCI.

Battista et al. (2017) conducted a quantitative study to assess cognitive measures and their potential in reducing the amount of neuropsychological tests used to improve the classification of AD patients, and at an early stage of impairment. They aimed to explore using more subdomains that are concerned with long-term memory and recognition memory. While the study had also mapped the features to DSM-5 cognitive domains, each feature was only limited to one domain when certain tasks could involve more than one. It would have been helpful to understand the method they used to assign the domains. The study utilised neuropsychological tests and features obtained from the ADNI database. Two feature reduction approaches were used to improve computational performance: (a) a computational approach, and (b) a clinician's understanding based on their expertise and experience. With the reduced set of features, a classification algorithm based on SVMs was used to train the dataset into predicting the likelihood of being diagnosed with AD. Overall, using the computational feature reduction approach, the best predictors were Q1, Q4, and Q8 in ADAS-Cog13. Similarly, using the clinician's reduction approach, ADAS-Cog13: Q1, Q4,and Q8 was found to be the best predictor, and due to the FAQ items having overlapping measures, they have been excluded.

A study conducted by Lemoine et al. (2010), focused on the accuracy of classifying normal, MCI, and AD as an individual diagnosis. The authors studied the ADNI-Merge dataset, including both PET imagery and clinical data such as the assessment scores. While our research does not cover PET imagery and only focuses on the cognitive methods, their research presents useful information of data mapping, data integration, and feature selection techniques. SVM classifiers (Cortes & Vapnik, 1995) were fitted and derived the best accurate classification for detecting AD. It is interesting to note that in their findings based on the ADNI data analysis, they have classified the FAQ test as one of the top clinical methods, along with ADAS-Cog13 amongst five clinical assessments to have better weighting in diagnosing AD.

2.6.3 Probabilistic Classification

Probabilistic classifiers approximate a joint probability using product distribution (Friedman et al., 1997). These algorithms compute the conditional probabilities of a target class, i.e. y, using a product of conditional probabilities of the attributes' values of the test instance and their probabilities in the training dataset as shown in Equation 2.4.

$$P\ (y|X) = \frac{P(X|y)P(y)}{P(X)} \qquad (2.4)$$

Where X is an instance of *n* attributes values to be classified (X =$((x_1, x_{2,\ldots\ldots,} x_n))$, and y is a target class to be classified

The Bayes-based classification algorithms assume that the features are independent from each other given the target class—not realistic in real applications (Hand & Yu, 2001). Common Bayes-based algorithms such as Naïve Bayes (Duda & Hart, 1973) predict the target class of test data using the computed product of the conditional probabilities and then assign the class that has the largest likelihood to the test data (Zhang, 2004). The Bayesian Network (Bayes Net) algorithm illustrates a set of variables and its conditional dependencies, and it searches the network that "best describes" the probability distribution over the training data (Friedman et al., 1997).

McCombe et al. (2020) investigated the problem of a missing diagnosis in dementia classification using data subjects who had taken the CDR-SB cognitive assessment. The data subjects were adopted from the ADNI-Merge and the cohort containing 1185 cases and controls with various diagnostic class frequencies. The K-Nearest Neighbour method (Cover & Hart, 1967; Fix & Hodges, 1951) was used to deal with the missing values in the input dataset. Classification algorithms including Naïve Bayes, random forest, and SVM (Duda & Hart, 1973; Cortes & Vapnik, 1995; Breiman, 2001) were then applied on the imputed dataset to derive dementia classification models. The results showed that models derived by random forest and SVM were superior in terms of the Area Under ROC curve when cross-validation was used as a testing procedure.

Maheux et al. (2020) investigated the progression of AD based on data subjects from the ADNI, and other dementia data repositories, for individuals who had taken the MMSE cognitive assessment. The methodology followed by the authors is based on the Bayesian probabilistic method and the data used cover a timeframe of six years for each participant i.e., since their first diagnostic clinical assessment. Results obtained from the classification algorithm were compared with those produced by linear regression, logistic regression, and constant prediction (no change to the MMSE scores). The results pinpointed to a higher classification rate for the AD advancement early (within 1–3 years) which is promising for early detection of the disease.

Pereira at al. (2018) explored the different feature selection techniques in predicting the conversion from MCI to AD. The study aimed to derive subsets of vastly reduced features from neuropsychological tests, using a feature selection with ensemble learner technique that combines both stability and predictability. Pereira et al. (2018) combined seven feature selection methods based on different strategies to measure the impactful features, and then paired these with different classifiers to observe which combination attained the best classification performance. They found

that the Naïve Bayes algorithm was the strongest performer and Decision Tree the weakest in terms of predictive accuracy, at least on the dataset used. The training dataset was retrieved from ADNI and the Portuguese Cognitive Complaints Cohort, both achieving relatively good model performance at AUC above 87% and 82% respectively. For the ADNI dataset, the top selected features for ADAS were ADAS-Cog Total Score 13, ADAS-Cog Total Score 11, ADA-cog Q4, ADAS-Cog Q8, and ADAS-Cog Q1. The results were comparable to Battista et al. (2017), where the study had also identified Q4, Q1, and Q8 to be effective in identifying dementia.

Jammeh et al. (2018) applied machine learning technology to identify dementia from the National Health Service (NHS) data. A total of 26,483 data instances of individuals aged over 65 years were collected from 18 general practitioner surgeries. Machine learning techniques such as random forest, SVM, Naive Bayes, and linear regression were utilised to build classification models that were then evaluated using cross validation testing methods with n=10 folds, in terms of specificity, accuracy, and sensitivity, among others. The dementia classification models derived by the machine learning algorithms pinpointed that the model derived by the Naïve Bayes algorithm produced the highest performance. The classification model of Naïve Bayes was able to detect 295 instances with dementia who had not received a clinical diagnosis.

2.6.4 Artificial Neural Networks

An ANN is a set of connected nodes that are normally called neurons, which capture the relationships between the available features and the target class in the input dataset using the same mechanisms as the human brain (Gardner & Dorling, 1998; Elyan et al., 2018). Usually, neurons communicate and receive data from each other, assign them weights, and then transmit the sum to other neurons. This mechanism of adjusting the weights of the input and passing them on can help in updating the network structure (the model) during the training phase (Grossberg, 1988). One of the common ANN algorithms is Backpropagation, which computes how much each neuron contributes to the current network's error rate (Mohammad et al., 2013). The algorithm amends the weights to enhance the current predictive power of the network model backwards from the output.

Nagaraj and Duong (2020) utilised ANN to differentiate between MCI and AD patients using the ADNI Merge dataset. To train the ANN algorithm, the authors prepared a subset of cognitive features related to the cognitive assessments reported in the ADNI-Merge in addition to features related to biomarkers and demographics. Once the input dataset was prepared, a few selection methods were applied on the dataset to isolate important features and discard those that were irrelevant. This yielded a small subset of features that were then processed by a Multi-layer Perceptron Neural Network algorithm to generate classification models. The results showed that

when processed by the classification algorithm, the accuracy of the CDR-SB score increased when compared with processing using other features.

Albright (2019) set out to predict the progression of AD with a neural network model based on the ADNI dataset, using longitudinal analysis methodology where the data was pre-processed for missing values or misaligned data type, and sorted into three data sets (LB1, LB2, and LB4) based on criteria established by the TADPOLE Challenge. LB1 was used as a training and validation dataset, processed using the all-pairs technique, and then classified using various machine-learning classifiers. Each classifier was evaluated using 7-fold cross-validation. A receiver operating characteristic area under the curve (ROC-AUC) was also used to measure the effectiveness of each classifier. Albright (2019) found that multilayer perceptron (MLP) neural networks and recurrent neural networks had the best performance in cross-validation studies. This model was 86.6% effective at predicting the progression of AD, either from a state of clinically normal (CN) or MCI. While this study illustrated the success of predicting disease progression, it did not explain specifically how the three sets of the features training group were derived.

Choi and Jin (2018) exploited ADNI data, aiming to predict longitudinal changes of cognitive decline in MCI patients by using FAQ; in their study they used data related to a biomarker known as PET. The authors obtained data about the PET images of 139 patients with AD, 171 with MCI, and 182 with a CN. They managed to achieve 84.2% accurate prediction using a convolutional neural network (CNN)-based approach (LeCun et al., 2015) for the conversion of MCI to AD. ROC analyses were carried out to reveal that the achieved performance was significantly higher than the conventional feature-based approaches. The authors used Pearson Correlation on subjects having MCI as a baseline diagnosis to seek FAQ score attribute behaviour. The results of the analysis showed that the FAQ score attribute is positively correlated with longitudinal changes, and the correlation was noticeably significant after three years following the initial MCI diagnosis, compared with one year following the initial MCI diagnosis.

Youn et al. (2018) applied a machine learning model that was trained using the TensorFlow package for the detection of cognitive impairment (CI), the Korean Dementia Screening Questionnaire (KDSQ), and MMSE scores. Data was obtained from the South Korea Clinical Research Center for Dementia (Park et al., 2011). A total of 9,885 and 300 instances respectively from the dataset were randomly assigned to train and test the learning models. The algorithm was trained using TensorFlow and its predictive power evaluated on independent data. The performance of the model in detecting CI, based on KDSQ was 84.3%; based on MMSE was 88.3%; and based on a combination of both was 86.3%. For KDSQ, sensitivity for detecting CI was 91.50% and the sensitivity of MMSE was 94.35%. When KDSQ and MMSE were integrated, the sensitivity rate for

predicting CI was 91.5%. Accuracy of the algorithm predicting CI based on MMSE was much better than when using KDSQ.

So et al. (2017) contrasted a number of classification algorithms such as Naive Bayes, Bayes Net, Bagging, logistic regression, random forest, SVM and MLP (Rish, 2001; Friedman et al., 1997; Quinlan, 1996; Hosmer et al., 2013; Liaw & Wiener, 2002; Kotsiantis, 2007; Gardner & Dorling, 1998), to classify dementia aiming to improve classification accuracy. The authors used data collected from patients who visited the Gangbuk-Gu Dementia Centre in the Republic of Korea from 2008–2013 to receive dementia screening. During the data cleaning process, they removed all missing values and errors whilst incomplete or incorrect data were replaced. Chi-Squared and Information Gain (Liu & Setiono, 1995; Quinlan, 1986) methods were chosen to select influential features related to temporal order, memory function, and a language fluency test. The results showed that MLP and SVM achieved the best performance according to accuracy, at least on the dataset considered.

2.6.5 Regression and Statistical-based Models

One of the common ways to measure relationships between a pair of features in mathematics is by using the correlation coefficient (*r*) (Pearson, 1920) (See *r* Equation in Chapter 1). The correlation coefficient value of two attribute values (x, y) ranges from -1 to 1. Usually, regression analysis measures the degree of a relationship between input and output features by providing a linear model. This model reveals whether there is a relationship between the two features and the strength and direction of that relationship (Cox, 1968). The model is often obtained based on the linear equation (Equation 2.5) against the input dataset using methods such as Least Squares.

$$y = B0 + B1*x \tag{2.5}$$

where

x, and y are the input and output features respectively.

When there are multiple x features, the linear line becomes a hyperplane. In addition, the logistic regression algorithm uses a logistic function to model a binary problem based on dependent variables and to predict the likelihood of an outcome using a linear combination of independent variables (Hosmer et al., 2013). This is often used as a classification model where the results are divided into specific categories.

Using the ADNI cohort data, Kemp et al. (2020) examined connections between statin use and cognitive decline in subjects with CN, MCI, and AD. The Linear Mixed Effects model and Cox Proportional Hazards model were used to examine cognitive change over time for 1629 individuals aged from 48–91. These statistical tests determined that statins may contribute to a reduced decline in memory, but there was no correlation between statin use and cognitive change.

Correlations between lifetime occupation and late-life cognitive impairment were examined by Kim et al. (2020) on 1733 participants over 65 years old. These subjects, from the Korean Longitudinal Study of Aging, included people of different genders and socio-economic divisions. The study utilised the Korean version of the MMSE to obtain results showing that longer occupation time contributes to higher cognitive performance in later-life. Male participants showed no significant cognitive change in regards to socio-economic class, but female subjects showed significantly higher cognitive impairment if they held blue-collar occupations compared to white-collar jobs.

A study by Wang et al. was conducted in 2020 to find connections between white matter hyperintensities (WMH) and cognitive decline in people with AD. A direct correlation between increased WMH and risk of cognitive impairment was observed through statistical models from 818 individuals: 259 with normal conditions (NC), 448 with MCI, and 111 with AD. Subjects were found through ADNI datasets and tested by the following exams: MMSE, MoCA, CDR-SB, ADAS-Cog13, RAVLT, FAQ, ADNI-EF (executive function test), and ADNI-Mem (memory function test). Researchers noted that the severity of WMH ultimately affects the progression of AD. With relation to AD, there were significant correlations between high WMH volume and high ADAS-Cog13 scores. High WMH volume also correlated with lower ADNI-EF scores. The WMH change rate was greatly associated with CoCA score, where an increased change rate demonstrated decreasing MMSE, MoCA, ADNI-Mem, and ADNI-EF scores.

Cognitive-functional Composite (CFC) has been researched in regards to progression of early dementia by Jutten et al. (2020). They used both longitudinal and cohort methods to measure changes in early dementia with developing CFC. Results from a group of 148 participants with MCI symptoms indicated that CFC shows sensitivity to clinical cognitive decline. Therefore, CFC can be considered a meaningful test for those with AD, and capable of monitoring the progression of disease in those with MCI. It is important to note though, that ADAS-Cog scores are not usually sensitive to changing cognition in MCI patients because they focus on language, practice, and factors that are often unpredictable.

Moradi et al. (2017) created a model to predict the scores of the RAVLT (Schmidt, 1996) based on grey matter density features derived from MRI scans in the ADNI dataset. The authors removed all observations with missing RAVLT scores and several observations that had outlier scores. Elastic

net linear regression (enlr) (Zou, & Hastie, 2005) was then used to model *RAVLT immediate* and *RAVLT percent-forgetting* scores using whole brain grey matter density maps; these consisted of 29,852 features for each participant. The models were evaluated using 100 runs of 10-fold cross validation (Picard & Cook, 1984). Across all runs, the averages of the correlation score (*r*), coefficient of determination (Q^2), and mean absolute error (MAE) were found to be: r=0.50, Q^2= 0.25, MAE = 7.86 for *RAVLT immediate*, and r=0.43, Q^2= 0.185, MAE = 25.53 for *RAVLT percent-forgetting*. The author also considered data subsets that included only AD, or only MCI participants, as well as different combinations of the three categories. Interestingly, it was found that removing the MCI participants improved the performance of the model.

Podhorna et al. (2016) compared the performance of the 3-, 5-, 11-, and 13-item ADAS-Cog variant subscales using ADNI data that could best detect cognitive decline. The original 11-item version of ADAS-Cog was to measure cognition in patients with mild to moderate AD, but lacked the capability to detect change and measure cognitive domains known to cause impairment in the early stages of AD. Thus, the creation of new ADAS-Cog variants such as the 13-item subscale to improve its properties for early AD screening with the use of additional tests such as digit cancellation and delayed word recall. As a baseline comparison, the authors assumed the ADAS-Cog 11 score of 10 (out of 70) for subjects with MCI, and a score of 18 (out of 70) for subjects with mild AD. Based on their findings, the ADAS-Cog13 score of 15 (out of 85) was considered MCI, while a score of 30 (out of 85) was considered mild AD. Overall, they concluded that the impact of expansion or reduction of the ADAS-Cog was subtle, but noted that in mild AD, adding rather than removing items appeared to provide more benefit.

Wessels at al. (2015) aimed to identify a composite scale called the Integrated Alzheimer's Disease Rating Scale (iADRS) that can measure the impactful domains of AD by combining cognition and function through the evaluation of existing scales. The datasets utilised to assess the iADRS were from the longitudinal studies of ADNI, and clinical trials of antidementia drugs such as Solanezumab, Semagacestat, and Donepezil. Due to the difference in various scales for total points relative to decline, signal-to-noise ratios (SNRs) were calculated for comparability. Principal component analysis (PCA) was used to establish the psychometric properties of the iADRS, assessing the contributions of the two scales' total scores, and the contribution of individual item scores. Their results found that composites combining cognition and IADL items are better at detecting the disease than traditional cognitive-only or functional-only scales across MCI, mild AD, and moderate AD.

Marshall et al. (2015) suggested that while there is cognitive assessment that has been useful in detecting IADL during the transition from MCI to AD, it has not been successful in detecting the

subtle functional changes in earlier stages when it progresses from CN to MCI. The authors decided to focus on this phase of the disease by investigating which of the FAQ items are sensitive in discriminating and identifying the progression from CN to MCI. In their study, the authors utilised data from two separate cohorts, the ADNI and the Massachusetts Alzheimer's Disease Research Centre. In their methodology, the authors commented that there is no established cut-off score for IADL impairment on the FAQ, however they referred to a study where a score of ≥ 6 is suggestive of functional impairment (Teng et al. 2010). Using both datasets, a cross-sectional analysis was implemented. The results derived revealed that Personal Memory and Administration are the key features in distinguishing between CN and MCI. The authors also identified two additional features of the ADNI cohort, i.e., Engagement with media and Finance, and a single feature, i.e., 'Heating water and turning off the stove' in the MADRC cohort.

Moradi et al. (2017) investigated the relationship between structural atrophy in the brain represented by MRI biomarkers, and a neuropsychological feature represented by RAVLT scores using machine learning (elastic netLR). The authors used a sample of 806 instances from the ADNI dataset which consisted of individuals with AD or MCI and others who were healthy. The authors' key question to be answered was whether RAVLT scores can be predicted using MRI features. The focus was on two RAVLT attributes: *RALVT immediate,* and *RAVLT Percentage Forgetting*. The results obtained on the MRI data and the two RAVLT variables using elastic net logistic regression with ten-fold cross-validation showed a strong relationship between RAVLT variable scores and brain structural atrophy of AD.

Balsis et al. (2015) studied the problem of mapping the scores of three different cognitive dysfunction methods for dementia particularly ADAS-Cog, CDR-SOB, and MMSE and (Rosen et al., 1984; Folstein et al., 1975). The aim of their research was to provide a meaningful way for clinicians to view how each cognitive dysfunction method corresponds to the other within a simple table. The authors used a real dataset downloaded in 2014 from ADNI with 1709 instances enrolled across all ADNI phases. Three variables corresponded to the scores of MMSE, CDR-SOB, and ADAS-Cog respectively which were modelled using the Item Response Theory approach (RTA) (Wright, 1992). The authors performed exploratory factor analyses to evaluate unidimensionality before conducting analysis on the considered variables. The reported results of the RTA analysis showed that scores obtained for CDR-SOB and ADAS-Cog are high at greater levels of cognitive dysfunction. More importantly, for a given MMSE score, multiple inflections have been observed for scores of both CDR-SOB and ADAS-Cog, suggesting consistency in measuring cognitive dysfunction for these two methods. This study can be extended to evaluate how CDR-SOB and ADAS-Cog differ in measuring AD severity besides using more advanced computational

intelligence or artificial intelligence methods to capture correlations between the scores of these methods and the progression of MCI and AD.

Teng et al. (2010) investigated the capability of the FAQ test to clinically distinguish between MCI and very mild AD. They utilised the National Alzheimer's Coordinating Centre (NACC) Uniform Data Set (UDS) (Beekly et al., 2007), a different cohort to ADNI and MADRC. In their study, they noted that only 66% of participants had completed all FAQ items, and thus the FAQ performance was evaluated using two separate methods to deal with incomplete data. One method had valid scores on all items, and the other used average scores across FAQ items with valid responses to allow for better analysis. The stepwise logistic regression method (Hosmer et al., 2013) was used in this study to determine which FAQ items were independently associated with an AD diagnosis. Their findings discussed the cut-off points for diagnosis indicating that a total FAQ score ≥ 6 is most consistent with a clinical diagnosis of AD - similar to the findings of Marshall et al. (2015). They identified that bill paying, shopping, tracking current events, and playing games were the FAQ items that distinguished AD from MCI. Apart from paying bills, the other FAQ items varied from other studies that used a different dataset, suggesting that their trial methods could be different.

2.6.6 Discussion

2.6.6.1 Advanced Learning Techniques for Diagnostic Decisions

Most of the considered data-driven research studies used statistical and conventional machine learning techniques to investigate the various types of dementia indicators including pathology, neuroimaging, biomarkers, and neuropsychological, and their combinations, i.e. Balsis et al. (2015; Podhorna et al. 2016; Moradi et al. 2017; Choi and Jin 2018; Kim et al. 2020). However, data-driven studies that considered neuropsychological elements such as functional and cognitive are highly beneficial to the medical community as these could provide systems that are cost effective, and potentially can be run using technological platforms especially during a pandemic when resources are scarce.

More importantly, machine learning techniques including support vector machines, artificial neural networks, rule induction, decision trees, random forest, statistical, and probabilistic, among others have shown good performance regarding predictive accuracy, sensitivity, and specificity when compared to conventional scoring functions used, particularly in the neuropsychological assessments research studies. However, limited attempts have been made to design and implement a computer-aided AD progression tool for medical assessments that assess functional and cognitive domains together, at least using a data-driven approach. While the reviewed research studies have

shown the potential performance (accuracy, sensitivity, specificity, etc.) of AD detection in terms of applying learning algorithms on datasets related to dementia, few of them have tackled the AD progression problem or have implemented a complete computer-aided diagnostic tool that has been tested and deployed.

With the use of computational intelligence and artificial intelligence (AI) methods, medical systems can be equipped with the latest yet affordable technologies such as deep learning. These systems can discover useful concealed information within semi-structured and unstructured indicators related to MCI and dementia conditions that can be of high value during the clinical evaluations. For instance, conventional neural networks and the deep learning technique can deal with complex datasets related to dementia involving videos of a clinical session, PET scans, MRI, and audios to measure communication and language.

Using AI and machine learning techniques, the dementia progression decision is based on search methods guaranteeing a non-biased outcome that in turn can be assessed by the clinician. The model will not replace the clinician in the medical assessment process of AD, rather with the rich knowledge that it provides it will serve as a complementary tool. Recently, since they provide accurate and reliable outcomes, AI methods such as deep learning have been used successfully in medical screening systems and computer-aided diagnostic tools.

Operating within Industry 4.0, the role of machine learning and artificial intelligence is indispensable in increasing the effectiveness and efficiency of clinical diagnostics and prognostics. Currently, many clinicians and healthcare organisations rely on AI and machine learning applications to provide solutions in digital platforms to improve patients' lives. The application of data-driven approaches to streamline clinical assessments is well-established across many areas in medicine, public health, and epidemiology. A proposed screening tool that represents a step forward in the establishment of clinically validated screening for AD progression. This endeavour is now available at a time when investment for improving the lives of the elderly global population is a pressing need.

2.6.6.2 Dementia Data

There is a limited number of databases related to dementia conditions available for scholars to investigate cognitive, functional, genetic, and neuropathological indicators such as the Cognitive Function and Ageing Studies (CFAS) (Wharton, et al., 2011), South Korea Clinical Research Centre for Dementia (CREDOS) (Park et al., 2011; Lee et al., 2016), ADNI (ADNI, 2021), and the Sydney Memory and Ageing Study (MAS) (Sachdev et al., 2010) among others. ADNI is the most popular

dataset among these since it has wider scope including data (cases and controls) related to genetics, pathology, biomarkers, neuropsychological, neuroimaging, etc.; additionally, it contains validated data observations that are sufficient for studies that involve data and experimental analyses.

The ADNI project originated to collect, organise, analyse, validate, and use data related to dementia conditions (AD) for the early detection and tracking of AD (ADNI, 2021), particularly using technological solutions. ADNI is a longitudinal research project that was initiated in the United States in 2004 to collect clinical data related to CN, MCI, and AD. Examples of datasets in ADNI are: ADNI 1, ADNI 2, ADNI3, ADNI-GO, ADNI-Merge, ADNI-neuropsychological, ADNI-MMSE, ADNI-MoCA, ADNI-ADAS-11, ADNI-ADAS-13, etc. The central dataset that contains the diagnostic and baseline diagnosis is the ADNI-Merge. This dataset consists of 2,361 individuals with over 14,600 data observations of which the individuals are diagnosed with three possible target classes: CN, MCI, or Dementia (AD) at baseline. The ADNI dataset is longitudinal in nature since each individual can be associated with multiple clinical visits during which multiple neuropsychological and pathological assessments are performed and recorded. The datasets are thus compatible with data-driven studies using machine learning besides longitudinal and cross functional studies. Currently scholars utilise the ADNI data repository to access datasets that fit their own scope of research and can integrate these datasets with ADNI-Merge to access the diagnosis as well as other useful clinical and demographic information.

The key competency of ADNI is the availability of a large number of dementia indicators in a single repository; this enables multidisciplinary research collaboration among researchers from different disciplines including psychiatry, neurology, neuropsychology, artificial intelligence, psychology, computational intelligence, neuroimaging, behavioural science, etc., to perform applied projects that can be useful for the medical community and the stakeholders (patients, family members, clinicians, hospitals, etc.). Another advantage of using ADNI is that the process of accessing the data is quick and transparent if the research team provides a clear research proposal, and they accept to adhere with the data access, data ethics, and publication policies. These reasons help to explain the large number of published works by scholars from different research disciplines that have utilised ADNI-related datasets. Examples of studies that are based on ADNI datasets are: Grueso and Viejo-Sobera (2021), Wang et al. (2020), Nagaraj and Duong (2020), Kemp et al. (2020), Albright (2019), Choi and Jin (2018), Moradi et al. (2017), Podhorna et al. (2016), Balsis et al. (2015), Marshall et al. (2015), Wessels at al. (2015), and many others.

The CFAS is longitudinal dataset that concentrates on the pathology indicators of dementia conditions in the UK. The project has more than 500 participants who willingly donated their brains to the project after death so clinicians could conduct pathological investigations (Wharton et al.,

2011). The dataset does not contain cognitive indicators related to neuropsychological assessments, but pathological indicators (biomarkers) including beta-amyloid (Aβ) and tauopathy features, which limits it use to just biomarker-related studies. In addition, the dataset is limited in size and in the number of features which reduces its use. More importantly, the process of obtaining the data can be lengthy and difficult.

MAS was a 14-year project that was started in 2005 in Australia to investigate the predictors of ageing-related conditions, mainly MCI and dementia, and to provide descriptive analytics to Australians on the rates related to these conditions. The dataset related to MAS comprised cognitive, genetic, and neuroimaging indicators of healthy and cognitively impaired individuals for scholars to study to identify factors illness associated dementia. The participants were recruited from Eastern Sydney and every other year they were assessed cognitively using standard neuropsychological assessments besides pathological assessments—the MAS dataset consisted of N= 362 and their informants N = 358. There are published works on this dataset including Babiloni et al., (2021) and Reppermund (2021). Accessing the data involves contacting the research bank and completing an application process.

The CREDOS data project aims to improve clinical assessments related to dementia by investigating MCI and AD cases in South Korea. Started in 2005 and completed in 2015, it had 800 participants from South Korea. The dataset included indicators related to dementia (AD), and MCI such as biomarkers, functional and structural neuroimaging, and neuropsychological tests. The criteria used to diagnose individuals with AD was based on the DSM-5 framework or NINCDS -ADRDA. The dataset has been associated with some research works such as Yoon et al. (2014), Lee et al. (2016), Bang et al. (2017), and Yoon et al. (2018). For instance, Bang et al. (2017) studied the performance of different machine learning algorithms using a sample of data related to the CREDOS dataset, while Yoon et al. (2018) investigated elements of the KDSQ and MMSE methods using artificial neural networks based on data samples of the CREDOS. So et al. (2017) compared different machine learning algorithms using samples from the Gangbuk-Gu Dementia Centre in Korea.

Since the ADNI data repository has a large number of participants when compared with the available datasets, besides the fact that these participants had multiple visits (up to 22 for a single participants), enabling us to detect the disease advancement, we have selected ADNI for this research. For example, the assessments are recorded at ADNI for the functional and cognitive assessments per individuals every six months, whereas for MAS this happened every two years, which makes detecting the progression clearer in ADNI. Another significant reason for using ADNI is the fact that it contains detailed cognitive and functional assessments per participant visit unlike most of the available datasets. In addition, there are several publications on cognitive and functional

assessments, respectively, based on ADNI which was useful for comparison and critical analysis in the literature review. Lastly, since the ADNI process of obtaining the data observations was straightforward and easier than other data repositories, this was also a core advantage of adopting ADNI.

2.7 Mobile Dementia Assessment Methods

2.7.1 Single Assessment Methods

MOBI-COG (Nirjon et al., 2014), a screening app based on the Mini-Cog Dementia Screening Test (Borson et al., 2000) aims to automatically identify dementia through three different tasks. This app can be administered by either the patient or the caregiver and the test takes 3–5 minutes to complete. It generates a score ranging from 1–10 and requires an expert to interpret the results. According to Nirjon et al. (2014), the app is capable of screening dementia and has a rating of 8.5/10 on the Android App Store.

BrainTest® (BrainTest Inc, 2013) is based on the Self-Administered Gerocognitive Exam (Scharre et al., 2010) and available in both Android and iOS operating systems. The app is designed to be administered by the patient or the caregiver and the test takes 10–15 minutes to complete. According to Scharre et al. (2017), BrainTest® is capable of discriminating AD with a sensitivity and specificity of 71% and 90%, respectively.

ACE (Hodges & Larner, 2017) is an app developed to be administered specially by healthcare professionals and is the automated version of Addenbrooke's Cognitive Examination-3 (ACE-3) medical examination (Noone, 2015). ACE supports iPads and evaluates the attention, memory, fluency, language, and visuospatial functioning capabilities of the individual then presents a score out of 100. According to Bruno and Vignaga (2019), ACE is capable of screening AD with a sensitivity of 93%–100% and specificity of 96%–100% and has a rating of 4+ on iTunes, despite some negative reviews regarding its functionality.

CAIDE-DRS (Sindi et al., 2015) is an iOS app that uses the CAIDE risk score [cardiovascular risk factors, aging, and incidence of dementia] (Kivipelto et al., 2006) to evaluate an individual's cognition based on their biographic information, systolic blood pressure (BP), body mass index, cholesterol level, and level of physical activity. CAIDE-DRS has a test that takes 5–10 minutes to complete and produces a score out of 15. Butcher (2007) revealed that CAIDE-DRS can determine the user's risk of developing AD within the next 20 years with sensitivity and specificity of 77% and 63%, respectively. CAIDE-DRS gained a 4+ rating on iTunes. Similar to CAIDE-DRS, the

DRT app (Google, 2019b) is also based on the CAIDE risk score (Kivipelto et al., 2006) and is available in both Android and iOS formats making it more accessible than CAIDE-DRS. DRT has a rating of 4.8/5 on the Google Appstore with over 1000 recorded downloads.

MoCA (Apple Inc., 2019b; Nasreddine et al., 2005) is an iOS app based on the Montreal Cognitive Assessment which evaluates executive functions, language, attention, memory recall, concentration, and time and place awareness of the individual, to diagnose many conditions including AD. The test presents a score out of 30, where score <26 is interpreted as normal cognition. MoCA is used diagnose AD and other types of MCIs with a sensitivity and specificity of 81% and 86%, respectively (Julayanont et al., 2015). It has gained 4+ a rating on App Store.

eSLUMS (Chewy Logic, 2019) is the digital version of the SLUMS medical test and designed to be administered by a healthcare professional. The app evaluates the user's brain health within 7–10 minutes and generates a score out of 30 where score >27 indicates normal cognition and score <20 indicates severe impairment. A study by Kansagara and Freeman (2010) revealed that SLUMS identifies AD traits with sensitivity and specificity ranging between 98–100%.

Two Android apps: **MMSE** (Google, 2016), and the Dementia & Alzheimer's Memory Diagnosis Test [**DAMDT]** (Google, 2017), have been developed based on the MMSE medical test. Both apps generate a score out of 30 where a score >25 indicates normal cognition and score <25 indicates limitations on certain cognitive areas. Both the apps have a 3+ rating on the App Store, and DAMDT has over 1000 recorded downloads.

The **6CIT** app is based on the Six Cognitive Item Test medical assessment (Callahan et al., 2002). This test takes 5–10 minutes to complete and produces a score out of 28 where score <7 indicates normal cognition and a score >8 indicates an abnormality in cognitive processing and behavioural changes of the individual. The 6CIT app has a sensitivity range from 78–90% and specificity of 100% (Callahan et al., 2002).

2.7.2 Multiple Assessment Methods

DementiaScreener (Mundt et al., 2000) is an Android app based on the Symptoms of Dementia Screener [SDS] (Flaherty et al., 2019) and AD8 Dementia Screening Interview (Galvin et al., 2005) medical exams taking 5–10 minutes to complete. SDS generates a score based on the repetitive behaviours, memory capacity, emotions, attention, and problem-solving skills of the individual. A score <4 indicates low risk and a score >4 indicates a high risk of developing AD. AD8 evaluates the behavioural changes of the individual where a score of 0–1 indicates normal cognition, and a

score >2 indicates impairment. The app gained a 2.7 rating on the App Store with over 1000 recorded downloads.

CDD (Sangha et al., 2015) is a UK-NHS clinically accepted iOS app involving multiple medical assessment methods including the Confusion Assessment Method (CAM; Inouye et al., 1990), Six Cognitive Item Screener (Callahan et al., 2002), the AMT, MoCA, and MMSE to screen several MCIs and dementia conditions. The CDD app is specially designed to be practised on orthopaedic trauma patients and can be used in a clinical setup by clinicians. It has 4+ rating on the App Store.

BrainCheck (Ehrensperger et al., 2014) is an iOS app that comprises three tasks: answering three questions (Holsinger et al., 2007), the Clock Drawing Test [CDT] (Friedman et al., 1994), and seven informant questions from the IQCODE medical method (Jorm et al., 1989). The test takes 10–15 minutes to complete and has a sensitivity and specificity of 97.4% and 81.6%, respectively (Ehrensperger et al., 2014).

ALZ (Berauk et al., 2017) is an iOS app that uses the: Mini-Cog Dementia Screening Test, Memory Impairment Screen [MIS] (Buschke et al., 1999), CDT, General Practitioner Assessment of Cognition [GPCOG] (Brodaty et al., 2002), MoCA, IQCODE, AD8 (Galvin et al., 2005); SLUMS, FAQ (Tappen et al., 2009), Geriatric Depression Scale (Yesavage et al., 1982), Hachinski Ischemic Score [HIS] (Johnson et al., 2014), and the Katz Index of Independence in activities of daily living as a functional assessment medical test (Aske, 1990) to screen for AD. This app is suitable for thorough medical screening and diagnosis of dementia in a clinical setup and can be administered by a clinician.

DementiaTest (Thabtah et al., 2019) is an app based on the 6-item Cognitive Impairment (6-CIT) and the Structured Clinical Interview (SCIDS) medical tests (Callahan et al., 2002; Ouimette & Klein, 1995) and available in both Android and iOS formats. It consists of one self-administered questionnaire which produces a score out of 26 and a caregiver administered questionnaire which produces a score out of 36. Both questionnaires interpret a lower score as indicating a lower risk of developing AD, and a higher score as indicating a higher risk of developing AD. The test takes 5–10 minutes to complete and has a rating of 3+ on the App Store.

Cognitive Exams (Google, 2019a) is an Android app that uses MMSE, CDT, the Geriatric Depression Scale (Yesavage et al., 1982), Katz Examination, Walk Test of six minutes (Abbott et al., 2004) and Fluency Test (Caramelli et al., 2007) to evaluate various functional and cognitive capabilities of the patients. This test can only be administered by a medical professional, takes 45–60 minutes to complete, and has a rating of 4.4 on the App Store with over 10,000 recorded downloads.

2.7.3 Non-conventional Assessment Methods

DTRCT (Apple Inc., 2019a) is an iOS app that uses a self-administered questionnaire based on several neurological tests and interactive tools to evaluate the memory, attention, and logical thinking skills of the user. The test takes 5–10 minutes to complete and provides a few other disease management services other than screening (Comerford et al., 2002).

BCI (BrainCheck, 2019) is a U.S. Food and Drug Administration (FDA) registered class III medical app available on iTunes. The app uses five games: Flanker Task (Fan et al., 2003), the Digit Symbol Substitution Task (Monte et al., 2010), the Stroop Task (Melara & Algom, 2003), the Trail Making Test [TMT] (Cicerone & Azulay, 2002), and balance and coordination (Greenwald et al., 2001), The Immediate and Delayed Recall Tests were developed based on gold-standard neurocognitive tests to evaluate cognitive processing, executive functions, visual attention, immediate recall and the delayed recall abilities of the user. This test takes 10–15 minutes and has a sensitivity and specificity of 83% and 87%, respectively.

Cognity (Inoven, 2018) is the only app that uses artificial intelligence technology to screen for AD through analysing a photo of a clock drawing done by the user in conjunction with the Mental Status Examination [MSE] (ChewyLogic, 2019). A digital version of the St Louis (Snyderman and Rovner, 2009) interactive tool. The test takes 10–15 minutes to complete and produces a score out of 30. It has pre-recorded videos to interpret the results and sensitivity and specificity ranges between 71–92% and 52–96%, respectively, with a recorded rating of 3+.

The **DST** (Google, 2019c) app is available on both Android and iOS formats and used in dementia screening to evaluate the executive functions, memory, verbal fluency, attention, and orientation skills of the user. The test takes 10–15 minutes to complete and generates a score out of 30. A score >29 indicates normal cognition whereas a score <28 indicates impairment in some cognitive areas. According to Google (2019c), this app has a sensitivity of 96% and 3+ rating on the App Store.

Table 2.6 depicts the summary of the digital dementia screening methods that are built on a mobile platform

Table 2.5: Summary of all Considered Digital Dementia Screening Methods Available on a Mobile Platform

Mobile Digital screening method	Target	Rating	Downloads	Apple iTunes	Google play Store	Video	Image	No of Questions	Time	Free/Paid	Cost	Language	Reference
MOBI-COG	Anyone	8.5/10	N/A	X	√	X	√	3	3-5 mins	Paid	N/A	English	Borson et al., (2000)
BrainTest	Anyone	4.5/5	N/A	√	√	√	√	Unknown	10-15 mins	Paid	$39.99	English	Scharre, et al., (2017)
ACE	Health care professionals	4+	N/A	√	√	X	√	24	5-10 mins	Free	N/A	English	Noone, (2015)
CAIDE-DRS	Age 40 to 65	4+	1	√	X	X	X	8	5-10 mins	Free	N/A	English, Finnish, French, German, Russian, Spanish, Swedish	Kivipelto et al., (2006)
DRT	Age 40+	4.8	1000+	√	√	X	X	8	2-5 mins	Free	N/A	English	Kivipelto et al., (2006)
DementiaScreener	Anyone	2.7	1000+	X	√	√	X	11	5-10 mins	Free	N/A	English	Galvin et al., (2005) Flaherty et al., (2019)
MoCA	Anyone	4+	1	√	X	X	√	11	10-15 mins	Paid	$10.00	English, Danish, Dutch, Finnish, French, German, Italian, Polish, Portuguese, Spanish, Swedish	Nasreddine et al., (2005)
CDD	Orthopaedic trauma Patients	4+	N/A	√	X	X	X	10	10-15 mins	Paid	N/A	English	Inouye et al., (1990), Callahan et al., (2002), Hodkinson, (1972), Folstein et al., 1975)
BCI	12 years +	N/A	N/A	√	X	X	√	5	10-15 mins	free	N/A	English, Spanish	Fan et al., (2003), Monte et al., (2010), Melara, & Algom, (2003), Cicerone & Azulay, (2002), Greenwald et al., (2001), Comerford et al., (2002)
BrainCheck	17 years+	N/A	N/A	√	X	X	√	10	10 mins	Paid	$4.99	German, English, French, Italian, Spanish and 25 more	Holsinger et al., (2007), Friedman et al., (1994) Jorm et al., (1989)
ALZ	12 years +	5	N/A	√	X	X	√	100+	45 -60 mins	Free	N/A	English	Borson et al., (2000), Tariq et al., (2006), Tappen et al., (2009), Yesavage et al., (1982), Johnson et al., (2014), Buschke et al., (1999)
eSLUMS	Anyone	3.9	N/A	√	X	X	X	24	7-10 mins	Free	N/A	English and Chinese	Tariq et al., (2006)

DTRCD	12 years +	5	N/A	√	X	X	X	11	5-10 mins	Free	N/A	English, Czech, French, German, Portuguese, Spanish	Torre, (2004)
Cognity	12 years +	3+	500+	√	√	√	√	30	10-15 mins	Paid	$8.99	English and Turkish	Snyderman & Rovner, 2009) Tariq et al., (2006)
DST	Anyone	3+	100+	√	√	√	√	30	10-15 mins	Paid	$8.49	English	No Reference available
MMSE	Anyone	3.4	5000+	X	√	X	X	10	5-10 mins	Free	N/A	English	Folstein et al., (1975)
DementiaTest	Age 60+	3+	50+	√	√	X	X	7	5-10 mins	Free	N/A	English	Callahan et al., (2002) Ouimette & Klein, (1995)
MMSE	Anyone	3+	1000+	X	√	X	√	13	5-10 mins	Free	N/A	English	Folstein et al., (1975)
6 Cognitive Item Test (6CIT) App	12 years +	N/A	N/A	√	X	X	X	6	5-10 mins	Free	N/A	English	Callahan et al., (2002)
CognitiveExams	Age 45+	4.4	10,000+	X	√	X	√	68	45-60 mins	Paid	$3.59	English	Folstein et al., (1975) Yesavage et al., (1982)

2.7.4 Discussion on Digital Assessment Methods

2.7.4.1 DSM-5 Neurocognitive Domains Coverage

There are scientific bodies that set the guidelines for screening and diagnosis of dementia and other types of MCIs, such as the NINCDS-ADRDA, and the American Psychiatric Association (McKhann et al., 1984; American Psychiatric Association, 2013). The NINCDS-ADRDA updated the diagnosis criteria of neurocognitive disorders in 2011 to enhance the medical diagnosis of AD and promote future research. The APA introduced the DSM-5 in 2013 by replacing the term 'dementia' with Major ND focussing entirely on the clinical diagnosis of dementia and its types. It also addresses the severity levels of ND and presents guidelines for clinicians to differentiate between mild and major levels of impairment along with other neurocognitive disorders falling into the same category. Hence, in the DSM-5 guidelines, Major ND refers to dementia, and Mild ND refers to other types of MCIs. DSM-5 describes the six main domains where the impairment can be present for an individual to be diagnosed as a case of Major or Minor ND (American Psychiatric Association, 2013) including complex attention, executive functions, learning & memory, language, perceptual motor, and social cognition.

If an individual is experiencing considerable deterioration in at least one of the above domain capabilities, a screening medical test reported by an informant, caregiver, or a clinician is required to establish the ND symptomatology. If the problem is serious enough to interfere with the independence of the individual's daily operations, the diagnosis suggested by DSM-5 is Major ND; if the deterioration of domain capabilities is mild and has no impact on the individual's daily operations, then the diagnosis would be Minor ND (Sachdev et al., 2014). To make a sound diagnosis, a medical test should be comprehensive enough to capture multiple domains described under the umbrella of DSM-5.

In this thesis, the comprehensiveness of a medical test refers to how well it covers the domains listed under DSM-5 for neurocognitive disorders. To understand that, each dementia screening app discussed was downloaded and evaluated with respect to the six domains of DSM-5. This becomes more complex when looking more closely at the DSM-5 domains, and considering how performance in each domain can be measured. For example, is it possible to separate complex attention, a term used to describe processes such as selectively paying attention only to relevant stimuli in our environments, from executive functioning, a term used to describe processes that prevent interference from distracting stimuli? Such issues were recognised by the APA taskforce who developed the new criteria (Ganguli et al., 2011), but the consensus was to retain these categories for diagnostic purposes.

Furthermore, not only are there difficulties in delineating precise definitions of each domain, but many of the tasks that can be used to measure neurocognitive performance tap into multiple domains. For example, digit span has been proposed as a measure of complex attention, with forward recall requiring repetition of items in the same order they are presented to the patient. This also taps into memory, and when conducted orally, requires command of language. Moreover, because of the manipulation of items in memory demanded by the reverse digit span task (where items should be repeated in the reverse order to that in which they were presented), this variant of the task is often considered to measure working memory, which comes under the domain of executive functioning, but also requires complex attention and memory. As such, it may not always be clear from deteriorations in task performance precisely which domains are impaired. In terms of practical implications for screening, since referral only requires deterioration in one or more domains, despite issues with overlap, poor performance in any subtest will still result in referral for more in-depth assessment of the patient's needs. If some domains are not covered there are concerns about the sensitivity of the screening tool. As a result, we took a pragmatic approach, whereby some of the subtests within the dementia screening digital methods were considered to tap into more than one domain (see Table 2.7 for details).

In some cases, statistical analysis has been used to demonstrate a cognitive tests ability to measure specific cognitive domains, for example Guerrero-Berroa et al. (2009) examination of MMSE. However, for less common cognitive tests, verification studies are not commonly available and so in such cases a test reported cognitive domain coverage has been determined based on the coverage stated by the test authors and our own discretion using the definition of each cognitive domain provided by the Johns Hopkins Psychiatry Guide (Peters & Rabins, 2017).

Table 2.7 illustrates the number of domains covered by each considered digital screening method and shows that few existing apps cover most of the cognitive domains listed under DSM-5. Many of these apps fulfil 1–4 domains, and most of the apps essentially test the user's memory. The MoCA, ALZ, BrainTest®, and CognitiveExams evaluate most of the cognitive areas defined in DSM-5 since they combine many medical tests in which each covers one or more cognitive domains. Therefore, these apps can be seen to be more comprehensive for clinical testing of MCI, and AD despite the complete cognitive domains of DSM, are not evaluated.

One of the notable concerns of the available digital screening methods is that none of them evaluates the social cognition domain. Social cognition often involves behaviours that are not socially acceptable, decisions without considering safety, and insensitivity to social standards, and so by not covering this domain during the screening may lead to people who have deterioration in social cognition being undetected. The only screening apps that might partly cover social cognition are

ALZ and CognitiveExams using the Geriatric Depression Scale (GDS) test, which could indicates emotional behaviours by asking questions on life satisfaction, interests, perception towards life, hopefulness, happiness, problems, worries, social gatherings, etc. The DementiaScreener app also contains the Symptoms of Dementia Screener (SDS) test, which has several questions that could indicate an issue linked with social cognition. However, the questions in the above tests would not allow for a quantifiable measure of social cognition and do not cover all areas of social cognition so a patient exhibiting a decline in social cognition may not be detected by these tests.

Table 2.6: Number of Domains Covered by Digital Dementia Methods

No	Digital Screening Method	Sub-Test	Domains Covered						No of Domains Covered
			Complex Attention	Executive Function	Learning and Memory	Language	Perceptual Motor	Social Cognition	
1	MOBI-COG	Mini-Cog Dementia Screening Test	X	x		x		x	2
2	Brain-Test	Self-Administered Gerocognitive Exam (SAGE)						x	5
3	ACE	Addenbrooke's Cognitive Examination -3 (ACE-3)		x				x	4
4	CAIDE-DRS	Memory Statement	X	x	~	x	x	x	0
5	DRT	Memory Statement	X	x	~	x	x	x	0
6	Dementia-Screener	All tests							5
		Symptoms of Dementia Screener (SDS)							
		AD8 Dementia Screening Interview				x		x	
7	MoCA	Montreal Cognitive Assessment (MoCA)		x				x	4
8	CDD	All tests		x				x	4
		Confusion Assessment Method (CAM)	~	x		x		x	
		Six Item Screener	X	x		x	x	x	
		Abbreviated Mental Test (AMT)		x			x	x	
		Montreal Cognitive Assessment (MoCA)		x				x	
		Mini-mental State Examination (MMSE)		x				x	
9	BCI	All tests				x	x	x	3
		Flanker Task			x	x	x	x	

		Digit Symbol Substitution Task			x	x	x	x	
		Stroop Task			x	x	x	x	
		Trail Making Test (TMT)			x	x	x	x	
		Recall tests	X	x		x	x	x	
10	Brain-Check	All tests	X	x		x		x	2
		Clock Drawing	X	x	x	x		x	
		7 questions from the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE)	X	x		x	x	x	
11	ALZ	All tests						~	5
		Mini-Cog Dementia Screening Test	X	x		x		x	
		Saint Louis University Mental Status Exam (SLUMS)		x				x	
		Functional Activities Questionnaire (FAQ)				x		x	
		Geriatric Depression Scale	X	x	x	x	x	~	
		Memory Impairment Screen (MIS)	X	x		x	x	x	
12	eSLUMS	Saint Louis University Mental Status Exam (SLUMS)		x				x	4
13	DTRCD	Dementia Risk Calculator		x				x	4
14	Cognity	All tests	X	x		x		x	2
		Clock Drawing	X	x	x	x		x	
		Mental Status Examination	X	x		x	x	x	
15	DST	Dementia Screening Test					x	x	4
16	MMSE	Mini-mental State examination (MMSE)		x				x	4
17	Dementia Test	6-item Cognitive Impairment Test (6CIT)		x		x	x	x	2
18	6CIT	6-item Cognitive Impairment Test (6CIT)		x		x	x	x	2
19	Cognitive Exams	All tests		x				~	4
		Mini-mental state examination (MMSE)		x				x	
		Clock Drawing	X	x	x	x		x	
		Geriatric Depression Scale	X	x	x	x	x	~	
		Verbal Fluency Test	X	x	x		x	x	
		Katz Test	X	x	x	x	~	x	

2.7.4.2 Validity and Reporting

Validity defines the suitability of a dementia screening app for clinical and primary care settings. This section evaluates the validity of all the considered apps to determine those that are suitable for use in a clinical setting. Table 2.8 lists all the medical diagnostic procedures used in the considered apps along with their validity measures. The table shows two types of validation tools: those that are recommended by an accepted international body, and others that are validated through an exclusive experiment conducted in a clinical setting—most of the medical diagnostic procedures considered are validated through one or both methods. BrainCheck's patients' questions, and the DST app are not validated by a medical body's recommendation. Nevertheless, the complete BrainCheck app was validated in a study by Yang et al. (2019) using 30 patients, and 568 controls and showed acceptable performance (83% sensitivity and 87% specificity).

Medical diagnostic procedures including MMSE, Mini-COG, AD8 Dementia Screening Interview, 7-item version of the IQCODE, MIS, eSLUM, MoCA, e-SAGE, and GPCOG are recommended by the Alzheimer's Association and American Academy of Neurology (AAN). This means MoCA, MMSE, ALZ, MOBI-COG, eSLUM and BrainCheck can be considered validated apps that can be used in clinical settings by clinicians. According to the Alzheimer's Association's guidelines, a patient should be assessed for the health risk and possible symptomatology before the annual wellness visit which can possibly occur in the primary care setting. Thus, most of the apps with tests that are brief (less than 10-15 minutes), can be administered in a clinical setting according to the Alzheimer's Association and can be considered as valid screening tools for cognitive assessment (Cordella et al., 2013). The performance of these tools is validated in terms of sensitivity and specificity through several research studies (Foster et al., 2019; Arevalo-Rodriguez et al., 2015; Borson et al., 2006; Kansagara & Freeman, 2010; Brodaty et al., 2002).

Reporting corresponds to providing meaningful results electronically to aid the clinician's formal diagnosis. Out of all the considered dementia screening apps, MOBI-COG, DRT, DementiaScreener, DTRCD, DementiaTest, and 6CIT generate reports with scores. Sometimes, the score presented is briefly interpreted with basic recommendations such as the need for further medical assistance. Apps such as BrainTest®, ACE, MoCA, BCI, eSLUMS, Cognity, DST, ALZ, and CognitiveExams produce a formal medical report that can be printed and presented to the clinician during a medical follow up. Some apps have the option to share the reports immediately via e-mail or other electronic means. Most of the apps track the results history, and a few apps such as BCI and BrainTest® are capable of recording and tracking the results and progress of multiple patients. BrainTest® does not provide an interpretation of the score immediately or generate a

report, however, results interpreted by a physician within an instructional video will be e-mailed to the user within five working days of using the BrainTest® app.

Usability, also known as user friendliness, can be defined as a tool's ability to achieve a specified goal in an effective, efficient, and satisfactory manner within a specified context (Bolchini et al., 2009; Douglas et al., 2011). But the question is, what makes a tool effective and efficient in achieving the specified goals? When it comes to biomedical science dementia screening tools, usability is an important parameter that can be used to evaluate their performance. Therefore, each tool needs to be followed by an exclusive usability study to demonstrate its user-friendliness, effectiveness, efficiency, and accessibility—merely a few of the above discussed tools have had usability studies carried out to test their user friendliness for both patient (case) and the clinician.

Newman et al. (2018) used trainee clinical psychologists and post-graduate health science students as the sample population to administer the ACE app to evaluate the usability of the tool. The findings suggested that the ACE app had a considerable error in items pertaining to the language domain. An adjustment was made to remove the incorrect and confusing naming to enhance the usability of the app in a clinical setting. Similarly, Nirjon et al. (2014) conducted a study on eight healthy individuals (four males and four females) to demonstrate that the tool can perform all three tasks of remembering words, clock drawing, and recalling words effectively in a standard primary care setting. The authors used a separate questionnaire with a score range from 1–10 (10 indicating a high level of usability and 0 indicating a low level of usability) given to participants to evaluate the usability of the MOBI-COG app. The findings of the study revealed that the tool is suitable for dementia screening in both clinical and primary care settings.

Table 2.7: Validity Details of the Considered Dementia Digital Screening Methods

No	Digital Screening Method	Medical Examination	Validity		
			Clinically Validated	Recommended by	Reference
1	MOBI-COG	Mini-COG dementia screening test		The Alzheimer's Association	Cordella et al., 2013)
2	Brain Test	Self-Administrated Gerocognitve Exam (e-SAGE)		American Academy of Neurology (AAN)	Foster et al., (2019)
3	ACE	Addenbrooke's Cognitive Examination (ACE-III)		Department of Health and the Alzheimer's Society in the UK.	Bruno & Vignana, (2019)
4	CAIDE-DRS	Cardiovascular Risk factors, Aging and Incidents of Dementia risk score (CAIDE) risk score		_	Sindi et al., (2015)
5	Dementia Screener	Symptoms of Dementia Screener (SDS)		_	Flaherty at al., (2019)
		AD8 Dementia Screening Interview		The Alzheimer's Association	Cordella et al., (2013)
6	MoCA	Montreal Cognitive Assessment		American Academy of Neurology (AAN)	Foster et al., (2019)
7	CDD	Confusion Assessment Method (CAM)		_	Waszynsk, (2012)
		Six item screener		_	Carpenter et al., (2011)
		Abbreviated Mental Test (AMT)		_	Jitapunkul et al., (1991)
8	BCI	Flanker Task		Registered as a Class II medical device with the U.S. Food and Drug Administration (FDA)	BrainCheck Inc., (2019)
		The Digit Symbol Substitution Task			
		The Stroop Task			
		The Trail Making Test (TMT)			
		Balance and coordination			
		The Immediate and Delayed Recall Tests			
9	Brain Check	3 patient questions (BrainCheck)	X	X	X
		Clock drawing test (CDT)		The National Collaborating Centre for Mental Health (UK)	National Collaborating Centre for Mental Health, (2007)
		7-item version of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE)		The Alzheimer's Association	Cordella et al., (2013)
10	ALZ	Functional Activities Questionnaire (FAQ)		National Alzheimer's Coordinating Center (NACC)	Mayo, (2016)

		Hachinski Ischemic Score (HIS)	X	_	Gay et al., (2008)
		Memory Impairment Screen (MIS)		American Academy of Neurology (AAN)	Foster et al., (2019)
		General Practitioner Assessment of Cognition (GPCOG)		The Alzheimer's Association	Cordella et al., (2013)
11	eSLUMS	Saint Louis University Mental Status (SLUMS) Exam		American Academy of Neurology (AAN)	Foster et al., (2019)
12	DTRCD	Mental Status Examination (MSE) tool		U.S. Preventive Services Task Force	Snyderman & Rovner, (2009)
13	DST	Dementia Screening Test	X	X	X
14	MMSE	Mini Mental State Examination (MMSE)		American Academy of Neurology (AAN)	Foster et al., (2019)
15	Dementia Test	6CIT		Alzheimer's Society and the National Collaborating Centre for Mental Health (UK)	Ballard et al., 2013)
		Structured Clinical Interview (SCIDS)		American Society of Addiction Medicine	Gerdner et al., (2014)
16	Cognitive Exams	Geriatric Depression Scale		_	Mitchell et al., (2011)
		Katz basic activities of daily living as functional assessment		The Hartford Institute for Geriatric Nursing, New York University Rory Meyers College of Nursing	McCab, (2019)
		Walk Test of 6 minutes		Neurology section of American Physical Therapy Association's Multiple Sclerosis Taskforce (MSEDGE), Parkinson's Taskforce (PD EDGE), Spinal Cord Injury Taskforce, Stroke Taskforce, Traumatic Brain Injury Taskforce, and Vestibular Taskforce	AbilityLab, (2019)
		Fluency Test		_	Herrera-García et al., (2019)

2.7.4.3 Performance

There are many measures that are widely used to evaluate the performance of a medical screening tool, including sensitivity and specificity. These measures define the capability of a screening tool to recognise dementia-related symptomology and the degree to which it can distinguish dementia from other MCIs (Maxim et al., 2014). As discussed earlier, sensitivity defines the tool's ability to classify an individual with the disease as positive; specificity defines the tool's ability to classify a person without disease as negative (Goetzinger & Odibo, 2011).

Several research studies have been conducted to evaluate the performance of dementia diagnostic procedures such as MMSE, Mini-COG, eSLUMS, CPCOG, and MIS, among others, for different populations (Borson et al., 2006; Cordella et al., 2013; Arevalo-Rodriguez et al., 2015; Foster et al., 2019; Kansagara & Freeman, 2010; Tsoi et al., 2015). In this research, reported sensitivity and specificity figures are considered using multiple research studies. Table 2.9 summarises performance measures along with the associated study from which these values are obtained. According to the information in Table 2.9, the Mini-COG assessment, ACE-III, CAM, and eSLUMS assessments are the highest performing tools with both sensitivity and specificity values over 90% (at least on the sample instances used in their corresponding research studies). Generally, it is not the norm to achieve very high values in both sensitivity and specificity simultaneously. The trade-off value between both measures is often considered as the best.

Table 2.8: Reported Sensitivities and Specificities of Medical Examinations used by the App

Digital Screening Method	Medical Examination	Medical Exam Reference	Sensitivity %	Specificity %	Study Performance Reference
MOBI-COG	Mini-COG dementia screening test	Borson et al., (2000)	95.00–100.00	92.00–98.00	Borson et al., (2006)
BrainTest	Self-Administered Gerocognitve Exam (e-SAGE)	Scharre et al., (2010)	71.00	90.00	Scharre et al., (2017)
ACE	Addenbrooke's Cognitive Examination (ACE-III)	Noone, (2015)	93.00–100.00	96.00–100.00	Bruno & Vignaga, (2019)
CAIDE-DRS	Cardiovascular Risk factors, Aging and Incidents of Dementia risk score (CAIDE)	Kivipelto et al., (2006)	77.00	63.00	Butcher, (2007)
DementiaScreener	Symptoms of Dementia Screener (SDS)	Flaherty et al., (2019)	78.40	84.00	Flaherty et al., (2019)
	AD8 Dementia Screening Interview	Galvin et al., (2005)			
MoCA	Montreal Cognitive Assessment	Nasreddine et al., (2005)	81.00	86.00	Julayanont et al., (2015)
CDD	Confusion Assessment Method (CAM)	Inouye et al., (1990)	94.00–100.00	89.00–95.00	(Waszynsk, (2012)
	Six item screener	Callahan et al., (2002)	68.00–80.00	74.00–80.00	Carpenter et al., (2011)
	Abbreviated Mental Test (AMT)	Hodkinson, (1972)	70.00–80.00	74.00–90.00	Jitapunkul et al., (1991)
BCI	Flanker Task	Fan et al., (2003)	81.00	94.00	Groppell et al., (2019)
	The Digit Symbol Substitution Task	Monte et al., (2010)			
	The Stroop Task	Melara & Algom, (2003)			
	The Trail Making Test (TMT)	Cicerone & Azulay, (2002)			
	Balance and coordination	Greenwald et al., (2001)			
	The Immediate and Delayed Recall Tests	Comerford et al., (2002)			

BrainCheck	3 patient questions (BrainCheck)	Holsinger et al., (2007)	85.80	74.30	Ehrensperger et al., (2014)
	Clock drawing test (CDT)	Friedman et al., (1994)			
	7-item version of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE)	Jorm et al., (1989)	81.40	75.70	
ALZ	Functional Activities Questionnaire (FAQ)	Tappen et al., (2009),	80.30	87.00	Teng et al., (2010)
	Hachinski Ischemic Score (HIS)	Johnson et al., (2014),	89.00	89.30	Moroney et al., (1997)
	Memory Impairment Screen (MIS)	Buschke et al., (1999)	68.00–86.00	84.00–96.00	Tsoi et al., (2015)
	General Practitioner Assessment of Cognition (GPCOG)	Brodaty et al., (2002)	85.00	86.00	Brodaty et al., (2002)
eSLUMS	Saint Louis University Mental Status (SLUMS) Exam	Tariq et al., (2006)	98.00–100.00	98.00–100.00	Kansagara & Freeman, (2010)
Cognity	Mental Status Examination (MSE) tool	Snyderman & Rovner, (2009)	71.00–92.00	52.00–96.00	Snyderman & Rovner, (2009)
DST	Dementia Screening Test	Google (2019c)	96.00	N/A	Developer's note on Appstore
MMSE	Mini Mental State Examination (MMSE)	Folstein et al., (1975)	23.00–76.00	40.00–94.00	Arevalo-Rodriguez et al., (2015)
DementiaTest	6-item Cognitive Impairment Test (6CIT)	Callahan et al., (2002)	78.00–90.00	100	Callahan et al., (2002)
	Structured Clinical Interview (SCIDS)	Ouimette & Klein, (1995)	75.00–100.00	100	Gerdner et al., (2014)
CognitiveExams	Geriatric Depression Scale	Yesavage et al., (1982)	92.50	77.20	Mitchell et al., (2011)
	Katz basic activities of daily living as functional assessment	Aske (1990)	38.00	N/A	Hartigan (2006)
	Walk Test of 6 minutes	Abbott et al., (2004)	82.00	84.00	AbilityLab (2019)
	Fluency Test	Caramelli et al., (2007)	N/A	N/A	N/A

2.7.4.4 The Role of Intelligent Methods

Machine learning is a hot topic that has a direct impact on the prognosis of dementia and other related MCIs. Machine learning is a process that allows medical systems using search techniques to automatically learn from data to improve the accuracy and efficiency of the screening with little human involvement (Gupta & Katarya, 2020; Sammut & Webb, 2017). The potential of these technologies in dementia prognosis is vast, yet few studies have been carried out to discover ways to integrate them with conventional assessment tests (Chua et al., 2019; Shankle et al., 2005; Bennasar et al., 2014). The accuracy of AD screening tools is paramount as it ensures that no patient is left undiagnosed, and it also facilitates the speedy implementation of treatment plans. Hence, assessment tools should be automated to yield better performance and accuracy and to offer rich knowledge for clinicians to make diagnostic-related decisions. Further, patients involved are often elderly and likely to have other physical, social, and accessibility issues, therefore using technology for screening instruments makes the entire process simple, accurate, and convenient for them as well as for the clinicians and caregivers.

Out of all considered mobile screening apps, few employ artificial intelligence in dementia screening. According to Snyderman and Rovner, (2009), Cognity is the first mobile app to utilise artificial intelligence for dementia screening; this app analyses a photo of a clock drawing done by the user and then, based a large sample of patients, determines the probability of dementia risk. The ACE app also has an inbuilt mechanism for automated administration, scoring, and reporting using the human factor approach. Findings of Newman et al. (2018) suggest that the computerised version (ACE app) can capture measures more accurately than the original version of the ACE medical assessment. MOBI-COG, the digitalised version of the Mini-Cog test, is also a fully automated mobile app that uses machine learning techniques to recognise hand-written digits and characters. It uses a k-NN classifier (Gonzalez & Woods, 2002) to identify the characteristics of the clock drawing done by the individual being assessed. Findings of Nirjon et al. (2014) indicate that the automated mobile version is more capable of evaluating the correctness of a clock drawing to detect the presence of dementia and other types of MCIs than the conventional pen-and-paper based Mini-Cog test, with an accuracy of 99.5%.

2.8 Chapter Summary

The process of diagnosing dementia according to the DSM-5 criteria requires progressive impairment to be demonstrated in one of six cognitive domains. This is normally achieved by assessing the subject using a variety of cognitive diagnostic procedures, however the DSM-5 does not prescribe any specific test to be used, so when diagnosing AD according to the DSM-5, the clinician must use a selection of cognitive diagnostic procedures that cover the six domains.

After looking at a selection of commonly used cognitive diagnostic procedures, we have mapped each procedure's sub-sections or tasks onto the neurological domains specified in DSM-5 which the subsection or task aims to measure. We found that no one procedure covered all of the specified cognitive domains, and therefore cognitive and functional items would need to be considered with a medical procedure which included functional abilities as well as complex attention and language to cover most of the cognitive domains specified by the DSM-5.

While this chapter does not definitively settle the issue of selecting which cognitive diagnostic procedure to use when diagnosing dementia, it does provide a foundation on which diagnosticians can compare some of the commonly used procedures to determine how they fit into the DSM-5 criteria. Since no single medical procedure reviewed here covers all cognitive domains involved in diagnosis besides there being a certain unstructured activity that can cover multiple cognitive domains (overlapping of cognitive domains within the tests / activities of the cognitive diagnostic procedures), a selection of at least two procedures is required. Therefore, deciding which cognitive diagnostic procedures to use could be based on how many cognitive domains it covers besides sensitivity, specificity, or the time it takes to administer.

In this chapter, we also reviewed machine learning studies focusing on the cognitive aspect of the dementia assessment work. We studied functional abilities and cognitive domains and their correlations from a psychological perspective and given relevant recent literature on cognitive and functional abilities defined in neuropsychological assessment theories.

Lastly, although there are many digital assessment applications available to screen dementia, many of them just partly cover the neurocognitive domains defined in the DSM-5 gold standard. We filtered out the digital screening methods available on a mobile platform to 20 after excluding apps that are not used for dementia screening. We then introduced five new criteria: DSM-5 Coverage, Performance, Validity, and Reporting, and the Use of machine learning techniques, to critically analyse these digital methods. The literature review showed a lack of use of emerging technologies particularly the adaptation of intelligent techniques such as machine learning in developing digital dementia assessment methods. Therefore, to fill the gap, we introduce in Chapter 3 a new data driven

framework using machine learning that reveals insights of functional and cognitive elements and offer it to the clinicians within a digital platform to improve not only accuracy of the diagnosis but also accessibility and time.

Chapter Three

A Machine Learning Architecture for Alzheimer's Disease

This chapter introduces the proposed data-driven architecture: Machine Learning Architecture for Alzheimer's Disease Progression (MLA-ADP), and details its major components: data understanding, data preparation, data modelling, progression class labelling, classification system design, and validation. Most of this chapter's content is being considered for publication in the Journal of Biomedical Informatics.

3.1 Introduction

This chapter proposes the MLA-ADP that consists of multiple important phases particularly

- Modelled datasets with target class labels that have been constructed from multiple sources
- A feature assessment phase to pinpoint influential features
- A rule-based classification algorithm for learning rules that can be used by clinicians for decision making
- A novel method for building classification models for predicting disease progression
- A mobile platform for conducting assessments for individuals undergoing pre-diagnosis of any dementia stages or MCI.

The new datasets integrate multiple features obtained from dissimilar datasets including cognitive, functional, and demographics within the ADNI data repository. After careful investigation of the ADNI data repository, we identified the required features and combined multiple datasets. We then modelled the new datasets by creating a new class label based on the disease progression. Further details on the datasets, features, data pre-processing, and data modelling are given in Section 3.3. The datasets have been integrated and loaded into a Google firebase database after data modelling and data sampling.

The MLA-ADP contains a feature assessment phase in which multiple feature selection methods were employed to assess feature-feature and feature-class significance. We sought functional and cognitive features that could directly trigger the progression of AD, and any correlations. Section 3.4 discusses the methods used for feature assessment. More importantly, we propose within MLA-ADP a rule-based classifier called Alzheimer's Disease Class Rules (AD-CR) to derive a knowledge

base of understandable If-Then rules, and more importantly to learn models for predicting the progression status of anyone undergoing pre-diagnosis. Section 3.5 provides details of the classification algorithm details.

The MLA-ADP is designed and implemented on a cloud-based environment to enable easy and affordable access. Java has been used for the implementation of the AD-CR; the core data used for deriving the classification models was loaded into a Google firebase. The architecture partly can be accessed via a new proposed mobile application for dementia progression detection (Section 3.6 gives more details on the mobile application components and design).

This chapter is structured such that Section 3.2 sheds light on the MLA-ADP architecture and its main phases. In subsequent sections details on each phase are provided. Finally, a summary is presented in Section 3.7.

3.2 Proposed Machine Learning Architecture for Alzheimer's Disease Progression: MLA-ADP

As discussed earlier, integrating technological advancements into traditional dementia assessments is a promising topic that is yet to gain the attention of the relevant research bodies. Further research that focuses on improving the accuracy and efficiency of dementia screening and diagnosis tools by using advanced intelligent technology is required. There has been limited research on adapting machine learning techniques for use in these digital systems (web and mobile systems) to enhance functionalities, performance, accuracy, and accessibility, hence our proposal.

MLA-ADP consists of a cloud-based architecture comprised of multiple ADNI datasets (ADNI-Merge, ADAS-Cog-sheet, and FAQ-sheet) (ADNI, 2021), a feature assessment, an intelligent algorithm based on rules discovery, and a mobile interface of medical assessment for dementia screening. MLA-ADP replaces conventional scoring functions used in traditional dementia medical screening and diagnosis methods with more advanced intelligent classification systems—potentially a great help for clinicians. MLA-ADP derives classification models using a classification method that are objectively learnt from functional and cognitive features besides features related to the cases and controls making the models and decisions less biased. These models can be exploited by the clinician to aid and improve the accuracy of the decision-making process.

The main phases of the proposed architecture are depicted in Figure 3.1. The first involves forming a database that consists of dissimilar features of data collected from various datasets after careful investigation and analysis. In particular, we collected data related to cognitive, and functional areas besides patients' characteristics and medical clinic visits. Since one of the aims of the research

project is to develop an intelligent method for dementia progression that learns models from both functional and cognitive areas, we have included functional and cognitive medical assessment methods. Further information on the data collection and data access is given in Section 3.3.1.

Once the features (cognitive, functional, demographics, patients' visits, etc.) have been identified, the second phase prepares the data in terms of missing values treatment, normalization, discretization, integration, balancing, and verification of the scores and answers from the medical assessments of the data subjects, among others. More importantly, we integrate the different data sources into a centralised dataset using the patient ID (RID) and visit code (VISCODE). The processed datasets after data modelling and balancing are stored in a Firebase database to ensure that the MLA-ADP is accessible and cost effective and can run in real-time when accessed via a web or mobile environment. Further details on data integration and cleansing are given in Section 3.3.

Prior learning the models against the integrated data, there was no disease progression attribute in the ADNI data repository, and only the medical diagnosis (DX attribute) was provided in the ADNI-Merge dataset, and per the patient's medical visit. Therefore, we modelled the data and created a

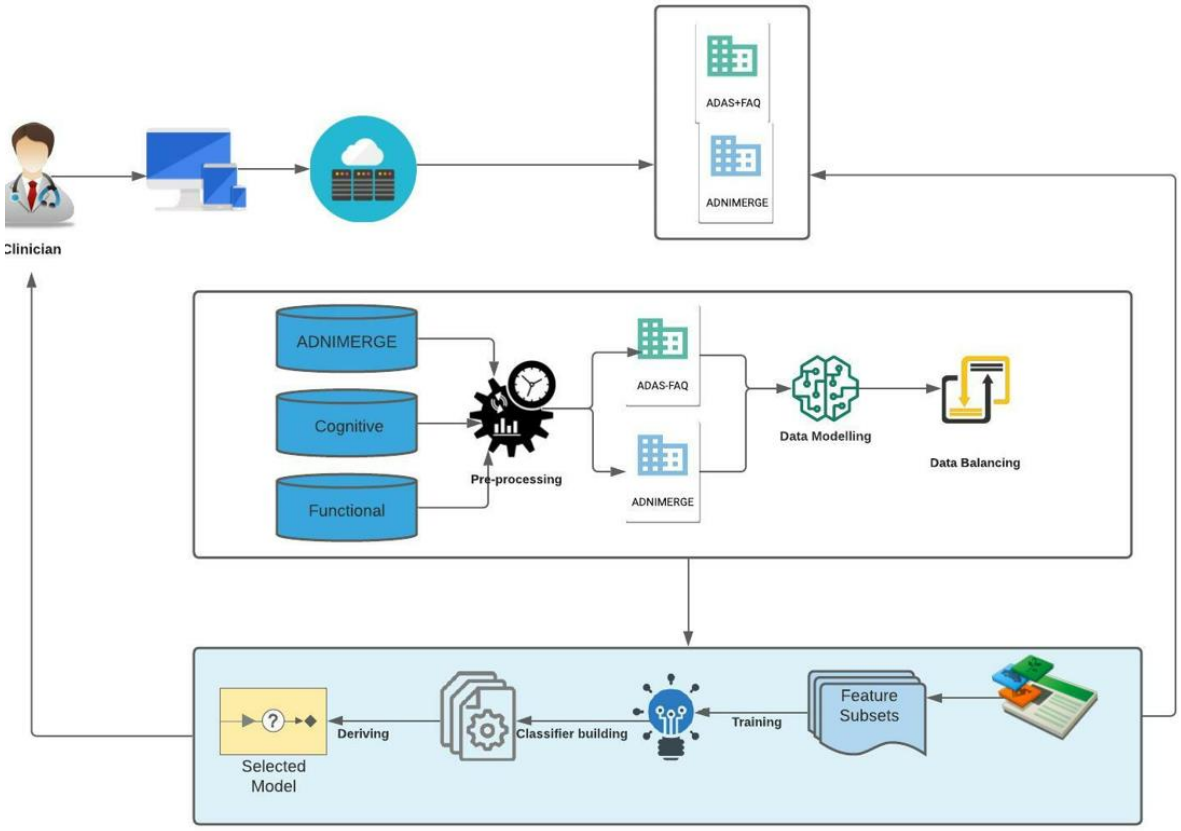

Figure 3.1: The Proposed Machine Learning Architecture for Alzheimer's Disease Progression (MLA-ADP)

new target class label for disease progression called (DX Progress) using a new method that we developed. Further details on data modelling and creating the new disease progression attribute are

given in Section 3.3.4. The final database obtained after modelling was imbalanced and favoured the "no progression" target class value thus it required balancing to avoid deriving biased classification models as discussed in Section 3.3.4.

One of the fundamental steps in the proposed architecture is to identify features that influence any sort of advancement of the disease and at any stage. Thus, in the MLA-ADP architecture, a feature selection phase on the processed data after data integration, modelling and balancing, is performed to identify the features that may trigger advancement of the AD. The features' subsets that will be identified during feature selection are then processed by classification algorithms to derive classification models. The classification approach has been used is based on a rule-based classification algorithm to discover rules that are humanly interpretable. The reason for selecting rule-based classification particularly class association rule is due to that the generated rules can be easily used by the clinicians to clarify any diagnostic decision and reveal associations among cognitive and functional features. Creating a rule-based knowledge foundation using the proposed rule algorithm provides clinicians with a digital information sheet to guide their decision-making process and so they can better understand each case's context. More crucially, the knowledge base will offer the patient and other stakeholders explanations on cognitive and functional items that have contributed to the disease progression, if any. Providing the rules complies with the patient's 'right for an explanation' as outlined in the General Data Protection Regulation (GDPR) regarding decision-making. Another reason is that class association rule approach derives higher predictive classification systems than most conventional classification techniques (Constantino et al., 2021; Ragab, 2019; Hadi et al., 2018; Abdelhamid, et al., 2012; Liu et al., 1998).

The training on the dataset uses ten-fold cross validation to minimise the chance of overfitting. Further details on the feature selection and the classification algorithm are provided in Sections 3.4 and 3.5, respectively.

In the proposed architecture, the clinician uses a mobile environment to record values related to the functional items in a questionnaire and activities-based scenarios. The design of the mobile interface follows the requirements required in designing mobile applications related to dementia (Kerkhof et al., 2017). Further details on the mobile interfaces are given in Section 3.6.

Once the clinician records the patient's information, the models learnt from the data stored in the Firebase on the cloud architecture are triggered, and a target class of whether there will be progression will be given and displayed on the mobile platform. The database gets updated periodically every three months and all classified cases become part of the training data subjects in the database. Communication between the classification algorithms, the dataset, and the mobile environment is implemented automatically on the cloud architecture. Lastly, various evaluation

measures that reveal the models' predictive ability for individuals undergoing the screening process, are also derived to reveal true performance. In the next subsection, we explain all the phases of the MLA-ADP in more detail.

3.3 Data Understanding, Access, Preparation, and Modelling

In this section, we describe the processes of data preparation and modelling including data access, data understanding, medical methods items, and data pre-processing as depicted in Figure 3.2. The detail of each activity is given in the following sub-sections.

3.3.1 ADNI Data Access

Initially, the unprocessed datasets are obtained from the ADNI project data repository—a longitudinal research project in the United States that started in 2004. To obtain the final diagnosis of the data subjects, most of the existing dataset's of ADNI link to a centralised dataset that combines the multi-phase project of ADNI (1, 2, 3, and GO) known as ADNI-Merge (ADNI, 2021). ADNI-Merge comprises a large number of features and 2,361 data subjects (cases and controls) who have been diagnosed with three possible target classes: CN, MCI, and AD. For each data subject, there could be multiple clinical visits, normally one every six months. Thus, each data subject is linked with at least one data instance (a visit) or multiple data instances. The total number of data instances (visits) in the ADNI dataset is 14,628. Further details on the ADNI project, its datasets, and data subject characteristics are provided in Chapter 4.

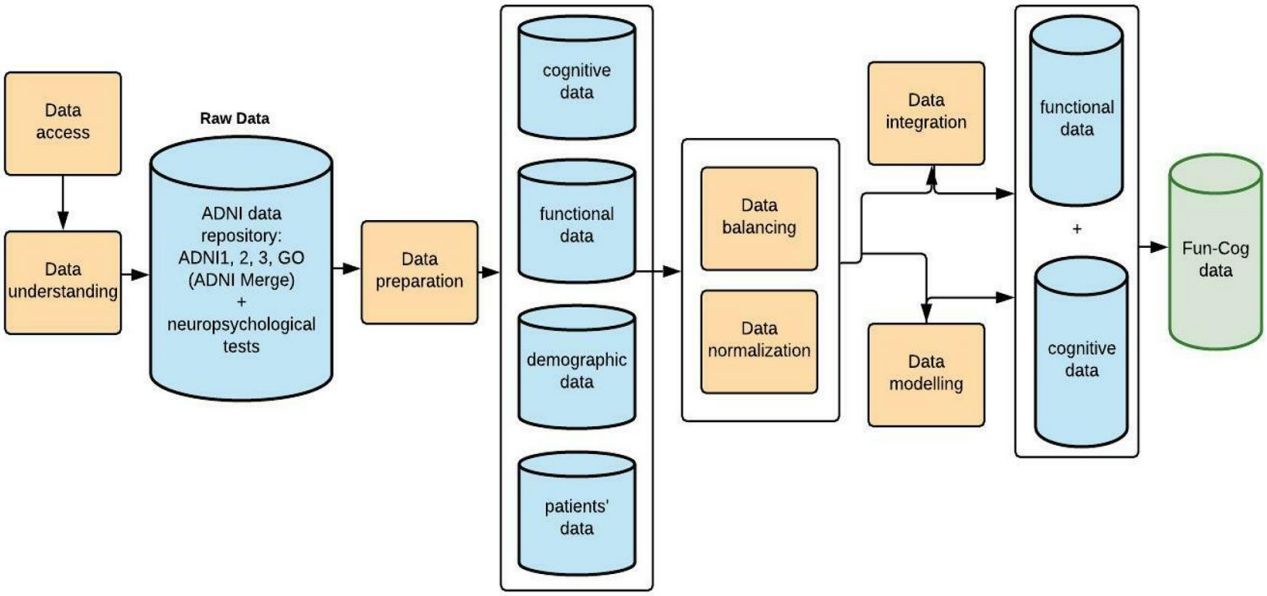

Figure 3.2: Data Preparation Processes of the MLA-ADP

Initially, we obtained data access to ADNI's data repository through a registration process that required relevant information about the researcher, the institution, the purpose of using the dataset(s), the research aims and objectives, and the research proposal details. The process is electronic, and the registration forms were completed online via ADNI's website (http://adni.loni.usc.edu). We submitted the relevant details needed for the data access application of ADNI, then some weeks later were granted access to the data repository by the ADNI data control and access team.

As part of this approval process we agreed to provide annual updates on the use of ADNI data in our research project. The ADNI team keeps track of applications that were approved; we can add additional information in regard to the progress of the research project including methodology, analysis, implementation, systems evaluation, and publications. The update process is conducted within the electronic workspace of the main researcher registered at the ADNI's web management system. Publications, if any, that have explored datasets related to ADNI are recommended to be approved by the Data and Publication Committee prior to submission.

3.3.2 Cognitive and Functional Data Understanding

The project scope is to detect any possible progression points that may occur for individuals in relation to AD and using neuropsychological assessments. We therefore investigated the ADNI's data repository thoroughly to determine features related to neuropsychological assessments including functional and cognitive, patients' medical history and visits, patients' demographics, and the diagnosis assigned by the clinicians, among others. These features were scattered over many datasets, posing a great challenge on which features are needed, and how the different features could be combined. Listed below are the challenges we encountered during the data understanding and data preparation:

- Various medical diagnosis methods for AD that are related to neuropsychological assessments have dissimilar performance so which one(s) to choose
- The medical diagnosis methods available cover partially cognitive criteria defined in the DSM-5 framework
- The medical diagnosis methods may overlap in the cognitive criteria defined in the DSM-5
- No dataset that contains all needed features is available
- No dataset that contains cognitive and functional features together is available
- No progression target class was found; the baseline and the final diagnoses only per patient's visit were available in the ADNI-Merge dataset
- Many neuropsychological assessments' datasets consist of a large number of missing values

- ADNI-Merge dataset contains the summative score of the neuropsychological methods and none of their features (items/activates) and values
- The neuropsychological assessments vary in their usage in detecting dementia since they can be used in detecting different stages of the disease

We have merged data subjects who participated in different phases of the ADNI project, in particular ADNI-1, ADNI-Go, ADNI-2 to obtain as many data subjects as possible to validate the proposed data-driven architecture. Chapter 2 facilitates domain understanding; besides understanding the items, familiarising ourselves with the metadata of the ADNI repository was vital to assist our data preparation processes. Three datasets: ADNI-Merge, ADAS-Cog-sheet, and FAQ-sheet datasets were retrieved from the ADNI database repository—these cover the scope of the research project and they include patients' cognitive and functional information as well as visits (ADNI, 2021). The ADNI-Merge dataset has a cohort of 2,260 participants, mainly based in the United States and Canada, in which each participant is being monitored on a six-monthly basis to track and study their AD progression, with multiple observations per patient at different points in time.

The ADAS-Cog-sheet dataset comprises cognitive tasks that assess learning and memory, language production, language comprehension, constructional praxis, ideational praxis, and orientation as discussed in Chapter 2. There are as many as 31 variants to the ADAS method; we will be analysing the ADAS-Cog13 version which is accessible in the ADNI data repository. It consists of approximately 120 attributes excluding the total score covering several activities with a few tasks having a slightly different scoring range. The total ADAS-Cog13 scores range from 0–85, with the largest score indicating significant impairment. Podhorna et al. (2016) suggested that a score of 15 be considered MCI, while a score of 30 be considered as mild AD. Table 3.1 displays the ADAS-Cog13 primary activities, their scores, and mapping to the DSM-5 framework.

The FAQ-sheet dataset in the ADNI data repository is based on the FAQ method, which assesses functional capability to carry out daily living activities. The FAQ dataset consists of 23 attributes, 10 of which are items with each assigned a 0–3 possible response: dependent = 3, requires assistance = 2, has difficulty but does by self = 1, normal = 0, never did the activity but could do now = 0, never did and would have difficulty now = 1. The total FAQ scores range from 0–30, with the largest score signalling significant impairment. According to Teng et al. (2010), the cut-off points for a total FAQ score < 6 is diagnosed as MCI, whereas ≥ 6 is most consistent with a clinical diagnosis of AD. FAQ items, their scores, their mapping to the DSM-5 platform, and the design and implementation of the FAQ within a mobile platform, are discussed in Section 3.6.

Table 3.1: ADAS-Cog Items and their Associated DSM-5 Cognitive Domains

Question No.	Column in Dataset	Tasks	Scoring	DSM-5 Cognitive domains
1	WORDRECALL	Word recall	0-10	Learning and memory Language
2	COMMAND	Commands	0–5	Language Executive function Perceptual motor function
3	CONSTRUCT	Constructional praxis	0–5	Learning and memory Executive function Perceptual motor function
4	DELAYWORD	Delayed word recall	0–10	Learning and memory
5	NAMING	Naming objects and fingers	0–5	Learning and memory Language
6	IDEATIONAL	Ideational praxis	0–5	Complex attention Learning and memory Executive function Perceptual motor function
7	ORIENT	Orientation	0–8	Learning and memory
8	WORDRECOG	Word recognition	0–12	Learning and memory
9	RMBRTESTINSTR	Remembering test instructions	0–5	Learning and memory Language
10	LANGUAGE	Comprehension of spoken language	0–5	Complex attention Language
11	WORDFIND	Word-finding difficulty	0–5	Learning and memory Language
12	SPOKENLG	Spoken language	0–5	Language
13	NUMBERCANCEL	Number cancellation (time limit: 45 seconds)	0–5	Complex attention Executive function Learning and memory
		TOTAL	0–85 points	

3.3.3 Data Integration

Table 3.2 presents the description of the retrieved datasets before pre-processing. Further details on the datasets and their characteristics including descriptive analysis are given in the next Chapter.

Features such as visit information related to the patient (PTID, VISCODE, SITE, COLPROT, ORIGPROT, etc) and biomarkers (AV45, ABETA, TAU, PTAU, etc) were removed from the ADNI-Merge dataset as they are out of the study's scope. In addition, features related that describe items and activities within the neuropsychological assessments were also excluded from the datasets as we were only interested in their values. Further details on the data modelling are given in the next sub-section.

All missing values such as incomplete individual task scores or missing diagnosis attribute (DX) values, were dropped. It was important to clean the datasets individually before integration to maintain decent computational performance. The reasons for ignoring missing values are:

1) The sensitivity of the application under consideration, so good data quality was required

2) Ensuring that data is cleansed before feature selection and model training and thus the expected outcome(s) are genuine and are not based on data approximation.

While the FAQTOTAL attribute in the FAQ-sheet dataset was within the 0–30 range, the scores of individual FAQ items in the same dataset ranged from -1–5, and when tallied, would exceed the

Table 3.2: General Statistics of the Datasets before Pre-processing

Dataset Name	# of Features	# of Patients	# of Data Observations (visits)	Missing Values in Key Attributes
ADNIMERGE	113	2,260	14,627	Class DX: 4,243 missing values
ADAS-Cog	121	1,751	6,770	100 missing values: 91 across ADAS tasks and 9 in VISCODE2 attribute
FAQ	23	2,267	10,905	FAQTOTAL: 131 invalid data (99 missing values; 32 incomplete data (-1))

maximum score of 30. We investigated the ADNI FAQ scoring procedure and identified that in the FAQ test used to execute the procedure in the ADNI study, there were six answers to choose from as given in Table 3.3. Each of the six answers has its own unique number labelled from 0–5, with -1 representing missing or incomplete data. We learned that the numbers were in fact labels to distinguish the responses clearly and were not representative of the FAQ scale. This data entry method was probably used to reduce data entry error as some of the answers share similar FAQ scores. To deal with this discrepancy, we mapped the six answers from the FAQ-sheet dataset to their corresponding FAQ scale values within the FAQ medical method to check whether the FAQTOTAL matched. We then executed the mapping procedure and compared both FAQTOTALs and verified that the data subjects within the FAQ-sheet are accurate and within the FAQ-scale

Table 3.3: Mapping Scores in the FAQ Dataset and FAQ Scale Values

ADNI FAQ	Representation	FAQ Scale
-1	Missing or incomplete data	n/a
0	Normal	0
1	Never did, but could do now	0
2	Never did, would have difficulty now	1
3	Has difficulty, but does by self	1
4	Requires assistance	2
5	Dependent	3

defined within the FAQ medical method. Data integration was performed to produce two datasets that would address cognitive (ADNI-Merge-ADAS) and functional (ADNI-Merge-FAQ) items individually. Once the relevant attributes were identified from the considered datasets in the ADNI study, we proceeded with data integration to create two new datasets. These new datasets combined ADNI-Merge with the ADAS-Cog-sheet and FAQ-sheet datasets, respectively. Since the ADAS-Cog-sheet and FAQ-sheet datasets have fewer data instances in comparison to the ADNI-Merge dataset, some data instances may not have a match due to the medical test not being taken, and will thus result in further missing values, and additional cleansing being required. In Chapter 4, we evaluate the proposed MLA-ADP on additional subsets of data related to ADAS-Cog-sheet and FAQ-sheet datasets including the dementia sub-categories from CN to MCI and MCI to AD respectively (for up to three years).

There are two common attributes within each dataset that form the basis of the data integration: the patient ID (RID) and the visit code (VISCODE), that will act as a compound primary key reference for the merger. The aim of the merging process is to capture individual cognitive and functional items together from their respective datasets and the diagnostic class (DX) from ADNI-Merge for each visit per patient. A cross-check was done on the individual scores of the ADAS-Cog and FAQ items in the medical methods and tallied to their total scores ADASTOTAL and FAQTOTAL to find any mismatches. As discussed above, FAQ items were identified as having a different score range to the original FAQ scale and did not tally to the FAQTOTAL, thus validation was required to ensure the scores were correct. There were also instances where a patient observation in the ADNI-Merge dataset was not integrated due to the ADAS-Cog-sheet and FAQ-sheet datasets not having a corresponding RID and visit code. This could be due to the assessments not being performed for the patient during a visit for various reasons, thus no merging occurs, resulting in fewer observations in the new datasets.

3.3.4 Data Modelling and Data Balancing

As the main element of our research is the progression of the disease, the diagnosis (DX) attribute originating from the ADNI-Merge dataset is the key to our data modelling process. However, no dataset in the ADNI data repository contains any progression attribute. To deal with this crucial issue, two new attributes are created to establish a target class attribute that will capture the diagnosis progress for each patient and their subsequent visits. The process we followed and have called ‘DX Progress’ is described in Figure 3.3.

Initially, we created an attribute called ‘DX Digit’ (Line #3) to encode the three possible diagnoses (CN:1, MCI:2, AD:3); this attribute will help us assign the appropriate values to the DX Progress.

The ‘DX Digit’ is filled based on the current diagnosis attribute (DX) in the original dataset (ADNI-Merge). The process of data modelling starts by iterating over the data subjects after they are ordered by patient number (RID), and then iterating over the patients’ visits. We always set the new attribute ‘DX Progress’ value to ‘0’ (no progression) for each patient’s first visit. The ‘DX Progress’ captures the change of diagnosis in a patient, establishing a new target class attribute.

The DX Progress attribute will model the changes of the DX digit from the matching patient ID and their subsequent visit with three possible class values. When there was a progression of diagnosis from CN ‘1’ to MCI ‘2’, or MCI ‘2’ to Dementia ‘3’, we labelled the change as ‘1’ in the ‘DX Progress’ attribute (Lines 8-9). If there was no progression, it was labelled as ‘0’; regression as ‘-1’

Input: D: a dataset of all patients’ information and visits
Output: D’ : A dataset with the new target variable ‘Dx Progress’

```
1.  D’ = D
2.  for each rid in D’ do
3.  D’.‘dx digit’  = 0
4.      for each viscode2  in D do
5.          D’.‘dx digit’  = d
6.          if (dn = dn-1)
7.              D’.‘dx progress’ = 0
8.          elseif (dn > dn-1)
9.              D’. ‘dx progress’ = 1
10.         else
11.             D’.‘dx progress’ = -1
12.     end
13. end
14. remove all data instances where ‘dx progress’ = -1
```

Figure 3.3: Modelling Process of the Data

(Lines 6-7). Once the new class (‘DX Progress’) was derived, we removed instances that had been assigned with regression ‘-1’ (Lines 11), as we focus only on classes that are either ‘1’ for progression or ‘0’ for no progression, with only two class values remaining.

A brief analysis is conducted after data modelling to ensure that the updated dataset (D’) after creating the new target class is not imbalanced to avoid any biased classification models. The result of the brief analysis shows that modelled data is imbalanced with respect to the ‘DX Progress’ and that the class set ‘DX Progress’ is linked with a high number of no progression (‘0’) class versus progression (‘1’). Proceeding with an imbalanced dataset to learn classification models produces skewed results that favour the majority target class and ignore the minority class. The primary challenge of dealing with datasets that are imbalanced is that most of the classification algorithms tend to produce poor models in terms of predictive performance on the low frequency class, despite the fact that in medical applications, such as detecting AD progression, the models’ performance on minority class, i.e. ‘progression’, is fundamental.

To deal with the imbalanced data situation, we oversampled the data by synthesized data observations of the low frequency class from existing data samples to add new information to the classification models that we expect to produce later via the classification algorithm. This approach of synthesizing data observations from existing data that belong to the minority class is based on an effective sampling approach in machine learning called Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2000). We arbitrarily generate new observations of the minority class to move the number of minority class items closer to the majority class. A randomisation technique is applied to ensure the minority class is randomly distributed throughout the dataset, and that the newly added data observations of 1s do not amalgamate in some of the folds. The choice of SMOTE is due to the fact that this data sampling method employs an intelligent machine learning algorithm, i.e. kNN (Aha, et al., 1991), to generate the synthetic data observations besides it has been utilised successfully for class data imbalance problems in dementia medical research, e.g. (El-Sappagh, et al., 2021; Yang et al., 2020).

The process of generating synthesized data observations of the minority class is done by initially setting up the size of the oversampling data observation and choosing a feature vector of the minority data observation randomly, i.e. *de*. Then, using the k-NN algorithm, 5 neighbours of the '*de*' are identified, and *X* of these are used to create the synthetic data observation. Often, the sampling algorithm employs any distance function to measure the difference in distance between neighbours and the feature vector. We embedded the revised data oversampling method of SMOTE into the MLA-ADP architecture. The mathematical notation of the oversampling method we used is illustrated below in Equation (3.1)

$$S = x + u.(x^r - x), \quad (3.1)$$

with $0 \leq u \leq 1$,

where u was randomly chosen from $U(0,1)$,

x is a set of variables,

x^r is randomly chosen among the 5 minority class nearest neighbours of x

3.4 Feature Selection

One of the vital steps in supervised learning to identify relevant items in the dataset is feature selection. Feature selection for classification problems in machine learning involves finding features that have high correlation with the target class in an automated manner. The relevancy of the feature to the target class can be found mathematically in a class of methods that we normally call filters. The process of feature selection via filters can simplify the data analysis by not only reducing the

input data dimensionality but also by pinpointing influential items. More useful knowledge is thereby offered to the domain expert and the quality of the models derived is potentially improved.

In the MLA-ADP architecture, a range of feature selection techniques including Information Gain (IG), Chi Squared Testing (CST), ReliefF, and Person Correlation (Mitchell, 1997; Liu & Setiono, 1995; Kononenko, 1994; Pearson, 1920) were employed to detect corelations from the datasets considered. These methods employ dissimilar mathematical models to define feature relevancy and have been successfully applied in previous dementia-related studies such as Pereira et al. (2018), and Zhu et al. (2020). More importantly, there could be discrepancies in the scores of features produced by the feature selection methods. Various metrics are used in computing the scores, so to reduce the vitality in the results, we employ more than one feature selection method. The results obtained from the feature selection methods each produce a different scale of weightings, as such normalising the results is necessary to maintain a common scale and simplify the analysis. The assessment of features to keep or omit was based on different criteria as discussed in the experimental analysis of Chapter 4.

Initially, we conducted a feature-to-feature assessment without the new target class attribute, to gain an understanding of the relationship between the cognitive and functional items. This is to establish their dependency towards the class attribute—a high correlation between two items will almost have the same effect on the dependent attribute, thus having the same properties towards the target class, so one of the two items can be removed. The cognitive and functional items are simultaneously referenced against their DSM-5 cognitive domains to understand their sensitivity towards AD diagnosis. This will further assist our analysis when performing feature selection to derive a variety of subsets that covers a mixture of DSM-5 cognitive domains.

The Pearson Correlation Coefficient was used to measure the linear correlation among the items excluding the class variable and the results were presented in a matrix model. This indeed pinpoints which cognitive activities are similar, allowing us to suggest key items needed for the progression of AD. The correlation coefficient is calculated using Equation (3.2) deriving a value in the range of [-1, +1]. The closer the coefficient number towards -1 or +1, the higher the dependency between the features.

$$r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \qquad (3.2)$$

Where $r = Pearson\ Coefficient,$

$n = number\ of\ variables$

$x_i = the\ values\ of\ the\ x - variables\ in\ a\ sample$

$\bar{x} = the\ mean\ of\ the\ values\ of\ the\ x - variable$

$y_i = the\ values\ of\ the\ y - variables\ in\ a\ sample$

$\bar{y} = the\ mean\ of\ the\ values\ of\ the\ y - variable$

After identifying pairs of features that were highly correlated with each other in both cognitive and functional domains, we were able to proceed to identify feature-class corelations. For this purpose, we used three methods: IG, CST, and ReliefF, to produce scores based on the correlations between the features and the class label using mathematical models. IG employs Shannon's Entropy to decide how informative the feature is based on Equations 3.3 and 3.4. IG also can be used to decide the ordering of features while building classifiers such as decision trees.

$$IG(T,X) = Entropy(T) - Entropy(T,X) \quad (3.3)$$

and $$Entropy(x) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i), \quad (3.4)$$

where $p(x_i) =$ probability of x_i in T, T is the input dataset, X is a subset of T that belongs to a particular feature.

CST is a statistical method that determines if two features are associated using the expected and actual frequencies in the dataset according to Equation (3.5)

$$x^2 = \sum_{i=1}^{n} \frac{(A_{ij} - E_{ij})^2}{E_{ij}} \quad (3.5)$$

Where:

A_{ij} = observed frequency; no. patterns in the i^{th} interval, j^{th} class,

E_{ij} = expected frequency of $A_{ij} = R_i * C_j/N$

R_j = no. patterns in the i^{th} interval = $\sum_{j=1}^{k} A_{ij}$,

C_j = no. patterns in the j^{th} class = $\sum_{i=1}^{2} A_{ij}$,

N = total no. patterns = $\sum_{i=1}^{2} R_i$,

Lastly, ReliefF is a method that handles noisy data and detects correlated features in a dataset; it gives a higher importance value to features that lead to better class separability. The mathematical notation of ReliefF is given below.

$$W[A] := W[A] - \sum_{j=1}^{k} \frac{diff(A,R_j,H_j)}{m.k} + \sum_{C \neq class(R_i)} \frac{\left[\frac{P(C)}{1 - P(class(R_i))} \sum_{j=1}^{k} diff\left(A, R_i, M_j(C)\right)\right]}{m.k}; \quad (3.6)$$

Where set all weights $W[A] := 0.0$;

For $i := 1\ to\ m\ do\ begin$

Randomly select an instance R_i;
Find $k\ nearest\ hits\ H_j$;
For each class $C\ \neq class(R_i)do$
From class $C\ find\ k\ nearest\ misses\ M_j(C)$;
For A: 1 $to\ a\ do$

The reasons for choosing these feature selection methods include:

- They use dissimilar models to define feature relevancy
- They were used successfully in previous medical research related to dementia such as by So et al. (2017), Trambaiolli et al., (2017), and Zhou et al. (2018), among others
- They are easy and efficient to implement
- They are filter-based so when applied against classification datasets such as ours they tend to not peak into the classification results by devising classifiers to measure their performance—they are thus considered less biased yet more efficient when compared to wrapping feature selection methods.

3.5 The Proposed Classification Algorithm

In this section, we discuss the machine learning algorithm that we developed as part of the MLA-ADP architecture: a knowledge-based algorithm consisting of If-Then rules and constructed using a class association rules approach, and it is used for prediction. Initially we introduce class association rule mining, and in the next two sub-sections we explain how the learning algorithm works.

3.5.1 Class Association Rule

Class association rule is a branch of classification methods which employs association rule mining to solve classification problems (Zhang, 2022). This method produces a classifier that comprises If-Then rules similar to conventional rule induction, and decision tree approaches (Abdelhamid et al., 2014). However, the way rules are discovered, trimmed, or produced is different in conventional rule-based induction approaches. For instance, a rule-based induction algorithm such as repeated Incremental Pruning to Produce Error Reduction (RIPPER) induces the rule and predicts the target class of test data in a multi-step process: rule discovery, rule pruning, and classification (Cohen, 1995). RIPPER uses the divide-and-conquer approach in which once a rule is discovered, the linked training data instances are discarded, and the process is repeated until the input dataset empties. For each rule, the algorithm keeps appending the attribute values that yield to the least error against a

validation subset of data into the current rule until the rule is generated. For each formed rule, the algorithm evaluates whether removing such a rule is possible by measuring its effect on the performance of the classifier in terms of error rate. Once all discovered rules are evaluated, then those that are not pruned form the classifier.

Most class association rule mining methods inherit from association rule mining the concepts of item support and rule confidence and use these during the process of rule discovery (definitions 6 and 8) (Ragab, 2019; Abdelhamid, et al., 2016). The item support for a classification problem in data science represents the frequency of the item and the target class in which that item most occurred in the dataset. When the item support fails to pass the required minimum support (definition 7) set by the user, then the item will be discarded during the rule induction process. The rule's confidence represents the fitness of the rule by capturing the proportion of the joint frequency of the item plus its largest occurring class label, and the frequency of the item by itself in the dataset. When the rule is associated with a confidence value greater than the minimum confidence value (definition 9) then it will be derived and become part of the classification system.

Class association rule methods have advantages over conventional classification methods as they produce simple-to-understand rules which make them highly favourable in applications that require simple-to-interpret classification systems such as dementia screening and diagnosis (Abdelhamid & Thabtah, 2014). Class association rule methods normally produce more hidden knowledge from the dataset resulting in higher predictive classification models than other rule-based classification approaches such as rule induction and decisions trees (Padillo et al., 2019; Liu et al., 1998). Nevertheless, without proper rule pruning the number of rules produced can be large.

Class association rule methods such as Classification-Based Association (CBA), Stock Market Associative Classification (SMAC), Active Pruning Rules (ARP), Hybrid Associative Classification (HAC), Multiclass Associative Classification (MAC) (Liu et al., 1998; Constantino et al., 2021; Ragab, 2019; Hadi et al., 2018; Abdelhamid, et al., 2012) often go through the number of steps below to learn the rules and produce the classification systems:

1) Discovering the rules: In this step, the class association algorithm generates frequent items (sometimes called large items). These are items with greater than the minimum support. The definition of items in the class association rule differs to that of association rule. In the former the attribute and the class values are considered together as an item, whereas in the latter only the attribute value or the item name is considered during the process of counting the frequencies of items in the input dataset. Once all frequent items are found, then the algorithm uses them to generate rules starting with rules of length 1 (rules with number of items in their body equal to 1), then rules of length 2 and so forth until the complete set of rules is extracted.

2) Rule ranking and pruning: Rules may have different confidence and support values, so a ranking step is implemented by most class association rule algorithms in which rules with higher confidence are favoured. The rule-ranking process precedes the rule pruning in which the algorithm tries to reduce any chance of overfitting by removing redundant rules or rules that yield to incorrect classification. There are many types of pruning implemented by class association rule algorithms such as lazy pruning, greedy pruning, and information gain-based pruning, among others (Veloso et al., 2011; Padillo et al., 2019; Cao et al., 2020). Often, during the rule pruning procedure the algorithm checks the applicability of each available rule on the training dataset or a validation dataset and removes any rule that has not classified any data observation or rules that have high number of misclassifications.

3) Classification: In this step, the rules retained after the rule pruning procedure are used to predict the target class labels in a test dataset. Usually, there are two known test data classification methods: one rule, and multi-rule. In the one rule classification method, one rule only is applied to the test case—usually the rule that has similar attributes to the test data and with the highest-ranking position. However, since more than one rule can be used to predict the class label, the one-rule approach has been criticised (Abdelhamid & Thabtah, 2014). For the multi-rule prediction methods, the class association rule algorithm employs more than one rule such as a voting approach in ensemble learning in which the class label belonging to the greatest number of rules is assigned to the test data. However, despite using more than one rule to predict the class these rules do not necessarily match the test data attribute values (Almnnaee et al., 2018).

Unlike classic class association rule mining algorithms that utilise the entire dataset at once to generate the rules so that a single data observation can be used to generate multiple rules, the proposed algorithm (AD-CR) employs an incremental rule learning process that generates a rule, discards the rule's data observation, and then generates the next rule from the updated training dataset. The algorithm ensures that each time a rule is derived the data observation of such rule is not to be used by the next rule during a pruning process therefore reducing the possible number of generated rules. The AD-CR algorithm's learning process deals with an inherited problem from association rule: that of rules overlapping in the training data observations, by disallowing rules to share training data like rule induction methods. However, unlike greedy induction algorithms such as PRISM that requires a rule to have 100% expected accuracy, the AD-CR permits the generation of rules if such rules pass the user's requirement of confidence and, without having to, continuously adds items to the rule's body to increase the rule's accuracy. By permitting the generation of these rules, the AD-CR algorithm minimises any chance of overfitting and allows each rule to have fewer items in its body hence cover a larger number of data observations.

The AD-CR algorithm presents a hybrid prediction method that initially considers the best rule when assigning the class label for a test data. However, when there is no rule matching the test data, the AD-CR algorithm invokes partial matching rules and then uses a voting mechanism for class prediction based on the class that belongs to the majority matching rules. Using this prediction method minimises the employment of the default class rule during the prediction phase and thus may improve the predictive power. The details on how the AD-CR algorithm generates the rules and predicts test data are explained in the next sub-sections.

One potential advantage of the training algorithm of AD-CR is that fewer rules are formed due to the assurance that whenever a potential rule is added to the candidate rule list, all related data examples are discarded similar to rule induction algorithms. The training process also ensures that a training data example is restricted to one rule only, and its associated data examples cannot be considered for generating other rules therefore cutting down the search space of potential rules. This indeed may result in a concise set of classification models that can be easily controlled and used by the clinician unlike conventional class association rule algorithms.

3.5.2 Class Association Terms

The rule-based algorithm that we propose is called Alzheimer's Disease Class Rules (AD-CR), which is based on a classification approach devised from association rule mining called class association rules. In AD-CR, the classification problem's dataset is represented as distinct items along with their corresponding values. Each distinct item is associated with the largest target class label in the dataset with which the item has occurred and such <item, class> representation is called a 'rule_item'. Below are the main definitions related to the AD-CR using *D* as an input classification dataset:

1. Data Observation: A collection of features with their values plus a target class value represented as $([(O_1, v_1), (O_2, v_2), (O_3, v_3) \ldots, (O_k, v_k)], Class_k)$

where O is an attribute or column in a dataset, v is the attribute value, and $Class$ is the target attribute in the dataset.

2. Training Dataset *D*: A combination of data observations each associated with a target class *c*

3. Feature in *D*: An attribute that relates to the individual undergoing the screening process of dementia such as age, gender, visit code, etc. The feature can be categorical (linked with a predefined set of values) or continuous (numeric or decimal). The algorithm assumes that the input

attributes are discretized, so any continuous attribute gets discretized before the learning process initiates.

4. Target Class in *D*: An attribute that represents the progression of AD stage presented in a multi-class categorical form (0,1,-1). We limit the problem to progression (1) or no progression (0).

5. 1-Rule_Item (1-RI_k) in *D*: Is represented as $([(O_k, v_k)], Class)$.

6. Support of the RI_k, i.e. supp (RI_k): Calculated from *D* as $\frac{|([(O_1,v_1),(O_2,v_2),(O_3,v_3)\ldots,(O_k,v_k)],Class_k)])|}{|D|}$. When Supp ($RI_k$) ≥ min_supp_threshold, the *RI* is considered frequent.

7. Minimum Support Threshold: Denoted as min_supp and is employed to differentiate among frequent and infrequent *RIs*.

8.Confidence of the rule_item (RI_k), i.e. Conf (RI_k): Calculated from *D* as $\frac{|([(O_1,v_1),(O_2,v_2),(O_3,v_3)\ldots,(O_k,v_k)],Class_k)])|}{|[(O_1,v_1),(O_2,v_2),(O_3,v_3)\ldots,(O_k,v_k)]|}$. When Conf ($RI_k$) ≥ min_conf_threshold, the *RI* is considered a potential rule.

9. Minimum Confidence Threshold: Denoted as min_conf and employed to measure the strength of a *RI*.

10. Potential Rule: Takes the form (I_1 ^ I_2 ^ … I_k) ⇾ Class.

11. Test Dataset: A combination of data observations each is associated with a true class *c*

3.5.3 Learning Phase

The proposed rules algorithm pseudocode is depicted in Figure 3.4. The inputs of the AD-CR algorithm are the classification dataset, the min_supp (See 6 and 8 above). The AD-CR algorithm deals with both categorical and continuous attribute values in the training dataset; any missing attributes values are treated as any other values by the AD-CR algorithm. The min_supp threshold is employed to determine rule_items that have sufficient frequency in the training data and mainly used to pinpoint the best frequent rule_item in any iteration. In doing so, only the best rule_item in terms of frequency is chosen each iteration by the AD-CR algorithm to start building a new rule or to append into the current rule's body. The algorithm keeps merging item(s)/attribute values into the rule until the current rule passes the min_conf threshold; when this happens, the current rule's is added into the candidate rules list. The confidence of the rule is calculated according to definition 10 and can be considered a key performance indicator that reflects the rule's position during the learning phase.

In discovering the rules (first phase), the AD-CR algorithm iterates over the training data instances to discover the best one rule_item (1-rule_item) in terms of support. It should be noted that only the

The Algorithm
Input: A classification dataset CD, the min-supp and min-conf thresholds
Output: CR: If-Then rules

1. For each rule_item (RI), i.e. $([(RI, v_k)], Class))$, in CD do
2. Best RI (BRI) ← Max (supp (RI))
3. if conf (BRI) <= min_conf
4. exit
5. elseif conf (BRI) >= min-conf
6. DS ← $Data\ Examples' of\ BRI$
7. CR ← BRI
8. CD ←CD – DS
9. end //elseif
10. else
11. begin
12. DS ← $Data\ Examples' of\ BRI$
13. CRF = Candidate_Rule (DS, BRI)
14. UDS ← Updated (DS, CRF)
15. CR ← CRF
16. CD ←CD – UDS
17. end // else
18. end for
19. Repeat 1-16
20. Exit when CD' is empty / checked
21. Produce the CR list
22. end
23. Order CDs by the confidence and support values

Figure 3.4: The AD-CR Algorithm

largest class label in the training data occurring with the attribute value of the rule_item is considered when counting the rule_item's support. Once the 1-ruleitem is identified, the AD-CR algorithm starts building the first potential rule as Best_Item ⇾ C. The algorithm evaluates the current rule's confidence, if the current rule has a confidence larger than the min-conf threshold then it will be added into the candidate rule list, and all data examples associated with it are removed from the training dataset. However, if the current rule's confidence is less than the min-conf threshold, then the learning algorithm isolates its training examples into a data structure: DS.

The learning algorithm then checks whether adding the best frequent item of DS into the current rule will improve its confidence value. If this check yields true, then the algorithm appends the frequent item found into the current rule's body and repeats the same process of potentially adding frequent items into the rule's body from DS, until the current rule's confidence passes the min-conf threshold. When this occurs, the rule will be generated and added into the candidate rule list. More importantly, all the rule's data examples will be removed from the original training dataset whenever the rule is generated ensuring that these examples are only used once during the training phase. The algorithm then starts creating the second potential rule from the updated training dataset after removing the first rule's data examples and repeats the same process described above until either

the training dataset empties, or no more potential rules can be discovered. Once all candidate rules are generated, they are sorted based on confidence and support values.

One potential advantage of the training algorithm of AD-CR is that fewer rules are formed due to the assurance that whenever a potential rule is added into the candidate rule list all related data examples are discarded. The training process ensures also that a training data example is restricted to one rule only, and its associated data examples cannot be considered for generating other rules therefore cutting down the search space of potential rules. This indeed may result in a concise set of classification models that can be easily controlled and used by the clinician.

Another possible advantage of the training algorithm is that it does not seek rules with 100% accuracy as found in classic Rule Induction and Covering classification algorithms such as PRISM, or for specific rules as found in enhanced rules algorithms. AD-CR permits the rules to be produced even when the rule's accuracy is not perfect at good confidence level thus reducing the chance of any overfitted predictive models. It should be noted that models that are generated by the proposed algorithm will be used as a knowledge base and for predicting the class of test data.

3.5.4 Classification Phase

The AD-CR algorithm proposes a simple yet influential classification method which uses the most suitable rule to assign an appropriate class label to the test data during the classification phase. A rule used to assign the class normally meets two conditions:

1) It has the best ranking among all other rules in terms of confidence and support values
2) The attribute values in its body are all contained within the test data thus ensuring attributes' values similarity.

During the classification phase, when a test data example is to be classified, the AD-CR seeks in the final set the rule that fully matches the test data's attribute values, allocates its class to the test data example, and then moves to the next test data example and so forth. However, if there is no rule in the final rules set that fully matches the test data example, then the AD-CR algorithm searches for a partially matching rule; such rules have at least one attribute data value similar to the test data example. The proposed algorithm then allocates the class linked with more partially matching rules to the test data example. In cases when no rules partially or fully match the test data example, the algorithm uses the default rule one which denotes the class label with most of the training data examples.

Using just one rule for classifying test data examples is not only a simple approach but it also provides good predictive power as seen later in Chapter 5 in the experimental analysis of the classification phase. In addition, the approach considers other rules that partially match the test data examples when no fully matching single rule is available rather than invoking the default class label. This reduces the number of arbitrary classifications and thus misclassifications.

We summarise the primary characteristics of the AD-CR algorithm:

1) Only rules with significant frequency and confidence are formed.
2) Fewer rules are formed thus smaller models are produced
3) Unlike association rule mining, no rules share data examples thus reducing the search space of items and potential rules
4) Rules can be associated with some degree of error to minimise overfitting
5) Simple and effective classification method is used in the prediction phase

3.6 Mobile Application Design and Content

A mobile application was designed to capture the necessary items related to IADLs for individuals undergoing the screening process. It is based on items collected from the FAQ and implemented according to the design requirements for dementia testing. We did not design mobile interface screens for the activities related to cognition since these activities require instructions, and detailed sub-activities that cannot be implemented on a mobile based platform so only these values are recorded by the user. Therefore, all detailed activities related to ADAS-Cog13 are recorded by the clinicians without a mobile medium.

Unlike screening mobile applications, the interface of the mobile app of MLA-ADP architecture adheres to most of the design requirements needed for dementia-related screening. Therefore, careful attention to the design of interfaces and functionalities was followed including colour, visual layout, item size, buttons, output format, text size and font, and user navigation, among others. The MLA-ADP app incorporates functionalities related to invoking the models for predicting the individual's possible progression. Table 3.4 displays the FAQ items in the app, their scores, and their mapping to the DSM-5 cognitive domains.

Table 3.4: FAQ Items and their Associated DSM-5 Cognitive Domains

Question No.	Column in Dataset	Tasks	Scoring	DSM-5 Cognitive Domains
1	FAQFINAN	Writing cheques, paying bills, balancing cheque book	0–3	Complex attention Executive functioning Learning and memory
2	FAQFORM	Assembling tax records, business affairs, or other papers	0–3	Complex attention Executive functioning Learning and memory
3	FAQSHOP	Shopping alone for clothes, household necessities, or groceries	0–3	Complex attention Executive functioning Learning and memory Perceptual motor function
4	FAQGAME	Playing game of skill, working on a hobby	0–3	Complex attention Executive functioning Learning and memory Perceptual motor function
5	FAQBEVG	Heating water, making a cup of coffee, turning off stove after use	0–3	Complex attention Executive functioning Learning and memory Perceptual motor function
6	FAQMEAL	Preparing a balanced meal	0–3	Executive functioning Perceptual motor function
7	FAQEVENT	Keeping track of current events	0–3	Learning and memory
8	FAQTV	Paying attention to, understanding, and discussing TV, book, magazine	0–3	Complex attention Executive functioning Learning and memory Social cognition Language
9	FAQREM	Remembering appointments, family occasions, holidays, medications	0–3	Learning and memory
10	FAQTRAVL	Traveling out of neighbourhood, driving, arranging to take public transportation	0–3	Complex attention Executive functioning Perceptual motor function
		TOTAL	0–30 points	

The navigation diagram of the proposed mobile application is shown in Figure 3.5. After launching the app the clinician can access the information sheet from the landing screen; this gives clear consent on the app's use. The 'terms of use' screen clearly states that the application is for research purposes only and any information or data collected during the pre-diagnosis process will be anonymous, and not shared with any party. The screen is available to the clinician before completing the pre-diagnosis in which he/she is required to agree to a disclaimer which elaborates on the aim of the app, use of the data, and privacy policy. We inform the user that data are collected

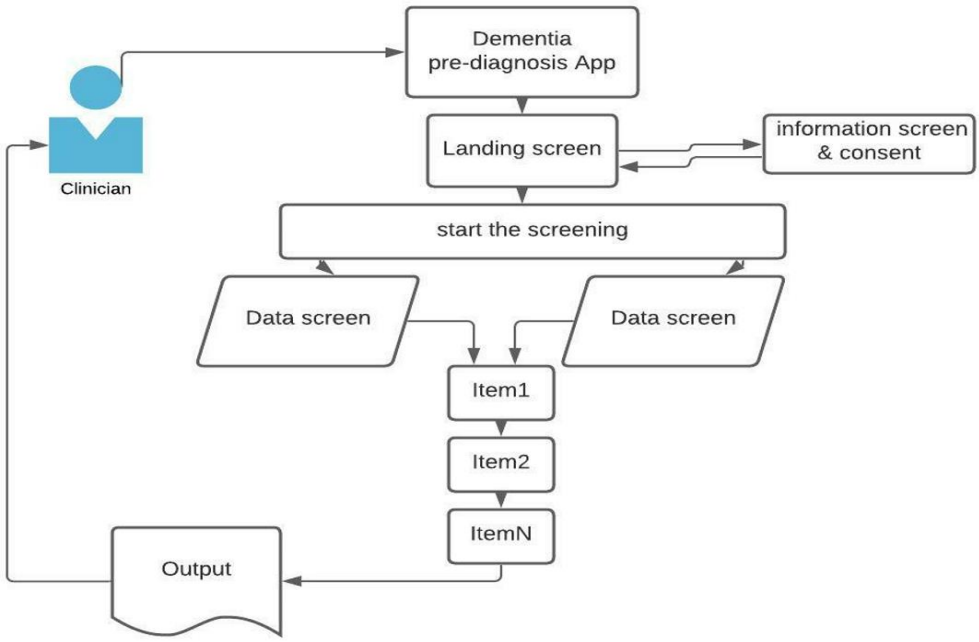

Figure 3.5: The Proposed Dementia Pre-diagnosis Application Navigation

anonymously and not shared—participants' identities are anonymous. In addition, no information related to the device is used through the process. From the landing screen (Figure 3.6a), two data screens (Figures 3.6b and 3.6c) can be accessed to collect data related to gender, age, ethnicity, marital status, and education level—similar to the demographic attributes that we used in the ADNI-Merge while building the classification model. Once these data screens are filled by the clinician, then the items of the FAQ method will appear on a separate screen in a sequential manner. Samples of two items are displayed in Figures 3.6d and 3.6e, respectively. For each item, there will be six possible responses that the clinician can choose each of which is associated with a score as follows: dependent = 3, requires assistance = 2, has difficulty but does by self = 1, normal = 0, never did the activity but could do now = 0, never did and would have difficulty now = 1.

The clinician can navigate through the app using 'Next' and 'Back' buttons. Once he/she completes the pre-diagnosis process then an output screen with either 'No sign of dementia progression' or

‘dementia progression’ will appear as shown in Figure 3.6f. The decision is based on models learned from a historical cases and controls that are stored within a database. The clinician also can have access to the knowledge base to access the rule produced by the AD-CR algorithm. The proposed app can be used in Android mobile applications or tablets.

Once the clinician clicks ‘submit’, a process is triggered at the backend architecture that will pass the items’ answers, and the demographics information to the classification system as test data. The classification system then uses the predictive models to guess whether there will be progression based on the test data characteristics and using rules. Based on the model decision, then the label will be replaced with the appropriate text that will appear in the final screen for the user.

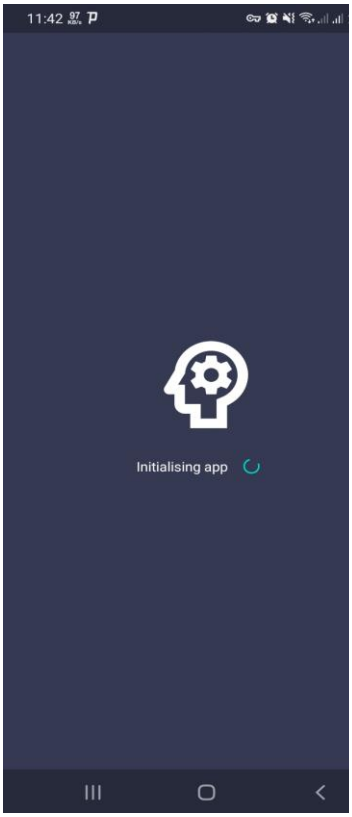

Figure 3.6a: Landing Screen

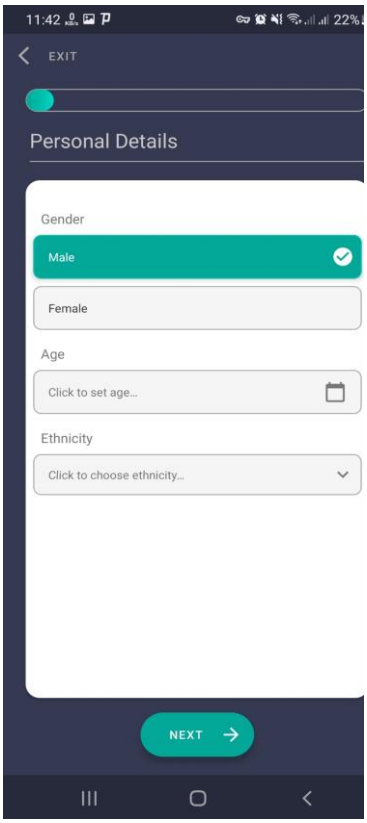

Figure 3.6b: Information Screen 1

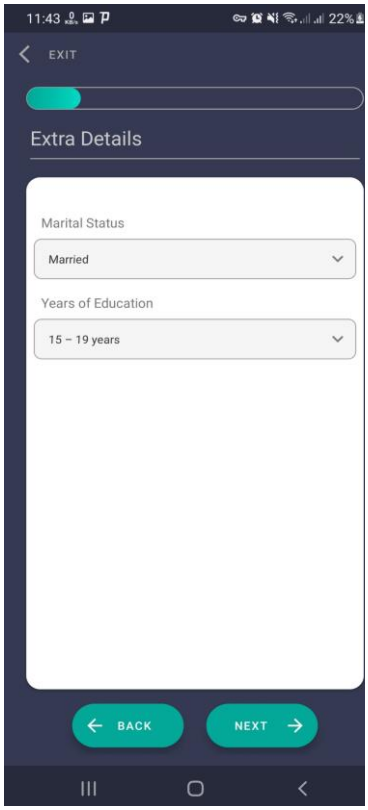

Figure 3.6c: Information Screen 2

For example, when the decision is ‘0’ a text stating that there is no dementia progression will appear in the results screen.

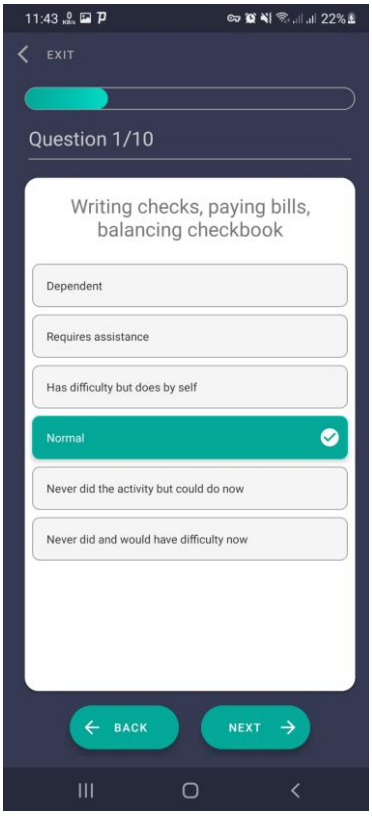

Figure 3.6d: Sample Question A

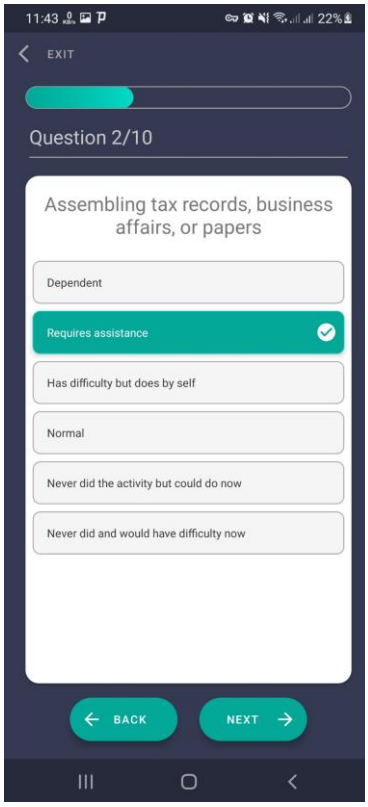

Figure 3.6e: Sample Question B

Figure 3.6f: Output Screen

3.7 Chapter Summary

One potential way to enhance the detection of AD progression and to empower clinicians is to use a data driven approach based on machine learning which is able to discover hidden patterns during screening. In this chapter, a new intelligent architecture called MLA-ADP has been proposed and implemented on a cloud-based environment—a mobile interface is provided for accessibility which is advantageous particularly during the difficult time of the Covid-19 pandemic. The proposed architecture comprises several major phases including data preparation and modelling, feature assessment, classification, and mobile application.

During the data collection, a number of cognitive and functional datasets were collected from the ADNI data repository then processed and modelled to create a new target class. In feature assessment, a number of mathematical-based feature selection methods were used to analyse the correlations among the features themselves and features with the target class label. In the classification phase, we used rules to predict the progression of AD using subsets of the features within the firebase in an automated manner; the clinician uses an easy-to-access-and-use mobile interface. In the next chapter, we perform in-depth experiments to build classification models for

AD progression and compare the models derived by the proposed classification system with the MLA-ADP with other models generated by common machine learning algorithms. These results will reveal the true performance of the MLA-ADP.

Chapter Four

Data, Methods Used, and Experimental Settings

This chapter introduces the unprocessed data, its meta data, the processed data, and detailed descriptive analytics of the data considered. In addition, the experimental platform and the methods used like computational intelligence and classification, are discussed. Most of this chapter's content is being considered for publication in the Journal of Biomedical Informatics, and Intelligent Decision Technologies.

4.1 Introduction

This chapter discusses the data, the classification methods, and platforms used in the experiments, and the experimental settings. The AD-CR algorithm was coded using high-level programming language (Java), and then integrated into Waikato Environment for Knowledge Analysis (WEKA) machine learning platforms prior to classification algorithms experiments being conducted to establish a fair comparison. The proposed AD-CR algorithm was embedded into the WEKA-Classifier package under folder 'Rules'—the algorithm generates rule-based classifiers in an 'If-Then' format. All experiments were conducted using ten-fold cross validation to ensure that any classifier derived was evaluated thoroughly. Additionally, we employed various evaluation measures (discussed later) including predictive accuracy, specificity, and sensitivity (see Equations 4.1–4.3 in Section 4.3.1).

In this chapter, we discuss the datasets used for the analysis and provide an in-depth descriptive analysis. The modelling process implementation is then highlighted and the revised datasets given the new class label 'DX Progress'. Furthermore, we show how special group datasets related to cognitive and functional features are derived besides their purposes and relevant statistics. For example, we show cognitive and functional datasets for data subjects with a baseline diagnosis 'CN' and then at 36 months to indicate whether they remain at 'CN' or have progressed to 'MCI'. We also show cognitive and functional datasets for other groups with a baseline diagnosis of 'MCI', and who for 36 months remained with 'MCI' or advanced to 'AD'.

Lastly, we describe the classification algorithms used in the experiments.

4.2 Datasets

4.2.1 Unprocessed Datasets

The unprocessed datasets used in this thesis were obtained from the ADNI project data repository. According to ADNI (2021), the primary aims of the ADNI project are:

- To find effective methods to track the progression of AD
- To identify potential cases early by developing affordable diagnostic methods using innovative technologies
- To manage and support the intervention of AD
- To support scholars and researchers worldwide by providing access to real subjects' dementia-related data to promote solutions for the intervention and treatment of AD.

The ADNI project is funded until 2022 (ADNI, 2021). Most of the participants have been recruited from various healthcare facilities in the USA and Canada, and represent a population with CN, MCI, and AD. They have completed several neuropsychological assessments and cognitive tests, besides clinical and pathological procedures when needed. There are several datasets in the ADNI data repository covering features related to genetics, neuroimaging, biomarkers, and neuropsychological and clinical methods, among others, as discussed earlier in Chapter 3. For example, for the neuropsychological section of the ADNI, there are cognitive assessments including, but not limited to, MoCA, Everyday Cognition study partner (ECogPT), Everyday Cognition patient reported (ECogSP), ADAS-Cog different versions, MMSE, CDR-SB, RAVLT different versions, etc. For neuroimaging, there are several brain-related features that can be captured through MRI and CAT images.

Since the scope of this thesis is limited to building AD classification systems from cognitive and functional items, we focus on three related major datasets:

- ADNI-Merge (a collection of multiple datasets: ADNI1, ADNI2, ADNI3, ADNI-GO)
- ADAS-Cog-sheet
- FAQ-sheet.

In Chapter 3, Table 3.2 displays the basic statistics related to the ADNI-Merge, ADAS-Cog-sheet, and FAQ-sheet datasets. The ADNI-Merge dataset is an amalgamation of data from the ADNI-1, ADNI-Go, ADNI-2, and ADNI-3 studies. The merged dataset contained 14,627 data observations and attributes related to patients' visits, cognitive tests, memory tests, functional questionnaires, genetics, demographics, and biomarkers, among others. There is one target class in the ADNI-Merge dataset, which is the diagnosis of the last examination visit (DX); another attribute that can be

considered important is the baseline diagnosis (DX_bl) which denotes the initial diagnosis given to the patient at the first visit. There are 10 and 4243 data subjects, respectively, with missing values for the DX_bl and DX attributes in the ADNI-Merge dataset. Table 4.1 gives data samples attributes of the ADNI-Merge dataset for four participants.

Table 4.1: Sample of Three Data Subjects of ADNI-Merge with Diagnostic Class (DX)

RID	VISCODE	SITE	COLPROT	ORIGPROT	EXAMDATE	DX_bl	AGE	PTMARRY	APOE4	ABETA	TAU	PTAU	CDRSB	ADAS11	ADAS13	ADASQ4	MMSE	MOCA	FAQ	..	DX
2	bl	11	ADNI1	ADNI1	8/9/2005	CN	74.3	Married	0				0	10.67	18.67	5	28		0	..	CN
3	bl	11	ADNI1	ADNI1	12/9/2005	AD	81.3	Married	1	741.5	239.7	22.83	4.5	22	31	8	20		10	..	Dementia
3	m06	11	ADNI1	ADNI1	13/03/2006	AD	81.3	Married	1				6	19	30	10	24		12	..	Dementia
3	m12	11	ADNI1	ADNI1	12/9/2006	AD	81.3	Married	1	601.4	251.7	24.18	3.5	24	35	10	17		17	..	Dementia
3	m24	11	ADNI1	ADNI1	12/9/2007	AD	81.3	Married	1				8	25.67	37.67	10	19		14	..	Dementia
4	bl	22	ADNI1	ADNI1	8/11/2005	LMCI	67.5	Married	0	1501	153.1	13.29	1	14.33	21.33	6	27		0	..	MCI
4	m06	22	ADNI1	ADNI1	2/5/2006	LMCI	67.5	Married	0				0.5	17.33	25.33	7	28		0	..	MCI
4	m12	22	ADNI1	ADNI1	14/11/2006	LMCI	67.5	Married	0	1176	159.7	13.3	1	15	22	7	26		0	..	MCI
4	m18	22	ADNI1	ADNI1	14/05/2007	LMCI	67.5	Married	0				1	20.33	28.33	7	27		1	..	MCI
4	m36	22	ADNI1	ADNI1	18/11/2008	LMCI	67.5	Married	0				1	18	25	7	25		0	..	MCI
5	bl	11	ADNI1	ADNI1	7/9/2005	CN	73.7	Married	0	547.3	337	33.43	0	8.67	14.67	4	29		0	..	CN
5	m06	11	ADNI1	ADNI1	9/3/2006	CN	73.7	Married	0				0	11	15	3	29		0	..	CN
5	m12	11	ADNI1	ADNI1	5/9/2006	CN	73.7	Married	0	472.8	334.1	34.04	1.5	5.67		3	30		0	..	CN
5	m24	11	ADNI1	ADNI1	7/9/2007	CN	73.7	Married	0				0	7	11	3	29		0	..	CN
5	m36	11	ADNI1	ADNI1	10/9/2008	CN	73.7	Married	0				1	6.67	11.67	4	30		0	..	CN

The age of the participants ranges from 54.4–94.4, and the average age is 73. Figure 4.1 depicts the age distribution of all participants—most are between 70 and 80 years of age. Figure 4.2 shows the frequency of the participants' medical visits with at least one per participant and up to 22 visits. To be exact, most of the participants had 2 medical visits followed by 1, 5, and 7 medical visits, respectively.

Initial descriptive analytics show that there are more males than females in the ADNI-Merge dataset with 1051 females and 1186 males. Figure 4.3 displays the number of data subjects in the ADNI-Merge dataset with respect to the baseline diagnosis (DX_bl) and the diagnosis after the final examination visit (DX). Based on the figures, it seems that at baseline diagnosis, most of the

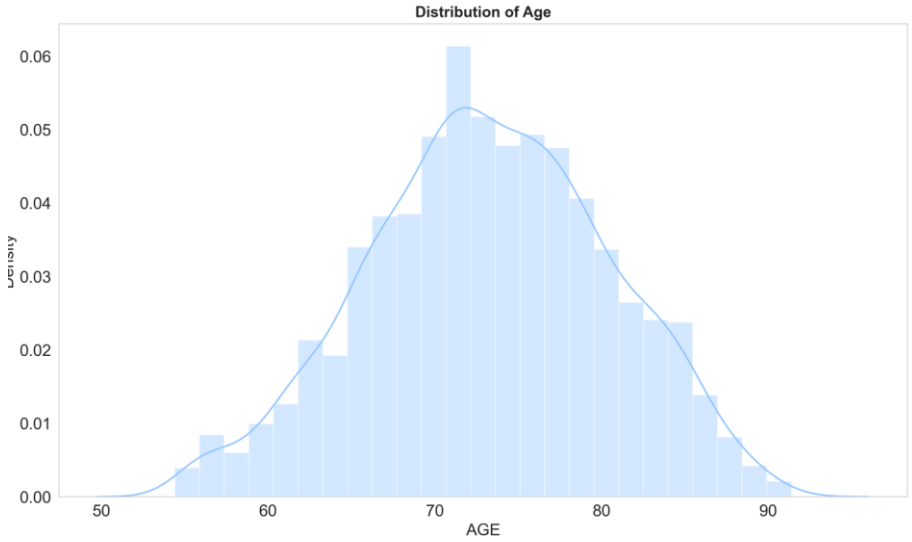

Figure 4.1: Age Distribution of the Participants

participants were diagnosed as MCI and mainly with Late MCI (LMCI), and the least number of participants were diagnosed as AD. To be exact, 1331 participants were diagnosed with different

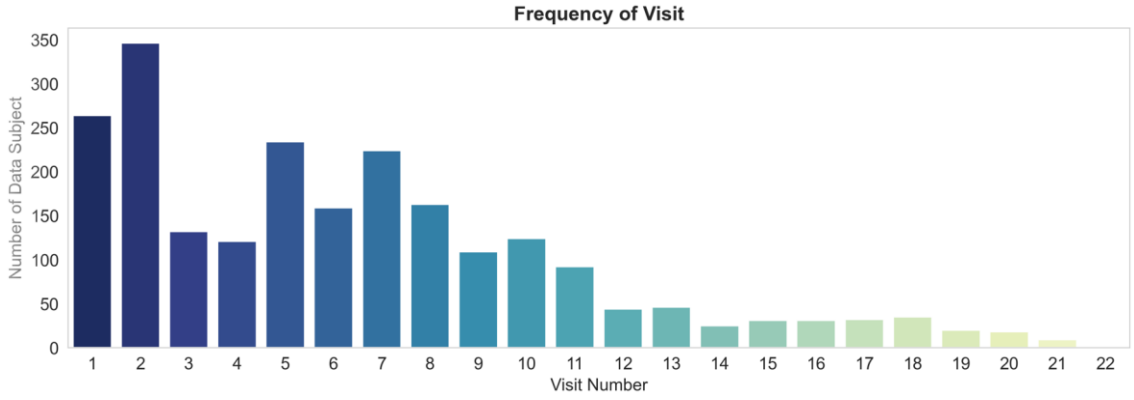

Figure 4.2: Frequency of the Patients' Visits

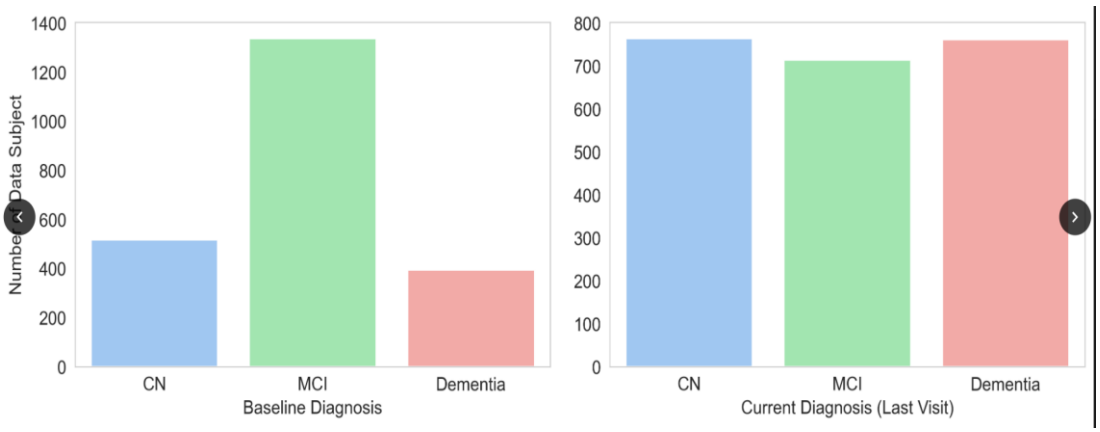

Figure 4.3: Class Labels' Distribution at Baseline and at Last Examination Visit

subcategories of MCI at baseline compared to 515 and 391 diagnosed with CN and AD, respectively. However, and for the final diagnosis (DX), the number of patients classified as CN or AD is larger than participants with MCI. To be exact, there are 763, 761, and 713 participants with a DX of CN, AD, and MCI, respectively, indicating that the data is somewhat balanced in terms of class distribution.

We further investigated the baseline diagnosis and the final diagnosis attributes by taking 'gender' into account as shown in Figure 4.4. Based on the figure, there are more males associated with MCI and AD than female for both baseline and final visit attributes, respectively. However, there are more female than male participants in the CN class category for both baseline and final visit attributes. We further investigated the 'age' attribute with respect to baseline and final diagnosis scenarios as depicted in Figure 4.5. Most of the participants diagnosed at baseline or subsequently with AD tend to be older than the participants belonging to the other class categories in both scenarios. Specifically, most participants who received an AD diagnosis at baseline or at their last medical examination were aged between 70 and 80 years, whereas participants diagnosed as CN at the final examination visit were normally younger than 80 years old.

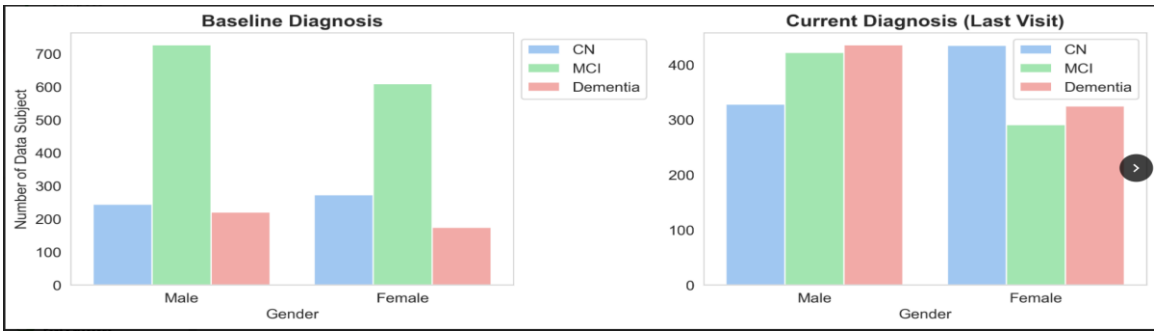

Figure 4.4: Gender Distribution at Baseline Diagnosis and at Last Examination Visit Diagnosis

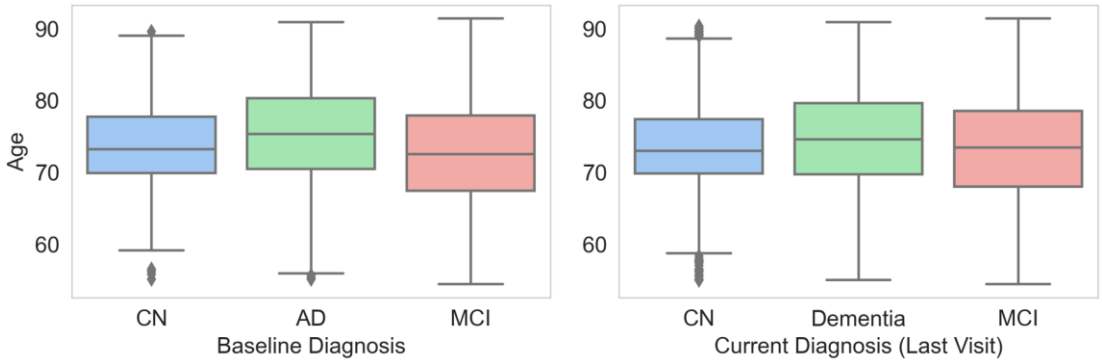

Figure 4.5: Age Distribution at Baseline Diagnosis and at Last Examination Visit Diagnosis

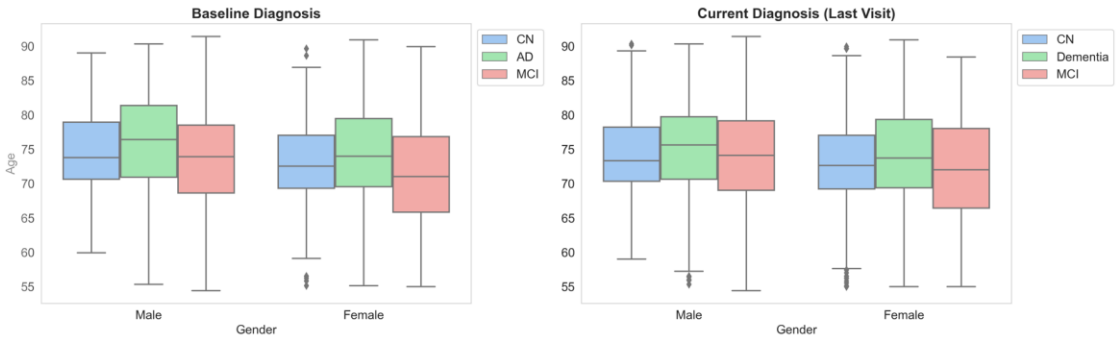

Figure 4.6: Gender and Age Distribution at Baseline Diagnosis and at Last Examination Visit Diagnosis Respectively

In addition, many of the participants diagnosed with early MCI (EMCI) or Significant Memory Concern (SMC) were in their late 60s to early 70s. We added the 'gender' with 'age' to visualize the class distributions at baseline diagnosis, and at the last medical examination diagnosis as shown in Figure 4.6. The figure reveals a balanced number of participants in terms of gender aged 70–80, and with an AD diagnosis at their last medical examination. More male than female participants with dementia (AD) diagnosis. For baseline diagnosis, there was more male at MCI (EMCI and LMCI respectively) than female.

The number of data subjects who progressed from CN to MCI or MCI to AD are depicted in Figure 4.7a. The figure reveals that there were more data subjects with baseline diagnosis MCI who progressed to AD than those with baseline CN who progressed to MCI. We further investigated these data subjects as shown in Figure 4.7b and discovered that in both progression stages there were more male subjects than female. In addition, Figure 4.8 depicts that most of the disease advancements occur between the ages of 70–79 in both the CN to MCI and MCI to AD stages. There were 39 participants aged between 70–79 who progressed from CN to MCI, while 174 MCI participants moved to AD stages. These figures represent 61.90% and 52.72%, respectively, of the total progression cases of all age categories and for both stages (CN to MCI and MCI to AD). There

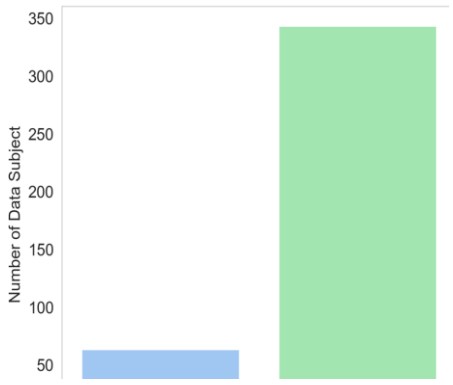

Figure 4.7a: Number of Data Subjects Progressed from CN to MCI or MCI to Dementia

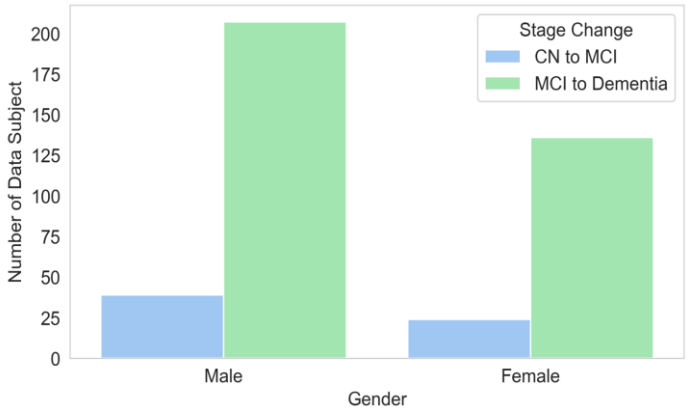

Figure 4.7b: Gender Distribution of the Data Subjects Progressed from CN to MCI or MCI to Dementia

was a notable result for patients younger than 60 years of age— 13 participants with a baseline diagnosis MCI progressed to a final diagnosis of AD.

The ADAS-Cog sheet dataset is based on the ADAS-Cog13 subscale used in the ADNI study, detailing the patient's score in each task and the total scores attained during the assessment. It contained 6,770 observations and 121 attributes. Examples of the attributes in the ADAS-Cog sheet dataset are patient ID (RID), time of visit (VISCODE2), Examination date, and the 13 main tasks

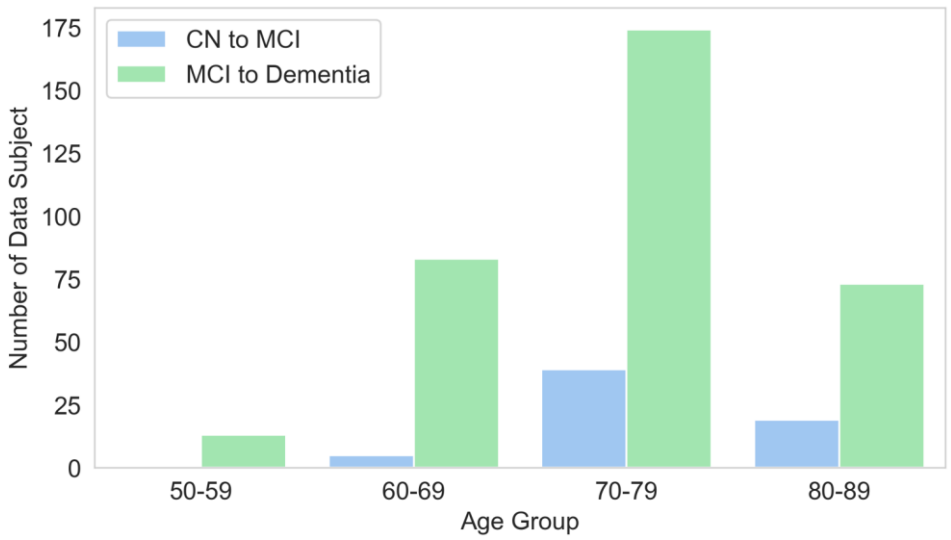

Figure 4.8: Age Distribution of the Data Subjects Progressed from CN to MCI or MCI to Dementia

along with their i.e. Q1SCORE, Q2SCORE, Q3SCORE, Q4SCORE, Q5SCORE, Q6SCORE, Q7SCORE, Q8SCORE, Q9SCORE, Q10SCORE, Q11SCORE, Q12SCORE, and Q13SCORE besides the total score of the test. Table 4.2 describes a sample of ADAS-Cog13 items for only 16 attributes.

The FAQ dataset contained 10,905 observations and 23 attributes with data related to the patient's information, individual FAQ item score, and total FAQ score attained during the assessment. Examples of the FAQ-sheet dataset attributes are the patient ID (RID), time of visit (VISCODE2), examination date, and the FAQ-items along with their answers of the ten items, i.e. FAQFINAN, FAQFORM, FAQSHOP, FAQGAME, FAQBEVG, FAQMEAL, FAQEVENT, FAQTV, FAQREM and FAQTRAVL. Table 4.3 shows a sample of two participants with 19 FAQ items.

Table 4.2: Sample of 3 Data Subjects from the ADAS-Cog-Sheet Dataset with 16 Attributes

Phase	ID	RID	SITEID	VISCODE	VISCODE2	USERDATE	USERDATE2	WORDLIST	Q1UNABLE	Q1TR1	Q1TR2	Q1TR3	Q1TRIT	Q1TR2T	Q1TRT
ADNIGO	108	2	8	m60	m60	9/24/2010	2/28/2011	1		0:1:2:8:9	2:4:6:8:9	0:2:5:7:8:9			
ADNI2	386	2	8	v06	m72	9/20/2011		1		0:1:6:8	0:2:4:6:7:8:9	2:3:4:5:6:7:8			
ADNI2	2808	2	8	v11	m84	9/28/2012		1		0:1:2:7:8:9	2:4:5:6:8:9	0:2:5:7:8:9			
ADNI2	5254	2	8	v21	m96	9/12/2013	9/12/2013	1		0:1:7:8:9	2:4:5:6:8:9	1:2:5:6:7:8:9			
ADNI2	8236	2	8	v41	m120	10/6/2015	7/31/2017	1		0:1:6:8	0:1:2:4:5:6:7:9	1:2:3:4:6:7:8:9			
ADNI2	9154	2	8	v51	m132	9/29/2016	2/1/2019	1		0:1:6:8:9	0:2:4:6:8	0:1:2:4:5:6:7:8:9			
ADNI3	16259	2	8	init	m144	10/13/2017	10/13/2017		0	0|6|7|9	2|4|6|7|8|9	0|2|5|7|8	4	6	6
ADNIGO	130	8	8	m60	m60	10/4/2010	2/28/2011	1		0:1:3:6:7:8:9	0:1:2:4:6:7:8:9	0:1:2:3:4:5:6:7:8:9			
ADNI2	422	8	8	v06	m72	9/29/2011	10/24/2011	1		3:5:6:7:8:9	0:1:2:4:6:8:9	0:1:2:3:4:5:7:8:9			
ADNI2	2802	8	8	v11	m84	9/27/2012		1		0:3:5:6:7:8:9	0:1:2:4:6:7:8:9	0:1:2:3:4:5:6:7:8:9			
ADNI2	5324	8	8	v21	m96	9/23/2013		1		0:1:5:7:9	2:4:5:6:7:8:9	0:1:2:3:4:5:7:8:9			
ADNI2	8330	8	8	v41	m120	11/6/2015		1		0:3:4:6:9	0:1:2:3:5:6:7:8:9	0:1:2:3:4:5:7:8:9			
ADNI2	598	15	40	v06	m72	11/3/2011		1		0:1:4:5:6	0:2:3:4:5:6:7:8:9	1:2:3:4:5:6:7:8:9			

Table 4.3: Sample of 3 Data Subjects from the FAQ-Sheet Dataset with 19 Attributes

Phase	ID	RID	SITEID	VISCODE	VISCODE2	USERDATE	USERDATE2	EXAMDATE	FAQSOURCE	FAQFINAN	FAQFORM	FAQSHOP	FAQGAME	FAQBEVG	FAQMEAL	FAQEVENT	FAQTV	FAQREM
ADNI1	4	2	107	bl	bl	9/8/2005		9/8/2005	1	0	0	1	0	0	1	0	0	0
ADNI1	214	2	107	m06	m06	3/7/2006		3/6/2006	2	0	0	0	0	0	0	0	0	0
ADNI1	5812	2	107	m36	m36	8/27/2008		8/27/2008	2	0	0	0	0	0	0	0	0	0
ADNIGO	86	2	8	m60	m60	9/22/2010			2	0	0	0	1	0	0	0	0	0
ADNI2	2830	2	8	v11	m84	10/1/2012			2	0	0	0	1	0	0	0	0	0
ADNI2	360	2	8	v06	m72	9/20/2011	10/24/2011		2	0	0	0	0	0	1	0	0	0
ADNI2	5276	2	8	v21	m96	9/11/2013	9/12/2013		1	0	0	0	1	0	0	0	0	0
ADNI2	7520	2	8	v31	m108	12/31/2014			1	0	0	0	1	0	0	0	0	0
ADNI2	8422	2	8	v41	m120	10/6/2015			2	0	0	0	1	0	1	0	0	0
ADNI2	9356	2	8	v51	m132	10/3/2016			2	0	0	0	0	0	1	0	0	0
ADNI3	18298	2	8	init	m144	10/26/2017	10/26/2017		2	0	3	3	1	0	1	0	3	0
ADNI1	8	3	107	bl	bl	9/13/2005		9/12/2005	1	2	3	2	3	1	1	4	3	3
ADNI1	234	3	107	m06	m06	3/13/2006		3/13/2006	1	4	4	2	4	0	0	3	3	3
ADNI1	1042	3	107	m12	m12	9/12/2006		9/12/2006	1	5	5	4	4	0	0	3	3	4
ADNI1	3818	3	107	m24	m24	9/12/2007		9/12/2007	1	3	5	5	3	0	0	3	3	3

4.2.2 Processed Datasets

To capture the progression of the disease in our datasets, we performed data modelling on the ADNI-Merge dataset by creating two new attributes called 'DX Digit' to label the current diagnostic class (DX) as a numeric, and 'DX Progress' to track the change in diagnosis also using numerical labels. The process of how the 'DX Progress' attribute (change of the diagnosis) was allocated values has been described earlier in Chapter 3 (Section 3.3.4). Table 4.4 illustrates the number of data examples for each class representation we derived after data modelling. Since the scope of our research only focuses on whether there is progression (1) or no change (0), we filter out regression (-1) as these cases are out of the scope. We end up with four general datasets as follows:

1) Processed ADNI-Merge
2) Processed ADNI-Merge-ADAS-Cog
3) Processed ADNI-Merge-FAQ
4) Processed ADNI-Merge-ADAS-Cog-FAQ (combined 3 datasets).

The main reason for creating the 'ADNI-Merge-ADAS-Cog sheet and 'ADNI-Merge-FAQ-sheet' datasets is two-fold:

- To assess cognitive and functional items separately
- Each of these datasets contains more data subjects than the 'ADNI-Merge-ADAS-Cog-FAQ', which indeed benefits the learning algorithm during building of the classification models.

Table 4.4: General Statistics after Data Pre-processing and Data Balancing

Dataset Name	# of Patients before Sampling	# of Data Observations (visits)	DX Progress - Class Distribution before Data Balancing	DX Progress - Class Distribution after Data Balancing
ADNI-Merge-ADAS-Cog dataset	1,710	6,330	Total observations: 6,330 '0': 6,020 (majority 95%) '1': 310 (5%)	Total observations: 11,943 '0': 6,020 (50.40%) '1': 5,923 (49.60%)
ADNI-Merge-FAQ dataset	2,244	10,265	Total observations: 10,265 '0': 9713 (majority 95%) '1': 552 (5%)	Total observations: 18,545 '0': 9,713 (52%) '1': 8,832 (48%)
ADNI-Merge-ADAS-Cog-FAQ dataset	1,710	6,330	Total observations: 6,330 '0': 6,020 (majority 95.10%) '1': 310 (48.90%)	Total observations: 11,724 '0': 6,020 (51.43%) '1': 5,704 (48.57%)

The 'ADNI-Merge-ADAS-Cog-FAQ' is an integrated dataset that captures cognitive, functional, demographic features, among others. Table 4.5 shows five participants from this integrated dataset with the two new modelled attributes (4^{th} and 5^{th} columns) and 16 attributes. It should be noted that the target class (DX Progression) is the 5^{th} column in Table 4.5 in which '0' denotes no progression

and ‘1’ denotes progression. Based on the DX class statistics presented in Table 4.5, there is a clear imbalance in class set with an overwhelming majority of 95% in the no change (0) class compared to 5% in the progression (1) class. Progressing with an imbalanced dataset would produce a skewed and biased result analysis, so to counter this issue, we implemented a sampling method based on SMOTE to randomly create additional minority class instances as we discuss in the next sub-section. The results of data balancing are presented in the final right-hand column of Table 4.5 earlier with a higher number of minority instances inserted to bring it closer to a 50-50 class ratio. There was 111 cases of regression within the ‘ADNI-Merge-ADAS-Cog-FAQ‘ dataset, which we did not consider in the initial analysis.

To obtain in-depth insight during data analysis, particularly feature assessment, and to differentiate between dementia stages we created four additional subsets of data from the integration of the ADNI-Merge, cognitive, and functional datasets as follows:

1) ‘ADNI-Merge-ADAS-Cog CN to MCI for up to 3 years’
2) ‘ADNI-Merge-ADAS-Cog MCI to AD for up to 3 years’
3) ‘ADNI-Merge-FAQ CN to MCI for up to 3 years’
4) ‘ADNI-Merge-FAQ MCI to AD for up to 3 years’.

These datasets represent multiple medical visits within 3 years from the baseline diagnosis for groups of participants who had FAQ or ADAS-Cog medical assessments to monitor the advancement of their disease—any participant who in the baseline visit had not undergone FAQ or ADAS-Cog has been excluded from the 36 months datasets shown in Table 4.6. We specifically would like to determine if dementia-related features vary during the progression from one dementia group to another, and for a specific period of time, i.e. 36 months from the baseline diagnostic visit. Table 4.6 depicts statistical information about these datasets.

Table 4.5: Sample of 5 Data Subjects from the Processed and Integrated Dataset with 16 Attributes

RID	VISCODE2	DX	DX_digit	DX_Progress	AGE	PTGENDER	PTEDUCAT	PTETHCAT	PTRACCAT	PTMARRY	Q1SCORE	Q2SCORE	Q3SCORE	Q4SCORE	Q5SCORE	Q6SCORE	Q7SCORE	Q8SCORE	Q9SCORE
8	m60	CN	1	0	84.5	Female	18	Not Hisp/L	White	Widowed	2	0	1	3	0	0	0	5	0
8	m72	CN	1	0	84.5	Female	18	Not Hisp/L	White	Widowed	3	1	2	3	0	0	0	2	0
8	m84	CN	1	0	84.5	Female	18	Not Hisp/L	White	Widowed	2	0	0	4	0	0	0	3	0
8	m96	CN	1	0	84.5	Female	18	Not Hisp/L	White	Widowed	3	0	1	3	0	0	0	0	0
8	m120	MCI	2	1	84.5	Female	18	Not Hisp/L	White	Widowed	2	0	0	4	0	0	1	2	0
15	m72	MCI	2	0	80.8	Male	18	Not Hisp/L	White	Married	2	0	0	3	0	0	0	1	0
21	m60	CN	1	0	72.6	Female	18	Not Hisp/L	Black	Divorced	1	0	1	2	0	0	0	0	0
21	m72	CN	1	0	72.6	Female	18	Not Hisp/L	Black	Divorced	2	0	0	2	0	0	0	1	0
21	m84	CN	1	0	72.6	Female	18	Not Hisp/L	Black	Divorced	1	0	0	0	0	0	0	0	0
21	m96	CN	1	0	72.6	Female	18	Not Hisp/L	Black	Divorced	0	0	0	1	0	0	0	4	0
21	m120	CN	1	0	72.6	Female	18	Not Hisp/L	Black	Divorced	2	0	0	2	0	0	0	1	0
21	m144	CN	1	0	72.6	Female	18	Not Hisp/L	Black	Divorced	1.33	0	0	1	0	0	0	8	0
23	m60	CN	1	0	71.7	Male	14	Not Hisp/L	Black	Widowed	4	0	1	4	0	0	0	0	0
23	m72	CN	1	0	71.7	Male	14	Not Hisp/L	Black	Widowed	4	0	1	3	0	0	0	1	0
23	m84	CN	1	0	71.7	Male	14	Not Hisp/L	Black	Widowed	4	0	1	3	0	0	0	0	0
23	m96	CN	1	0	71.7	Male	14	Not Hisp/L	Black	Widowed	4	0	1	4	0	0	0	0	0
31	m60	CN	1	0	77.7	Female	18	Not Hisp/L	White	Widowed	3	0	0	3	0	0	1	1	0
31	m72	CN	1	0	77.7	Female	18	Not Hisp/L	White	Widowed	2	0	0	2	0	0	0	2	0
31	m84	CN	1	0	77.7	Female	18	Not Hisp/L	White	Widowed	2	0	0	5	0	0	0	2	0
31	m96	CN	1	0	77.7	Female	18	Not Hisp/L	White	Widowed	3	0	0	5	0	0	0	0	0
31	m120	CN	1	0	77.7	Female	18	Not Hisp/L	White	Widowed	3	0	0	7	0	0	1	0	0
31	m144	CN	1	0	77.7	Female	18	Not Hisp/L	White	Widowed	3	0	0	4	0	0	0	4	0
31	m156	CN	1	0	77.7	Female	18	Not Hisp/L	White	Widowed	3	0	0	3	0	0	0	4	0

Table 4.6: General Statistics for the Groups of Participants within 36 months from the Baseline and for Different Dementia Stages

Dataset Name	# of Patients before sampling	# of Data Observations (visits)	DX Progress - Class Distribution before Data Balancing	DX Progress - Class Distribution after Data Balancing
ADNI-Merge-ADAS-Cog CN to MCI	287	1,695	Total observations: 1,695 '0': 1,622 (majority 95.70%) '1': 73 (4.30%) -1:20	Total observations: 3,198 '0': 1,622 (50.71%) '1': 1,576 (49.28%)
ADNI-Merge-ADAS-Cog MCI to AD	651	3,275	Total observations: 3,275 '0': 3,068 (majority 93.68%) '1': 207 (6.32%) -1:16	Total observations: 5,982 '0': 3,068 (51.29%) '1': 2,914 (48.71%)
ADNI-Merge-FAQ CN to MCI	328	1,134	Total observations: 1,134 '0': 1,120 (majority 98.77%) '1': 14 (1.23%) -1: 2	Total observations: 2,177 '0': 1,120 (51.45%) '1': 1,057 (48.55%)
ADNI-Merge-FAQ MCI to AD for	693	2,386	Total observations: 2,386 '0': 2,213 (majority 92.75%) '1': 173 (7.25%) -1:3	Total observations: 4,358 '0': 2,213 (50.78%) '1': 2,145 (49.22%)

Obviously, and based on Table 4.6, all datasets are imbalanced in terms of the diagnostic class with many more instances linked with 'No Progression'. To be exact and based on the figures in the third column of Table 4.6, there are just 73 and 14 instances in the 'ADNI-Merge-ADAS-Cog' and 'ADNI-Merge-FAQ' datasets, respectively, linked with participants who progressed from CN to MCI during 36 months from the baseline diagnosis. These numbers constitute just 4.30% and 1.23%, respectively, when compared with instances linked with no progression in both datasets. However, the number of instances who progressed from MCI to AD in the 'ADNI-Merge-ADAS-Cog' and 'ADNI-Merge-FAQ' datasets are 207 and 173, respectively, (6.32% and 7.25%, respectively, of the total). To treat the class imbalance issue we applied SMOTE sampling as shown in the last column of Table 4.6 to avoid biased classification systems in later stages of the analysis. We noticed some

regression instances within the datasets derived; particularly, and for the 'ADNI-Merge-ADAS-Cog' and 'ADNI-Merge-FAQ' datasets that contain 'CN to MCI' instances, there are 20 and 2 regression instances, respectively. For the 'ADNI-Merge-ADAS-Cog' and 'ADNI-Merge-FAQ' datasets that contain 'MCI to AD' instances, there are 16 and 3 regression cases, respectively.

4.3 Platforms, Experimental Setting, and Methods Used

4.3.1 Platforms Used and Settings

All experiments were conducted on a computing machine with an Intel® Core™ i7-6200U 2.8 Ghz with 8GB RAM, on a Windows 10 Home, 64-bit. The hyperparameters of all feature selection methods and classification algorithms remained unchanged in the Weka platform. Moreover, all experiments were conducted using open-sourced software—WEKA and Python, where all platforms have extensive data pre-processing, statistical and graphical tools, as well as machine learning algorithms for data analysis (Witten & Frank, 2002; Van Rossum & Drake Jr, 1995).

Using Python's Seaborn library, we assessed the feature-to-feature correlation within the datasets and to identify highly correlated items to derive influential features subsets from the 'ADNI-Merge-ADAS-Cog' and the 'ADNI-Merge-FAQ-Cog' datasets. The Python Corr function plots the graph of the correlation matrix with coefficients to signal the strength between two items. The function identifies highly correlated features by calculating the largest mean absolute correlation between each item to remove any redundant features. We used Pearson Correlation as the default correlation method within the function to generate a correlation matrix of the data's features as a vector of integers to reduce independent attributes' correlations. When two attributes are highly correlated, the function evaluates the correlation of the mean absolute value for each attribute and drops the one with the greatest value. The suggested Cut-off = 0.60 (Akoglu, 2018).

For the implementation of the rule-based classifier to feed in the knowledge base, we used Java and integrated the algorithm within WEKA version 3.8.4. The reason for selecting Java is that WEKA is implemented in Java and many of the functions used for rule generation and pruning can be re-engineered to design and implement a new algorithm. More importantly, for comparing all other techniques including feature selection and classification, WEKA contains a massive number of algorithms making the process of comparison straightforward. Therefore, for all other computations, we used WEKA version 3.8.4 to perform data sampling techniques, feature selection methods to derive subsets for the input of the classification models and used various classification methods to predict a diagnostic class.

Ten-fold cross validation was used during the experiments as a measure of testing to ensure less biased results. Using ten-fold cross validation, the input dataset is divided into 10 partitions arbitrary

with stratification (Witten & Frank, 2002). Nine partitions are then used for training and the remaining partition for testing; the procedure is repeated ten times to derive the performance measure results.

We applied SMOTE to further sample the minority class labels in the dataset. As discussed in Chapter 3, SMOTE is a data sampling technique that adjusts the class distribution by taking the entire dataset as input increasing the minority class using K nearest neighbours (KNN) (Aha et al., 1991). For feature selection, we used three implemented methods in WEKA: IG, CST, and ReliefF as these methods produce scores based on the correlations between the features and the class label using mathematical models.

To measure the performance of the AD progression models derived by the classifiers against the subsets of functional and cognitive items we used a number of standard evaluation metrics in machine learning including predictive accuracy, sensitivity, and specificity (Nogueira et al., 2018; Teng et al., 2010) as shown in Equations 4.1–4.3, respectively. Sensitivity is the measure of the proportion of actual positive cases predicted as positive. Specificity is the measure of how well a test can identify the true negatives, whilst accuracy is the measure of the correct classification of the instances based on models and measures.

$$Accuracy = \frac{TN+TP}{TN+FP+FN+TP} \quad (4.1)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (4.2)$$

$$Specificity = \frac{TN}{TN+FP} \quad (4.3)$$

where

TP (True Positive) = The model predicts a positive outcome among those with the positive class

FP (False Positive) = The model predicts a positive outcome among those with the negative class

TN (True Negative) = The model predicts a negative outcome among those with the negative class

FN (False Negative) = The model predicts a negative outcome among those with the positive class

4.3.2 Classification Methods Used

To measure the effectiveness of the proposed AD-CR algorithm models in the classification step, besides cognitive and functional features subsets' quality chosen by the feature selection methods, we used seven classification algorithms: Logistic Regression (LR), Multilayer Perceptron (MLP), Sequential Minimal Optimization (SMO), K-Nearest Neighbour (KNN; k=5), Naïve Bayes, Ripple-Down Rule learner (Ridor), and Non-nested generalised exemplars (Nnge), (le Cessie &

Houwelingen, 1992; Rumelhart et al., 1986; Platt, 1998; Aha et al., 1991; John & Langley, 1995; Gaines & Compton, 1995; Martin, 1995). The reasons for choosing these classification methods are due to

1) The dissimilar learning mechanisms used by these algorithms
2) The different types of classifier formats they offer
3) Many of them have been used in medical related research such as artificial neural network, SVM, statistical based classification, rule-based classification and KNN among others
4) To obtain a general conclusion when we compare with the proposed algorithms in terms of performance measures

MLP is a type of neural network that implements a feed forward mechanism in the way it models the problem; the structure of the MLP comprises three layers: input, output, and hidden. The input layer consists of a set of neurons that represent the features in the training dataset, and it receives the data input that require processing. The output layer performs the task of predicting the class label when the task is classification-based on computations made in the hidden layer. Thus, the heart of the neural network that is in charge with the computations is the hidden layers in which one or more hidden layers are created between the input and output layers to model and process the data input. Usually, the data are flowed from the input to the output later and the neurons are processed by a back propagation method during the training phase. The algorithm keeps adjusting the model derived by amending the weights of the neurons until it reaches a performance level that is acceptable

LR is a statistical algorithm that in its simplest form can describe the relationship between two features of data. The first feature is called the independent variable and the second is the dependent variable (class label). In its general form, LR can deal with multi-class problems in which the input dataset contains three or more class labels. LR uses a logistic function as shown in Equation 4.4 to model a class label with two possible values. Unlike linear regression, LR's range is restricted between 0 and 1, and it does not necessitate a linear relationship between the independent variables and the class label since it uses a nonlinear log conversion. In addition, unlike linear regression, LR employs a conditional probability loss function called 'maximum likelihood estimation'. When the probability is less than 0.50, the test data's class label will be predicted as 1; otherwise, 0.

$$Logistic\ function = \frac{1}{1+e^{-x}} \tag{4.4}$$

NB is a probabilistic-based algorithm which uses Bayes theorem to develop a strong feature-based assumption. The algorithm assumes that the target class is independent from all other features in the

dataset, and it computes the likelihood of each class given a test data based on the probabilities of the attributes' values of the test data within the training dataset. The algorithm assigns the class with the largest likelihood to the test data.

KNN is an instance-based learning algorithm in which for a test data to be classified the algorithm utilises the nearest neighbour's class information to assign the test data the appropriate class. The algorithm does not learn a model from the training dataset and then uses that model for predicting the class label as in conventional classification algorithms rather the algorithm employs the training dataset to make the class assignment. The selected K nearest neighbours is often determined by KNN using distance functions such as Manhattan distance or Euclidean distance in which the algorithm selects the closest points to the test data point to make the prediction, therefore determining the ideal K and which distance function to use are crucial. In classifying a test data, the algorithm assigns a class label that belongs to the largest group of neighbours.

Nnge is a generalisation of instance-based learning algorithms with an incremental function in which it utilises non-nested generalised hyperrectangles that can be represented as simple If-Then rules. Every time a new data instance is inserted into the training dataset, Nnge forms a hyperrectangle by integrating the new data instance with a group of neighbours with a similar class label. Nnge disallows hyperrectangles to overlap by using post-pruning based on heuristic. The algorithm employs a modified Euclidean distance function that processes the features, hyperrectangles, and weights.

Ridor is a rule-based induction algorithm that initially produces a default rule and then all possible exceptions for that rule with the smallest expected error rates. Exceptions of the default rule are other rules that forecast the class labels which are dissimilar with that of the default rule. Afterward, the algorithm finds the ideal exceptions for each produced exception and repeats the process until it reaches the best performance. Ridor expands the search of exceptions like decision tree expansion.

SMO is a support vector machine (SVM) type of algorithm disseminated to deal with an important issue that appears during the learning phase of SVM known as a quadratic programming problem. The SMO is a repetitive algorithm that reduces the problem into a set of optimisation tasks and solves each in an analytical manner.

In all classification experiments these algorithms were run in WEKA and without amending the algorithms' hyperparameters. We have used the settings provided in the literature for class association rule as well as some warming up experiments to establish the minimum support and confidence thresholds. Since the minimum support threshold controls the number of rules generated, for the AD-CR algorithm we used a minimum support of between 0.25% and 2% based on the

experimental analysis conducted and previous research studies (Zhang, 2022; Abdelhamid et al., 2014; Liu et al., 1998; Li et al., 2001). These settings often balance between the classification system's size in terms of the number of rules produced and the computing resources used during the learning phase. For the minimum confidence threshold, it has less impact on the performance, hence it has been set to 50% similar to previous studies (Zhang, 2022; Abdelhamid et al., 2014; Liu et al., 1998; Li et al., 2001).

4.4 Chapter Summary

In this chapter, we have discussed the different datasets used for the experimental analysis, the classification methods, the experimental settings, and the evaluation measures used. We also performed in-depth descriptive analytics on the cognitive, functional, and ADNI-Merge datasets besides showing the statistics related to special groups of data subjects: those who progressed from CN to MCI and the ones who progressed from MCI–AD from the baseline diagnosis in a 36-month timeframe.

In the next chapter, we show the feature selection experiments, the results of the classification algorithms, and the results analysis based on the evaluation measures discussed. Specifically, we compare the proposed classification algorithm with the common classification algorithms against the datasets discussed in this chapter.

Chapter Five

Experiments and Results Analysis

This chapter covers the implementation and experimental evaluation of the proposed algorithm with the data-driven architecture. To be exact, we show the results of the proposed algorithm and other dissimilar machine learning algorithms in terms of several evaluation measures and on the datasets we considered. In addition, the chapter discusses the experimental platform and the methods used including computational intelligence and classification. Most of this chapter's content is being considered for publication in the Journal of Biomedical Informatics, and the Journal of Intelligent Decision Technologies.

5.1 Introduction

This chapter discusses the experiments, the results, and analysis of the AD-CR algorithm along with the selected classification algorithms used for real cases and controls of multiple datasets. All experiments related to feature selection methods were conducted and then specific subsets of features identified after analysing the results. To be more specific, we used three different feature selection methods to derive results on datasets related to dementia. We then conducted multiple experiments using a number of classification algorithms to derive dementia progression models that in turn are compared with the proposed AD-CR algorithm's models in terms of predictive accuracy, specificity, and sensitivity measures.

Furthermore, we reveal the performance of the predictive models obtained by AD-CR and the remaining classification algorithms on two dementia groups: data subjects with a baseline diagnosis of CN who for 36 months remained with CN or progressed to MCI, and data subjects with a baseline diagnosis of MCI, who for 36 months remained with MCI or advanced to AD. The analysis is from a neuropsychological item perspective to reveal if any of the cognitive and functional items change as the disease stage changes.

5.2 Feature Selection Experiments

Multiple major sets of experiments have been conducted using feature selection methods against the 'ADNI-Merge-ADAS' and the 'ADNI-Merge-FAQ' datasets to evaluate and identify the cognitive and functional items that can be utilised to diagnose AD progression. The analysis criteria were to

ascertain potential effective subsets of cognitive and functional features that could trigger the progression of AD, and their association with the DSM-5 diagnostic areas related to dementia. The first subset of features in each component contains all medical test items to serve as a baseline for performance comparison against other subsets. Initially, we assessed cognitive and functional parameters separately to capture as many instances as possible per participants as each of these datasets contains more instances than an integrated version of both.

We also assessed an integrated version of both cognitive and functional items ('ADNI-Merge-ADAS- FAQ') to seek any possible correlations and to validate the results obtained against the 'ADNI-Merge-ADAS' and the 'ADNI-Merge-FAQ' datasets. The classification results analysis details the predictive performance of the subsets and the process for selecting the best performing classification model. Each of the neuropsychological components is addressed separately concluding with an evaluation of whether AD progression is best assessed on cognitive or functional items, or whether a combination of both produces better results.

In Section 5.3, we compare the results obtained from the 'ADNI-Merge-ADAS' and the 'ADNI-Merge-FAQ' datasets and their combined set ('ADNI-Merge-ADAS- FAQ') for 36 months from the baseline diagnosis with those derived from specific groups of participants to seek any possible changes in the results. The feature assessment of the neuropsychological items to keep or omit was based on dissimilar criteria including:

1. High-ranked features derived using the scores calculated by the feature selection methods
2. Clusters identified by observing any large drops in % of the scores between subsequent features' scores, from top to bottom ranking order as per the below mathematical formula:

   $$Score\ Drop\ in\ \%\ (Si, Si+1) = \frac{(Si - Si+1)}{Si} \qquad (5.1)$$

   Where Si corresponds to the score of feature i, and $Si + 1$ denotes the score of the next in rank feature (feature $i+1$)
3. Similarity of features is identified based on feature-to-feature assessment where a low intercorrelation is preferred.
4. Common features and their position ranking among the results obtained by the feature selection methods

Each experiment derived unique subsets of features using the approaches summarised in Table 5.1.

The experiments evaluating the neuropsychological items each result in five unique subsets. More details on these selected features' sets are given in Sections 5.3 and 5.4. We anticipate that the

derived subsets can provide clinicians an indication of the sensitive features and their association with the DSM-5 cognitive domains to assist in early screening of AD progression.

Table 5.1: Summary of the Methods Used to Derive Each Neuropsychological Subset

Feature Subset	Analysis Approach Used
1	All Items in each diagnosis method (ADAS, FAQ)
2	Pearson correlation
3	Cluster # 1 – A composite of normalised average scores of the feature selection methods
4	Cluster #1 + #2 – A composite of normalised average scores of the feature selection methods
5	Common features identified by the feature selection methods.

5.3 Cognitive Feature Selection Results & Analysis

Unique subsets of cognitive items were derived from the 'ADNI-Merge-ADAS' dataset (Table 5.2) using the methods and criteria described in Table 5.1. 'Cog-subset2' was derived based on the feature-feature coefficient matrix excluding the class label (Figure 5.1). Based on the correlation matrix figures, 'word recall', 'word-finding', and 'language comprehension' had the highest correlation against other features. For example, 'word recall' and 'word delay' had a strong correlation, both having the same influence on the diagnostic class attribute. Accordingly, one of the items could be ignored —for instance, 'word recall' has a larger mean absolute correlation than the other cognitive items and is thus it can be a candidate for removal.

The properties of the 'word recall', 'word-finding', and 'language comprehension' have the highest contributing factor to a high correlation when compared with each cognitive item in the dataset. While these three features can be made redundant, the remaining features will still have the same effect on the class attribute. The removal of these items may reduce overlapping of feature properties. The redundant items also have overlapping DSM-5 cognitive domains which are learning and memory, language, and complex attention—the remaining items also cover these domains.

Table 5.2: Summary of the Cognitive Items for Each Data Subset

Subset	Items Description	Criteria used
Cog-subset1	All Cog items	-
Cog-subset2	COMMAND, CONSTRUCT, DELAYWORD, NAMING, IDEATIONAL, ORIENT, WORDRECOG, RMBRTESTINSTR, SPOKENLG, NUMBERCANCEL	Remove highly correlated items based on the feature-feature correlation matrix
Cog-subset3	WORDRECALL, DELAYWORD, WORDRECOG	Cluster analysis based on the drop score %
Cog-subset4	WORDRECALL, DELAYWORD, WORDRECOG, ORIENT, COMMAND, WORDFIND	Cluster analysis based on the drop score %
Cog-subset5	WORDRECALL, DELAYWORD, WORDRECOG	Common items in the feature selection results

However, ‘Cog-subset2’ has only three fewer features than the original cognitive items in ADAS, thus does not significantly reduce assessment and computational time.

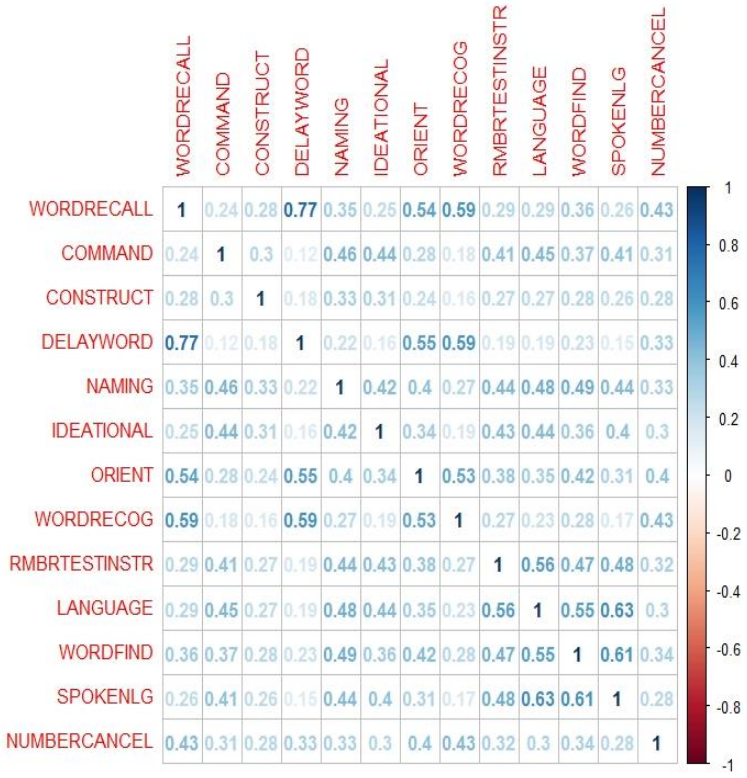

Figure 5.1: Correlation Coefficient Matrix of Cognitive Items

‘Cog-subset3’ and ‘Cog-subset4’ are derived from a composite normalised average weighting of scores computed by the considered feature selection methods. The computed scores are normalised, then averaged as illustrated in Table 5.3a, and then ranked in Table 5.3b from highest to lowest to measure the % drop between the features. We then identified three clusters by observing a distinct drop pattern—in this case when there was a drop of >30% in the scores of two successive features. With the identified cluster groups, we utilise the features within cluster 1 as ‘Cog-subset3’, which consists of ‘word recall’, ‘delayed word recall’, and ‘word recognition’, covering the two DSM-5 cognitive domains of learning and memory, and language. For ‘Cog-subset4’, we combined all features within clusters 1 and 2, bringing the subset to six items by adding ‘orientation’, ‘command’, and ‘word finding’. ‘Cog-subset4’ covers four out of six of the cognitive domains, which are learning and memory, language, executive function, and perceptual motor function. The remaining clusters were made redundant due to their lower effect on the class attribute.

'Cog-subset5' produced similar results to 'Cog-subset3', whereby the three feature selection methods individually ranked similar items in the top three when computed separately. Thus 'Cog-subset5' will not be tested in the classification process.

The feature selection methods used identified three items that occurred the greatest number of times within each of the derived subsets: 'word recall', 'delayed word recall', and 'word recognition', signalling their significant influence on predicting AD progression. These three items have the common association of retrieving words which requires a patient to read, remember, recall, and recognise, and thus taps into their learning and memory, and language cognitive domains. During an assessment, a clinician can pay attention to the performance of these tasks by the patient and determine whether there are signs of disease advancement and if the patient needs early intervention.

We performed feature assessment on the cognitive items for specific dementia groups from baseline diagnosis up to 36 months: 'Cog-CN-MCI' – participants with a baseline diagnosis of CN who then advanced to MCI; 'Cog-MCI-AD' – participants with a baseline diagnosis of MCI who then advanced to AD. In the 'Cog-CN-MCI' group, we treated participants with CN who had not progressed to MCI as negative cases, and those who progressed to MCI as positive cases. For the 'Cog-MCI-AD' group, participants with a baseline diagnosis of MCI who remained with this diagnosis were treated as negative cases, while those who progressed from MCI to AD were treated as positive cases. Figures 5.3a and 5.3b represent the correlations among cognitive items for the 'Cog-CN-MCI' and 'Cog-MCI-AD' sub-groups, respectively.

The results show a high correlation between 'word recall' and 'delayed word recall' with correlations of 70% or more in the 'Cog-CN-MCI' and 'Cog-MCI-AD' sub-groups. In addition, 'word recall', and 'delayed word recall' have high correlations with 'word recognition' activity

Table 5.3A: Cognitive Items with Computed Scores and Normalised Scores Derived by the Feature Selection Methods (ADAS-subset3 and ADAS-subset4)

Feature	IG		CST		ReliefF		Average Scores
	Score	Normalised	Score	Normalised	Score	Normalised	
WORDRECALL	0.135	1.000	1956.268	1.000	0.144	0.683	0.894
COMMAND	0.049	0.246	699.447	0.242	0.026	0.100	0.196
CONSTRUCT	0.032	0.093	475.613	0.107	0.030	0.121	0.107
DELAYWORD	0.084	0.551	1327.623	0.621	0.208	1.000	0.724
NAMING	0.032	0.101	472.299	0.105	0.022	0.080	0.095
IDEATIONAL	0.038	0.150	534.736	0.143	0.015	0.047	0.113
ORIENT	0.044	0.201	691.697	0.237	0.081	0.372	0.270
WORDRECOG	0.062	0.364	975.468	0.408	0.172	0.826	0.533
RMBRTESTINSTR	0.021	0.002	298.131	0.000	0.005	0.000	0.001
LANGUAGE	0.028	0.059	389.760	0.055	0.009	0.016	0.043
WORDFIND	0.042	0.182	629.659	0.200	0.035	0.147	0.176
SPOKENLG	0.025	0.033	347.811	0.030	0.009	0.020	0.028
NUMBERCANCEL	0.021	0.000	314.056	0.010	0.060	0.268	0.092

pinpointing to overlapping of these activities. To be more specific, the correlations between 'word recall' & 'word recognition', and 'delayed word recall' & 'word recognition' derived from the 'Cog-

Table 5.3B: Cognitive Clusters Identified Based on the % Drop Between the Ranked Items According to their Normalised Average Scores (ADAS-subset3 and ADAS-subset4)

Feature By Rank	Normalised Average Scores	% Drop after Normalisation	Cluster No.
WORDRECALL	0.894		Cluster #1
DELAYWORD	0.724	19.07%	Cluster #1
WORDRECOG	0.533	26.41%	Cluster #1
ORIENT	0.270	49.29%	Cluster #2
COMMAND	0.196	27.47%	Cluster #2
WORDFIND	0.176	10.17%	Cluster #2
IDEATIONAL	0.113	35.66%	Cluster #3
CONSTRUCT	0.107	5.45%	Cluster #3
NAMING	0.095	10.90%	Cluster #3
NUMBERCANCEL	0.092	3.13%	Cluster #3
LANGUAGE	0.043	53.22%	Cluster #4
SPOKENLG	0.028	36.03%	Cluster #5
RMBRTESTINSTR	0.001	97.89%	Cluster #6

CN-MCI' and 'Cog-MCI-AD' sub-groups are 0.35 and 0.56, and 0.34 and 0.61, respectively thus signalling higher correlations in the participants who progressed from MCI to AD.

Results derived by the feature selection methods (IG, CST, and ReliefF) against the dementia sub-groups showed that 'delayed word recall', 'word recognition', and 'word recall' maintained high correlation with the diagnostic class. This supports earlier results obtained by the feature selection methods from the 'ADNI-Merge-ADAS' dataset in which these cognitive activities appeared to have high correlation with the class label (DX Progress), ranking top when compared with the remaining cognitive activities. Therefore, when one or more of these features is noticed by the clinicians during the diagnosis process of a patient, this should be taken into consideration.

The results obtained on the 'Cog-CN-MCI' sub-group suggest that 'ideational praxis', 'naming of objects', and 'commands' cognitive features form a cluster since these have little overlapping with other cognitive features in the dataset. For example, 'command', which measures impairment in receptive speech, has low negative correlations with 'word recall', 'delayed word recall', 'naming',

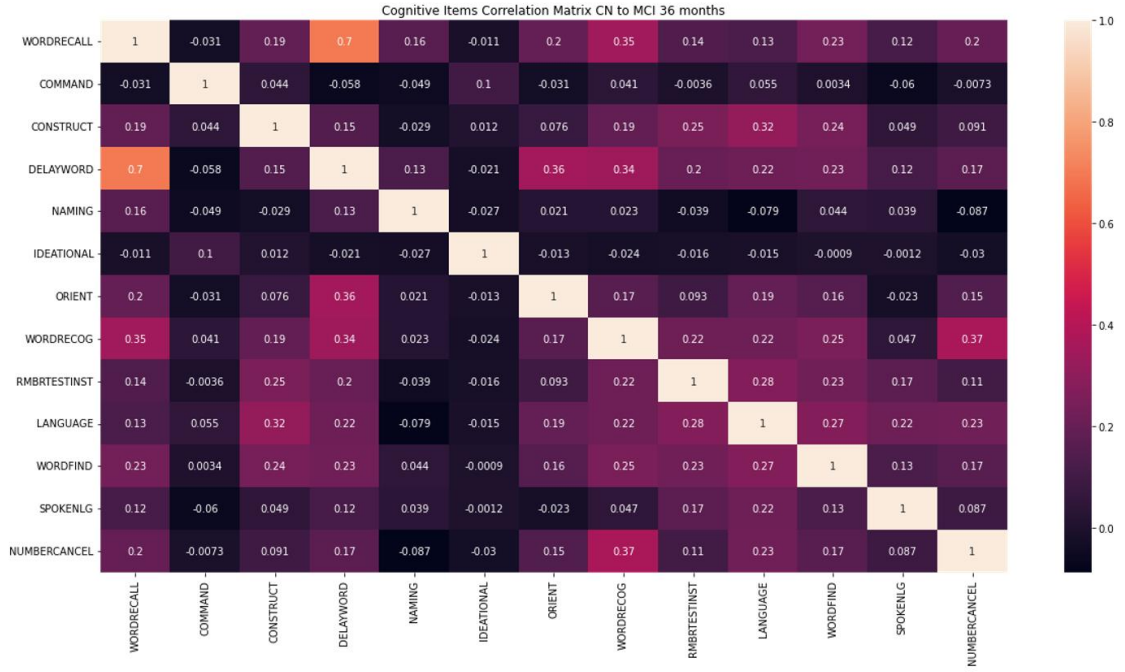

Figure 5.2a: Feature-Feature Correlation Matrix of Cognitive Items for CN to MCI Group

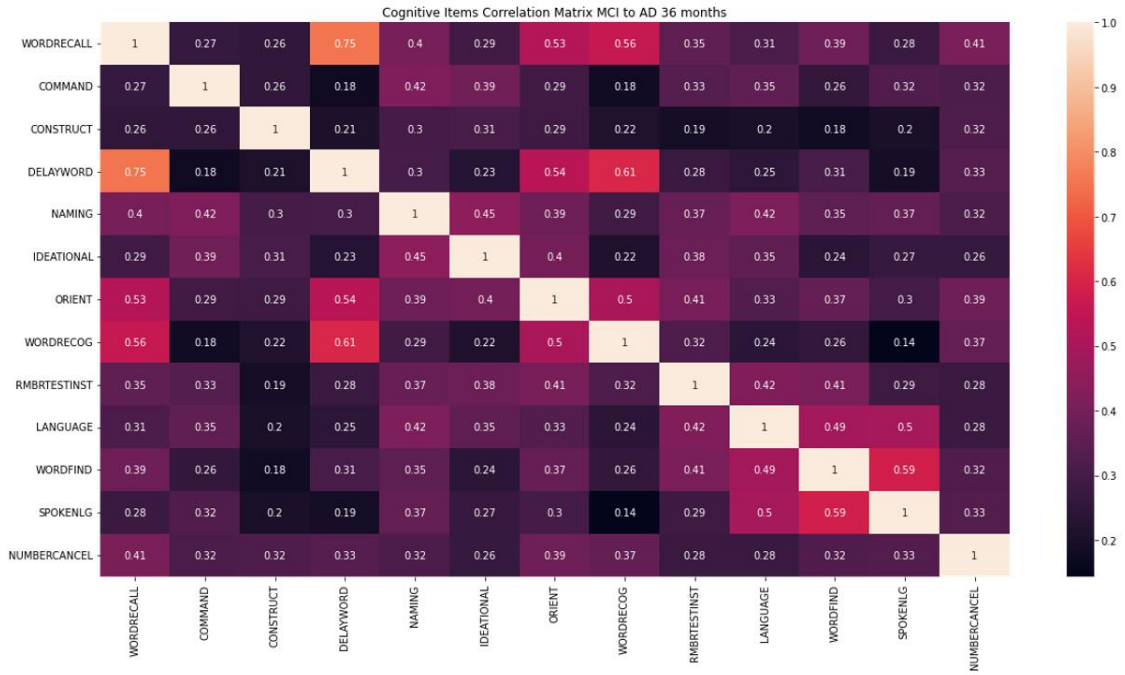

Figure 5.2b: Feature-Feature Correlation Matrix of Cognitive Items for MCI to AD Group

orientation', 'remembering instructions', 'spoken language comprehension', and 'number cancellation tasks'. In addition, 'naming of objects', which measures learning and memory besides language by asking patients to name 12 random objects besides the fingers on their dominant hand has low negative correlations with 'command', 'constructional praxis', 'ideational praxis', 'remembering instructions', 'language' and' number cancellation tasks'. While 'ideational praxis', which measures the ability of the participant in performing a complex sequence of actions, has very little negative correlations with all cognitive items except two: 'command' and 'constructional praxis,' making it the most different item. The three cognitive activities ('command', 'naming of objects', 'ideational praxis') cover five cognitive areas in the DSM-5 framework (language, executive function, perceptual motor skills, learning and memory, complex attention) making them good candidates for representing cognitive items in neuropsychological assessments including ADAS especially for individuals that exhibit MCI traits.

Overall, the results on assessing the cognitive activities show non-overlapping features derived from the group of participants who had a baseline diagnosis of CN and progressed to AD, which are 'command', 'naming of objects', and 'ideational praxis'. These cognitive activities when considered as a group have very little correlations with all remaining cognitive items assessed. There was more overlapping among the features for the 'Cog-MCI-AD' sub-groups suggesting a difficult task to separate the participants who remained MCI from those who may advance to light dementia since feature-feature correlations were all above 14%. For example, the least correlation found from the sub-group of 'Cog-MCI-AD' was between 'spoken language' and 'word recognition', which supports previous finding from the 'ADNI-Merge-ADAS' dataset, possibly pinpointing that these two can be investigated further by the clinicians when clinically assessing patients with MCI.

5.4 Functional Feature Selection Results Analysis

From the 'ADNI-Merge-FAQ' dataset, five unique subsets of functional items were devised (Table 5.4) using the feature selection methods based on criteria described in Table 5.1. Using the Pearson's correlation matrix shown in Figure 5.3, six features including 'finance', 'shopping', 'travel', 'remembering an occasion', 'completing forms' and 'current events' have been identified as having high correlation with other functional items, all scoring above the cut-off coefficient of 0.6. For example, 'finance' and 'completing forms' activities have a strong correlation thus one of the two items can be ignored. Both items have high mean absolute correlation against the rest of the items when the entire correlation matrix is analysed. This is also due to having overlapping DSM-5

domains with the remaining functional items ('playing a game', 'making a beverage', 'watching TV', 'remembering an occasion') which makes up 'Func-subset2'. These four functional items tapped into all six of the prescribed DSM-5 criteria to diagnose AD, and therefore serve as a good measure to investigate if this subset when processed by a classification algorithm can provide models with the best prediction accuracy for AD progression.

Func-subset3' and 'Func-subset4' were also derived by identifying cluster groups using the composite normalised average weighting of the scores derived by the considered feature selection methods. 'Func-subset3', chosen using cluster #1 as shown in Table 5.5a, consists of 'remembering

Table 5.4: Summary of the Derived Functional Items for Each Subset of FAQ

Subset	Functional Items	FS Criteria
Func-subset1	All FAQ items	Remove highly correlated items based on the feature-feature correlation matrix
Func-subset2	GAME, BEVERAGE, MEAL, TV	Cluster analysis based on the drop score %
Func-subset3	REMEMBER, SHOP, FINANCE, FORM, TRAVEL, EVENT	Cluster analysis based on the drop score %
Func-subset4	REMEMBER, SHOP, FINANCE, FORM, TRAVEL, EVENT, TV, GAME, MEAL	Common items in the feature selection results
Func-subset5	REMEMBER, SHOP, FINANCE	Remove highly correlated items based on the feature-feature correlation matrix

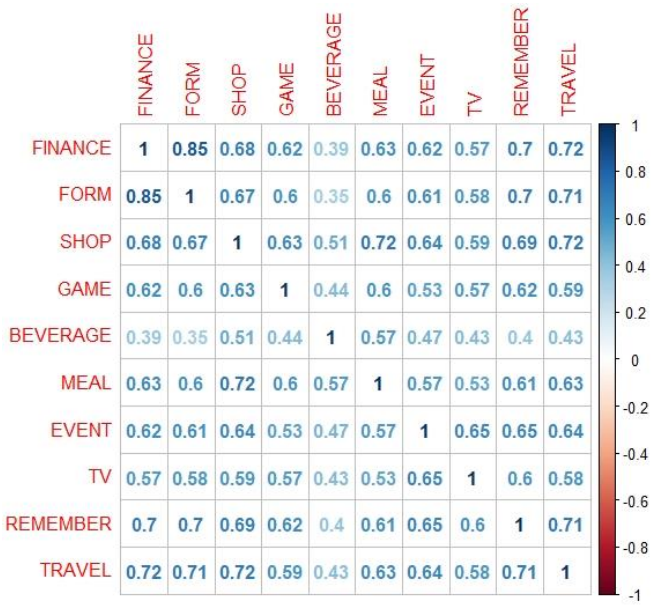

Figure 5.3: Correlation Coefficient Matrix of Functional Items

an occasion', 'shopping', 'finance', 'completing forms', 'travel', and 'current events' activities, covering four DSM-5 cognitive domains. 'Func-subset4' is a combination of both clusters 1 and 2, adding three additional functional items: 'watching TV', 'playing a game', and 'preparing a meal'. 'Func-subset4' covers more DSM-5 domains in comparison to 'Func-subset4' with two additional domains which are language and social cognition associated with 'watching TV' activity. However, 'Func-subset4' has only one feature less than the original functional items, and while it may provide better performance, it does not dramatically reduce assessment time. Moreover, 'Func-subset5' was

derived based on the common top-ranked features derived by the feature selection methods' scores as shown in Table 5.5b, i.e., 'finance', 'shopping', and 'remembering an occasion'. The domains that these functional items tapped into are executive function, perceptual motor skills, complex attention besides learning and memory.

Within each of the subsets derived, 'remembering an occasion', 'shopping,' and 'finance' are the functional activities that occurred the greatest number of times. They tapped into the cognitive domains of executive function, perceptual motor skills, complex attention besides learning and memory. Learning and memory is the only common domain that is covered across the three items, suggesting the significance of this domain on AD progression. During an assessment, a clinician can specifically focus on the patient's performance of these tasks and determine whether early intervention is required.

Table 5.5A: Functional Items with Computed Scores and Normalised Scores Derived by the Feature Selection Methods (Func-subset3 and Func-subset4)

Feature	IG		CST		ReliefF		Average Scores
	Score	Normalised	Score	Normalised	Score	Normalised	
FINANCE	0.064	0.968	1610.665	0.977	0.050	0.698	0.881
FORM	0.061	0.905	1542.824	0.920	0.050	0.706	0.844
SHOP	0.065	0.998	1618.804	0.984	0.053	0.808	0.930
GAME	0.049	0.623	1229.347	0.654	0.050	0.698	0.658
BEVERAGE	0.022	0.000	456.453	0.000	0.033	0.000	0.000
MEAL	0.047	0.572	1168.080	0.603	0.051	0.722	0.632
EVENT	0.054	0.738	1313.117	0.725	0.056	0.963	0.809
TV	0.053	0.711	1289.698	0.705	0.050	0.682	0.699
REMEMBER	0.065	1.000	1637.527	1.000	0.053	0.812	0.937
TRAVEL	0.053	0.711	1328.007	0.738	0.057	1.000	0.816

Table 5.5B: Cluster Items Identified Based on the % Drop Between the Functional Items According to their Normalised Average Scores (Func-subset3 and Func-subset4)

Feature by Rank	Normalised Average Scores	% drop after Normalisation	Cluster No.
REMEMBER	0.937		Cluster #1
SHOP	0.930	0.79%	Cluster #1
FINANCE	0.881	5.28%	Cluster #1
FORM	0.844	4.23%	Cluster #1
TRAVEL	0.816	3.26%	Cluster #1
EVENT	0.809	0.88%	Cluster #1
TV	0.699	13.57%	Cluster #2
GAME	0.658	5.85%	Cluster #2
MEAL	0.632	3.96%	Cluster #2
BEVERAGE	0.000	100.00%	Cluster #3

We further examined the correlation among functional features for two dementia sub-groups from baseline diagnosis up to 36 months: 'Func-CN-MCI' – participants with a baseline diagnosis of CN who then progressed to MCI or remained CN; and 'Func-MCI-AD' – participants with a baseline

diagnosis of MCI who then progressed to AD or remained as MCI. The feature-feature correlations for the 'Func-CN-MCI' and 'Func-MCI-AD' sub-groups are depicted in Figures 5.4a and 5.4b, respectively.

The results obtained from the 'Func-CN-MCI' reveal that there are positive correlations although less than the cut-off at 0.60 among five cognitive activities, which are 'finance', 'completing forms', 'playing a game', 'current events', and 'remembering an occasion'. A feature-feature correlation was observed although it was less than 0.60 between 'current events ', and 'playing a game' as both activities overlap in the learning and memory cognitive domain. This correlation is more obvious for the participants who progressed to MCI or remained CN from a baseline CN when compared with the previous results of the original 'ADNI-Merge-FAQ' dataset. To be exact, the corelation between the 'current events' and 'playing a game' is 34.40% higher than the second highest correlations obtained among the functional items from the 'Func-CN-MCI' group. Additionally, 'current events' has a 0.34 correlation with 'preparing a meal', and 'remembering an occasion' activities which aligns with previous findings derived from the 'ADNI-Merge-FAQ' dataset using feature selection methods.

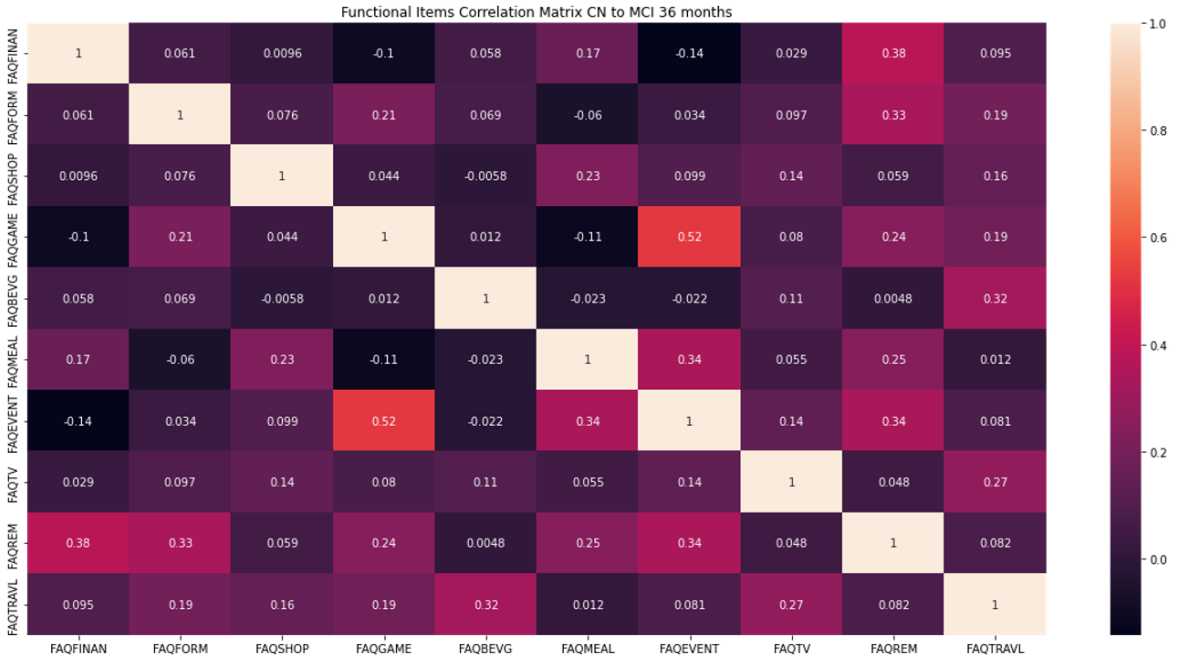

Figure 5.4a: Feature-Feature Correlation Matrix of Functional Items for CN to MCI Group

The results derived by the correlation matrix show that despite that there are some associations among functional items in the FAQ method, these are limited to few functional items and often less than the cut-off score of 0.60, at least for participants in the 'Func-CN-MCI' group. For example, the 'finance' activity, which deals with preparing finances and involves cognitive domains including complex attention, executive functioning, and learning and memory has none to minimal

associations with most functional activities except 'remembering an occasion'. The same pattern is observed for the 'completing forms' activity since such an activity involves similar cognitive functions. The least functional activity that has associations with other functional features is 'watching TV' as although it covers five cognitive areas, the overlapping of such areas with the remaining functional items is not high with one exception which is its association with 'travel' activity. When we evaluated the subset of 'Func-CN-MCI' using a correlation feature subset (CFS) method, which considered both feature-feature and feature-class associations (Hall, 1999), 'watching TV' was selected along with 'remembering an occasion'—the latter was the top-ranked feature in IG, CST, and ReliefF feature selection results.

The results obtained from the 'Func-MCI-AD' show that in general it is hard to distinguish cases of MCI from AD using just functional items, especially in the early stages up to 36 months from a baseline diagnosis of MCI. This is since the correlations obtained among the functional items are noticeable and with close proximities making the process of differentiating specific functional items that can trigger advancements to AD a difficult task. It seems that there is high overlapping in

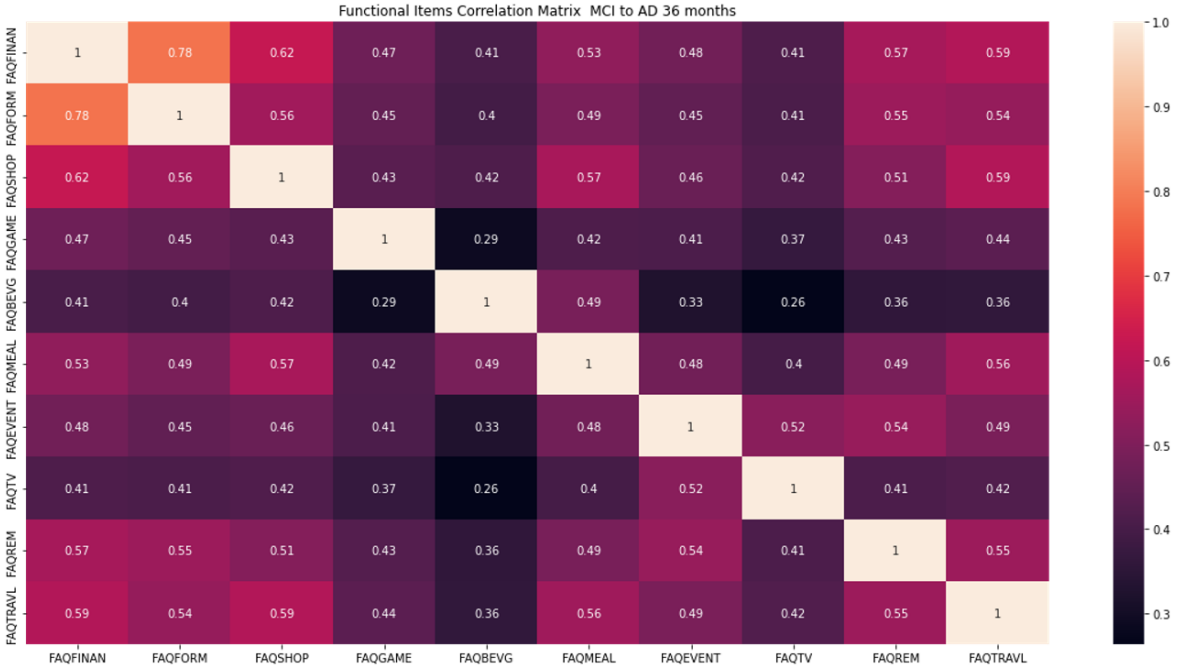

Figure 5.4b: Feature-Feature Correlation Matrix of Functional Items for MCI to AD Group

functional activities and cognitive traits for participants with a baseline diagnosis of MCI who then advance to AD or remain MCI, at least when using functional and cognitive neuropsychological assessments. The fine lines that differentiate cases with MCI from AD, especially in the early stages,

are difficult to obtain at least using the data subjects and the medical assessment methods we considered.

The top-ranked features obtained by the feature selection methods from the 'Func-MCI-AD' dataset are 'finance', 'completing forms', 'shopping', and 'remembering an occasion' (Func-subset5), which supports the results derived from the general functional data, i.e. 'ADNI-Merge-FAQ' dataset, by the same methods. The same features also appeared to be significant based on the correlation feature set results.

5.5 Combined Cognitive-Functional Feature Selection Results Analysis

Most of the existing data-driven approaches in cognitive-related studies have investigated cognitive domains in dementia screening and diagnosis, at least using the ADNI data repository or other scarce data repositories related to dementia conditions. Limited previous research works have shown the assessment of cognitive and functional items together in a single composite to seek the progression of AD, for instance Wessels et al. (2015). However, their approach to measure cognitive and functional items was primarily based on statistical visualisation with principal component analysis (PCA) over a period of time (longitudinal) and did not use advanced techniques such as computational intelligence and machine learning. Therefore, their study did not capture the similarity or dissimilarity of items in both domains, mapped these items to the degenerative domains defined in the DSM-5 manual, and constructed classification models based on data modelling and training processes to predict AD progression. Our research augmented previous research works, not only by measuring cognitive and functional domains separately, but also by assessing key performance indicators in both domains that may trigger advancements of the disease using feature selection with classification algorithms. The approach proposed may provide different stakeholders, such as the clinicians, with rich information related to the disease's progression, and cognitive and functional areas that need more attention.

The relationships among functional and cognitive items together have been investigated using feature selection. Figure 5.5 displays cognitive and functional items correlations from the processed 'ADNI-Merge-ADAS-FAQ' dataset, and for all baseline diagnostic class labels. This dataset combined 'ADAS-sheet', 'FAQ-sheet', and 'ADNI-Merge' retaining only the cognitive and functional items. The results obtained support the previous results derived by the feature selection

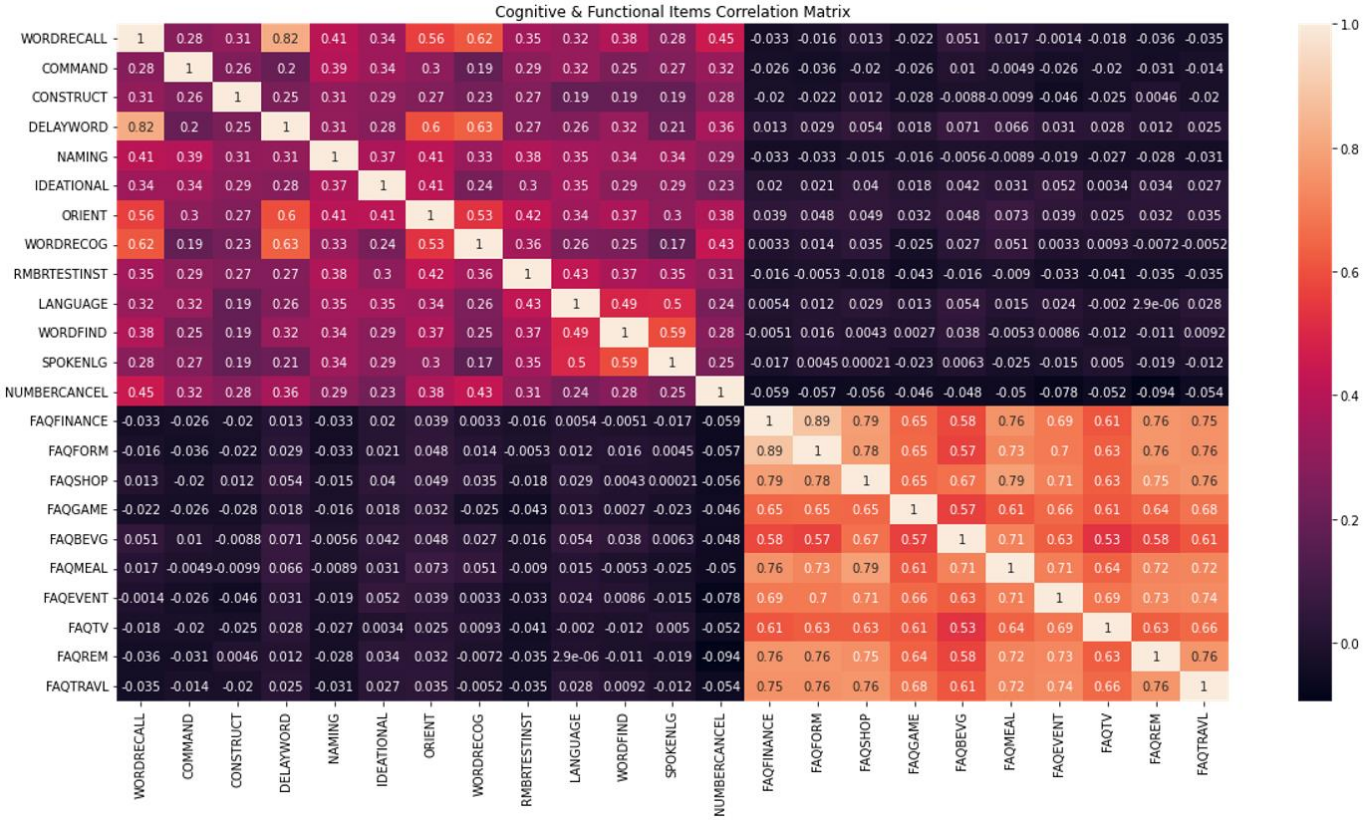

Figure 5.5: Feature-Feature Correlation Matrix of Combined Cognitive and Functional Items for all Dementia Stages

methods and the correlation matrices' results obtained from the 'ADNI-Merge-ADAS' and 'ADNI-Merge-FAQ' datasets.

5.6 Classification Results Analysis

Using the derived distinct subsets of features from the 'ADNI-Merge-ADAS' and 'ADNI-Merge-FAQ' datasets, we evaluated their goodness by building predictive models using dissimilar classification algorithms including Logistic Regression (LR), Multilayer Perceptron (MLP), Sequential Minimal Optimization (SMO), K-Nearest Neighbour (KNN; k=5), Naïve Bayes (NB), Ripple-Down Rule learner (Ridor), and Non-nested generalised exemplars (Nnge) (le Cessie & Houwelingen, (1992); Rumelhart et al., (1986); Platt, (1998); Aha et al., (1991); John & Langley, (1995); Gaines & Compton, (1995); Martin, (1995)), besides the AD-CR. The reason for choosing

these algorithms is because they adopt different learning schemes, and they produce a various format of classification models. Other reasons have been discussed in Chapter 4.

The difference in the number of features in each subset and their DSM-5 cognitive domain coverage will assist us when comparing the classification model performance measures such as predictive accuracy, sensitivity, and specificity. 'Cog-subset1' and 'Func-subset1' in Tables 5.2 and 5.3, respectively, act as a baseline performance guide for both cognitive and functional items, to assess the quality of the performance of the other subsets with reduced features. More importantly, we assess predictive models of data subjects that belong to specific dementia stages including the progression of the disease for up to 36 months for two sub-groups (CN-MCI) and (MCI-AD).

5.6.1 Cognitive Classification Results Analysis

Table 5.6 (left side) illustrates the performance of the classification algorithms against each cognitive subset using only the derived features and the class. The results from the feature assessment analysis and without using any demographic features showed the superiority of the proposed AD-CR algorithm. Specifically, the models derived by the AD-CR algorithm all have acceptable performance in terms of sensitivity, specificity, and predictive accuracy even when only three items were processed ('Cog-subset3').

Moreover, the AD-CR algorithm was superior to the other algorithms when processing the baseline and 'Cog-subset2' in terms of predictive accuracy. For instance, it was able to produce predictive models from the baseline cognitive subset with a higher accuracy than these of LR, MLP, SMO, KNN, NB, Ridor, and Nnge with 13.12%, 7.36%, 14.18%, 0.87%, 23.93%, 2.56%, and 0.36%, respectively. In fact, the AD-CR algorithm was able to derive a model from just three cognitive items of the 'Cog-subset3' (word recall, delayed word recall, and word recognition) with 83.80% sensitivity rate which is relatively good. This rate is just 3.10% less than when the same algorithm processed the complete cognitive items (Cog-subset1). It seems that using three items only of ADAS that belong to the learning and memory and language DSM cognitive domains can indeed help clinicians in detecting possible disease progression, at least using models generated by the AD-CR and KNN algorithms. These three items also commonly appear in the other derived subsets, underlining their importance in predicting AD progression.

Furthermore, the results revealed that albeit the KNN algorithm derived predictive models that were not the most accurate from the complete cognitive items and 'Cog-seubset2', it was able to derive models from 'Cog-subset3' and 'Cog-subset4' that were slightly better than the AD-CR. This supports the view that the cognitive items in these subsets are important for detecting any possible

Table 5.6: Performance of the Classification Methods from Different Subsets of the Cognitive Items

Subset	Algorithm	Excluding Demographic Features			Including Demographic Features		
		Accuracy %	Sensitivity %	Specificity %	Accuracy %	Sensitivity %	Specificity %
Cog-subset1 (baseline)	LR	74.88	76.90	72.90	83.15	87.10	79.30
	MLP	80.64	83.60	77.80	87.03	87.70	86.40
	SMO	73.82	77.00	70.70	82.59	88.60	76.70
	KNN	87.13	92.10	82.30	88.78	93.50	84.10
	NB	64.07	94.40	34.30	70.54	94.50	35.18
	Ridor	85.44	79.70	91.10	90.32	86.10	94.50
	Nnge	87.64	86.90	88.40	92.22	91.60	92.80
	AD-CR	88.00	86.90	89.10	92.38	91.30	93.50
Cog-subset2	LR	71.28	69.90	72.70	90.26	93.30	87.30
	MLP	77.29	76.90	77.60	86.97	87.70	86.30
	SMO	69.72	63.00	76.30	81.81	88.30	75.40
	KNN	85.64	89.80	81.50	88.52	92.90	84.30
	NB	63.15	93.50	33.30	70.81	93.20	48.80
	Ridor	89.32	90.20	88.50	89.41	86.70	92.10
	Nnge	83.68	82.60	84.70	92.16	91.80	92.50
	AD-CR	85.21	83.80	86.60	91.90	90.80	93.00
Cog-subset3	LR	61.11	58.60	63.60	77.03	80.90	73.30
	MLP	62.42	64.60	60.30	83.78	85.30	82.30
	SMO	61.96	58.90	64.90	77.20	84.20	70.30
	KNN	82.81	88.90	76.80	89.10	91.90	86.30
	NB	59.59	61.90	57.30	78.51	85.90	71.30
	Ridor	79.46	77.30	81.50	88.59	82.90	94.20
	Nnge	77.01	76.80	77.20	91.77	91.50	92.00
	AD-CR	80.53	83.80	77.40	91.25	89.50	92.90
Cog-subset4	LR	79.40	84.10	74.80	90.35	93.10	87.60
	MLP	74.10	84.20	65.90	85.97	87.80	84.20
	SMO	68.26	66.90	69.60	80.80	88.10	73.60
	KNN	85.91	90.30	81.60	89.61	92.70	86.60%
	NB	68.78	93.20	44.80	77.63	90.50	65.00
	Ridor	83.06	75.80	90.10	89.63	86.30	92.90%
	Nnge	83.34	83.50	83.20	92.12	92.00	92.30
	AD-CR	84.77	85.40	84.20	91.93	90.70%	93.20

disease advancement. Specifically, evaluating three additional cognitive items of 'orientation', 'command', and 'word finding' in addition to the items of 'Cog-subset3' resulted in predictive models with a 3.10% and 4.24% increase in accuracy for the KNN and the AD-CR algorithms, respectively, when compared with models derived by the same algorithms from just three items ('Cog-subset3'). In addition, the Ridor and Nnge algorithms, which produce classifiers with rules, were able to produce classification models for AD progression that were good in terms of accuracy

from all cognitive subsets except ‘Cog-subset3’. This suggests that rule-based classification is an appropriate machine learning approach for the problem of predicting AD progression.

We investigated the confusion matrix results to understand how these classification models behave in terms of performance. For instance, for the AD-CR model produced from ‘Cog-subset1’, out of the 5,923 positive instances that were supposed to have AD progression, the AD-CR algorithm was able to correctly predict 5,147 instances (true positive) and misclassified 776 instances as ‘no progression’ (false negative), when these instances were in fact having AD progression. The ability of the model to predict AD progression correctly is significant especially in the medical field.

While using the AD-CR algorithm, models derived from ‘Cog-subset1’ and ‘Cog-subset2’ produced the best accuracy measure; surprisingly, the NB probabilistic algorithm derived models from these two cognitive subsets with the highest sensitivity. NB algorithm derived models from the ‘Cog-subset1’ and ‘Cog-subset2’ with 93.50% and 94.40%, respectively—superior to the remaining classification algorithms, at least on the sensitivity metric. Nevertheless, the specificity rates derived by the NB algorithm from these datasets were unacceptable at 33.30% and 34.30%, respectively. In other words, the NB algorithm can detect disease advancement but with a high number of false positives. For example, by analysing the confusion matrices of the NB model against ‘Cog-subset1’, we discovered that out of 5,923 positive instances, only 333 instances were misclassified by the NB algorithm as ‘no progression’. However, there were 3,957 false positives out of 6,020 cases which contributed to the low specificity rate. In general, the sensitivity rate of the models derived by NB deteriorated significantly to 61.90% when the number of cognitive items was reduced to just three (‘Cog-subset3’) as shown in Table 5.6.

The proposed algorithm was able to balance between specificity and sensitivity rates across all models derived from the distinct cognitive feature subsets. Specifically, the AD-CR derived models with sensitivity and specificity rates of 86.90%, 89.10%, and 83.80%, 86.60%, respectively from the ‘Cog-subset1’ and ‘Cog-subset2’. When comparing the sensitivity rates of the models produced by the AD-CR algorithm it is evident that the model derived from the ‘Cog-subset1’ had the highest; the difference in percentages of AD-CR models across all other measures against ‘Cog-subset3’ is not significant, where only its specificity (77.40%) is below 80.00%. The fact that ‘Cog-subset3’ when processed by the proposed algorithm can produce good results with only three key cognitive items, means less assessment and computation time is required. Furthermore, models derived from ‘Cog-subset3’ achieved this despite this subset of data covering only two DSM-5-prescribed cognitive domains, in comparison to ‘Cog-subset2’ and ‘Cog-subset4’ which cover five and four cognitive domains, respectively. Interestingly, the overall performance of the classification techniques produced similar results as other validation studies where the accuracy ranged from

82%–99.6%, sensitivity from 58%–74% and specificity from 91%–98% (Nogueira et al., 2018; Yang et al., 2019).

Reducing the number of items in assessments related to dementia conditions, such as AD, can enable diagnosticians to evaluate specific elements related to a patient's condition based on the state of the disease, the patient's medical history, and the characteristics. Focus can then be on the important cognitive elements detected for the progression of the disease for individualised intervention and management plans to be designed to suit the patient and their family members. This will have positive impacts on their lives as well as on the healthcare system.

We investigated the impact of adding demographic features (age, gender, level of education, race category, marital status) to the cognitive items (right side of Table 5.6). The performance across all classification algorithms improved immediately, particularly evident in the SMO, LR, and MLP algorithms with noticeable improvement across all performance measures. For example, the improvement in accuracy, sensitivity, and specificity rates of the LR models after considering demographics with cognitive items of 'Cog-subset2' increased 18.98%, 23.40%, and 14.60%, respectively, when compared with the model derived from 'Cog-subset2' without geographics. Overall, the performance metrics' results improved for all the classification algorithms besides AD-CR when including demographics in the cognitive subsets.

While the AD-CR algorithm also improved, though not as significantly (at least when processing 'Cog-subset1'), it was still superior in comparison to the other algorithms overall as it further elevated reaching a performance above 92%, 92.38% accuracy, 91.30% sensitivity, and 93.50% specificity. The fact that the proposed algorithm derived classification models from all distinctive cognitive subsets with predictive accuracy above 90.00% is evidence that this algorithm is suitable for detecting dementia stage progression. The AD-CR algorithm produced models from all distinctive cognitive subsets with sensitivity and specificity rates of over 90.00%—except a specificity of 89.50% from 'Cog-subset3', which contains only three cognitive items plus demographics.

In general, the Nnge algorithm derived competitive classification models on most cognitive subsets and above 90.00% accuracy rate. The AD progression models' performance derived by the AD-CR and Nnge algorithms from just six features ('Cog-subset4') particularly their accuracy, sensitivity, and specificity, are close to those derived by the same algorithm from the complete cognitive items set ('Cog-subset1'). One notable result was observed by the models produced by the LR algorithm from 'Cog-subset4' including demographics in which the accuracy, sensitivity, and specificity rates improved by 7.20%, 6.00%, and 8.30%, respectively, when compared with those derived by the same algorithm from 'Cog-subset1' with demographics. The same pattern was also observed on the

accuracy and specificity rates of the NB algorithm which improved by 7.09% and 29.82%, respectively, making the six cognitive items in 'Cog-subset4' significant, especially when investigated with demographic attributes such as age and gender. In fact, age was the highest-ranked feature among geographics that correlated with the class label followed by gender and education level, at least using the feature assessment results.

Even though the AD-CR algorithm also improved when it processed 'Cog-subset2', the subset still involves more cognitive features and does not meet our aim of identifying key cognitive items that can predict disease progression, and will not result in time and resource saving for the clinicians. The model derived by the C4.5 algorithm from 'Cog-subset4' can achieve better performance with just six cognitive items covering four out of six of the DSM-5 cognitive domains which are learning and memory, language, executive function, and perceptual motor function. From the 5,923 positive instances that were supposed to have AD progression, the AD-CR algorithm was able to correctly predict 5,371 instances to have AD progression (true positive) and misclassified 552 instances as no progression (false negative), when these instances actually had AD progression.

Overall, models produced from 'Cog-subset4' with additional demographic features by the machine learning algorithms generate good performance and save time by only requiring assessment in 6 activities. While the AD-CR algorithm may not have produced the best results across all measures, its overall results are above 90%, falling within the performance range of other ADAS-cog validation studies (Nogueira et al., 2018; Yang et al., 2019) and required the use of fewer features, thus it can be classified as a good performing model.

We investigated the predictive models generated by the classification algorithms from two subsets of participants for 36 months from the baseline diagnosis. Figure 5.6a depicts the accuracy, sensitivity, and specificity rates derived by the considered classification algorithms from the 'Cog-CN-MCI' and 'Cog-MCI-AD' dementia sub-groups, respectively. The left side images in each of these two figures represent the performance metrics generated by the classification algorithm from the cognitive items without demographics, while the right side images represent the models derived with demographics.

The performance metrics' rates show that the AD-CR algorithm is superior to the remaining algorithms, at least for models derived from the 'Cog-CN-MCI', as it generated models with the best accuracy and specificity rates. To be exact, the AD-CR algorithm generated models from the 'Cog-CN-MCI' dataset with 96.06% accuracy and 98.01% specificity. The results were (21.06%, 20.11%, 15.95%, 16.61%, 20.30%, 21.60%, 8.48%) and (12.50%, 25.36%, 14.50%, 3.73%, 6.20%, 12.39%, 19.90%) higher in accuracy and specificity rates than those of the LR, MLP, SMO, KNN, NB, Ridor, and Nnge algorithms, respectively.

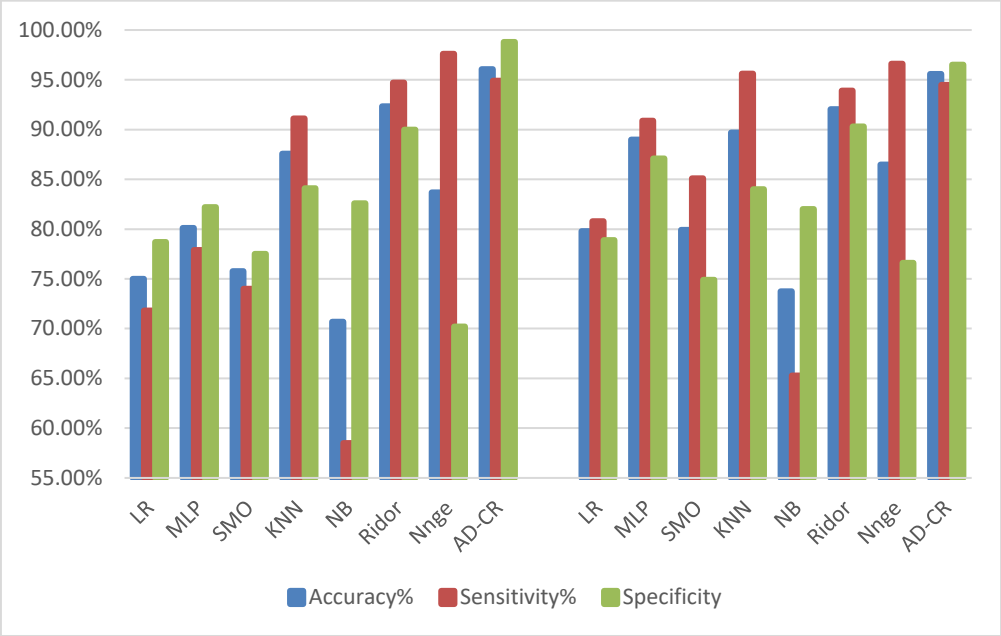

Figure 5.6a: Performance of the Classification Methods against CN-MCI Cognitive Groups with and without Demographics, Respectively

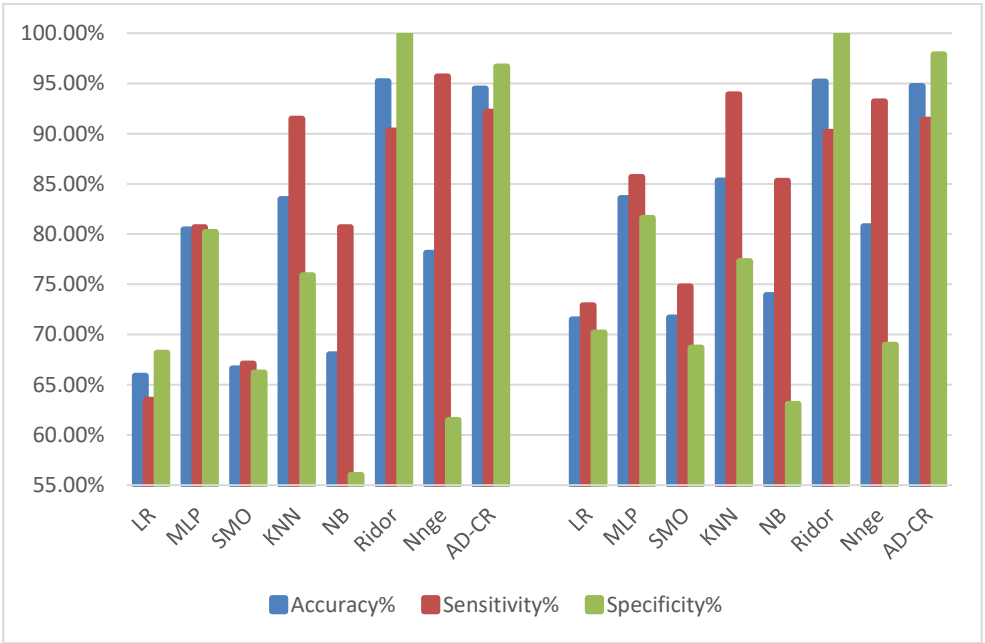

Figure 5.6b: Performance of the Classification Methods against MCI-AD Cognitive Groups with and without Demographics Respectively

The sensitivity rate, which denotes the proportion of visits that showed progression, is high for the models derived by the AD-CR algorithm from the 'Cog-CN-MCI' dataset at 94.90%; the Nnge algorithm alone produced models with slightly higher sensitivity than AD-CR. For the models produced, at least from the 'Cog-MCI-AD' dataset, rule-based algorithms particularly Ridor and ours produced better performance than the statistical, neural network, SVM, and probabilistic-based algorithms. Ridor produced predictive models with 95.23% accuracy, 1.25% higher than the AD-CR algorithm. However, the sensitivity rate of the AD-CR's algorithm models derived from the

‘Cog-MCI-AD’ dataset was 0.75% higher than Ridor making AD-CR better at detecting cases with positive class labels which are potentially progressing from the MCI stage to dementia (AD).

The highest sensitivity rate of 95.70% was produced from the ‘Cog-MCI-AD’ dataset when associated with Nnge’s model; however, the specificity and accuracy rates of this algorithm are unacceptable at 61.50% and 78.14%, respectively. Despite the Nnge algorithm being able to detect positive cases correctly it has high numbers of false positives. To be specific, the Nnge algorithm misclassified 1,182 instances from 3068 to ‘AD progression’, which is relatively high. The AD-CR algorithm had only 102 cases of false positives for the same group of participants (MCI-AD).

We further investigated the impact when demographic attributes are coupled with the cognitive items of the special groups; the results show little to no impact (see right side images of Figure 5.6a and 5.6b. Possible reasons are that the age ranges within these groups are limited, and the number of instances in these datasets is small when compared with the original cognitive dataset.

We want to evaluate whether the results represented by the evaluation metrics (see Figure 5.6a) may deviate. Specifically, if the accuracy, sensitivity, and specificity results, which were obtained based on the mean evaluation, are real or due to statistical chance. This assumption can be achieved using the t-test, so we test the data sample of the CN-MCI dementia sub-group. Initially, and to determine whether these populations have equal variances, we conducted an F-test with alpha significant level, i.e. $\alpha = 0.05$, to identify the ideal t-test type. The results showed that there was a significant difference in the population variances between metrics, i.e., accuracy and sensitivity rates. as the two-tailed p-value was not large enough. The large difference explains the significance in the variance between the two populations of these metrics. We repeated the same F-test for pairs of the sensitivity + specificity, and accuracy + specificity populations, and the results were consistent and indicated that the two-tailed p value was not large enough, and hence there is a significant difference in population variances between sensitivity + specificity and accuracy + specificity, respectively.

The results of the variance analysis led us to conduct a t-test for the considered metrics. The results of the t-test using the critical value show that the one-tailed p-value is larger than 0.05, so we cannot support the alternate hypothesis that the means of the metrics’ results have been created by statistical chance. Since the test statistics obtained are smaller than the critical value, there is no statistical evidence that these metrics’ results will deviate when different samples are used. We repeated the t-test on the other pairs of evaluation metrics’ results shown in the same figure and the results were consistent.

Lastly, we show top rules produced by the AD-CR algorithm from the ‘Cog-subset1’, ‘Cog-MCI-AD’, and ‘Cog-CN-MCI’ datasets in Table 5.7. The AD-CR algorithm generated 31 rules from the

‘Cog-subset1’ dataset including the default class Rule 7 (displayed at the top of Table 5.7). The number of rules derived by the models of AD-CR when 2% minimum support is used is smaller than those of the other rule-based classifiers (Ridor, Nnge). For instance, Ridor and Nnge derived 57 and 2,772 rules, respectively from the ‘Cog-subset1’ making the Nnge an unsuitable algorithm for AD progression as the number of rules is large thus must be controlled by clinicians.

The proposed algorithm was able to offer controllable-sized classification models that can be exploited during the screening process of dementia. The model of the proposed algorithm is equipped with interpretable classification rules that can easily be understood by clinicians and novice users. These models offer the patient, and their family answers as to which components have triggered dementia progression. This is aligned with the patient’s right outlined in the General Data Protection Regulation (GDPR), particularly the section on decision-making using automated algorithmic methods and the “right for an explanation” (GDRP, n.d.; Goodman & Flaxman, 2016). The GDRP requires that for any data collected from a subject (the individual undergoing medical screening) in an automated decision process (the screening process using the machine learning) that the subject should have the right to be given the rationale behind the decision-making process. Consequently, having an intelligent and easy to understand screening system for the patients and their family members, besides clinicians, is valuable. The system can explain to the different stakeholders useful information answering questions such as:

- What are the cognitive features that relate to the progression of AD?
- What are the functional features that relate to the progression of AD?
- Why does the output of the screening show no progression or potential progression?
- Why is the patient being screened for AD, MCI, or CN?
- What further assessment can be made based on the outcome?

In addition, the AD-CR rules can be used by the clinicians during the clinical session to identify factors than may impact the disease or the disease advancement.

Table 5.7: Sample Cognitive Rules Derived by the AD-CR Algorithm from the Cognitive Data Subsets

The AD-CR rules derived from the 'Cog-subset1' (support, confidence)
1. (241, 0.96) Label = 0 when IDEATIONAL in (0.5, 3.4028235E38) , WORDRECOG in (-3.4028235E38, 6.5) , DELAYWORD in (-3.4028235E38, 8.5)
2. (243, 0.98) Label = 1 when DELAYWORD in (9.5, 3.4028235E38) , WORDRECOG in (3.5, 5.5) , NAMING in (-3.4028235E38, 0.5) , WORDFIND in (-3.4028235E38, 0.5) , NUMBERCANCEL in (-3.4028235E38, 2.5) , ORIENT in (0.5, 4.5) , COMMAND in (-3.4028235E38, 0.5)
3. (566, 0.99) Label = 0 when WORDRECALL in (1.165, 2.5) , DELAYWORD in (0.5, 2.5) , WORDRECOG in (-3.4028235E38, 9.5)
4. (263, 0.96) Label = 0 when WORDFIND in (0.5, 1.5) , WORDRECOG in (-3.4028235E38, 4.5) , WORDRECALL in (-3.4028235E38, 5.5)
5. (408, 0.79) Label = 0 when WORDRECALL in (0.165, 1.5) , DELAYWORD in (-3.4028235E38, 5.5)
6. (261, 0.94) Label = 0 when SPOKENLG in (0.5, 3.4028235E38) , WORDRECOG in (2.5, 3.4028235E38) , DELAYWORD in (7.5, 3.4028235E38) , WORDFIND in (-3.4028235E38, 3.4028235E38) , LANGUAGE in (-3.4028235E38, 3.4028235E38) , RMBRTESTINST in (-3.4028235E38, 3.4028235E38) , IDEATIONAL in (-3.4028235E38, 3.4028235E38) , COMMAND in (-3.4028235E38, 3.4028235E38)
7. (253, 0.97) Label = 1 when WORDRECALL in (5.835, 6.165) , WORDRECOG in (4.5, 5.5) , COMMAND in (-3.4028235E38, 0.5) , DELAYWORD in (6.5, 3.4028235E38) , WORDFIND in (-3.4028235E38, 0.5) , NAMING in (-3.4028235E38, 0.5) , NUMBERCANCEL in (-3.4028235E38, 2.5) , SPOKENLG in (-3.4028235E38, 3.4028235E38) , LANGUAGE in (-3.4028235E38, 3.4028235E38) , RMBRTESTINST in (-3.4028235E38, 3.4028235E38) , IDEATIONAL in (-3.4028235E38, 3.4028235E38)
The AD-CR rules derived from the 'Cog-CN-MCI' without demographic attributes (support, confidence)
1. (1016, 1.00) Label = 1 when CONSTRUCT in (3.5E-6, 0.9999995)
2. (176, 1.00) Label = 1 when LANGUAGE in (4.755E-4, 0.994774)
3. (122, 1.00) Label = 1 when ORIENT in (5.0E-7, 0.998617)
4. (77, 1.00) Label = 1 when NUMBERCANCEL in (1.004509, 1.988689)
5. (580, 0.99) Label = 0 when DELAYWORD in (0.971643, 1.0146215) or DELAYWORD in (1.9801745, 2.006372)
6. (156, 0.99) Label = 0 when DELAYWORD in (-3.4028235E38, 3.0446) , WORDRECOG in (1.5, 3.4028235E38) , WORDRECALL in (1.835, 3.835)
7. (77, 0.97) Label = 0 when WORDRECALL in (-3.4028235E38, 1.432405) , WORDFIND in (-3.4028235E38, 0.5)
The AD-CR rules derived from the 'Cog-MCI-AD' without demographic attributes (support, confidence)
1. (1284, 1.00) Label = 1 when NAMING in (1.0E-6, 0.9999325)
2. (666, 1.00) Label = 1 when CONSTRUCT in (7.0E-6, 0.9999985)
3. (318, 1.00) Label = 1 when ORIENT in (6.9E-5, 0.9998905)
4. (181, 1.00) Label = 1 when COMMAND in (0.0055745, 0.999991)
5. (350, 0.99) Label = 0 when WORDRECOG in (0.9589805, 1.0379485) , WORDRECALL in (-3.4028235E38, 5.5)
6. (179, 1.00) Label = 0 when WORDRECOG in (-3.4028235E38, 0.03653) , WORDRECALL in (1.5, 5.5) , COMMAND in (-3.4028235E38, 2.5) , NUMBERCANCEL in (-3.4028235E38, 1.5) , DELAYWORD in (-3.4028235E38, 6.5)
7. (165, 0.99) Label = 0 when DELAYWORD in (4.911518, 5.005151) , WORDRECALL in (4.165, 7.5) , SPOKENLG in (-3.4028235E38, 2.5)
8. (189, 0.95) Label = 0 when DELAYWORD in (-3.4028235E38, 1.5)
9. (165, 0.97) Label = 0 when DELAYWORD in (2.967585, 3.0101662) , CONSTRUCT in (-3.4028235E38, 1.3679985) , WORDRECALL in (2.835, 3.4028235E38) , LANGUAGE in (-3.4028235E38, 0.5)
10. (155, 0.95) Label = 0 when DELAYWORD in (4.911518, 5.005151) , SPOKENLG in (-3.4028235E38, 0.5) , COMMAND in (-3.4028235E38, 3.0) , WORDRECALL in (1.665, 3.4028235E38)
11. (144, 0.90) Label = 0 when WORDRECOG in (7.981304, 8.0100975)
12. (120, 0.87) Label = 0 when WORDRECALL in (4.991171, 5.006675) , LANGUAGE in (-3.4028235E38, 1.5) , SPOKENLG in (-3.4028235E38, 2.5)

5.6.2 Functional Classification Results Analysis

Similar to the models produced by the considered classification algorithms on the cognitive subsets, the AD-CR algorithm showed dominancy when processing the distinct functional subsets, generating the best performing classification model overall, at least from three functional subsets. For comparison against the complete functional item set of ('Func-subset1') excluding demographic features (left side of Table 5.7), the performance results of the models produced from 'Func-subset4' by the AD-CR algorithm, provide the closest predictive accuracy, sensitivity, and specificity measures, at 75.88%, 59.0%, and 91.20%, respectively. However, this is due to 'Func-subset4' having only one less feature.

While the performance may be fair in terms of accuracy for the AD-CR algorithm, it does not serve our purpose of identifying key functional items to identify disease progression and save assessment

Table 5.8: Performance of the Classification Methods from Different Subsets of the Functional Items

Subset	Algorithm	Excluding Demographics Features			Including Demographics Features		
		Accuracy %	Sensitivity %	Specificity %	Accuracy %	Sensitivity %	Specificity %
Func-subset1 (baseline)	Logistic	71.29	61.5	80.1	77.22	77.7	76.8
	MLP	72.79	60.9	83.6	80.33	81.1	79.7
	SMO	65.75	49.2	80.8	73.24	76.6	70.2
	KNN	75.33	64.6	85.1	88.78	93.8	84.2
	NB	68.92	63.7	73.7	73.23	77.4	69.4
	Ridor	73.85	55.8	90.3	87.46	86.4	88.5
	Nnge	74.32	67.1	81	89.52	89.3	89.7
	AD-CR	76.07	58.6	92	89.74	88	91.4
Func-subset2	Logistic	66.29	53.1	78.3	68.82	72.3	65.6
	MLP	66.06	46.4	83.9	76.84	75.9	77.7
	SMO	62.16	39.4	82.9	66.49	79.2	55
	KNN	67.41	52.4	81	87.53	92.6	83
	NB	51.64	92.1	14.9	65.12	85.7	46.4
	Ridor	66.96	50.3	82.1	85.6	83.8	87.3
	Nnge	57.89	57.2	58.5	89.24	89.4	89.1
	AD-CR	66.85	50	82.2	86.59	85.7	87.4
Func-subset3	Logistic	64.71	52	76.2	70.78	72.6	69.2
	MLP	69.49	60.2	77.9	79.14	79.3	79
	SMO	62.1	46.9	76	70.5	75.7	65.9
	KNN	74.21	62.3	85.1	88.7	93.3	84.3
	NB	63.59	53.3	73	68.19	62.4	73.5
	Ridor	71.95	54.1	88.2	86.89	80.6	92.6
	Nnge	70.73	60.10%	80.4	88.89	88.9	88.9
	AD-CR	73.03	58.2	86.6	88.75	87	90.3
Func-subset4	Logistic	70.56	60.7	79.5	72.1	73.6	70.7
	MLP	71.11	60.9	80.4	80.11	77.9	82.2
	SMO	64.15	45.8	10	71.91	75	69.1
	KNN	75.4	64.5	85.3	88.9	94	84.3
	NB	65.5	56.4	73.7	69.52	63.8	74.8
	Ridor	73.9	56.2	90	86.64	85	88.2
	Nnge	73.97	66.9	80.4	89.24	89.4	89.1
	AD-CR	75.88	59	91.2	89.88	88.3	91.3
Func-subset5	Logistic	68.07	59.4	75.9	70.25	72.3	68.4
	MLP	67.22	56.2	77.2	77.06	78.1	76.1
	SMO	59.2	37.9	78.6	70.48	76.1	65.4
	KNN	69.04	58.1	79	88.13	92.2	84.4
	NB	62.09	46.7	76.1	69.98	67.4	72.3
	Ridor	68.51	57.7	78.3	86.16	81.1	90.7
	Nnge	57.98	56	59.8	87.22	86.9	87.5
	AD-CR	68.22	58.8	76.8	87.57%	86.7	88.3

time and cost. The second-best performing models were generated from 'Func-ubset3' using the AD-CR algorithm, consisting of six functional items, which are 'remembering an occasion', 'shopping', 'finance', 'completing forms', 'travelling' and 'current events'. With four fewer items than the 'Func-subset1', the classification algorithm (AD-CR) did not sacrifice much in terms of performance, with 73.03% accuracy, 58.20% sensitivity, and 86.60% specificity. While these results are close, they are not ideal in terms of model performance to be applied in a medical setting, in particular for sensitivity where out of the 8,832 positive instances that were supposed to have AD progression, the AD-CR was able to correctly predict 5,1637 instances to have AD progression (true

positive) and misclassified 3,695 instances as no progression (false negatives), when these instances were in fact having AD progression.

Based on the results generated from the functional item subsets by the considered classification algorithms, on the datasets used, it seems that evaluating solely the functional activities may not provide the clinicians with a comprehensive picture of all areas needed to predict the disease progression and hence measuring additional features like demographics and cognitive items using other cognitive assessments is useful for a detailed screening.

The specificity measure also appears to perform much better than the sensitivity measure across most of the considered classification algorithms when processing functional subsets, with differences up to 32.20% in the case of processing 'Func-subset2', and 'Func-subset4' by the AD-CR algorithm. This could be due to the data sampling technique performed during the data balancing process, whereby synthetic instances of the minority class (progression) were generated to minimise the gap to the majority class (no progression) within the dataset. As a result of this, the classification algorithms may encounter hardship in their ability to identify significant patterns, which ultimately caused them to predict a higher number of false positives and false negatives.

Overall, the specificity rate produced by the AD-CR algorithm was good for all the functional subsets reaching more than 92.00% when processing 'Func-subset2', and 'Func-subset4', except models produced from 'Func-subset5' ('remembering an occasion', 'shopping', 'finance'). The same pattern was also noticed by the models derived by the remaining classification algorithm concluding that more functional, demographic, and possibly cognitive activities are required to assess the patient's probable disease progression during clinical assessments.

To improve the AD progression models' performance, we experimented by including demographic features (age, gender, level of education, race category, marital status), shown on the right side of Table 5.8, with the functional subsets. Immediately, the overall results improved significantly across the subsets and algorithms, particularly for the AD-CR algorithm where the model performance measures were all above 86.59%, meeting the benchmark of the study conducted by Teng et al. (2010), and others. The gap that previously caused the large difference between the sensitivity and specificity measures has also been substantially minimised and the differences now only range from 1.60%–3.40.00% for the models generated by the proposed algorithm, suggesting the classification models are able to produce better results when demographic features are included.

While the AD-CR algorithm was able to produce good results across the derived subsets, it is remarkable to note that when it processed 'Func-subset5' which only consists of three functional items, the algorithm was still able to generate close to identical results of the baseline subset ('Func-

subset1') with a difference in sensitivity of 1.30%. These results, if limited, show that functional activities related to learning and memory, executive function, complex attention, and perceptual motor function when assessed with age, gender, and education level are the important items for dementia advancement, at least when using the dataset and the classification methods we considered. The results also show that choosing common features across the top-ranked features of the feature selection methods considered is impactful. By checking the confusion matrix produced by the AD-CR algorithm when processing the 'Func-subset5', out of the 8,832 positive instances that were supposed to have AD progression, the algorithm was able to correctly predict 7,660 instances to have AD progression (true positive) and misclassified 1,172 instances as no progression (false negatives), when these instances were in fact having AD progression.

The rule-based classification algorithms (Ridor, Nnge), KNN, and the proposed algorithm generated good performance from the 'Func-subset2' with demographics albeit using a different set of functional items ('game', 'beverage', 'meal', and 'watching TV'). This shows that the criteria used to form 'Func-subset2', which is removing redundant features based on the feature-to-feature correlation matrix, is influential as only dissimilar functional features that have high correlations with the progression (the class label) are kept. With one item more than 'Func-subset5', the AD-CR algorithm was able to generate models from the 'Func-subset2' with demographics with good sensitivity, specificity, and accuracy. The functional items used to derive the classification models of the AD-CR algorithm cover all six of the DSM-5 cognitive domains, which may prove to be a more well-rounded model for clinicians in predicting AD progression.

To ensure a better model performance, it is highly recommended that when using machine learning to assess functional or cognitive activities during the screening or diagnosis of dementia conditions such as AD to include demographic features. Functional activities when assessed with demographic features seem partly effective in detecting the disease progression, at least when a data driven approach is used such as machine learning. This is due to the good sensitivity obtained by the KNN, Nnge, Ridor and AD-CR classification algorithms.

Overall, models generated from the 'Func-subset2' and 'Func-subset5' which comprise four and three activities, respectively, appear to meet the aim of identifying key functional items that can predict AD progression. This is achieved by reducing assessment time without sacrificing too much model performance. However, both subsets use different functional items for building the classification models of the machine learning algorithms. A suggestion for clinicians is that they trial both models in parallel and observe the results for further validation of which model will be more accurate.

A major advantage of having fewer functional or cognitive activities for measuring AD level or detecting AD, is minimising the time taken within a clinical setting for clinicians to perform any diagnostic or screening assessments. Medical assessments of function, cognition, or both can be time-consuming as multiple degenerative domains are assessed such as communication, memory and learning, orientation, and executive functions, among others. Thus, enhancing the time of the screening process of these assessments is crucial due to the resources needed, especially in low social economic nations or during the pandemic time when resources are limited.

We further investigated the models derived by the classification algorithms from the functional items of 'Func-CN-MCI' and 'Cog-MCI-AD' dementia sub-groups over 36 months from the baseline diagnosis. Figures 5.7a and 5.7b show the accuracy, sensitivity, and specificity rates derived by the considered classification algorithms from these two sub-groups with and without considering demographic attributes. The images on the left side in these two figures represent the performance metrics generated by the classification algorithm from the functional items without demographics; the right side images represent the models derived with demographics.

The models produced by the machine learning algorithm from the functional items of the 'Func-CN-MCI' dementia sub-group indicate that the AD-CR algorithm is superior in terms of predictive accuracy and specificity. All the classification algorithms derived predictive models with good sensitivity, specificity, and accuracy rates from the 'Func-CN-MCI' sub-group. In fact, the AD-CR algorithm built predictive systems from the 'Func-CN-MCI' dataset and without considering demographic features produced 97.97% accuracy, 96.60% sensitivity, and 99.30% specificity. When demographic attributes were added to the functional items the predictive models produced by the AD-CR algorithm slightly enhanced by 0.92%, 1.70%, and 0.20% for accuracy, sensitivity, and specificity rates, respectively, showing that demographic features do not greatly impact the 'Func-CN-MCI' sub-group. When analysing the confusion matrix of the classification models of the AD-CR algorithm, we discovered that only 36 instances from 1,057 were misclassified into 'not progressing', minimising the false negatives. However, false positives numbered just 8 out of 1,120 displaying the high ability of the proposed algorithm in distinguishing cases not progressing correctly, at least for participants with a baseline diagnosis of CN.

The models derived from the 'Func-MCI-AD' revealed that rule-based classification algorithms including Ridor and AD-CR are superior in terms of sensitivity, accuracy, and specificity rates. These classification algorithms derived predictive models with measures above 90.00% from the 'Func-MCI-AD'. These rates did not improve when the demographic attributes were added into the 'Func-MCI-AD' by the rule-based classifiers in addition to the Nnge algorithm's model. However, the SVM (SMO), ANN (MLP), statistical (LR), and probabilistic (NB) models were improved when

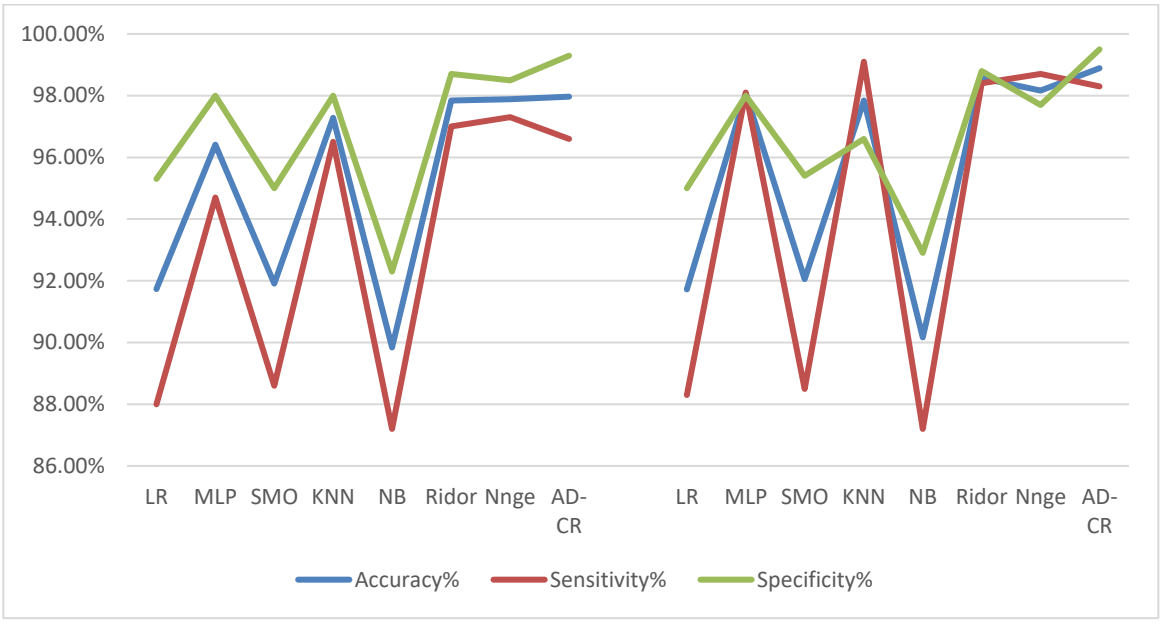

Figure 5.7a: Performance of the Classification Methods against CN-MCI Functional Groups with and without Demographics, Respectively

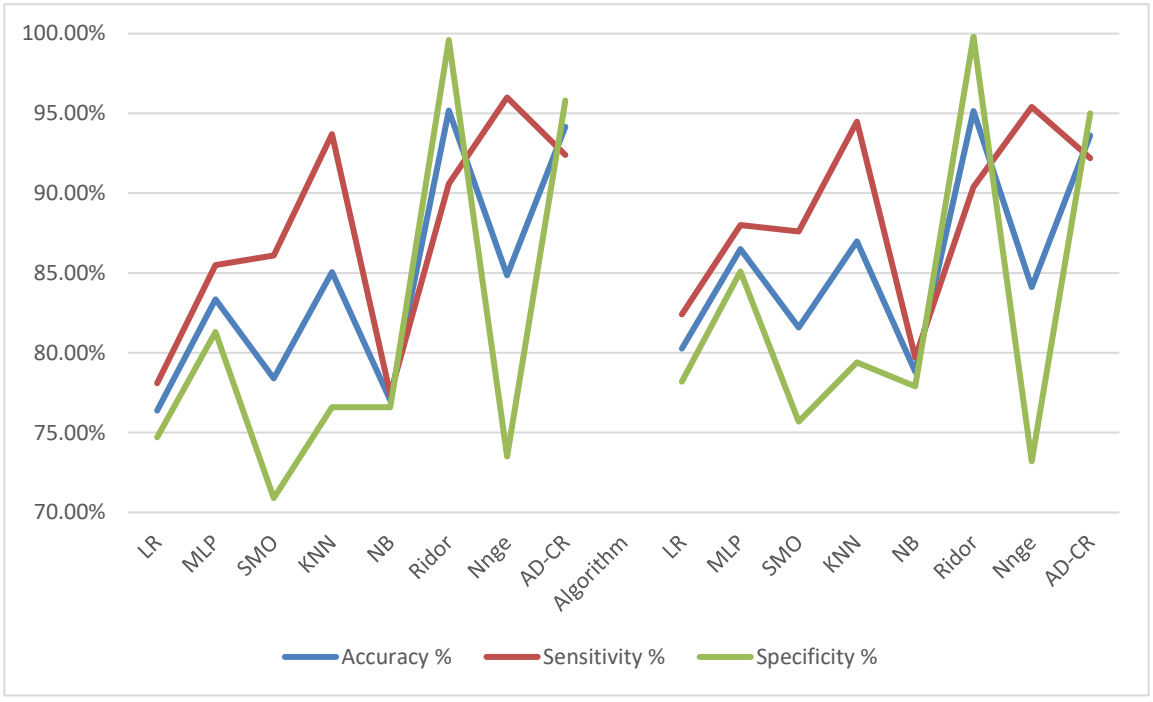

Figure 5.7b: Performance of the Classification Methods against MCI-AD Functional Groups with and without Demographics, Respectively

demographic attributes were considered during the learning phase. Nevertheless, these algorithms' general performance was below the required medical standard when processing the 'Func-MCI-AD' dataset without demographic attributes except MLP, and KNN (at least for sensitivity and accuracy rates).

Table 5.9: Sample Cognitive Rules Derived by the AD-CR Algorithm from the Cognitive Data Subsets

The AD-CR rules derived from the 'Func-subset1' (support, confidence)
1. (284, 0.98) Label = 0 when FAQBEVG in (2.5, 3.4028235E38) , FAQGAME in (0.5, 3.4028235E38) , FAQTRAVL in (2.5, 3.4028235E38)
2. (379, 0.92) Label = 0 when FAQEVENT in (2.5, 3.4028235E38) , FAQFINAN in (2.5, 3.4028235E38) , FAQFORM in (0.5, 3.4028235E38)
3. (46, 1.00) Label = 0 when FAQTV in (2.5, 3.4028235E38) , FAQREM in (2.5, 3.4028235E38) , FAQTRAVL in (1.5, 3.4028235E38) , FAQGAME in (-3.4028235E38, 2.5)
4. (483, 0.93) Label = 1 when FAQREM in (1.5, 2.5) , FAQMEAL in (0.5, 1.5) , FAQSHOP in (0.5, 1.5) , FAQTRAVL in (-3.4028235E38, 1.5) , FAQGAME in (-3.4028235E38, 2.5) , FAQEVENT in (-3.4028235E38, 1.5) , FAQBEVG in (-3.4028235E38, 0.5) , FAQFINAN in (0.5, 3.4028235E38) , FAQTV in (-3.4028235E38, 1.5)
5. (390, 0.92) Label = 1 when FAQFINAN in (1.5, 2.5) , FAQEVENT in (0.5, 1.5) , FAQSHOP in (0.5, 1.5) , FAQBEVG in (-3.4028235E38, 0.5) , FAQGAME in (-3.4028235E38, 2.5) , FAQFORM in (0.5, 3.4028235E38) , FAQTV in (-3.4028235E38, 1.5)
6. (748, 0.60) Label = 0 when FAQFORM in (0.5, 2.5) , FAQTRAVL in (-3.4028235E38, 0.5)
7. (620, 0.68) Label = 0 when FAQBEVG in (0.5, 3.4028235E38)
8. (47, 0.90) Label = 0 when FAQTV in (2.5, 3.4028235E38) , FAQFORM in (1.5, 3.4028235E38) , FAQMEAL in (0.5, 3.4028235E38)
The AD-CR rules derived from the 'Func-CN-MCI' without demographic attributes (support, confidence)
1. (537 , 1.00) Label = 1 when FAQMEAL in (4.5E-6, 0.999633)
2. (460 , 1.00) Label = 1 when FAQFORM in (5.82E-4, 0.99904597) or FAQREM in (1.00161, 1.9998425) or FAQREM in (0.2227545, 0.9999925)
3. (1053 , 0.98) Label = 0 when FAQREM in (-3.4028235E38, 0.5) , FAQMEAL in (-3.4028235E38, 0.5) , FAQTRAVL in (-3.4028235E38, 3.4028235E38) , FAQTV in (-3.4028235E38, 3.4028235E38) , FAQEVENT in (-3.4028235E38, 3.4028235E38) , FAQBEVG in (-3.4028235E38, 3.4028235E38) , FAQGAME in (-3.4028235E38, 3.4028235E38) , FAQSHOP in (-3.4028235E38, 3.4028235E38) , FAQFORM in (-3.4028235E38, 3.4028235E38) , FAQFINAN in (-3.4028235E38, 3.4028235E38)
4. (52 , 0.93) Label = 0 when FAQREM in (-3.4028235E38, 1.5) , FAQGAME in (-3.4028235E38, 0.0398595)
The AD-CR rules derived from the 'Func-MCI-AD' without demographic attributes (support, confidence)
1. (1509 , 1.00) Label = 1 when FAQMEAL in (2.5E-6, 0.999997) or FAQMEAL in (1.0000005, 1.999998)
2. (210 , 1.00) Label = 1 when FAQREM in (1.0003265, 1.9994905)
3. (90 , 1.00) Label = 1 when FAQTV in (1.27E-4, 0.98143554)
4. (957 , 0.98) Label = 0 when FAQREM in (-3.4028235E38, 0.0289375) , FAQSHOP in (-3.4028235E38, 1.5) , FAQMEAL in (-3.4028235E38, 1.5) , FAQTRAVL in (-3.4028235E38, 3.4028235E38) , FAQTV in (-3.4028235E38, 3.4028235E38) , FAQEVENT in (-3.4028235E38, 3.4028235E38) , FAQBEVG in (-3.4028235E38, 3.4028235E38) , FAQGAME in (-3.4028235E38, 3.4028235E38) , FAQFORM in (-3.4028235E38, 3.4028235E38) , FAQFINAN in (-3.4028235E38, 3.4028235E38)
5. (321 , 0.98) Label = 0 when FAQFORM in (-3.4028235E38, 0.0956455) , FAQFINAN in (-3.4028235E38, 0.5) , FAQTRAVL in (-3.4028235E38, 2.5) , FAQREM in (-3.4028235E38, 2.5) , FAQTV in (-3.4028235E38, 3.4028235E38) , FAQEVENT in (-3.4028235E38, 3.4028235E38) , FAQMEAL in (-3.4028235E38, 3.4028235E38) , FAQBEVG in (-3.4028235E38, 3.4028235E38) , FAQGAME in (-3.4028235E38, 3.4028235E38) , FAQSHOP in (-3.4028235E38, 3.4028235E38)
6. (103 , 0.87) Label = 0 when FAQMEAL in (2.999885, 3.4028235E38) , FAQTRAVL in (-3.4028235E38, 3.4028235E38) , FAQFORM in (-3.4028235E38, 3.4028235E38)
7. (87 , 0.89) Label = 0 when FAQFORM in (1.9882281, 2.0041676) , FAQEVENT in (-3.4028235E38, 0.192287) , FAQBEVG in (-3.4028235E38, 1.5) , FAQREM in (-3.4028235E38, 2.5)
8. (88 , 0.87) Label = 0 when FAQFINAN in (1.9883344, 2.0217185) , FAQMEAL in (-3.4028235E38, 1.5) , FAQBEVG in (-3.4028235E38, 1.5) , FAQREM in (0.5, 3.4028235E38)
9. (88 , 0.81) Label = 0 when FAQSHOP in (1.9508796, 2.0022464) , FAQMEAL in (-3.4028235E38, 2.2250066) , FAQTRAVL in (0.5, 3.4028235E38)

Lastly, we show the top-ranked rules produced by the proposed AD-CR algorithm from 'Func-subset1', 'Func-MCI-AD', and 'Func-CN-MCI' in Table 5.9. The number of rules produced by the proposed algorithm when the min-support threshold is set to 2% is smaller than the other rule-based classifier. For example, the AD-CR algorithm was able to produce just 9 rules from the Func-subset1' dataset whereas Ridor generated a model with 77 rules which is difficult for the clinicians to manage during clinical assessments. The fact that the AD-CR algorithm generated models with far fewer rules, and with higher predictive accuracy than Ridor, reveals the effectiveness of such a classification algorithm not only in detecting progression, but also in providing clinicians valuable, easy-to-understand information. More importantly, a digital information sheet can be used by the clinician when conducting clinical sessions to pinpoint crucial functional factors than may impact

the progression. Such information can be articulated to the patients and their families so that they can understand the reasons behind the screening decision.

5.7 Chapter Summary

This chapter discussed the experiments conducted related to features assessment, classification, and provided in-depth results analysis and discussion from two neuropsychological perspectives: cognitive and functional. We evaluated the cognitive and functional items for multiple subsets of real data obtained by thorough analysis using the results obtained from three feature selection methods. The ADAS13-Cog and FAQ complete activities were used as a baseline during the experiments of the dissimilar classification algorithms including ANN, SVM, statistical, probabilistic, instance-based learning, and rule-based induction.

We showed the empirical analysis on cognitive and functional items and the predictive models derived from these subsets of items by the considered classification algorithms. The bases of comparison of the classification algorithms are sensitivity, accuracy, and specificity besides classification rules. We also investigated empirically specific dementia sub-groups of cognitive and functional items including participants who progressed from CN to MCI and others who progressed from MCI to AD using the feature selection and classification algorithms. The analysis aimed to determine whether neuropsychological items change from one stage of dementia to another.

The results derived by the feature selection and classification algorithms, including ours, pinpointed that the proposed AD-CR algorithm was able to build classification systems for AD progression that are highly competitive with the considered classification algorithms and from most of the cognitive and functional subsets. In addition, we were able to identify influential cognitive and functional subsets that are uniquely associated with certain dementia stages. More importantly, we showed that the AD-CR algorithm was able to derive models with easy-to-interpret rules which can be exploited by clinicians during the process of dementia assessment and when using neuropsychological tests. These models align with the GDPR standard and the patients' right to know the reason behind the diagnosis, especially when using automated decision-making systems as described.

This chapter showed that the proposed architecture, equipped by the AD-CR classification algorithm, derived competitive classification results for detecting AD advancement. In the next chapter, conclusions, future works, and the study limitations are discussed.

Chapter Six

This chapter discusses the conclusions of the thesis including the response to the research questions raised in Chapter 1. In addition, the major findings are highlighted and how these are useful for the medical community. Lastly, the chapter sheds light on the thesis's limitations, and possible future works. Some of this chapter's content is under consideration for publication in the Journal of Biomedical Informatics, and others have been disseminated in the Journal of Health and Technology, the Journal of Behavioural and Healthcare Research, and the Journal of Intelligent Decision Technologies.

Conclusions and Implications

The primary dementia condition is AD, which is typically diagnosed by a specialised physician or clinician based on a set of criteria including factors such as: cognitive decline reported by a patient or their family members, a patient's or their family's medical history, and the scores of neuropsychological assessments that are designed to measure the patient's cognitive abilities. One of the challenging problems related to dementia is determining the points when the disease advances and the cognitive items that may trigger such an advancement. This thesis investigated this problem based on real data related to neuropsychological assessments' items by using machine learning to identify functional and cognitive items that influence the progression of AD—this to produce fast, accurate, and accessible classification models.

The thesis proposes a Machine Learning Architecture called Alzheimer's Disease Progression (MLA-ADP) that has been implemented in a cloud-based environment to make operation more sustainable and cost effective. The MLA-ADP contains an Alzheimer's Disease Class Rules (AD-CR) algorithm that derives predictive models that can be exploited by clinicians during the dementia screening process. The AD-CR algorithm is used to predict the progression of the disease for individuals undertaking a neuropsychological assessment based on intelligent models consisting of 'If-Then' rules that have been derived from real data (cases and controls). These classification models are easy to interpret by the clinicians as they indicate associations among assessed items. They can thus be utilised as a digital information sheet to guide clinicians in the diagnostic-related decisions such as disease advancement. More specifically, the models are aligned with the patient's right outlined in the GDPR since they offer stakeholders information about items that may advance AD.

Empirical results obtained by dissimilar classification and feature selection techniques against combinations of datasets related to cognitive and functional items from the ADNI repository revealed that the AD-CR algorithm produced classification models that are competitive in terms of accuracy, specificity, and sensitivity rates. Moreover, models derived by the classification algorithms from different dementia sub-groups (CN-MCI or MCI-AD) reveal that the AD-CR algorithm's models are highly competitive to ANN, SVM, statistical, rule induction, and probabilistic approaches.

The results demonstrated that both cognitive and functional assessments are needed to predict AD progression; using just functional elements, irrespective of the cognitive elements, do not provide a clear distinction between progression or no progression of AD. Thus, functional elements should be used alongside cognitive items as the criteria to diagnose disease progression in patients. The results suggest that integrating influential cognitive elements and functions of IADLs in a single composite assessment may, a) improve the prediction of AD progression, and to b) fulfil the DSM-5 framework's diagnostic criteria for Major/Minor neurological disorder.

After carefully analysing the classification models derived, there were some associations, albeit low, between cognitive items and functions of IADL that can be captured during the progression of AD. For example, orientation and delayed word recall, which are related to memory and learning cognitive domains, have low correlations with functional activities. In addition, there are associations between cognitive elements, for example, word recall, word delay, word-finding, and language comprehension. These correlations can be utilised within neuropsychological methods to retain fewer cognitive items with the least overlapping—beneficial for clinicians conducting clinical assessments for dementia progression.

Analysing the classification models in regard to the impact of demographics attributes revealed that age contributes the most critical factor followed by gender and then education level. When age and education level were combined with cognitive or functional items, the AD progression models derived by the machine learning algorithms improved. The results showed that younger individuals possibly will show signs of disease progression quickly since cognitive decline and the initiation of MCI often unfold in the early stages of the disease.

In assessing dementia sub-groups (CN to MCI or MCI to AD) based on cognitive items, the results obtained by the feature selection methods were similar to those obtained from the general population (ADNI-Merge-ADAS), i.e. regardless of the dementia-subgroups. The results indicate that delayed word recall, word recognition, and word recall maintain good correlation with the diagnostic class. Further, results on the 'Cog-CN-MCI' sub-group revealed that ideational praxis, naming of objects, and commands cognitive features make a good cluster since these have low overlapping with other

cognitive elements. However, the results obtained on the 'Cog-MCI-AD' sub-group reveal that there was overlapping among the cognitive features revealing that it is a difficult task to isolate subjects who remain MCI from those who may progress to early dementia. However, for the functional items of the dementia sub-groups, the results obtained by the feature selection methods were in general similar to those obtained from the general population (ADNI-Merge-FAQ). Specifically, the results on the 'Func-MCI-AD' dataset show that finance, completing forms, shopping, and remembering an occasion activity are significant.

The proposed approach is one of the rare models that has attempted to not only capture the advancement of AD, but also to map responsible items into their degenerative areas as defined in the DSM-5 framework. We identify a few, yet impactful, cognitive and functional items, while maintaining a high DSM-5 domain coverage if possible. The observed findings carry significant practical implications in DSM-5 mapping of degenerative dementia domains. For instance, in the cognitive items (word recall, delayed word recall, and word recognition) demonstrating adequately reliable results when processed by the AD-CR algorithm, and only two cognitive domains featured relevance: learning and memory, as well as language. Regarding the functional items, remembering, shopping, and finance appeared to cover four domains of the DSM-5 concerning degenerative dementia: learning and memory, executive function, complex attention, and perceptual motor function. Note that the three functional features need to be considered in conjunction with demographic attributes for the machine learning algorithms to generate reliable results.

There are a few cognitive tests available online for detecting dementia-related conditions however, the availability of AD progression systems in a digital platform such as mobile to detect AD progression is scarce. The proposed approach can improve healthcare accessibility around the world; at a time when healthcare costs have skyrocketed, especially during the COVID-19 global pandemic, accessible, and validated medical screening tools such as ours are imperative. Face-to-face interactions have dwindled because of the pandemic, so virtual medical assessments have proliferated globally. Virtual-assisted clinical evaluations have proved effective, as well as being efficient at serving those in highest need. The availability of these tools on mobile applications with certified translations make them accessible to the clinicians in the neediest populations worldwide. Clinicians in the developing world, or even in industrialised nations, can readily access validated tools to help screen people with AD and MCI advancements so as to provide them with health care access to improve their daily lives.

Providing the clinicians with a simple to use, accurate, and validated model using a few functional and cognitive items can indeed be useful for AD assessments as it provides rapid access, a secure environment, and accurate results. This model can serve clinicians with a knowledge-based system

(when models are used) as described earlier, providing the clinicians with key information. Since the AD-CR provides a concise set of rules with just a few cognitive items, the clinicians could utilise some of the rules in Cognitive Behavioural Therapy (CBT) interventions for their patients, especially in the early dementia stages. Most patients with MCI or AD (mild/severe) are impacted at some time during the disease progression by neuropsychiatric signs. Understanding the patient's cognitive, emotional, and functional status can assist the therapists in understanding the patients thus to form individualised treatment plans to make them feel better and to slightly adjust their behaviour and daily functions so they can become their own therapists.

While there was no single cognitive or functional item that stood out, our research was able to identify sets of key cognitive and functional items within neuropsychological assessments. Our research discovers few effective cognitive items for AD progression that cover learning and memory, and language. On the other hand, and for functional items, demographic features must be included when using machine learning to build models for AD progression. Cognitive items appear to be more impactful seeing that when they are processed the AD-CR algorithm generates models within the standard medical research. This can partly be attributed to that patient with dementia risks exhibit cognitive decline at an early stage (pre-dementia such as MCI and mild dementia) and before they experience deficiencies in daily functioning.

Models derived by machine learning techniques such as AD-CR can reduce time, especially when the neuropsychological assessment for dementia is lengthy—this can be a burden for the patients and possibly cause a loss of concentration and fatigue. Additionally, these models can be embedded in a mobile platform to ease the interaction with the patient in a clinical setting, especially in the era of pandemics. These benefits are aligned with the European Task Force which concentrated on a few cognitive and functional activities that could potentially trigger the progression of AD in the early stages.

One of the limitations of this research project is transforming the cognitive tasks into a mobile-based environment; automating the process of scoring is challenging and may require decomposing the tasks such as Constructional Praxis, Naming, etc. into subtasks and then hard coding each within the digital platform. In addition, cognitive tasks such as Ideational Praxis may not be suitable to decompose, model, and code within a digital platform such as mobile. Another possible challenge in the data is some participants have not taken the neuropsychological assessments during medical visits which creates a challenge, especially during data modelling to record the progression of the disease, if any, between two subsequent clinical visits. More importantly, data in which a participant 'X' has 15 medical visits where he/she undertakes cognitive assessment 'A' from which he/she undertakes assessment 'B' in just 8 visits may be obtained. So, when the items of both assessments

are merged many fields will contain null values—this would demand further pre-pre-processing. Lastly, one of the challenges observed in the datasets we considered is the imbalanced ratio between the positive and the negative class labels. There were more instances associated with negative class labels in which without designing a proper treatment, the results derived would be biased towards the majority class. This class imbalance problem was clearly observed when dealing with the dementia sub-categories datasets.

One of the challenging tasks in detecting progression for the different dementia conditions such as AD involves considering the time elapsing between two or more consecutive medical assessments for every participant. Since AD is a progressive disease in which patients usually exhibit a decline in cognitive abilities, it is imperative for the diagnosticians and clinicians to be able to approximate the likely deterioration in terms of time based on cognitive indicators. Discovery of the disease's progression is difficult especially in the early stages of dementia, or its precursors such as MCI, since early dementia traits can be vague, and people wrongly assume that they are exhibiting normal signs of ageing.

The problem becomes more difficult when the clinicians aim to identify the ideal period to predict the dementia progression as this requires collecting data related to the individual's medical history and their cognitive scores for each medical visit over a specific time window, i.e. 2 years, 3 years, 5 years, etc. A longitudinal data analysis using computational intelligence is then conducted to examine any changes in the collected attributes' values and their impact on the diagnostic class. In other words, questions such as, 'Has the cognitive score increased or decreased in certain cognitive activities or areas?' will result in a target class change (the dementia diagnosis). In addition, features from different periods can be contrasted in terms of their change values when the target class changes and to determine the best features set. In this thesis, we have investigated the feature-class changes for three years and for two types of dementia sub-groups—adding more time periods could be a future research direction that may reveal more interesting information on how cognitive features correlate with the diagnostic class for all visits of patients within multiple time windows. Cognitive elements are recorded during each episode using neuropsychological assessments such as the ones we considered in this thesis (ADAS-Cog, FAQ), so machine learning based on longitudinal data analysis can provide potential solutions when time periods are considered between patients' multiple medical episodes.

Another future work direction would be integrating neuropsychological elements and neuropathological indicators besides biomarkers together before applying machine learning techniques to seek any impact on dementia detection or progression. There are large numbers of pathological indicators that can be captured for cases and controls through invasive medical

procedures to pinpoint to signs of dementia. Examples of pathological features related to dementia conditions such as AD are tauopathy (tau) pathology, cerebral amyloid angiopathy (CAA), the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) assessments including Beta-Amyloid deposition (Aβ), neocortical neuritic plaques, Braak neurofibrillary tangle (NFT) stage, Thal phase, and primary age-related tauopathy (PART) among others (Wharton et al., 2019). Analysing these features by neuropathologists and clinicians can reveal indicators of dementia. However, it will be difficult to capture most of these pathological features during the progression of dementia because of resource availability including cost and personal factors. For example, some of the above pathological procedures may require hospitalisation, and others are not cost effective to conduct such as PET scans and MRI, which require expensive equipment. Thus, determining affordable biomarkers related to dementia is vital, especially in countries that have a shortage of resources.

Some research has considered elements of biomarkers or cognitive assessments to reveal whether these indicators would improve the prediction of dementia or dementia level, i.e. Mattsson-Carlgren et al. (2016), Hadjichrysanthou et al. (2020), Vasanthakumar et al. (2020), and Chang et al. (2021). However, little research has been directed to integrating cognitive and cost-effective biomarkers to deal with the problem of dementia progression. Examples of pathological features that can be measured during the AD progression are tau-related biomarkers such as phosphorylated tau (P-tau), and total tau (T-tau). These have been shown to be good dementia indicators in recent neuropathological research studies such as by Wharton et al. (2019; 2011) as they correlate with dementia higher than other pathological procedures like CERAD. P-tau is a more specific indicator since it can be measured by the levels of the tau protein in the neurofibrillary tangles using cerebrospinal fluid (CSF) or PET scans. The former is cheaper to conduct if the individual does not require hospitalisation. Other pathological indicators related to dementia conditions that are now available in ADNI and other datasets for participants are Aβ, CAA, Hippocampus, and others.

In the near future, we will also investigate ways to integrate both classification and feature assessments into a single phase using deep neural networks (DNN). Initially, we will investigate DNN through the Keras library as a python interface to Tensorflow backend. These algorithms are powerful as they provide feedforward models where information flow from the input layer into the last layer (output layer) without iterating back. Initially, the DNN algorithm can generate several logical neurons and allocate weights (numerical values) in an arbitrary manner to link them together. Then, weights and input variables are multiplied to produce an output value ranging between 0–1, that sequentially will be returned. If the DNN is unable to identify the outcome with some degree of accuracy, then it will reconfigure the network by amending the weights until it reaches a

termination condition. The input data items can be modelled by the DNN, and then encoded using a method such as ‘one-hot’ to ensure that the learning algorithm can process and correlate these variables with the target class correctly. Then, the data items are fed into the DNN algorithm, which is composed of several hidden layers with many filters plus a pooling layer for sampling the maps of the variables after each layer. We can configure hyperparameters of the DNN using a conventional grid search algorithm after evaluating multiple models’ configurations and hyperparameters statistically.

References

[1] Abbott, R., White, L., Ross, G., Masaki, K., Curb, J., & Petrovitch, H. (2004). Walking and dementia in physically capable elderly men. *American Medical Association*, 1447–1454. https://doi.org/10.1001/jama.292.12.1447

[2] Abdelhamid N., Abdul Jabbar A., Thabtah (2016) Associative Classification Common Research Challenges. ICPP Workshops 2016: 432-437.

[3] Abdelhamid, N., Ayesh, A., & Hadi, W. (2014). Multi-label rules algorithm based associative classification. *Parallel Processing Letters, 24*(01). https://doi.org/10.1142/S0129626414500017

[4] Abdelhamid N., Thabtah F., (2014) Associative Classification Approaches: Review and Comparison. Journal of Information and Knowledge Management (JIKM). Vol. 13, No. 3 (2014) 1450027.

[5] Abdelhamid, N., Ayesh, A., Thabtah, F., Ahmadi, S., Hadi, W. (2012). MAC: A multiclass associative classification algorithm. *Journal of Information and Knowledge Management (JIKM), 11*(2), pp. 1250011-1–1250011-10. https://doi.org/10.1142/S0219649212500116

[6] AbilityLab. (2019). *6 minute walk test*. https://www.sralab.org/rehabilitation-measures/6-minute-walk-test

[7] Aha D., Kibler D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning, 6*, 376–6. https://doi.org/10.1007/BF00153759

[8] Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine, 18*(13), 91-93. https://doi.org/10.1016/j.tjem.2018.08.001

[9] Albright, J. (2019). Forecasting the progression of Alzheimer's disease using neural networks and a novel pre-processing algorithm. *Alzheimer's & Dementia: Translational Research & Clinical Interventions, 5*, 483–491. https://doi.org/10.1016/j.trci.2019.07.001

[10] Alexopoulos, G. S., Abrams, R. C., Young, R. C., & Shamoian, C. A. (1988). Cornell Scale for Depression in Dementia. *Biological Psychiatry, 23*(3), 271–284. https://doi.org/10.1016/0006-3223(88)90038-8

[11] Allen, S. C., Warwick-Sanders, M., & Baxter, M. (2009). A comparison of four tests of cognition as predictors of inability to learn to use a metered dose inhaler in old age. *International Journal of Clinical Practice, 63*(8), 1150–1153. https://doi.org/10.1111/j.1742–1241.2009.02060.x

[12] Almnnaee, M. S., Thabtah, F., & Lu J. (2018). *An improved associative classification algorithm based on incremental rules*. 27th International Conference on Information Systems Development (Isd2018 Lund, Sweden), pp. 1-11.

[13] Alzheimer's Association (n.d.). *Alzheimer's Disease and dementia. What is dementia?* https://www.alz.org/alzheimers-dementia/what-is-dementia

[14] Alzheimer's Disease Neuroimaging Initiative [ADNI]. (2021). http://adni.loni.usc.edu

[15] American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. https://www.psychiatry.org/psychiatrists/practice/dsm

[16] Apple Inc. (2019a). *6CIT*. https://apps.apple.com/us/app/6cit/id1442965681

[17] Apple Inc. (2019b). *Dementia test - risk calculator of dementia*. https://apps.apple.com/us/app/dementia-test-risk-calculator-of-dementia/id1014958634

[18] Apple Inc. (2019c). *MoCA App*. https://apps.apple.com/us/app/moca-app/id1206246590

[19] Arevalo-Rodriguez, I., Smailagic, N., Roqué i Figuls, M., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., Pedraza, O. L., Bonfill Cosp, X., & Cullum, S. (2015). Mini-mental state examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI) (Review). *The Cochrane Library*, 1–62. https://doi.org/10.1002/14651858.CD010783.pub2

[20] Aske, D. (1990). The correlation between mini-mental state examination scores and Katz ADL status among dementia patients. *Rehabilitation Nursing*, *15*(3), 140–146. https://doi.org/10.1002/j.2048-7940.1990.tb01456.x

[21] Baldwin, S., & Farias, S. T. (2009). Neuropsychological assessment in the diagnosis of Alzheimer's disease. *Current Protocols in Neuroscience, 10,* Unit 10–3.

[22] Ballard, C., Burns, A., Corbett , A., Livingston, G., & Rasmussen, J. (2013). *Helping you to assess cognition - a practical toolkit for clinicians.* Alzheimer's Society Research.

[23] Balsis, S., Benge, J. F., Lowe, D. A., Geraci, L., & Doody, R. S. (2015). How do scores on the ADAS-Cog, MMSE, and CDR-SOB correspond? *The Clinical Neuropsychologist, 29*(7), 1002–1009. https://doi.org/10.1080/13854046.2015.1119312

[24] Bang, S., Son, S., Roh, H., Lee, J., Bae, S., Lee, K., & Hong, C., & Shin, H. (2017). Quad-phased data mining modeling for dementia diagnosis. *BMC Medical Informatics and Decision Making, 17*(1), Article 60. https://doi.org/10.1186/s12911-017-0451-3

[25] Barberger-Gateau, P., Commenges, D., Gagnon, M., Letenneur, L., Sauvel, C., & Dartigues, J. F. (1992). Instrumental activities of daily living as a screening tool for cognitive impairment and dementia in elderly community dwellers. *Journal of the American Geriatrics Society, 40*(11), 1129–1134. https://doi.org/10.1111/j.1532-5415.1992.tb01802.x

[26] Battista, P., Salvatore, C., & Castiglioni, I. (2017). Optimizing neuropsychological assessments for cognitive, behavioral, and functional impairment classification: A machine learning study. *Behavioural Neurology, 2017*, Article 1850909. https://doi.org/10.1155/2017/1850909

[27] Beekly, D., Ramos, E., Lee, W., Deitrich, W., Jacka, M., Wu, J., Hubbard, J., Koepsell, T., Morris, J., & Kukull, W. (2007). The National Alzheimer's Coordinating Center (NACC) database: The uniform data set. *Alzheimer Disease and Associated Disorders, 21*(3), 249–258.

[28] Bennasar, M., Setchi, R., & Hicks , Y. (2014). *Cascade classification for diagnosing dementia*. IEEE International Conference on Systems, Man, and Cybernetics, (pp. 25352–540). San Diego, CA, USA.

[29] Berauk, V., Murugiah, M., Soh, Y., Sheng, Y., Wong, T., & Ming, L. (2017). Mobile health applications for caring of older people: Review and comparison. *Therapeutic Innovation & Regulatory Science, 52*(3), 374–382. https://doi.org/10.1177/2168479017725556

[30] Bolchini, D., Finkelstein, A., Perrone, V., & Nagl, S. (2009). Better bioinformatics through usability analysis. *Bioinformatics, 25*(3), 406–412. https://doi.org/10.1093/bioinformatics/btn633

[31] Borson, S., Scanlan, J., Brush, M., Vitaliano, P., & Dokmak, A. (2000). The Mini-Cog: A cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. *International Journal of Geriatric Psychiatry, 15*(11), 1021–1027. https://doi.org/10.1002/1099-1166(200011)15:11<1021::aid-gps234>3.0.co;2-6

[32] Borson, S., Scanlan, J., Watanabe, J., Tu, S., & Lessig, M. (2006). Improving identification of cognitive impairment in primary care. *International Journal of Geriatric Psychiatry, 21*(4), 349–355. https://doi.org/0.1002/gps.147

[33] Bossers, W. J. R., van der Woude, L. H. V., Boersma, F., Scherder, E. J. A. & van Heuvelen, M. J. G. (2012). Recommended measures for the assessment of cognitive and

physical performance in older patients with dementia: A systematic review. *Dementia & Geriatric Cognitive Disorders Extra, 2*(1), 589–609. https://doi.org/10.1159/000345038

[34] Boyle, P. A., Malloy, P. F., Salloway, S., Cahn-Weiner, D. A., Cohen, R., & Cummings, J. L. (2003) Executive dysfunction and apathy predict functional impairment in Alzheimer disease. *American Journal of Geriatric Psychiatry, 11*(1), 214–221. https://pubmed.ncbi.nlm.nih.gov/12611751/

[35] BrainCheck. (2019). *About us.* https://braincheck.com/about

[36] BrainTest Inc. (2013). *Take the brain test*. https://braintest.com/

[37] Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

[38] Brodaty, H., Pond, D., Kemp, N., Luscombe, G., Harding, L., Berman, K., & Huppert, F. (2002). The GPCOG: A new screening test for dementia designed for general practice. *Journal of the American Geriatric Society*. https://doi.org/10.1046/j.1532-5415.2002.50122.x.

[39] Brown, T. G., Jolliffe, L., & Fielding, L. (2014). Is the mini mental status examination (MMSE) associated with inpatients' functional performance? *Physical & Occupational Therapy in Geriatrics, 32*(3), 228–240. https://doi.org/10.3109/02703181.2014.931504

[40] Bruno, D., & Vignaga, S. (2019). Addenbrooke's cognitive examination III in the diagnosis of dementia: A critical review. *Neuropsychiatric Disease and Treatmen*t, 441–447.

[41] Burns, A., & Iliffe, S. (2009). Alzheimer's disease. *BMJ, 338*(b158). https://doi.org/10.1136/bmj.b158

[42] Buschke, H., Kuslansky, G., Katz, M., Stewart, W., Sliwinski, M., Eckholdt, H., & Lipton, R. (1999). Screening for dementia with the memory impairment screen. *Neurology*. https://doi.org/10.1212/WNL.52.2.231.

[43] Butcher, J. (2007). CAIDE dementia risk score validated in study. *Clinical Neurology News,* 22a–22b.

[44] Cahn-Weiner, D. A., Malloy, P. F., Boyle, P. A., Marran, M., & Salloway, S. (2000). Prediction of functional status from neuropsychological tests in community-dwelling elderly individuals. *The Clinical Neuropsychologist, 14*(2), 187–195. https://doi.org/10.1076/1385-4046(200005)14:2;1-Z;FT187

[45] Callahan, C., Unverzagt, F., Hui, S., Perkins, A., & Hendrie, H. (2002). *Six-item screener to identify cognitive impairment among potential subjects for clinical research.* Lippincott Williams & Wilkins.

[46] Caramelli, P., Carthery-Goulart, M., Porto, C., Charchat-Fichman, H., & Nitrini, R. (2007). Category fluency as a screening test for Alzheimer disease in illiterate and literate patients. *Alzheimer Disease & Associated Disorders, 21*(1), 65–67. https://doi.org/10.1097/WAD.0b013e31802f244f

[47] Carrion, C., Folkvord, F., Anastasiadou, D., & Aymerich, M. (2018). Cognitive therapy for dementia patients: A systematic review. *Dementia & Geriatric Cognitive Disorders, 46*,(1), 1–26. https://doi.org/10.1159/000490851

[48] Carpenter, C., DesPain, B., Keeling, T., Shah, M., & Rothenberger, M. (2011). The six-item screener and AD8 for the detection of cognitive impairment in geriatric emergency department patients. *Annals of Emmergency Medicine,* 653–661.

[49] Cao W., Zhong Q., Li H., and Liang S. (2020) A Novel Approach for Associative Classification Based on Information Entropy of Frequent Attribute Set. in IEEE Access, vol. 8, pp. 140181-140193, 2020, doi: 10.1109/ACCESS.2020.3013141.

[50] Ceriani, L., & Verme, P. (2012). The origins of the Gini index: Extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *Journal of Economic Inequality, 10,* 421–443. https://doi.org/10.1007/s10888-011-9188-x

[51] Cicerone, K., & Azulay, J. (2002). Diagnostic utility of attention measures in post-concussion syndrome. *Journal the Clinical Neuropsychologist, 16*(3), 280–289. https://doi.org/10.1076/clin.16.3.280.13849

[52] Chang, C. H., Lin, C. H., & Lane, H. Y. (2021). Machine Learning and Novel Biomarkers for the Diagnosis of Alzheimer's Disease. International journal of molecular sciences, 22(5), 2761. https://doi.org/10.3390/ijms22052761

[53] Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, P. (2000). SMOTE: Synthetic Minority Over-sampling Technique. In International Conference of Knowledge Based Computer Systems, pp. 46–57. National Center for Software Technology, Mumbai, India, Allied Press.

[54] Chewy Logic. (2019). *A digital version of St Louis University Mental Status Exam (SLUMS)*. https://www.eslumstest.com/

[55] Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Translational Vision Science & Technology*, *9*(14). https://tvst.arvojournals.org/article.aspx?articleid=2762344.

[56] Choi, H., & Jin, K. H. (2018). Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behavioural Brain Research, 344*, 103–109. https://doi.org/10.1016/j.bbr.2018.02.017

[57] Chua, S., Tan, N., Wong, W., Allen Jr, J., Quah, J., Malhotra, R., & Østbye, T. (2019). Virtual reality for screening of cognitive function in older persons: Comparative study. *Journal of Medical Internet Research, 21*(8). https://doi.org/10.2196/14821.

[58] Cipriani, G., Danti, S., Picchi, L., Nuti, A., & Fiorino, M. D. (2020). Daily functioning and dementia. *Dementia & Neuropsychologia, 14*(2), 93–102. https://doi.org/10.1590/1980-57642020dn14-020001.

[59] Clark, P., & Niblett, T. (1989). The CN2 induction algorithm. *Machine Learning, 3*, 261–283. https://doi.org/10.1007/BF00116835

[60] Clemmensen, F. K., Hoffmann, K., Siersma, V. Sobol, N., Andersen, B. B., Vogel, A., Lolk, A., Gottrup, H., Høgh, P., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S. (2020). The role of physical and cognitive function in performance of activities of daily living in patients with mild-to-moderate Alzheimer's disease – a cross-sectional study. *BMC Geriatrics, 20,* Article 513 . https://doi.org/10.1186/s12877-020-01926-9

[61] Cohen, W. (1995). *Fast effective rule induction*. Proceedings of the 12th International Conference on Machine Learning, 115–123.

[62] Collin, C., Wade, D. T., Davis, S., & Horne, V. (1988). The Barthel ADL Index: A reliability study. *International Disability Study, 10*(2), 61–63. https://doi.org/10.3109/09638288809164103

[63] Comerford, V. E., Geffen, G. M., May, C., Medland, S., & Geffen, L. (2002). A rapid screen of the severity of mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology,* 409–419. https://doi.org/10.1076/jcen.24.4.409.1044

[64] Constantino, M., Pistori, H., Zardo, T., Zanoni, D.A., Belete, N.A., Rorigues Jr, J.F., Tetila, E.C., & Machado, B.B. (2021). Associative Classification Model for Forecasting Stock Market Trends. International Journal of Business Intelligence and Data Mining, 1, 1.

[65] Cordella, C., Borson, S., Boustan, M., Chodosh, J., Reuben, D., Verghese, J., Thies, W., & Fried, L. (2013). Alzheimer's Association recommendations for operationalizing the detection of cognitive impairment during the Medicare annual wellness visit in a

primary care setting. *Alzheimer's & Dementia, 9*(2), 141–150. https://doi.org/10.1016/j.jalz.2012.09.011

[66] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

[67] Cover, M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21–27.

[68] Cox, D. R. (1968). Notes on some aspects of regression analysis. *Journal of the Royal Statistical Society. Series A (General), 131*(3), 265–279. https://doi.org/10.2307/2343523

[69] Crisis Prevention Institute [CPI]. (n.d.). *Major neurocognitive disorder: The DSM-5's new term for dementia*. https://www.crisisprevention.com/Blog/Major-Neurocognitive-Disorder-Dementia

[70] Das, D., Ito, J., Kadowaki, T., & Tsuda, K. (2019). An interpretable machine learning model for diagnosis of Alzheimer's disease. *PeerJ, 7*, Article e6543. https://doi.org/10.7717/peerj.6543

[71] Dementiastatistics (2021) https://www.dementiastatistics.org/statistics/cost-and-projections-in-the-uk-and-globally-3/. [accessed May 21 2022].

[72] De Paula, J. J., & Malloy-Diniz, L. F. (2013). Executive functions as predictors of functional performance in mild Alzheimer's dementia and mild cognitive impairment elderly; *Estudos De Psicologia, 18*(1), pp. 117–124.

[73] DiBenedetti, D. B., Slota, C., Wronski, S. L. Vradenburg, G., Comer, M., Callahan, L. F., Winfield, J., Rubino, I., Krasa, H. B., Hartry, A., Wieberg, D., Kremer, I. N., Lappin, D., Martin, A. D., Frangiosa, T., Biggar, V., & Hauber, B. (2020). Assessing what matters most to patients with or at risk for Alzheimer's and care partners: A qualitative study evaluating symptoms, impacts, and outcomes. *Alzheimer's Research Therapy, 12,* Article 90. https://doi.org/10.1186/s13195-020-00659-6

[74] Douglas, C., Goulding, R., & Atkinson-Grosjean, J. (2011). Socio-cultural characteristics of usability of bioinformatics databases and tools. *Interdisciplinary Science Reviews, 36*(1), 55-71.

[75] Duda, R. & Hart, P. (1973). *Pattern classification and scene analysis*. John Wiley & Son.

[76] Eggermont, L. H., Gavett, B. E., Volkers, K. M., Blankevoort, C. G., Scherder, E. J., Jefferson, A. L., Steinberg, E., Nair, A., Green, R. C., & Stern, R. A. (2010). Lower-extremity function in cognitively healthy aging, mild cognitive impairment, and

Alzheimer's disease. *Archives of Physical Medicine and Rehabilitation, 91*(4), 584–8. https://doi.org/10.1016/j.apmr.2009.11.020

[77] Ehrensperger, M., Taylor, K., Berres, M., Foldi, N., Dellenbach, M., Bopp, I., Gold, G., von Gunten, A., Inglin, D., Müri, R., Rüegger, B., Kressig, R. W., Monsch, A. (2014). BrainCheck – A very brief tool to detect incipient cognitive decline: Optimized case-finding combining patient- and informant-based data. *Alzheimer's Research & Therapy, 6,* Article 69. https://doi.org/10.1186/s13195-014-0069-y.

[78] El-Sappagh SH., Abuhmed T., Alouffi B., Sahal R., Abdelhade N., Saleh, H., (2021) The Role of Medication Data to Enhance the Prediction of Alzheimer's Progression Using Machine Learning. Computational Intelligence and Neuroscience, vol. 2021, Article ID 8439655, 8 pages, 2021. https://doi.org/10.1155/2021/8439655

[79] Elyan, E., Moreno-Garcia, C. F., Jayne, C. (2018). *Symbols classification in engineering drawings*. In: International Joint Conference on Neural Networks (IJCNN).

[80] Fan, J., Flombaumb, J., McCandliss, B., Thomas, K., & Posner, M. (2003). Cognitive and brain consequences of conflict. *Neuroimage, 18*(1), 42–57. https://doi.org/10.1006/nimg.2002.1319

[81] Farias, S., Mungas, D., Reed, B. R., Cahn-Weiner, D., Jagust, W., Baynes, K., & Decarli, C. (2008). The measurement of everyday cognition (ECog): Scale development and psychometric properties. *Neuropsychology, 22*(4), 531-544. https://doi.org/10.1037/0894-4105.22.4.531.

[82] Fix, E., & Hodges, J. L. Jr. (1951). *Discriminatory analysis-nonparametric discrimination: Consistency properties*. Technical report, DTIC Document.

[83] Flaherty, L., Midden, A., & Mast, B. (2019). Psychometric evaluation of the symptoms of dementia screener (SDS) in a geriatric primary care population. *Clinical Gerontologist*, 504–511. https://doi.org/10.1080/07317115.2018.1453906

[84] Folstein, M., Folstein, S. E., & McHugh, P. (1975). "Ini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

[85] Foster, N., Bondi, M., Das, R., Foss, M., Hershey, L., Koh, S., Logan, R., Poole, C., Shega, J. W., Sood, A., Thothala, N., Wicklund, M., Yu, M., Bennett, A., & Wang, D. (2019). Quality improvement in neurology mild cognitive impairment quality measurement set. *Neurology, 93*(16). https://doi.org/10.1212/WNL.0000000000008259.

[86] Fountoulakis, K. N., Tsolaki, M., Chantzi, H., & Kazis, A. (2000). Mini Mental State Examination (MMSE): A validation study in Greece. *American Journal of Alzheimer's Disease*, *15*(6), 342–345. https://doi.org/10.1177/153331750001500604

[87] Frank, E., & Witten, I. H. (1998). *Generating accurate rule sets without global optimization*. pp. 144–151. Morgan Kaufmann.

[88] Friedman, B., Heisel, M. J., & Delavan, R. L. (2005). Psychometric properties of the 15-item geriatric depression scale in functionally impaired, cognitively intact, community-dwelling elderly primary care patients. *Journal of the American Geriatric Society*, *53*(9), 1570–6. https://doi.org/10.1111/j.1532-5415.2005.53461.x

[89] Friedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K., & Delis, D. (1994). *DC: Clock drawing: A neuropsychological analysis.* Oxford University Press.

[90] Friedman, N., Geiger, D., & Goldszmidt, M. (1997) Bayesian network classifiers. *Machine Learning - Special Issue on Learning with Probabilistic Representations*, *29*(2–3), pp.131–63.

[91] Gaines B. R., & Compton P. (1995). Induction of ripple-down rules applied to modeling large databases. *Journal of Intelligent Information Systems*, *5*(3), 211–228. https://doi.org/10.1007/BF00962234

[92] Galasko, D., Bennett, D., Sano, M., Ernesto, C., Thomas, R., Grundman, M., & Ferris, S. (1997). An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. *Alzheimer Disease and Associated Disorders*, *11*(2), S33–S39.

[93] Galvin, J., Roe, C., Powlishta K., Coats, M., Muich S., Grant E., Miller, J. P., Storandt, M., & Morris, J. (2005). The AD8: A brief informant interview to detect dementia. *Neurology*, *65*(4), 559–64.

[94] Ganguli, M., Blacker, D., Blazer, D. G., Grant, I., Jeste, D. V., Paulsen, J. S., Petersen, R. C., & Sachdev, P. S. (2011). Classification of neurocognitive disorders in DSM-5: A work in progress. *The American Journal of Geriatric Psychiatry*, *19*(3), 205–210. https://doi.org/10.1097/jgp.0b013e3182051ab4

[95] Gardner, M., & Dorling, S. (1998). Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment*, *32*(14–15), 2627–2636. https://doi.org/0.1016/S1352-2310(97)00447-0

[96] Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., & Cummings, J. L. (2006). Mild cognitive impairment. *The Lancet*, *367*(9518), 1262–1270.

[97] Gay, B., Taylor, K., Hohl, U., Tolnay, M., & Staehelin, H. (2008). The validity of clinical diagnoses of dementia in a group of consecutively autopsied memory clinic patients. *The Journal of Nutrition Health and Aging, 12*(2), 132–137. https://doi.org10.1007/BF02982566

[98] Gelinas, I., Gauthier, L. & McIntyre, M. (1999). Development of a functional measure for persons with Alzheimer's disease: The Disability Assessment for Dementia. *American Journal of Occupational Therapy, 53*(5), 471-481. https://doi.org/10.5014/ajot.53.5.471

[99] General Data Protection Regulation (GDPR). *General Data Protection Regulation (GDPR)*. https://gdpr-info.eu/

[100] Gerdner, A., Kestenberg, J., & Edvinsson, M. (2014). Validity of the Swedish SCID and ADDIS diagnostic interviews for substance use disorders: Sensitivity and specificity compared with a LEAD golden standard. *Nordic Journal of Psychiatry, 69*(1). https://doi.org/10.3109/08039488.2014.926987

[101] Ghafar, M. Z. A. A., Miptah, H. N., & O'Caoimh, R. (2019). Cognitive screening instruments to identify vascular cognitive impairment: A systematic review. *International Journal of Geriatric Psychiatry, 34*(8), 1114–1127.

[102] Gheware, S. D., Kejkar, A. S., & Tondare, S. M. (2014). Data mining: Task, tools, techniques and applications. *International Journal of Advanced Research in Computer and Communication Engineering, 3*(10), 8095–8098. https://doi.org/10.17148/IJARCCE.2014.31003

[103] Goetzinger, K., & Odibo, A. (2011). Statistical analysis and interpretation of prenatal diagnostic imaging studies, Part 1: Evaluating the efficiency of screening and diagnostic tests. *Journal of Ultrasound Medicine*, 1121–1127.

[104] Golden, C. J. (1978). *Stroop color and word test: A manual for clinical and experimental uses*. Wood Dale: Stoelting Company.

[105] Gonzalez, R., & Woods, R. (2002). *Digital image processing, 2nd edition.* Prentice Hall.

[106] Goodman, B. & Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine, 38*(3). https://doi.org/10.1609/aimag.v38i3.2741.

[107] Google. (2016). *MMSE*. https://play.google.com/store/apps/details?id=com.yasintanriverdi.mmse

[108] Google. (2017). *Dementia & Alzheimer's memory diagnosis test: MMSE*. https://play.google.com/store/apps/details?id=com.alzheimers_mme

[109] Google. (2019a). *Cognitive exams*. https://play.google.com/store/apps/details?id=br.com.digos.examescognitivos

[110] Google. (2019b). *Dementia risk tool*. https://play.google.com/store/apps/details?id=com.dementiarisktool

[111] Google. (2019c). DST - *Dementia screening test, Alzheimer test*. https://play.google.com/store/apps/details?id=com.dementiascreeningtest

[112] Greenwald, B., Cifu, D., Marwitz, J., Enders, L., Brown, A., Englander, J., & Zafonte, R. (2001). Factors associated with balance deficits on admission to rehabilitation after traumatic brain injury: A multicenter analysis. *Journal of Head Trauma Rehabilitation, 16*(3), 238–252. https://doi.org/10.1097/00001199-200106000-00003

[113] Groppell, S., Soto-Ruiz, K., Flores, B., Dawkins, W., Smith, I., Eagleman, D., & Katz, Y. (2019). A rapid, mobile neurocognitive screening test to aid in identifying cognitive impairment and dementia (BrainCheck): Cohort Study. *JMIR Aging, 2*(1), Article e12615, 1–14. https://doi.org/10.2196/12615

[114] Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks, 1*(1), pp. 17–61.

[115] Grueso, S., Viejo-Sobera, R. (2021) Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review. Alz Res Therapy 13, 162 (2021). https://doi.org/10.1186/s13195-021-00900-w.

[116] Guerrero-Berroa, E., Luo, X., Schmeidler, J., Rapp, M. A., Dahlman, K., Grossman, H. T., Haroutunian, H., & Beeri, M. S. (2009). The MMSE orientation for time domain is a strong predictor of subsequent cognitive decline in the elderly. *International Journal of Geriatric Psychiatry, 24*(12), 1429–1437. https://doi.org/10.1002/gps.2282

[117] Gupta, A., & Katarya, R. (2020). Social media based surveillance systems for healthcare using machine learning: A systematic review. *Journal of Biomedical Informatics, 108,* Article 103500.

[118] Hadi, W.M., Al-Radaideh, Q.A., & Alhawari, S. (2018). Integrating associative rule-based classification with Naïve Bayes for text classification. Appl. Soft Comput., 69, 344-356.

[119] Hadjichrysanthou, C., Evans, S., Bajaj, S. et al. The dynamics of biomarkers across the clinical spectrum of Alzheimer's disease. Alz Res Therapy 12, 74 (2020). https://doi.org/10.1186/s13195-020-00636-z

[120] Hall, K., Hamilton, B., Gordon, W., & Zasler, N. (1993). Characteristics and comparisons of functional assessment indices: Disability rating scale, functional independence measure and functional assessment measure. *Journal of Head Trauma Rehabilitation 8*(2), 60–74.

[121] Hand, D. J., & Yu, K. (2001). Idiot's Bayes—not so stupid after all? *International Statistical Review, 69*(3), 385–398.

[122] Harrison, J., Dgetluck, N., Gawryl, M., Moebius, H., & Hilt, D. (2013). Validation of a novel cognitive composite assessment for mild and prodromal Alzheimer's disease. *Alzheimer's and Dementia, 9*, p. 661.

[123] Hartigan, I. (2006). A comparative review of the Katz ADL and the Barthel Index in assessing the activities of daily living of older people. *International Journal of Older People Nursing*, *2*(3), 204–212. https://doi.org/10.1111/j.1748-3743.2007.00074.x

[124] Herrera-García, J., Rego-García, I., Guillén-Martínez, V., Carrasco-García, M., Valderrama-Martín, C., Vílchez-Carrillo, R., López-Alcalde, S., & Carnero-Pardo, C. (2019). Discriminative validity of an abbreviated Semantic Verbal Fluency Test. *Dementia & Neuropsychologia*, *13*(2). https://doi.org/10.1590/1980-57642018dn13-020009 .

[125] Hodges, J., & Larner, A. (2017). Addenbrooke's cognitive examinations: ACE, ACE-R, ACE-III, ACEapp, and M-ACE. *Cognitive Screening Instruments, 15*, 109–137. https://doi.org/10.2147/NDT.S151253

[126] Hodkinson, H. (1972). Evaluation of a mental test score for assessment of mental impairment in the elderly. *Age and Aging*, *1*(4), 233–238. https://doi.org/10.1093/ageing/1.4.233

[127] Hoffmann, K., Frederiksen, K. S., Sobol, N. A., Beyer, N., Vogel, A., Simonsen, A. H., Johannsen, P., Lolk, A., Terkelsen, O., Cotman, C. W., Hasselbalch, S. G., & Waldemar, G. (2013). Preserving cognition, quality of life, physical health and functional ability in Alzheimer's disease: The effect of physical exercise (ADEX trial):

Rationale and design. *Neuroepidemiology, 41,* 198–207. https://doi.org/10.1159/000354632

[128] Holsinger, T., Deveau, J., Boustani, M., & Williams, J. (2007). Does this patient have dementia? *Journal of American Medical Association, 297*(21), 2391–2404. https://doi.org/10.1001/jama.297.21.2391.

[129] Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied logistic regression* (3rd ed., 398). John Wiley & Sons. https://doi.org/10.1002/9781118548387

[130] Hsu, J. -L., Hsu, W. -C., Chang, C. -C., Lin, K. -J., Hsiao, I. -T., Fan, Y. -C., & Bai, C. -H. (2017). Everyday cognition scales are related to cognitive function in the early stage of probable Alzheimer's disease and FDG-PET findings. *Scientific Reports, 7,* Article 1719. https://doi.org/10.1038/s41598-017-01193-6

[131] Hughes, C. P., Berg, L., Danziger, W., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *The British Journal of Psychiatry, 160*(6), 566–572. https://doi.org/10.1192/bjp.140.6.566

[132] Hühn, J., & Hüllermeier, E. (2009). FURIA: An algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery, 19*, 293–319. https://doi.org/10.1007/s10618-009-0131-8

[133] Ihl, R., Lutz, F., Dierks, T., Martin, E.-M., & Maurer, K. (1992). Differential validity of psychometric tests in dementia of the Alzheimer type. *Psychiatry Research, 44*(2), 93–106. https://doi.org/10.1016/0165-1781(92)90044-4

[134] Inouye, S., Dyck, C., Alessi, C., Balkin, S., Siegal, A., & Horwitz, R. (1990). Clarifying confusion: The confusion assessment method - a new method for detection of delirium. *Journal of Annals of Internal Medicine*, 113(12), 941–948. https://doi.org/10.7326/0003-4819-113-12-941

[135] Inoven. (2018). *Cognity.* https://cognity.app

[136] Jammeh, E. A., Carroll, C. B., Pearson, S. W., Escudero, J., Anastasiou, A., Zhao, P., Chenore, T., Zajicek, J., & Ifeachor, E. (2018). Machine-learning based identification of undiagnosed dementia in primary care: A feasibility study. *BJGP Open, 2*(2), https://doi.org/10.3399/bjgpopen18X101589.

[137] Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electron Markets, 31*, 685–695. https://doi.org/10.1007/s12525-021-00475-2

[138] Jitapunkul, S., Pillay, I., & Ebrahim, S. (1991). The Abbreviated Mental Test: Its use and validity. *Age and Ageing, 20(5),* 332–336. https://doi.org/10.1093/ageing/20.5.332 .

[139] Joachims, H. (1999). *Large-scale support vector machine learning practical, advances in kernel methods: Support vector learning*. MIT Press.

[140] John, G. H., & Langley P. (1995). *Estimating continuous distributions in bayesian classifiers*. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 338–345, 1995.

[141] Johnson, L., Cushing, B., Rohlfing, G., Edwards, M., Davenloo, H., D'Agostino, D., Hall, J. R., & O'Bryant, S. (2014). The Hachinski Ischemic scale and cognition: The influence of ethnicity. *Age and Ageing, 43*,(3), 364–369. https://doi.org/10.1093/ageing/aft189

[142] Jorm, A., Scott, R., & Jacomb, P. (1989). Assessment of cognitive decline in dementia by informant questionnaire. *International Journal of Geriatric Psychiatry*, *4*(1). https://doi.org/10.1002/gps.930040109.

[143] Julayanont, P., Tangwongchai, S., Hemrungrojn, S., Tunvirachaisakul, C., Phanthumchinda, K., Hongsawat, J., Suwichanarakul, P., Thanasirorat, S., & Nasreddine, Z. (2015). The Montreal cognitive assessment—basic: A screening tool for mild cognitive impairment in illiterate and low-educated elderly adults. *Journal of the American Geriatrics Society*, *63*(12), 2550–2554. https://doi.org/10.1111/jgs.13820

[144] Jutten, R. J., Harrison, J. E., Brunner, A. J., Vreeswijk, R., Deelen, R. A. J., de Jong, F. J., Opmeer, E. M., Ritchie, C. W., Aleman, A., Scheltens, P., & Sikkes, S. A. M. (2020). The Cognitive-Functional Composite is sensitive to clinical progression in early dementia: Longitudinal findings from the Catch-Cog study cohort. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *6*(1). https://doi.org/10.1002/trc2.12020

[145] Jutten, R. J., Harrison, J. E., de Jong, F. J., Aleman, A., Ritchie, C. W., Scheltens, P., & Sikkes, S. A. M. (2017). A composite measure of cognitive and functional progression in Alzheimer's disease: Design of the Capturing Changes in Cognition study. *Alzheimer's and Dementia, 3*(1), 130–8. https://doi.org/10.1016/j.trci.2017.01.004

[146] Jutten, R. J., Harrison, J. E., Kjoe, L. M., Ingala, S., Vreeswijk, R., van Deelen, R. A. J., Jan de Jong, F., Opmeer, E. M., Aleman, A., Ritchie, C. W., Scheltens, P., & Sikkes, S. A. M. (2019). Assessing cognition and daily function in early dementia using the cognitive-functional composite: findings from the Catch-Cog study cohort. *Alzheimer's Research and Therapy*, *11*, Article 45. https://doi.org/10.1186/s13195-019-0500-5

[147] Kansagara, D., & Freeman, M. (2010). *A systematic evidence review of the signs and symptoms of dementia and brief cognitive tests available in VA*. Department of Veterans Affairs (US).

[148] Kemp, E. C., Ebner, M. K., Ramanan, S., Godek, T. A., Pugh, E. A., Bartlett, H. H., McDonald, J. W., Mecca, M. C., van Dyck, C. H., & Mecca, A. P. (2020). Statin use and risk of cognitive decline in the ADNI cohort. *The American Journal of Geriatric Psychiatry*, *28*(5), 507–517. https://doi.org/10.1016/j.jagp.2019.11.003

[149] Kerkhof, Y., Bergsma, A., Graff, M., & Droes, R. (2017). Selecting apps for people with mild dementia: Identifying user requirements for apps enabling meaningful activities and self-management. Journal of Rehabilitation and Assistive Technologies Engineering, 4, 1-21. doi:https://doi.org/10.1177/2055668317710593

[150] Kim, H.-J., Min, J.-Y., & Min, K.-B. (2020). The association between longest-held lifetime occupation and late-life cognitive impairment: Korean longitudinal study of aging (2006–2016). *International Journal of Environmental Research and Public Health, 17*(17), 6270. https://doi.org/10.3390/ijerph17176270

[151] Kim, M. J., Tsutsumimoto, K., Doi, T., Nakakubo, S., Kurita, S., Makizako, H., & Shiada, H. (2019). Relationships between cognitive leisure activities and cognitive function in older adults with depressive symptoms: A cross sectional study. *BMJ Open, 10*, Article e032679. https://doi.org/10.1136/ bmjopen-2019-032679

[152] Kira, K., & Rendell, L. A., (1992). *A practical approach to feature selection*. In: Proceedings of the Ninth International Workshop on Machine Learning. pp. 249–256.

[153] Kivipelto, M., Ngandu, T., Laatikainen, T., Winblad, B., Soininen, H., & Tuomilehto, J. (2006). Risk score for the prediction of dementia risk in 20 years among middle aged people: A longitudinal, population-based study. *The Lancet Neurology*, 735–741.

[154] Knopman, D. S., Boeve, B. F., & Petersen, R. C. (2003, October). *Essentials of the proper diagnoses of mild cognitive impairment, dementia, and major subtypes of dementia*. In Mayo Clinic Proceedings (Vol. 78, No. 10, pp. 1290–1308). Elsevier.

[155] Kononenko I. (1994) Estimating attributes: Analysis and extensions of RELIEF. In: Bergadano F., De Raedt L. (eds) *Machine learning: ECML-94. ECML 1994*. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), 784. Springer. https://doi.org/10.1007/3-540-57868-4_57

[156] Kotsiantis, S. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, *31*(3), 249–268.

[157] Krall, J. R., Carlson, M. C., Fried, L. P., & Xue, Q. L. (2014). Examining the dynamic, bidirectional associations between cognitive and physical functioning in older adults. *American Journal of Epidemiology*, *180*(8), 8388–46. https://doi.org/10.1093/aje/kwu198

[158] Kueper, J., Speechley, M., & Montero-Odasso, M. (2018). The Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog): Modifications and responsiveness in pre-dementia populations. A narrative review. *Journal of Alzheimer's Disease, 63*(2), 423-444. https://doi.org/10.3233/JAD-170991

[159] Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist*, *9*(3), 179–186. https://doi.org/10.1093/geront/9.3_Part_1.179

[160] le Cessie, S., & van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Applied Statistics*, *41*(1), 191–201. https://doi.org/10.2307/2347628

[161] Lecky, G. S., & Beatty, W. W. (2002). Predicting functional performance by patients with Alzheimer's disease using the Problem in Everyday Living (PEDL) test: A preliminary study. *Journal of the International Neurological Society*, *8*(1), 48–57.

[162] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–44. https://doi.org/10.1038/nature14539

[163] Lee YM, Park JM, Lee BD, Moon E, Jeong HJ, Chung YI, Kim JH, Kim HJ, Mun CW, Kim TH, Kim YH (2016) Gray matter volumes and treatment response of psychotic symptoms to risperidone in antipsychotic-naïve Alzheimer's disease patients. J Clin Psychiatry. 2016 Jan;77(1):e8-13. doi: 10.4088/JCP.14m09740.

[164] Lemoine, B., Rayburn, S., & Benton, R. (2010*). Data fusion and feature selection for Alzheimer's diagnosis.* International Conference on Brain Informatics. 6334, pp. 320–327. Toronto, ON, Canada: Springer-Verlag. https://doi.org/10.1007/978-3-642-15314-3_30

[165] Li, F., Guo, Q., Qin, W., & Hao, D. (2017). Discriminative ability of everyday cognition (ecog) scale for patients with subjective cognitive decline (scd) in memory clinics. *Alzheimer's and Dementia, 13*(7), 809–810. https://doi.org/10.1016/j.jalz.2017.06.1106

[166] Li, W., Han, J., and Pei, J. (2001). *CMAR: Accurate and efficient classification based on multiple-class association rule*. Proceedings of the IEEE International Conference on Data Mining –ICDM, 369-376.

[167] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18–22.

[168] Lim, K. B., Kim, J., Lee, H. J., Yoo, J., You, E. C., & Kang, J. (2018). Correlation between Montreal cognitive assessment and functional outcome in subacute stroke patients with cognitive dysfunction. *Annals of Rehabilitation Medicine, 42*(1), 26–34. https://doi.org/10.5535/arm.2018.42.1.26

[169] Liu, B., Hsu, W., and Ma, Y. (1998) Integrating classification and association rule mining. Proceedings of the Knowledge Discovery and Data Mining Conference- KDD, 80-86. New York.

[170] Liu, H., & Setiono, R. (1995). *Chi2: Feature selection and discretization of numeric attributes.* Proceedings of the IEEE 7th International Conference on Tools with Artificial Intelligence (pp. 388–391). IEEE.

[171] Liu-Seifert, H., Siemers, E., Price, K., Han, B., Selzler, K. J., Henley, D., Sundell, K., Aisen, P., Cummings, J., Raskin, J., & Mohs, R. (2015). Cognitive impairment precedes and predicts functional impairment in mild Alzheimer's disease. *Journal of Alzheimer's Disease, 47*(1), 205–14. https://doi.org/10.3233/JAD-142508.

[172] Loewenstein, D. A., Amigo, E., Duara, R., Guterman, A., Hurwitz, D., Berkowitz, N., Wilkie, F., Weinberg, G., Black, B., & Gittelman, B. (1989). A new scale for the assessment of functional status in Alzheimer's disease and related disorders. *Journal of Gerontology, 44*(4), 114–121. https://doi.org/10.1093/geronj/44.4.p114.

[173] Maglogiannis, I. G., & Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica, 31*(2007), 249–268.

[174] Maheux, E., Koval, I., Archetti, D., Redolfi, A., & Durrleman, S. (2020). Prediction of the MMSE up to 6 years ahead with cross-cohort replications. *Alzheimer's & Dementia, 16*(S5). https://doi.org/10.1002/alz.043541

[175] Mahyoub M., Randles M., Baker T. and Yang, P. (2018)Comparison Analysis of Machine Learning Algorithms to Rank Alzheimer's Disease Risk Factors by Importance. 2018 11th International Conference on Developments in eSystems Engineering (DeSE), 2018, pp. 1-11, doi: 10.1109/DeSE.2018.00008.

[176] Makizako, H., Shimada, H., Park, H., Doi, T., Yoshida, D., Uemura, K., Tsutsumimoto, K., & Suzuki, T. (2013). Evaluation of multidimensional neurocognitive function using a tablet personal computer: test-retest reliability and validity in community-dwelling older adults. *Geriatrics and Gerontology International, 13*(4), 860–6. https://doi.org/10.1111/ggi.12014

[177] Marinescu, R. V., Oxtoby, N. P., Young, A. L., Bron, E. E., Toga, A. W., Weiner, M. W., Fox, N. C., Golland, P., Klein, S., & Alexander, D. C. (2019). TADPOLE challenge: Accurate Alzheimer's disease prediction through crowdsourced forecasting of future data. *Predictive Intelligence in Medicine*, 11843, pp. 1–10. https://doi.org/10.1007/978-3-030-32281-6_1

[178] Marshall, G. A., Zoller, A. S., Lorius, N., Amariglio, R. E., Locascio, J. J., Johnson, K. A., Sperling, R. A., & Rentz, D. M. (2015). Functional Activities Questionnaire items that best discriminate and predict progression from clinically normal to mild cognitive impairment. *Current Alzheimer Research, 12*(5), 493–502. https://doi.org/10.2174/1567205012051505261150003

[179] Martin, B. (1995). *Instance-based learning: Nearest neighbour with generalisation*. (Thesis) University of Waikato Hamilton, New Zealand.

[180] Martyr, A., & Clare, L. (2012). Executive function and activities of daily living in Alzheimer's disease: A correlational meta-analysis. *Dementia and Geriatric Cognitive Disorders, 33*(2-3), 189–203. https://doi.org/10.1159/000338233

[181] Maxim, L., Niebo, R., & Utel, M. (2014). Screening tests: A review with examples. *Inhalation Toxicology*, 811–828.

[182] Mayo, A. (2016). Use of the Functional Activities Questionnaire. *Best Practices in Nursing Care to Older Adults with Dementia, D13,* 323–329.

[183] Mayo Clinic. (2021). *Dementia - symptoms and causes.* https://www.mayoclinic.org/diseases-conditions/dementia/symptoms-causes/syc-20352013

[184] McCab, D. (2019). *Katz Index of Independence in Activities of Daily Living (ADL).* University Rory Meyers College of Nursing.

[185] McCombe, N., Liu, S., Ding, X., Prasad, G., Bucholc, M., Finn, D. P., Todd, S., McClean, P. L., Wong-Lin, K., & Alzheimer's Disease Neuroimaging Initiative (ADNI). (2020). *Practical strategies for extreme missing data imputation in dementia diagnosis*. MedRxiv, Article 20146118. https://doi.org/10.1101/2020.07.13.20146118

[186] Mattsson, N., Zetterberg, H., Janelidze, S., Insel, P. S., Andreasson, U., Stomrud, E., Palmqvist, S., Baker, D., Tan Hehir, C. A., Jeromin, A., Hanlon, D., Song, L., Shaw, L. M., Trojanowski, J. Q., Weiner, M. W., Hansson, O., Blennow, K., & ADNI Investigators (2016). Plasma tau in Alzheimer disease. Neurology, 87(17), 1827–1835. https://doi.org/10.1212/WNL.0000000000003246

[187] McGough, E. L., Kelly, V. E., Logsdon, R. G., McCurry, S. M., Cochrane, B. B., Engel, J. M. & Teri, L. (2011). Associations between physical performance and executive function in older adults with mild cognitive impairment: Gait speed and the timed 'up & go' test. *Physical Therapy, 91*(8), 1198–207. https://doi.org/10.2522/ptj.20100372

[188] 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), 939–939. https://doi.org/10.1212/wnl.34.7.939

[189] Melara,, R., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110*(3), 422–471. https://doi.org/10.1037/0033-295X.110.3.422.

[190] Mennella, H., & Heering, H. (2015). *Dementia assessment: Using the clinical dementia rating scale*. https://www.ebscohost.com/assets-sample-content/Dementia_Assessment_-_Using_the_Clinical_Dementia_Rating_Scale.pdf

[191] Mioshi, E. (2011). Cognitive screening in dementia: Does it tell the whole story? *Occupational Therapy News, 19*(12), 20–21.

[192] Mitchell, A., Bird, V., Rizzo, M., & Meader, N. (2011). Which version of the geriatric depression scale is most useful in medical settings and nursing homes? Diagnostic validity meta-analysis. *American Journal of Geriatric Psychology*, 1066–1077.

[193] Mitchell, T. (1997). *Machine learning*. ISBN 0070428077. McGraw Hill.

[194] Mohammad, R. M., Thabtah, F., & McCluskey, L. (2013). *Predicting phishing websites using neural network trained with back-propagation.* Las Vegas, World Congress in Computer Science, Computer Engineering, and Applied Computing, pp. 682–686.

[195] Mohs, R. C., Knopman, D., Petersen, R. C., Ferris, S. H., Ernesto, C., Grundman, M., Sano, M., Bieliauskas, L., Geldmacher, D., Clark, C., & Thal, L. J. (1997). Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Disease and Associated Disorders, 11*(2), S13–21.

[196] Monllau, A., Pena-Casanova, J., Blesa, R., Aguilar, M., Bohm, P., Sol, J. M., & Hernandez, G. (2007). Diagnostic value and functional correlations of the ADAS-Cog scale in Alzheimer's disease: Data on NORMACODEM project. *Neurologia, 22*(8), 493–501.

[197] Monte, V., Geffen , G., May, C., & McFarland, K. (2010). Improved sensitivity of the rapid screen of mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *32*(1), 28–37. https://doi.org/10.1080/13803390902806519

[198] Moradi, E., Hallikainen, I., Hänninen, T., Tohka, J., & Alzheimer's Disease Neuroimaging Initiative. (2017). Rey's auditory verbal learning test scores can be predicted from whole brain MRI in Alzheimer's disease. *Neuroimage: Clinical*, *13*, 415–427.

[199] Moroney , J., Bagiella , E., Desmond , D., Hachinski , V., Mölsä , P., Gustafson , L., Brun, A., Fischer, P., Erkinjuntti, T., Rosen, W., Paik, M. C., & Tatemichi, T. (1997). Meta-analysis of the Hachinski Ischemic Score in pathologically verified dementias. *Neurology*, *49*(4), 1096–105. https://doi.org/10.1212/wnl.49.4.1096

[200] Mundt, J. C., Freed, D. M., & Greist, J. H. (2000). Lay person-based screening for early detection of Alzheimer's disease: Development and validation of an instrument. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 55*(3), pp.163–170.

[201] Mura, T., Proust-Lima, C., Jacqmin-Gadda, H., Akbaraly, T. N., Touchon, J., Dubois, B., & Berr, C. (2014). Measuring cognitive change in subjects with prodromal Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 85*(4), 363–70. http://dx.doi.org/10.1136/jnnp-2013-305078

[202] Nagaraj, S., & Duong, T. Q. (2020). *Risk score stratification of Alzheimer's disease and mild cognitive impairment using deep learning*. https://doi.org/10.1101/2020.11.09.20226746

[203] Najar, J., Östling, S., Gudmundsson, P., Sundh, V., Johansson, L., Kern, S., Guo, X., Hällström T., Skoog, I. (2019). Cognitive and physical activity and dementia A 44-year longitudinal population study of women. *Journal of Neurology*, *92*(12). https://doi.org/10.1212/WNL.0000000000007021.

[204] Nall, R. M. (2017). *10 types of dementia*. https://www.healthline.com/health/types-dementia#alzheimers

[205] Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–9. https://doi.org/10.1111/j.1532-5415.2005.53221.x

[206] National Collaborating Centre for Mental Health. (2007). *Dementia: A NICE-SCIE Guideline on supporting people with dementia and their carers in health and social care.* British Psychological Society.

[207] Newman, C., Bevin, A., Zajicek, J., Hodges, J., Vuillermoz, E., Dickenson, J., Kelly, D. S., Brown, S., & Noad, R. (2018). Improving the quality of cognitive screening assessments: ACEmobile, an iPad-based version of the Addenbrooke's Cognitive Examination-III. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 10*(2018), 182–187. https://doi.org/10.1016/j.dadm.2017.12.003

[208] Nirjon, S., Emi, I. A., Sayeed Mondol, M. A., Salekin, A., & Stankovic, J. (2014). *MOBI-COG: A mobile application for instant screening of dementia using the mini-cog test*. 5th Conference on Wireless Health. Bethesda, United States: Association for Computing Machinery, Inc.

[209] Nogueira, J., Freitas, S., Duro, D., Almeida, J., & Santana, I. (2018). Validation study of the Alzheimer's disease assessment scale - cognitive subscale (ADAS-Cog) for the Portuguese patients with mild cognitive impairment and Alzheimer's disease. *The Clinical Neuropsychologists, 32*(1), 46–59. https://doi.org/10.1080/13854046.2018.1454511

[210] Noone, P. (2015). Addenbrooke's cognitive examination-III. *Occupational Medicine, 65*(5), 418–420.

[211] O'Bryant, S. E., Lacritz, L. H., Hall, J., Waring, S. C., Chan, W., Khodr, Z. G., Massman, P. J., Hobson, V., & Cullum, C. M. (2010). Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the National Alzheimer's Coordinating Center database. *Archives in Neurology and Neuroscience, 67*(6), 746–749. https://doi.org/10.1001/archneurol.2010.115

[212] Ong, K. (Dec, 2017). *Challenges in dementia studies*. https://www.intechopen.com/chapters/58494

[213] Ouimette, P., & Klein, D. (1995). Test-Retest stability, mood-state dependence, and informant-subject concordance of the SCID-AXIS II questionnaire in a non-clinical sample. *Journal of Personality Disorders, 9*(2), 105–111. https://doi.org/10.1521/pedi.1995.9.2.105

[214] Padillo, F., Luna, J. & Ventura, S. Evaluating associative classification algorithms for Big Data. Big Data Anal 4, 2 (2019). https://doi.org/10.1186/s41044-018-0039-7

[215] Panegyres, P., Berry, R., & Burchell, J. (2016). Early dementia screening. *Diagnostics*, *6*(1), 6. https://doi.org/10.3390/diagnostics6010006.

[216] Park, H. K., Na, D. L., Han, S. H., Kim, J. Y., Cheong, H-K., Kim, S. Y., Kim, S. Y., Hong, C. H., Kim, D-K., Ku, B. D., Moon, S. Y., Lee, J-Y., Shim, Y. S., Youn, Y. C., Kim, E-J., Kim, B-C., Park, K. H., Cha, K. R., Seo, S. W., & Lee, J-H. (2011). Clinical characteristics of a nationwide hospital-based registry of mild-to-moderate Alzheimer's disease patients in Korea: A CREDOS (Clinical Research Center for Dementia of South Korea) study. *Journal of Korean Medical Science, 26*(9), 1219–1226. https://doi.org/10.3346/jkms.2011.26.9.1219

[217] Payne, J. L., Lyketsos, C. G., Steele, C., Baker, L., Galik, E., Kopunek, S., Steinberg, M., & Warren, A. (1998). Relationship of cognitive and functional impairment to depressive features in Alzheimer's disease and other dementias. *Journal of Neuropsychiatry and Clinical Neurosciences, 10*(4), 440–7. https://doi.org/10.1176/jnp.10.4.440.

[218] Pearson, K. (1920). Notes on the history of correlation. *Biometrika, 13*(1), 25–45. https://doi.org/10.2307/2331722

[219] Pereira, T., Ferreira, F., Cardoso, S., Silva, D., de Mendonca, A., Guerreiro, M., & Madeira, S. (2018). Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease: A feature selection ensemble combining stability and predictability. *BMC Medical Informatics and Decision Making, 18*(2018), Article 137. https://doi.org/10.1186/s12911-018-0710-y

[220] Peters, M. E., & Rabins, P. V. (2017). *Cognitive impairment*. In Johns Hopkins psychiatry guide. https://www.hopkinsguides.com/hopkins/view/Johns_Hopkins_Psychiatry_Guide/787027/all/Cognitive_Impairment

[221] Pfeffer, R. I., Kurosaki, T. T. Harrah, C. H. Jr., Chance, J. M., & Filos, S.(1982). Measurement of functional activities in older adults in the community. *Journal of Gerontology, 37*(3), 323–329. https://doi.org/10.1093/geronj/37.3.323

[222] Picard, R. R., & Cook, R. D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association, 79*(387), 575–583.

[223] Pickett, J., Bird, C., Ballard, C., Banerjee, S., Brayne, C., Cowan, K., Clare, L., Comas-Herrera, A., Corner, L., Daley, S., & Knapp, M., (2018). A roadmap to advance dementia research in prevention, diagnosis, intervention, and care by 2025.

International Journal of Geriatric Psychiatry, 33(7), 900–906. https://doi.org/10.1002/gps.4868

[224] Platt, J. (1998). Fast training of SVM using sequential optimization. In B. Scholkopf, C. Burges, and A. Smola (eds) *Advances in kernel methods – support vector learning*, pp. 185–208. MIT Press.

[225] Podhorna, J., Krahnke, T., Shear, M. E., & Harrison, J. for the Alzheimer's Disease Neuroimaging Initiative. (2016). Alzheimer's Disease Assessment Scale – Cognitive subscale variants in mild cognitive impairment and mild Alzheimer's disease: Change over time and the effect of enrichment strategies. *Alzheimer's Research and Therapy, 8*(8). https://doi.org/10.1186/s13195-016-0170-5

[226] Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

[227] Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: A systematic review and meta-analysis. *Alzheimer's & Dementia, 9*(1), 63–75. https://doi.org/10.1016/j.jalz.2012.11.00

[228] Qin, H. Y., Zhao, X. D., Zhu, B. G., & Hu, C. P. (2020). Demographic factors and cognitive function assessments associated with mild cognitive impairment progression for the elderly. *Biomed Research International,* Article 3054373. https://doi.org/10.1155/2020/3054373

[229] Quinlan, J. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106.

[230] Quinlan, J. (1993). C4.5: Programs for machine learning. Morgan Kaufmann.

[231] Quinlan, J. (1996). Bagging, boosting, and C4.5. *AAAI/IAAI, 1,* 725–730.

[232] Quinlan, J. (2002). *Data mining tools See5 and C5.0*. http://www.rulequest.com/see5-info.html.

[233] Rajab K. D. (2019) New Associative Classification Method Based on Rule Pruning for Classification of Datasets. in IEEE Access, vol. 7, pp. 157783-157795, 2019, doi: 10.1109/ACCESS.2019.2950374.

[234] Razani, J., Wong, J. T., Dafaeeboini, N., Edwards-Lee, T., Lu, P., Alessi, C., & Josephson, K. (2009). Predicting everyday functional abilities of dementia patients with the Mini-Mental State Examination. *Journal of Geriatric Psychiatry and Neurology, 22*(1), 62–70. https://doi.org/10.1177/0891988708328217

[235] Reppermund S (2021), 'The interplay between depressive symptoms, cognitive function, activities of daily living and cognitive reserve in older adults', International Psychogeriatrics, pp. 1 - 6, http://dx.doi.org/10.1017/S1041610221000508.

[236] Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie, 28*, 286–340.

[237] Rish, I. (2001). *An empirical study of the naive Bayes classifier*. Proceedings of IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence. New York, USA: IBM.

[238] Robin, L. A., Wang, C., Katz, M. J., Derby, C. A., Buschke, H., & Lipton, R. B. (2012). Predicting Alzheimer's disease: Neuropsychological tests, self-reports, and informant reports of cognitive difficulties. *Journal of the American Geriatric Society, 60*(6), 1128–34. https://doi.org/10.1111/j.1532-5415.2012.03956.x

[239] Rosen, W., Mohs, R., & Davis, K. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry, 141*(11), 1356–1364. https://doi.org/10.1176/ajp.141.11.1356

[240] Royall, D. R., Cordes, J. A., & Polk, M. (1988) CLOX: An executive clock drawing task. *Journal of Neurology and Neurosurgical Psychiatry, 64*(5), 588–94. https://doi.org/10.1136/jnnp.64.5.588

[241] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. Parallel distributed processing: Explorations in the microstructure of cognition. *Foundations, January*(1), pp. 318–362.

[242] Sachdev, P. S., Blacker, D., Blazer, D. G., Ganguli, M., Jeste, D. V., Paulsen, J. S., & Petersen, R. C. (2014). Classifying neurocognitive disorders: The DSM-5 approach. *Nature Reviews Neurology, 10*(11), 634. https://doi.org/10.1038/nrneurol.2014.181

[243] Sachdev, P., Brodaty, H., Reppermund, S., Kochan, N., Trollor, J., Draper, B., . . . Lux, O. (2010). The Sydney Memory and Ageing Study (MAS): Methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70–90 years. International Psychogeriatrics, 22(8), 1248-1264. doi:10.1017/S1041610210001067

[244] Sammut, C., & Webb, G. (2017). *Encyclopedia of machine learning and data mining.* Springer.

[245] Sangha, S., George, J., Winthrop, C., & Panchal, S. (2015). Confusion: Delirium and dementia - a smartphone app to improve cognitive assessment. *British Medical Journal, 4*(1). https://doi.org/10.1136/bmjquality.u202580.w1592.

[246] Scharre, D., Chang, S., Murden, R., Lamb, J., Beversdorf, D., Kataki, M., Nagaraja, H. N., & Bornstein, R. (2010). Self-administered gerocognitive examination (SAGE): A brief cognitive assessment instrument for mild cognitive impairment (MCI) and early

dementia. *Journal of Alzheimer Disease and Associated Disorders*, *24*(1), 64–71. https://doi.org/10.1097/WAD.0b013e3181b03277

[247] Scharre, D., Chang, S., Nagaraja, H., Vrettos, N., & Bornstein, R. (2017). Digitally translated self-administered gerocognitive examination (eSAGE): Relationship with its validated paper version, neuropsychological evaluations, and clinical assessments. *Alzheimer's Research & Therapy*, *9*(44). https://doi.org/10.1186/s13195-017-0269-3

[248] Schmidt, M. (1996). *Rey auditory verbal learning test: A handbook.* Los Angeles, CA., US: Western Psychological Services.

[249] Shahbaz, M., Niazi, A., Ali, S., Guergachi, A., & Umer, A. (2019). *Classification of Alzheimer's disease using machine learning techniques*. 8th International Conference on Data Science, Technology and Applications, (pp. 296–303). https://doi.org/10.5220/0007949902960303

[250] Shankle, W., Mani, S., Pazzani, M., & Smyth, P. (2005). Detecting very early stages of dementia from normal aging with machine learning methods. *Artificial Intelligence in Medicine*, 71–85.

[251] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

[252] Sikkes, S. A., de Lange-de Klerk, E. S., Pijnenburg, Y. A., Gillissen, F., Romkes, R., Knol, D. L., Uitdehaag, B. M. J., & Scheltens, P. (2012). A new informant-based questionnaire for instrumental activities of daily living in dementia. *Alzheimers and Dementia*, *8*(6), 536–43. https://doi.org/10.1016/j.jalz.2011.08.006

[253] Sindi, S., Calov, E., Fokkens, J., Ngandu, T., Soininen, H., Tuomilehto, J., & Kivipelto, M. (2015). The CAIDE dementia risk score app: The development of an evidence-based mobile application to predict the risk of dementia. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *1*(3), 328–333. https://doi.org/10.1016/j.dadm.2015.06.005

[254] Smith, A. (1982). *Symbol digit modalities test (SDMT). Manual (revised*). Los Angeles: Western Psychological Services.

[255] Snyderman, D., & Rovner, B. (2009). Mental status examination in primary care: A review. *American Family Physician*, *80*(8), 809–814.

[256] So, A., Hooshyar, D., Park, K., & Lim, H. (2017). Early diagnosis of dementia from clinical data by machine learning techniques. *Applied Sciences*, *7*(7), 651. https://doi.org/10.3390/app7070651

[257] Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack Jr., C. R., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., … & Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia, 7*(3), 2802–92. https://doi.org/10.1016/j.jalz.2011.03.003

[258] So, A., Hooshyar, D., Park, K., & Lim, H. (2017). Early diagnosis of dementia from clinical data by machine learning techniques. *Applied Sciences, 7*(7), 651. https://doi.org/10.3390/app7070651

[259] Stamate, D., Alghamdi, W., Ogg, J., Hoile, R., & Murtagh, F. (2018). *A machine learning framework for predicting dementia and mild cognitive impairment.* 17th IEEE International Conference on Machine Learning and Applications (ICMLA)2018. p. 671–8

[260] Tabatabaei-Jafari H., E Shaw M., Walsh E., Cherbuin N. (2020) for the Alzheimer's Disease Neuroimaging Initiative (ADNI), Cognitive/Functional Measures Predict Alzheimer's Disease, Dependent on Hippocampal Volume, The Journals of Gerontology: Series B, Volume 75, Issue 7, September 2020, Pages 1393–1402, https://doi.org/10.1093/geronb/gbz011

[261] Tappen, R., Rosselli, M., & Engstrom, G. (2009). Evaluation of the functional activities questionnaire (FAQ) in cognitive screening across four American ethnic groups. *Journal of the Clinical Neuropsychologist, 24*(4), 646–661. https://doi.org/10.1080/13854040903482855

[262] Tariq, S., Tumosa, N., Chibnall, J., Perry, M., & Morley, J. (2006). Comparison of the Saint Louis University mental status examination and the mini-mental state examination for detecting dementia and mild neurocognitive disorder—a pilot study. *The American Journal of Geriatric Psychiatry, 14*(11), 900–910.

[263] Teng, E., Becker, B. W., Woo, E., Knopman, D. S., Cummings, J. L., & Lu, P. H. (2010). Utility of the functional activities questionnaire for distinguishing mild cognitive impairment from very mild Alzheimer's disease. *Alzheimer Disease & Associated Disorders, 24*(4), 348–353. https://doi.org/10.1097/WAD.0b013e3181e2fc84

[264] Thabtah, F. (2108). Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Informatics for Health and Social Care, 43*(2), 1–20. https://doi.org/10.1080/17538157.2017.1399132

[265] Thabtah, F., Mampusti, E., Peebles, D., Herradura, M., & Varghese, J. (2019). A mobile-based screening system for data analyses of early dementia traits detection. *Journal of Medical Systems, 44*(1). https://doi.org/10.1007/s10916-019-1469-0

[266] Thabtah, F., Spencer, R., & Ye, Y. (2020). The correlation of everyday cognition test scores and the progression of Alzheimer's disease: A data analytics study. *Health Information Science and Systems, 8*(1). https://doi.org/10.1007/s13755-020-00114-8.

[267] Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatrics Society, 40*(9), 922–935. https://doi.org/10.1111/j.1532-5415.1992.tb01992.x

[268] Torre, J. (2004). Is Alzheimer's disease a neurodegenerative or a vascular disorder? Data, dogma, and dialectics. *The Lancet Neurology, 3(3),* 184–190. https://doi.org/10.1016/s1474-4422(04)00683-0

[269] Tsoi, K., Chan, J., Hirai, H., Wong, S., & Kwok, T. (2015). Cognitive tests to detect dementia: A systematic review and meta-analysis. *JAMA Internal Medicine, 175*(9), 1450–1458. https://doi.org/10.1001/jamainternmed.2015.2152

[270] Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

[271] Vasanthakumar, A., Davis, J.W., Idler, K. et al. Harnessing peripheral DNA methylation differences in the Alzheimer's Disease Neuroimaging Initiative (ADNI) to reveal novel biomarkers of disease. Clin Epigenet 12, 84 (2020). https://doi.org/10.1186/s13148-020-00864-y

[272] Vellas, B., Andrieu, S., Sampaio, C., Coley, N., Wilcock, G., & European Task Force Group (2008). Endpoints for trials in Alzheimer's disease: A European task force consensus. *The Lancet Neurology, 7*(5), 436–50. https://doi.org/10.1016/S1474-4422(08)70087-5

[273] Veloso A., Meira W., Zaki M., Goncalves M. and Mossri H. (2011). Calibrated Lazy Associative Classification. Information Sciences: an International Journal, Volume 13 (181), 2656-2670.

[274] Wang, Y., Chen, W., Cai, W., Hu, H., Xu, W., Wang, Z., Cao, X-P., Tan, L., Yu, J., & Alzheimer's Disease Neuroimaging Initiative. (2020). Associations of white matter

hyperintensities with cognitive decline: A longitudinal study. *Journal of Alzheimer's Disease, 73*(2), 759–768. https://doi.org/10.3233/jad-191005

[275] Waszynsk, C. (2012). *The Confusion Assessment Method (CAM).* New York University College of Nursing.

[276] Weakley, A., Williams, J. A., Schmitter-Edgecombe, M., & Cook, D. J. (2015). Neuropsychological test selection for cognitive impairment classification: A machine learning approach. *Journal of Clinical and Experimental Neuropsychology, 37*(9), 899–916. https://doi.org/10.1080/13803395.2015.1067290

[277] Wessels, A., Siemers, E., Yu, P., Andersen, S., Holdridge, K., Sims, J., Sundell, K., Stern, Y., Rentz, D. M., Dubois, B., Jones, R. W., Cummings, J., & Aisen, P. (2015). A combined measure of cognition and function for clinical trials: The integrated Alzheimer's Disease Rating Scale (iADRS). *The Journal of Prevention of Alzheimer's Disease, 2*(4), 227–241. https://doi.org/10.14283/jpad.2015.82

[278] West, S., McCue, R., & Golden, C. (2012). Does memory predict decline in activities of daily living in older adults with Alzheimer's disease? *Archives of Assessment Psychology, 2*(1), 32–43.

[279] Wharton, S.B., Wang, D., Parikh, C. et al. Epidemiological pathology of Aβ deposition in the ageing brain in CFAS: addition of multiple Aβ-derived measures does not improve dementia assessment using logistic regression and machine learning approaches. acta neuropathol commun 7, 198 (2019). https://doi.org/10.1186/s40478-019-0858-4

[280] Wharton S.B., et al., (2011) Medical Research Council Cognitive Function and Aging Study, Epidemiological neuropathology: the MRC Cognitive Function and Aging Study experience, J. Alzheimers. Dis. 25 (2011) 359–372.

[281] Wilkinson, I. M., & Graham-White, J. (1980). PGDRS: A method of assessment for use by nurses. *The British Journal of Psychiatry*, *137*, pp. 558–565. https://doi.org/10.1192/bjp.137.6.558

[282] Wimo, A., Guerchet, M., Ali, G. C., Wu, Y. T., Prina, A. M., Winblad, B., Jönsson, L., Liu, Z., & Prince, M., (2017). The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimer's & Dementia, 13*(1), 1-7. https://doi.org/10.1016/j.jalz.2016.07.150

[283] Wimo, A., Jönsson, L., Bond, J., Prince, M., Winblad, B., & Alzheimer's Disease International. (2013). The worldwide economic impact of dementia 2010. *Alzheimer's & Dementia, 9*(1), 1-11.e3. https://doi.org/10.1016/j.jalz.2012.11.006

[284] Wimo, A., Winblad, B., Aguero-Torres, H., & von Strauss, E. (2003). The magnitude of dementia occurrence in the world. *Alzheimer Disease & Associated Disorders, 17*(2), 63–67. https://doi.org/10.1097/00002093-200304000-00002

[285] Wimo, A., Winblad, B., & Jönsson, L. (2007). An estimate of the total worldwide societal costs of dementia in 2005. *Alzheimer's & Dementia, 3*(2), 81–91. https://doi.org/10.1016/j.jalz.2007.02.001

[286] Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., Nordberg, A., Bäckman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K,, de Leon, M., DeCarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., … & Petersen, R. (2004). C. Mild cognitive impairment-beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine, 256*(3), 240–6. https://doi.org/10.1111/j.1365-2796.2004.01380.x.

[287] Wind, A. W., Schellevis, F. G., Van Staveren, G., Scholten, R. P., Jonker, C., & Van Eijk, J. T. (1997). Limitations of the Mini-Mental State Examination in diagnosing dementia in general practice. *International Journal of Geriatric Psychiatry, 12*(1), 101–108. https://doi.org/10.1002/(sici)1099-1166(199701)12:1<101::aid-gps469>3.0.co;2-r

[288] Witten, I., & Frank, E. (2002). *Data mining: Practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann.

[289] Wittenberg, R., Hu, B., Barraza-Araiza, L., & Rehill, A. (2019). *Projections of older people with dementia and costs of dementia care in the United Kingdom, 2019-2040.* The London School of Economics and Political Science, Care Policy and Evaluation Centre. https://www.alzheimers.org.uk/sites/default/files/2019-11/cpec_report_november_2019.pdf

[290] World Health Organization. (2020, September 21). *Dementia.* https://www.who.int/news-room/fact-sheets/detail/dementia

[291] Wright, B. D. (1992). IRT in the 1990s: Which models work best? *Rasch Measurement Transactions, 6*(1), 196–200.

[292] Xefteris, S., Konstantinidis, E., Billis, A. S., Antoniou, P. E., Styliadis, C., Paraskevopoulos, E., Kartsidis, P. E., Frantzidıs, C. A. & Bamidis, P. D., (2020). Early detection of dementia: Advances, challenges, and future prospects. In *Data analytics in medicine: Concepts, methodologies, tools, and applications* (pp. 1963–1988). IGI Global.

[293] Xue, M., & Zhu, C. (2009). *A study and application on machine learning of artificial intelligence*. 2009 International Joint Conference on Artificial Intelligence, 272–274.

[294] Yamagata, C., & Kowto, M. (2013). *Mobile app development and usability research to help dementia and Alzheimer patients*. Systems, Applications and Technology Conference (LISAT). Long Island, NY.

[295] Yang, H. and Bath, P.A. orcid.org/0000-0002-6310-7396 (2020) The use of data mining methods for the prediction of dementia : evidence from the English longitudinal study of aging. IEEE Journal of Biomedical and Health Informatics, 24 (2). pp. 345-353. ISSN 2168-2194.

[296] Yang, H., Cheng, Z., Li, Z., Jiang, Y., Zhao, J., Wu, Y., Gu, S., & Xu, H. (2019). Validation study of the Alzheimer's Disease Assessment Scale-Cognitive Subscale for people with mild cognitive impairment and Alzheimer's disease in Chinese communities. *International Journal of Geriatric Psychiatry, 34*(11), 1658–1666. https://doi.org/10.1002/gps.5179

[297] Yesavage, J., Brink, T., Rose, T., Lum, O., Huang, V., Adey, M., & Leirer, V. (1982). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research, 17*(1), 37–49. https://doi.org/10.1016/0022-3956(82)90033-4

[298] Yoon B, Shim YS, Cheong HK, Hong YJ, Lee KS, Park KH, Ahn KJ, Kim DJ, Kim YD, Choi SH, Yang DW. White matter hyperintensities in mild cognitive impairment: clinical impact of location and interaction with lacunes and medial temporal atrophy (2014) J Stroke Cerebrovasc Dis. 2014 May-Jun;23(5):e365-72. doi: 10.1016/j.jstrokecerebrovasdis.2013.12.040. Epub 2014 Feb 28.

[299] Youn, Y. C., Choi, S. H., Shin, H. W., Kim, K. W., Jang, J. W., Jung, J. J., Hsiung, G-Y. R., & Kim, S. (2018). Detection of cognitive impairment using a machine-learning algorithm. *Neuropsychiatric Disease and Treatment, 14*, 2939–2945. https://doi.org/10.2147/NDT.S171950

[300] Yu, S. T. M., Yu, M-I., Brown, T., & Andrews, H. (2018). Association between older adults' functional performance and their scores on the Mini Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA). *Irish Journal of Occupational Therapy*, *46*(1), 4–23. https://doi.org/10.1108/IJOT-07-2017-0020

[301] Zahodne, L. B., Manly, J. J., MacKay-Brandt, A., & Stern, Y. (2013). Cognitive declines precede and predict functional declines in aging and Alzheimer's disease. *PLOS One, 8*(9), Article e73645. https://doi.org/10.1371/journal.pone.007364

[302] Zhang, H. (2004). The optimality of naive Bayes. *Association for the Advancement of Artificial Intelligence*, *1*(2), 3.

[303] Zhang H. (2022) Nursing Diagnosis of Urology Operating Room Based on New Association Classification Algorithm", Journal of Healthcare Engineering, vol. 2022, Article ID 4674959, 13 pages, 2022. https://doi.org/10.1155/2022/4674959

[304] Zhu, F., Li, X., Haipeng, T., He, Z., Zhang, C., Hung, G.-U., Chiu, P-Y., & Zhou, W. (2020). Machine learning for the preliminary diagnosis of dementia. *Scientific Programming*. https://doi.org/10.1155/2020/5629090

[305] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology*), *67*(2), 301–320.

[306] Zucchella, C., Bartolo, M., Bernini, S., Picascia, M., Malinverni, P., & Sinforiani E. (2017). Modeling Alzheimer disease through functional independence and participation. *Alzheimer Disease and Associated Disorders, 31*(3), 218–224.