Detection of Dementia Progression from Functional Activities Data using Machine Learning Techniques

Fadi Thabtah
ASDTests
Auckland, New Zealand

Swan Ong
Manukau Institute of Technology
Auckland, New Zealand

David Peebles
Department of Psychology
University of Huddersfield
Huddersfield, UK

"for the Alzheimer's Disease Neuroimaging Initiative*"[1]

Abstract

Early screening for Alzheimer's disease (AD) is crucial for disease management, intervention, and healthcare resource accessibility. Medical assessments of AD diagnosis include the utilisation of biological markers (biomarkers), positron emission tomography (PET) scans, magnetic resonance imaging (MRI) images, and cerebrospinal fluid (CSF). These methods are resource intensive as well as physically invasive, whereas neuropsychological tests are fast, cost effective, and simple to administer for providing early AD diagnosis. However, neuropsychological assessments contain elements related to executive functions, memory, orientation, learning, judgment, and perceptual motor function (among others) that overlap, making it difficult to identify the key elements that trigger the progression of dementia or mild cognitive impairment (MCI). This research investigates the elements of the Functional Activities Questionnaire (FAQ) an early screening method using a data driven approach based on feature selection and classification. The aim is to determine the key items in the FAQ that may trigger AD advancement. To achieve the aim, real data observations of the Alzheimer's Disease Neuroimaging Initiative (ADNI) project have been processed using the proposed data driven approach. The results derived by the machine learning techniques in the proposed approach on data subsets of the FAQ items with demographics show models with accuracy, sensitivity, and specificity all exceeding 90%. In addition, FAQ elements including Administration and Shopping related activities showed correlations with the progression class; these elements cover four out of the six Diagnostic and Statistical Manual's (DSM-5's) neurocognitive domains.

Keywords: Alzheimer's disease**,** classification**,** clinical informatics, data analysis, FAQ**,** ADNI**,** machine learning

1. Introduction

Dementia as a clinical condition characterised by the progressive deterioration of cognitive functions which is associated with loss of functional and behavioural abilities to the extent that daily life is impeded [1]. A prevalence of neuropsychiatric symptoms also occurs including apathy, agitation, hallucinations, and depression [2]. AD is the most common form of dementia affecting almost two thirds of dementia sufferers [3]. A person diagnosed with dementia loses their ability to think and act coherently, is incapable of controlling their emotions, and unable to perform day-to-day tasks independently [4].

[1] *Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

According to the Dementia Economic Impact Report 2016, it is estimated that 70,000 people are living with dementia in New Zealand and this number is expected to increase to over 170,000 by 2050 [5]. Globally, it is noted that the economic and societal costs associated with dementia are estimated to exceed 1 trillion USD due to high costs of care and loss of productivity [6]. Although this disease is not considered a normal part of aging, the risk of the disease increases with age. This sparks an urgent call for action to curb this rising disease due to population growth and increasing longevity.

While there are currently many cognitive assessments available to diagnose the disease, the exact causes are not yet fully understood and there is no treatment available [7]. It is imperative that more resources are invested into raising awareness, improving research, and establishing clinical trials to understand the disease so as to provide effective treatment and better care [8]. General practitioners (GPs) may conduct a series of medical exams to diagnose possible or probable AD, in the hope of achieving a level of certainty in the etiological diagnosis. A personal medical records assessment would be the first step, where a GP reviews a patient's medical history, medication history and symptoms over time to assesses any changes in the patient's overall health, including mood, memory, and behaviours [9]. More importantly, the patient's mental status would be evaluated by utilising neuropsychological tests that evaluate the patient's memory, cognition, learning, problem solving, and executive functions among others [7].

The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) describes the conditions for the diagnosis of neurocognitive disorders such as AD (possible AD/ probable AD) [10]. To be more specific, major neurocognitive disorder requires that the individual exhibits a significant decline over time in at least one of the six cognitive domains: complex attention, executive function, learning and memory, language, perceptual-motor, or social cognition. This decline can be reported either by the individual, an informant, or the GP and then validated by conducting a cognitive assessment that is associated with the domain.

There are many neuropsychological tests to screen for MCI and dementia conditions such as the Montreal Cognitive Assessment (MOCA), the Clinical Dementia Rating Scale Sum of Boxes (CDR-SB), the Mini Mental State Examination (MMSE), the Alzheimer's Disease Assessment Scale - Cognitive Subscale (ADAS-Cog), the Functional Activities Questionnaire (FAQ), and others [11-15]. These tests evaluate one or multiple cognitive areas for patients and according to study conducted by [16], have demonstrated good performance in detection dementia). However, there are limited studies on how individual items or activities within these tests measure the impact of the progression of the AD [17, 18]. Items within the same test may fully or partly cover certain cognitive domains, so according to the DSM-5, it is imperative to identify the few items within a neuropsychological test that may trigger any changing point in the disease for patients. For instance, the first two items in the FAQ test cover domains including complex attention, and partly, learning and memory. In addition, the MMSE test contains multiple questions related to memory and recall.

This research aims to identify key functional items in the FAQ test that may trigger a change in the condition of the diagnostic class of AD. More precisely, we would like to identify the impactful FAQ items that can help clinicians in detecting any progression of dementia, i.e., from Cognitively Normal (CN) to MCI or MCI to AD. We also want to investigate whether functional activities are enough indicator for the disease progression. To achieve the aim, we propose a data driven approach based on machine learning that integrates multiple datasets from ADNI (ADNI-Merge and FAQ data sheets) and model the new dataset to capture any progression of the subjects' diagnostic class. The modelled dataset is processed using feature selection methods and machine learning algorithms to derive classification models that help in detecting a progression of the AD at any stage. Machine learning techniques were used medical arenas, such as dementia detection, to improve the performance of medical methods [19-23]. Our research expands existing research on dementia progression by trying to explain influential functional elements which can help in detecting early change of AD, therefore quick intervention [24, 25].

The research question that this study seeks to answer is:

What are the features of the FAQ test that may trigger the progression of AD?

The scope of the research is limited to functional assessments—results related to other assessments including neuroimaging, pathological, or genetics are excluded. In this research, we contribute the following:

1. The ability to detect the AD advancement with few functional features using a data driven model based on classification algorithms that can be utilised for early screening

2. To reveal the true performance of predictive machine learning algorithms on the problem of AD progression, which can help future researchers in developing problem specific techniques

3. To improve the performance of current medical screening methods related to neuropsychological tests with reference to predictive accuracy

4. To determine a few functional items related to FAQ that are key to discover AD progression so medical professionals can use them as a digital information sheet and map them to DSM-5 framework during the screening process

We hope that this research offers more insight to improve existing assessment protocols for the purpose of screening and diagnosing AD. Our research can help in the design and implementation of a cost-effective AD progression system using innovative technologies such as artificial intelligence. By identifying fewer impactful functioning ability items within the FAQ, potentially clinicians can use such knowledge within information sheets to aid the early screening process of AD. This is since functional items necessitates cognitive abilities to perform functional tasks like paying a bill, shopping, and cooking among others. It is the firm belief of the authors that little research has been undertaken using machine learning on ways to detect progression of AD, at least for the FAQ items, and therefore this research fills the gap by providing more insights on how FAQ items are correlated with the diagnostic changes. Moreover, this research also pinpoints the items of FAQ that are linked within the DSM-5 neurodegenerative areas as illustrated in Table 1. The FAQ items have been associated with the relevant cognitive domains as according to the nature of the task to be performed by the patient. For example, FAQFINAN consists of financial tasks such as writing checks and paying bills, it taps into three cognitive domains which are: complex attention that is required for mental calculation; executive functioning for planning and organising; and learning and memory to recall recent events [10]. While FAQGAME and FAQBEVG also tap into the same three cognitive domains to play games or make a hot drink, they also involve the perceptual motor function domain that carries out routine actions and the familiar use of tools and appliances. More analysis on the FAQ items and its associated cognitive domains will be discussed in the results analysis in Section 6 based on the results derived from this research.

Table 1: the FAQ items and its associated DSM-5 cognitive domains

Question No.	**Column in Dataset**	**Tasks**	**Task Type**	**DSM-5 Cognitive Domains**
1	FAQFINAN	Writing checks, paying bills, balancing check book	Finance	Complex attention Executive functioning Learning and memory
2	FAQFORM	Assembling tax records, business affairs, or other papers	Administration	Complex attention Executive functioning Learning and memory

3	FAQSHOP	Shopping alone for clothes, household necessities, or groceries	Shopping	Complex attention Executive functioning Learning and memory Perceptual motor function
4	FAQGAME	Playing game of skill, working on a hobby	Leisure and hobbies	Complex attention Executive functioning Learning and memory Perceptual motor function
5	FAQBEVG	Heating water, making a cup of coffee, turning off stove after use	Beverage-making	Complex attention Executive functioning Learning and memory Perceptual motor function
6	FAQMEAL	Preparing a balanced meal	Meal preparation	Complex attention Executive functioning Learning and memory Perceptual motor function
7	FAQEVENT	Keeping track of current events	Current affairs	Learning and memory
8	FAQTV	Paying attention to, understanding, discussing TV, book, magazine	Engagement with media	Complex attention Executive functioning Learning and memory Social cognition
9	FAQREM	Remembering appointments, family occasions, holidays, medications	Personal memory	Learning and memory
10	FAQTRAVL	Traveling out of neighbourhood, driving, arranging to take public transportation	Travel	Complex attention Executive functioning Perceptual motor function

2 Literature Review

The FAQ is an instrument for screening dementia that focuses on measuring the instrumental activities of daily living (IADLs) [26]. The FAQ is useful to monitor functional changes to determine which ADLs are impacted to help clinicians to predict potential progression of dementia. The FAQ differentiates between MCI and mild AD and provides a sensitivity of 85% and inter-rater reliability of 97.00% [27]. The FAQ consists of 10 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. This research is limited to the FAQ items within the ADNI data repository. The remaining part of this section critically analyses research that is related to the ADNI data repository and focused on the progression of AD and machine learning techniques.

[26] 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 clinically normal (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 (MADRC, n.d.) [28]. In their methodology, the authors commented that there is no established cut-off score for IADL impairment on the FAQ, however they have referred to a study where a score of ≥ 6 is suggestive of functional impairment [29]. 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 at 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.

[29] investigated the capability of the FAQ test to clinically distinguish between MCI and very mild AD. They have utilised the National Alzheimer's Coordinating Centre (NACC) Uniform Data Set (UDS) [30], 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. Stepwise logistic regression method [31] was used in this study to determine which FAQ items are 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 - like 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.

A study conducted by [32], 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 FAQ test, their research presents useful information of data mapping, data integration, and feature selection techniques. Support vector machine (SVM) classifiers [33] 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 test, along with ADAS amongst five clinical assessments, to have better weighting in diagnosing AD.

[34] utilised machine learning techniques to minimize any degree of subjectivity arising in some clinical assessments of dementia. The authors believed that there are many machine learning methods used to study neuroimaging, more than there is for cognitive, behavioural, and functional studies. Thus, they consider that if more focus is given to functional and cognitive assessments, this would lead to optimising or even reducing the quantity of tests needed to diagnose AD patients at an early stage of impairment. Consequently, this may reduce cost, time required, and cognitive stress of the assessments. Two different feature reduction approaches have been used by the authors: (a) a computational approach, based on the mathematical discriminatory power of features among classes, and (b) an approach based on our basic understanding of the redundancy of features. The author also stated that in the latter approach, neuropsychologists' understanding of the disease guided the classification model, highlighting that there was strong domain expertise when conducting this approach. The study also uses SVMs [35] to generate a predictive model for classifying the subjects into different groups based on clinical datasets. Their findings show that FAQ scores, along with ADAS scores, were some of the best predictors for the classification of AD; this shares the same findings as [32].

[36] exploited ADNI data, aiming to predict longitudinal changes of cognitive decline in MCI patients; 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 normal condition / control (CN). They managed to achieve 84.2% accurate prediction using a convolutional neural network (CNN)-based approach [37] for the conversion of MCI to AD. Receiver operating characteristic (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 FAQ score attribute is positively correlated with longitudinal changes, and the correlation was noticeably significant after 3 years following the initial MCI diagnosis, compared with 1 year following the initial MCI diagnosis.

[38] contrasted a number of classification algorithms - Naive Bayes, Bayes Network, Bagging, Logistic Regression, Random Forest, SVM and Multilayer Perceptron (MLP) [31, 39-44], 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 [45, 46] 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.

3. Methodology

Figure 1 illustrates the methodology followed to answer our research question. The process is divided into three phases, beginning with data integration, followed by data modelling and pre-processing, and finally data analysis using feature assessment and classification.

Data integration involved the retrieval of two datasets, ADNI-Merge and FAQ-sheet from the ADNI data repository, where a literature review as per Section 2 was conducted in parallel to assist our understanding of the datasets. The two datasets are merged using two attributes which are the patient ID (RID) and the visit code as a primary key reference. The aim of the merging process is to capture individual FAQ items' scores from the FAQ-sheet and the diagnostic class (DX) from ADNI-Merge for each visit and per patient. The merging of the two datasets also involves a comparison of the FAQTOTAL scores between the two datasets to check for any score mismatches. For some individuals, mis-recorded FAQ item scores were identified by comparing the FAQ total scores between the datasets and screening the FAQ item scores for those outside the possible range of 0 to 3. Thus a validation was required to ensure the FAQTOTAL scores were correct between the two datasets. There are also instances when a data observation (a patient visit) in the ADNI-Merge has no corresponding FAQ-sheet data thus no merging occurs. Only the data observations that have presence in both datasets (ADNI-Merge and FAQ-sheet) have been integrated tin the new dataset (ADNIFAQ).

For data modelling, since the key element we want to capture for the analysis is the progression of the disease, the diagnosis (DX) for each data observation is labelled and modelled into a new class (DX Progress), according to the progress of diagnosis from the previous visit of the same patient. The process we followed to derive the DX Progress class is as follows:

1. The dataset is multi-sorted first by patient ID (RID) and then the visit code (VISCODE2).
2. A new attribute called 'DX Digit' is added to map the four types of diagnosis from the DX column, which are labelled '1' for CN, '2' for MCI, '3' for dementia and '0' for those not defined.
3. A second attribute called 'DX Progress' was created to capture the change of diagnosis in the patient's subsequent visits, modelling the labelled data in 'DX digit'. When there is 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. If there is no progression, it is labelled as '0'. Regression as '-1', and for undefined change as '2'.
4. For each patient's baseline visit, the DX Progress column is labelled as '0' to ensure that the class does not model the diagnosis of a previous patient's observation.

Once the DX Progress attribute has been derived, we proceed with cleansing the data to focus only on values that are either 1 (progression of AD at any stage) or 0 (no progression). Prior to any feature assessment or classifier construction phases, we removed rows in the FAQTOTAL column with missing values or with values -1 due to incomplete individual scores of some functional ability items in the FAQ test. In addition, rows with No DX Found in the DX column were deleted. We then filter out the irrelevant attributes such as the patient information and biomarkers, which are not the focus of our research. The refined dataset (ADNIFAQ) consists of 16 features including DX Class as shown in Section 4 (Table 5), which is then used for further analysis as described later in Section 5.

Upon deriving the new class, i.e., DX Progress, we discovered that the class set was highly imbalanced (more data observations are linked with 0 - no progression than with 1 - progression) and when processed by machine learning techniques, this can result in a bias and skewed classifier that favours

the majority class. Consequently, we used the data balancing method known as the Synthetic Minority Oversampling Technique (SMOTE) to achieve a balanced and randomly distributed dataset in preparation for the data analysis phase [47]. SMOTE randomly generates new observations to bring the minority class of 1 closer to the number of the majority class of 0. The randomisation technique is also applied to ensure that the newly generated minority class, which will sit at the end of the dataset, is distributed randomly so that during the cross-validation process, the 1s do not bunch up in some of the folds.

The final phase of the methodology (data analysis) involves feature selection and classification techniques aiming to derive the sensitive FAQ items for the class variable (DX Progress) to predict a probable progression of diagnosis. In our experiments in the data analysis phase, subsets of distinct feature sets were derived based on in-depth feature assessment results derived by dissimilar feature selection techniques including Chi-square testing (CST), Information Gain (IG), and correlation analysis (CA) [41, 45, 48]. More importantly, the Pearson correlation matrix model [49] was used to gain an understanding of the relationship between the FAQ items, without the class attribute, which indeed can help in determining functional items similarities of FAQ. In addition, we measured the relevancy of each FAQ item using the Leave One Out Cross Validation (LOOCV) method with a Naïve Bayes classifier [39]. The reasons for choosing these feature selection methods are that they use dissimilar mathematical models to define feature relevancy, and they were utilised successfully in previous medical research related to dementia such as by [38, 50, 51].

We proceed by normalising the results obtained by the feature selection methods to maintain a similar scale for all the results and to simplify the analysis. The feature assessment analysis was based on different criteria including:

1) Top-ranked features based on the scores assigned by the feature selection methods
2) Common features among the feature selection methods' results
3) Similarity of features to obtain highly influential yet dissimilar features

To achieve the above outcomes, a few distinct features subsets have been obtained for further data processing using machine learning techniques. These subsets are included in addition to the FAQ-items' demographic features (Age, Gender, Level of education, PTRACCAT, Marital status). More details on these features' sets are given in Section 6.

With the derived subsets, the next stage of the experimental methodology is to process the cognitive features (FAQ-items) using various classification algorithms to derive predictive models for AD progression. This necessitates processing the dissimilar datasets of the distinctive features sets with and without considering the demographical features. Three different classification techniques, including Bayesian Network (Bayes Net), Logistic Regression (LR), and C4.5 (Decision Tree) have been used for deriving the AD progression models. These algorithms employ dissimilar learning methods to construct the classifier and have been widely used in medical screening and diagnosis [29, 38, 51, 52]. More details on how these classification algorithms work is given in Section 6. The performance of the classification models has been measured using predictive accuracy, sensitivity, and specificity.

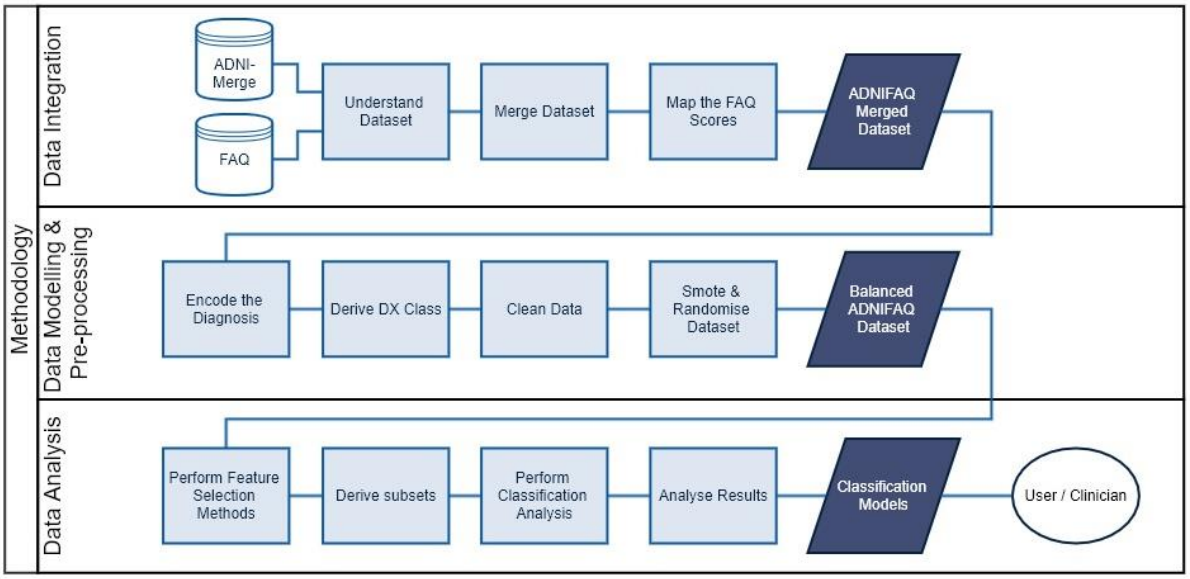

Figure 1. Methodology followed

4. Data and Features

Data used in this research has been obtained from ADNI[2]. ADNI is a longitudinal multi-centre study with a cohort of 1,900 participants designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD [53]. It consists of four different phases: ADNI1, ADNIGO, ADNI2, and ADNI3, with each phase conducted with a different primary goal, timeframe, funding, and group of participants. The participants in the ADNI project are primarily based in the United States and Canada and each phase has a different start date and trial duration. The study is longitudinal as the participants are monitored on a six to twelve monthly basis, and thus have multiple observations in the metadata but at different points in time.

The datasets used in this research, ADNI-Merge and FAQ-sheet, are obtained from the ADNI data repository [54] after the approval of data access. The FAQ-sheet contains FAQ items and their scores, whereas the ADNI-Merge dataset combines key features from four different phases: ADNI1, ADNIGO, ADNI2, and ADNI3. These two datasets are usually used for data analysis regarding improving early detection and tracking of AD. The ADNI-Merge dataset contained 14,627 observations and 113 attributes with data related to patient information, clinical, genetic, MRI image, PET image, and biospecimen results. The FAQ-sheet dataset contained 10,905 observations and 23 attributes with data related to the patients' information, individual FAQ items' score, and total FAQ score.
We extract 6 attributes of interest from the ADNI-Merge as described earlier into the FAQ-sheet dataset which is used as a base. As the FAQ-sheet dataset has 3,722 observations, fewer than the ADNI-Merge dataset, the excess rows from the ADNI-Merge are excluded so no merging occurred. The newly merged dataset is called ADNIFAQ, with 10,905 observations and 30 attributes remaining. Table 2 summarises the data description before pre-processing.

Table 2: General statistics of the datasets before pre-processing

[2] Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

Dataset Name	# of Features	# of Patients	# of Data Observations (visits)	Missing Values in Key Attributes
ADNI-Merge	113	2,260	14,627	DX: 4,243 missing values
FAQ-Sheet	23	2,267	10,905	FAQTOTAL: 131 invalid data (99 missing values; 32 incomplete data (-1))
ADNIFAQ	30	2,267	10,905	FAQTOTAL: 131 invalid data (99 missing values; 32 incomplete data (-1)) DX: 383 'No DX' Found

While the FAQTOTAL was within the 0–30 range, the individual FAQ items had a score range from -1 to 5, and when tallied, would exceed the 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 [55], there were six answers to choose from as given in Table 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 perhaps used to reduce data entry error as some of the answers share similar FAQ scores. As a result, we mapped the six answers from the FAQ dataset to their corresponding FAQ scale values to check whether the FAQTOTAL matched.

Table 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

Table 4: ADNIFAQ DX class statistics after data modelling

DX Class	Class Representation	# of Observations
1	Progression	558
0	No change	9,899
-1	Regression	141
-2	Invalid class due to missing diagnosis	307

As the scope of our research focuses on whether there is a progression (1) or no change (0) in diagnosis (the new modelled class DX Progress), we removed observations with remission (-1) or missing class (null). Sixteen attributes that were unnecessary for the modelling were filtered out of the analysis dataset. This resulted in 10,265 data observations and 16 attributes retained in the dataset: FAQFINAN, FAQFORM, FAQSHOP, FAQGAME, FAQBEVG, FAQMEAL, FAQEVENT, FAQTV, FAQREM, FAQTRAVL, AGE, PTGENDER, PTEDUCATE, PTRACCAT, PTMARRY, and the DX Progress. Table 1 maps each of these functional attribute to the task and FSM-5 cognitive domain.

From the outcome of the DX Progress, we have detected massive class-disproportion with the 'no change' (0) having a large majority at 95% to those with 'progression' (1) at only 5%. Due to the probability of bias from using a class-imbalanced dataset, we used the SMOTE method to balance the ADNIFAQ dataset resulting in a revised version of 18,545 observations ready for analysis. The DX Progress distribution before and after data balancing is displayed in Table 5.

Table 5: ADNIFAQ general statistics after data pre-processing

Dataset Name	# of Features	# of Patients	# of Data Observations (visits)	DX Class Distribution before Data Balancing (1:0)	DX Class Distribution after Data Balancing (1:0) using SMOTE

ADNIFAQ	16	2,244	10,265	Total observations: 10,265 '0': 9713 (majority 95%) '1': 552 (5%)	Total observations: 18,545 '0': 9713 (52%) '1': 8832 (48%)

5. Experimental Setting and Methods Used

All experiments have been conducted using the Waikato Environment for Knowledge Analysis (Weka)—an open-sourced software that provides a comprehensive collection of machine learning algorithms and data pre-processing tools [56]. We have utilised Weka version 3.8.4 to perform data sampling techniques, multiple feature selection methods to derive subsets which become the input of the classification models, and also used various classification methods to predict a diagnostic class. Ten-fold cross validation has been utilised during the experiments as a measure of testing to ensure less biased results are derived. Using ten-fold cross validation, the input dataset is divided into 10 partitions arbitrary with stratification [57]. Then nine partitions are used for training and the remaining partition for testing; the procedure is repeated ten times to derive the performance measure results.

Since the ADNIFAQ dataset is imbalanced, SMOTE was used as a data sampling method with cross validation to balance the class label, i.e., DX Progress. SMOTE is a statistical technique that adjusts the class by taking the entire dataset as input increasing the minority class only [58]. It takes n features, then considers its K nearest neighbours (KNN) [59] in the oversampling process as shown in Equation (1) [60].

$$S = x + u.(x^r - x), \quad \text{with } 0 \le u \le 1, \tag{1}$$

where u was randomly chosen from $\mathrm{U}(0,1)$,

x is a set of variables,

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

In SMOTE, the randomize function randomly shuffles the order of the observations within the dataset. The aim is to ensure the new minorities are distributed randomly across the dataset to prevent clusters in the cross-validation folds.

For feature assessment we used CA, IG, and CST. The PC Coefficient was used to measure the linear correlation among the FAQ items excluding the class variable and the results were presented in a matrix model. This indeed pinpoints to which activities are similar, allowing us to suggest fewer FAQ items needed for the progression of AD. The correlation coefficient is calculated using Equation (2) and derived a value in the range of [-1, +1] [49]. The closer the coefficient number towards -1 or +1, the higher the dependency between the features.

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

Where $r = Pearson\ Coefficient$,

$n = number\ of\ variables$

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

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

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

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

IG is a key feature selection method that decides the ordering of features in the nodes of a Decision Tree (DT) classifier [61]. IG can be calculated as

$$IG(T,X) = Entropy(T) - Entropy(T,X) \quad (3.1)$$

$$\text{and} \quad Entropy(x) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i), \quad (3.2)$$

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 determines if two features are associated using the expected and actual frequencies in the dataset according to Equation (5) [45].

$$x^2 = \sum_{i=1}^{2} \sum_{i=1}^{n} \frac{(A_{ij}-E_{ij})^2}{E_{ij}} \quad (4)$$

Where:

k= number of (no.) classes,

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

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 = $\sum_{i=1}^{2} R_i$,

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

We used the LOOCV method with a Naïve Bayes classifier to evaluate the relevancy of a feature in the ADNIFAQ dataset when that feature is present versus when it is absent. In other words, the Naïve Bayes classifier will process the ADNIFAQ dataset when feature X is present (Case A) and when feature X is absent (Case B). Then feature X's relevancy is measured by computing the difference in predictive accuracy between Case A and Case B.

Several dissimilar classification algorithms have been utilized to build classification models for AD progression including Bayes Net, LR, and C4.5. Bayes Net illustrates a set of variables and its conditional dependencies. [40] explain it as a search algorithm for a network that "best describes" the probability distribution over the training data. On the other hand, LR 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 [31]. It is often used as a classification and prediction model where the results are divided into specific categories [38]. Lastly, C4.5 is a DT classifier that calculates the expected information value for features in the training dataset to construct a tree-based classification model for prediction [62]. Once the tree is constructed, then C4.5 simplifies it by trimming unnecessary sub-trees that may overfit the derived classification model. In the tree, each path from the root node to the leaf denotes a classification rule in which its body represents attribute values and its consequent class variable.

All experiments have been conducted on a computing machine with Intel® Core™ i5-6200U 2.3 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.

To assess the FAQ items in the models derived by the classification algorithms, we have evaluated the performance of the classification models by measuring their accuracy, sensitivity, and specificity percentages using the confusion matrix. Accuracy is to assess the model's ability to correctly classify

an observation as true, whether there was progression or no change. Sensitivity is to assess the ability to determine progression correctly. Specificity is to assess the ability to determine no change correctly.

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

$$Sensitivity = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Specificity}=\frac{TN}{TN+FP} \quad (7)$$

where

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

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

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

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

6. Results Analysis

Feature Selection Results Analysis

To identify the FAQ items that are more sensitive towards the progression of AD, we applied multiple feature selection techniques to filter out attributes that are not useful to the predictive models for AD progression. Table 6 depicts the scores computed by the feature selection methods on the ADNIFAQ dataset along with the drop/increase in the accuracy metric for each feature using the LOOCV method with a Naïve Bayes classifier. We have normalised the scores computed by each feature selection method as the three methods produced figures that varied in range and scale. Normalising and averaging the results enabled ease of comparison for FAQ items which were then ranked accordingly.

Table 6 illustrates that Finance, Shopping, Administration and Personal memory as the top ranked features, at least using the feature selection methods considered. These items tap into multiple cognitive domains including executive function, complex attention, memory, and perceptual motor function as illustrated in Table 1, where they involve everyday tasks that consist of finance, shopping, administration, and remembering, which require multi-tasking, high level of concentration, recalling recent events and performing routine skills. These results are comparable with the findings of the correlation matrix model shown in Figure 2, with all four items having a coefficient of ≥ 0.70 with at least one other FAQ item. Similarly, the tasks performed in some of these top four items all involved overlapping cognitive domains. For example, two of the top features based on the average scores produced by the feature selection methods are Finance and Administration; these are mainly linked with complex attention, executive function, and memory according to the DSM-5 cognitive domains, thus have high similarity. According to Figure 2, the correlation between these two features is 0.85 which is relatively high.

To gain further understanding of the relationship between the FAQ items and to identify similar ones, we conducted a feature-to-feature analysis using a correlation matrix derived by the Pearson correlation coefficient method as shown in Figure 2. Multiple FAQ functional activities including Finance, Administration, Meal preparation, Personal memory and Travel have high correlation coefficients of ≥ 0.70 with at least one other item. The FAQ items' similarities were also confirmed when mapped against their respective DSM-5 cognitive domains (see Table 1) as multiple FAQ items are overlapping in certain neurocognitive domains particularly complex attention, executive functioning, memory, and perceptual motor function. These identified cognitive domains match the FAQ items derived from the feature selection methods and the LOOCV method as per Table 6.

Table 7 describes the subsets of FAQ features that we have decided to form for predictive analysis based on the feature assessment outcomes. The choice of producing these distinctive items subsets is based on the following criteria:

a) Top ranked items based on the average scores of the feature selection methods and cognitive coverage of the DSM-5
b) Items that when they are absent from the dataset impact the classifier negatively in terms of accuracy according to LOOCV results

Table 6: FAQ items with computed scores and normalized scores derived by the feature selection methods

	IG		CA		CST		Average Scores	LOOCV with Naïve Bayes Accuracy %
Feature	Score	Normalised	Score	Normalised	Score	Normalised Score		
FAQFINAN	0.064	0.968	0.217	1.000	1610.665	0.977	0.982	0.00464
FAQFORM	0.061	0.905	0.214	0.974	1542.824	0.920	0.933	0.00563
FAQSHOP	0.065	0.998	0.207	0.912	1618.804	0.984	0.965	0.00496
FAQGAME	0.049	0.623	0.201	0.860	1229.347	0.654	0.712	0.00164
FAQBEVG	0.022	0.000	0.103	0.000	456.453	0.000	0.000	0.00550
FAQMEAL	0.047	0.572	0.163	0.526	1168.080	0.603	0.567	0.00318
FAQEVENT	0.054	0.738	0.179	0.667	1313.117	0.725	0.710	0.00245
FAQTV	0.053	0.711	0.203	0.877	1289.698	0.706	0.764	0.00375
FAQREM	0.065	1.000	0.194	0.798	1637.527	1.000	0.933	-0.00111
FAQTRAVL	0.053	0.711	0.202	0.868	1328.007	0.738	0.772	0.00677

Based on the results obtained by the LOOCV with a Naïve Bayes classifier, Travel, Administration, Beverage-making, and Shopping have shown to have the highest impact in terms of predictive accuracy. Travel and Beverage-making were newly identified functional ability items by LOOCV method and have 0.43 correlation as shown in the correlation matrix results. There is an overlapping between these two items in one DSM-5 cognitive domain: perceptual motor function, which can be a reason for 0.43 correlation result. Therefore, we create a new feature subset: Subset 3 which consists of 6 items as shown ion Table 7 to measure whether we could expand the domain coverage of DSM-5 by including the perceptual motor function. In general, activities involving Administration and Shopping were featured across all the four subsets, highlighting that complex attention, executive functioning, and learning and memory are the key cognitive domains to look out for to detect progression of the disease.

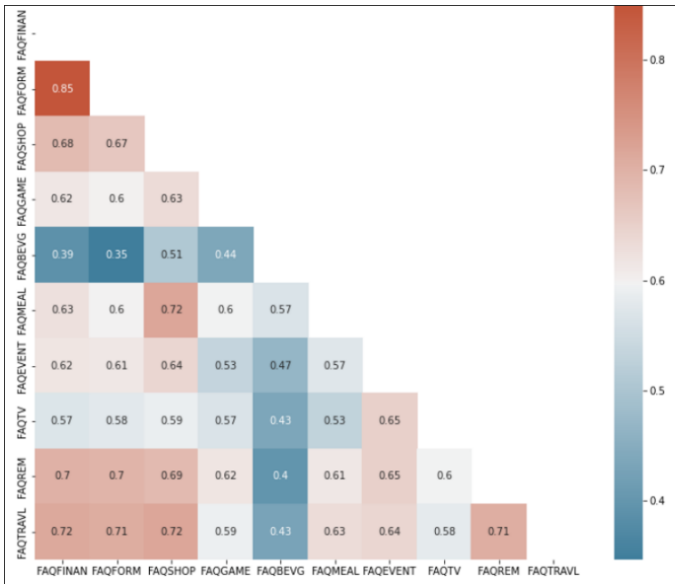

Figure 2: Pearson Correlation Coefficient Matrix of FAQ items

Table 7: Summary of the derived FAQ items for each subset

Subset	Derived FAQ Items
1	All FAQ Items (ten items)
2	Finance, Shopping, Administration, Personal memory
3	Travel, Administration, Beverage-making, Shopping
4	**Subset 2 & Subset 3:** Finance, Shopping, Administration, Personal memory, Travel, Beverage-making

In general, the results indicated that there is no clear dominant item in the FAQ test that stands out for the detection of AD progression. However, after investigating the drop in % between two successive average scores of the FAQ items according to Equation (8) after ranking the average scores in descending order, it was clear that there are two clusters of features:

- Cluster 1: (FAQFINAN, FAQSHOP, FAQFORM, FAQREM)
- Cluster 2: (FAQTRAVL, FAQTV, FAQGAME, FAQEVENT).

$$Relative\ Difference\ (S_i, S_{i+1}) = \frac{(S_i - S_{i+1})}{S_i} \quad (8)$$

$$where\ \ S_i is\ the\ a\ score\ of\ feature\ S\ at\ rank\ i\ and\ \ S_{i+1}\ is\ the\ score\ of\ feature\ at\ rank\ i+1$$

Cluster 1 contains items that are related mainly to executive functions, complex attention, and memory. Poor performance in any of the FAQ task from the cluster is a sign for clinicians to take further action on whether to perform additional assessments or to provide the patient with treatment. In the next section, we will isolate this cluster and others shown in Table 7 and process them using classification techniques to seek which specific FAQ items are impactful for AD stage conversion.

Classification Results Analysis

With the subsets of FAQ items being derived based on the feature selection methods, multiple classification techniques including Bayes Net, LR, and C4.5 have processed these subsets to assess the classification models derived for the AD progression in terms of accuracy, sensitivity, and specificity metrics. Additionally, we conducted a second round of experiments that included demographics such as age, gender, education, race, and marital status, along with the subset of FAQ items, to evaluate their impact on the classification models. As a guideline for the reliability of the diagnostic classification, [29] suggest a sensitivity ranging from 85%–98% and specificity ranging from 71%–91% when discriminating between normal control and demented subjects.

Table 8 illustrates the results of sensitivity, specificity, and accuracy obtained on the different subsets of FAQ items that we derived from the ADNIFAQ dataset by Bayes Net, LR, and C4.5 classification algorithms. The left half of the table contains the results against the distinctive subsets of features and without considering demographics, whereas the right half of the table contains the results of the classification algorithms when including demographics within each subset of features. The results show that the C4.5 algorithm was able to produce the best classification models for AD progression when demographics were not included and from merely six features (subset 4). To be specific, processing six FAQ features only by a decision tree algorithm, i.e., C4.5, achieved 74.73% accuracy and 86.00% specificity.

The sensitivity of the models produced from Subset 4 by C4.5 was relatively low, i.e., 62.40%. In fact, most of the classification algorithms' performance in terms of sensitivity on all subsets including Subset 1 (the complete FAQ items) has been low, i.e., between 56.90% and 63.40%. Therefore, we

investigated the confusion matrix results by checking the TPs and FNs. It is apparent that the C4.5 algorithm has misclassified 3,324 data observations that should be 'AD progression' (DX progress = 1) as 'no change' (DX progress = 0). This has indeed increased the False Negatives which explains the high number of misclassifications by the considered classification algorithms which have contributed to low sensitivity rates. Conversely, the number of FPs is low for C4.5 especially when processing Subset 1 and Subset 4, which have contributed to high specificity rate. For instance, the number of data observations that have been misclassified into 'AD progression' and supposed to be 'no change' by the C4.5 algorithm from Subset 4 was 1,362, which is a relatively small number if we compare it with the FNs (3,324).

Overall, the performance of the classification algorithms considered on the distinctive features of FAQ is not ideal, particularly the sensitivity results as they were far from meeting the guideline range as described by [29], even when all FAQ items were included (Subset 1). Perhaps that measuring functional activities does not tell us everything we need to know to be able to predict progression and including more cognitive items of the other clinical assessment methods are needed for a comprehensive assessment battery. It was interesting to note that processing Subset 3 by the classification algorithms did not provide significantly better performing models despite having an enhanced coverage of cognitive domains than Subset 2, with the addition of the perceptual motor function domain needed to 'make a hot drink' and 'travel' which refers to routine activities that have been well-learned. This possibly indicates that perceptual motor function has been covered partly by other items as isolating each cognitive domain based on the FAQ items is a difficult task due to the fact that an item may overlap in multiple cognitive domains. The results also suggest that motor function is less of a marker of progression at least on the datasets and methods we considered.

In the second round of experiments, we used the same subsets of features (1–4) in addition to adding the demographics to each subset. The results are shown in the right half of Table 8. After running the same classification algorithms on the subsets of features with demographics we realised an improvement of the AD progression models' performance particularly when using the C4.5 algorithm. The classification models produced by C4.5 from Subset 4 were the best in regard to accuracy, sensitivity, and specificity rates. In fact, the C4.5 algorithm derived a classification model with 92.85%, 93.40%, and 92.40% of accuracy, sensitivity, and specificity, respectively. When comparing these results with those derived by the same algorithm on the same dataset but without demographics an increase of 18.12%, 31%, and 6.4% on accuracy, sensitivity, and specificity respectively has been achieved. These results, if limited, show that demographics (age in particular) play a critical role in the progression of dementia. A significant improvement can be seen in the classification of FNs which decreased 82.34% from 3,324–587.

The results obtained from the second experiment showed that the DT algorithm (C4.5) derives high predictive classifiers from all data subsets including Subsets 2 and 3 with each containing just four features when coupled with demographics. There is a higher confidence level in the reliability of the model where the results all exceed 90% with the largest difference being only 2.4% between the Subset 1 baseline and Subset 3 on specificity. Overall, Subset 2 increased by 21.72%, 32.50%, 11.90% on accuracy, sensitivity, and specificity respectively, while Subset 3 increased by 21.60%, 34.70%, 9.60% on the same measures. When demographics are taken into consideration, Subsets 2 and 3 are also able to close the performance gap on Subset 4, with a difference of 1.7% between Subset 3 and Subset 4 on specificity. Without demographics, the largest difference is 6.1% between Subset 2 and Subset 4 on specificity. The C4.5 algorithm is a powerful classification method that derives AD progression models associated with reliable performance metric results even with fewer FAQ features.

The significant improvement in results demonstrated the crucial impact of demographics on the classification algorithms. However, only C4.5 was able to greatly improve the results to align with the suggested sensitivity and specificity range described by [29]. Bayes Net and LR only achieved a minimum gain and did not reach the same level of progress as C4.5, even when all FAQ items were included. Similarly, in the first experiment, while Subset 3 has an enhanced coverage of more cognitive

domains than Subset 2, with the addition of the perceptual motor function, the model performance was still slightly lower. No specific FAQ item stood out as the main feature that could predict progression due to the overlapping of multiple cognitive domains in particular complex attention, executive functioning, and memory which have the most coverage amongst the FAQ items (see Table 1).

Table 8. Performance of the classification methods when using different subsets of the FAQ items

		Excluding Demographics Features			Including Demographics Features		
Subset	**Algorithm**	**Accuracy %**	**Sensitivity %**	**Specificity %**	**Accuracy %**	**Sensitivity %**	**Specificity %**
Subset 1 (baseline)	BayesNet	68.48	61.40	74.90	70.78	63.90	77.00
	LR	71.30	61.50	80.20	76.46	77.10	75.80
	C4.5	79.48	63.40	94.10	93.42	93.80	93.10
Subset 2	BayesNet	67.82	63.20	72.10	70.48	65.30	75.20
	LR	68.57	57.00	79.10	75.23	77.40	73.20
	C4.5	70.68	60.60	79.90	92.40	93.10	91.80
Subset 3	BayesNet	68.28	56.90	78.60	70.97	62.40	78.70
	LR	68.53	61.90	74.50	74.55	76.10	73.10
	C4.5	70.26	58.40	81.10	91.86	93.10	90.70
Subset 4	BayesNet	68.09	63.10	72.70	71.43	66.00	76.30
	LR	69.81	60.30	78.50	76.10	78.10	74.30
	C4.5	74.73	62.40	86.00	92.85	93.40	92.40

7. Conclusions & Future Work

Pathological assessments, and neuropsychological tests are common methods to detect dementia. While the former provides supporting evidence for cognitive impairment, it can be costly, intrusive, and stressful for both patient and family to undergo numerous tests. Furthermore, time taken to perform pathological tests may delay early detection and intervention. Hence, our research into applying innovative technology such as machine learning techniques to predict AD progression by using only the functional activities test data, in this case the FAQ test, is promising as the process is quick, less resource intensive, and an easier approach for patients. This research aimed at identifying few items of the FAQ test that may trigger AD progression. The aim was achieved by a data driven approach that comprised of supervised machine learning techniques with feature selection.

Empirical results on datasets from the ADNI repository related to patients' diagnosis and FAQ test sheets using different feature selection methods including CA, CST, IG, and LOOCV, revealed that the Finance and Shopping activities in FAQ were featured in all the data subsets used as part of the classification experiments. These functional activities, thought to tap into cognitive domains of executive functioning, complex attention, and memory, are the key indicators of disease progression. The LOOCV method derived FAQTRAVL – 'travel' and FAQBEVG – 'making a hot drink' as impactful features providing additional coverage of perceptual motor function, thus ensuring an more coverage of the DSM-5 domains, and slightly improving the classification results, notably in data Subset 4. More importantly, the machine learning techniques, especially DT (C4.5 algorithm) produced 62.40% and 86.00% sensitivity and specificity rates respectively when processing data Subset 4 that contains six FAQ items (Finance, Shopping, Administration, Personal memory, Travel, Beverage-making). When adding demographic features into data Subset 4, the performance of models derived by the C4.5 algorithm improved substantially to reach 93.40% and 92.40% sensitivity and specificity rates, respectively. Overall, C4.5 was superior in all evaluation metrics used when contrasted with LR and

Bayes Net. It showed good performance in detecting the progression of AD, specifically when demographic features are added to the features of FAQ all achieving an accuracy level of over 90%.

In conclusion, our findings can be useful for predicting the probable progression of the disease. We were able to extract functional features that can provide us with a high-performing classification model potentially providing patients early intervention. This process of detecting progression can be automated via a computational model where clinicians can use to derive a quick assessment of the patient's condition where key features can be identified and then mapped to the degenerative domain. Although there was no specific FAQ item that stood out as the main feature that could predict progression due to the overlapping of multiple cognitive domains, clinicians could focus and evaluate the patient's ability to carry out everyday tasks that specifically involve complex attention, executive functioning, and memory, to have an estimated measure of the progression of the disease. Poor performance in these tasks may serve as a red flag for clinicians to decide on whether to proceed with conducting further diagnostic assessments, or if in worrying cases, to proceed with intervention immediately.

A key challenge we faced is the mapping process between the individual FAQ items and its associated DSM-5 cognitive domains. While each task often has multiple domains overlapping, a further study can be carried out to measure the level of DSM-5 degenerative domains required to perform each individual FAQ item. For example, Shopping may require a higher percentage input of memory and executive functioning, than complex attention and perceptual motor function. And from there, we can weigh out more clearly which domain has a more significant impact on the progression of dementia. More importantly, a combination of brief cognitive tests along with the most effective FAQ items would provide additional sensitivity.

While we deduced that demographics play a role in diagnosis, another future work could involve the independent investigation of each demographic feature, evaluating which sociology aspect bears the highest influence towards the diagnosis. For example, the FAQ items extracted during feature selection, were mainly associated with complex attention, executive functioning, and memory, thus suggesting that education could have an influential role in comparison to the other demographics. Other possible future work could involve longitudinal analysis, where a patient's progress across time is analysed, such as predicting the speed of progression of the disease and if age is a crucial factor.

References

[1] van der Flier W & Scheltens, P. Epidemiology and risk factors of dementia. *Journal of Neurology, Neurosurgery & Psychiatry;* 2005; *76*(5). https://doi.org/10.1136/jnnp.2005.082867

[2] Finkel S, Costa e Silva J, Cohen G, Miller S & Sartorius N. Behavioral and psychological signs and symptoms of dementia: A consensus statement on current knowledge and implications for research and treatment. *International Pscyhogeriatrics;* 1996; *8*(3), 497–500. https://doi.org/10.1017/s1041610297003943

[3] Alzheimer's New Zealand. (n.d.). *What is dementia?* https://www.alzheimers.org.nz/

[4] Sachdev PS, Blazer DG, Blacker D & Ganguli M. *Classifying neurocognitive disorders: The DSM-5 approach*; (2014, September). https://pubmed.ncbi.nlm.nih.gov/25266297/

[5] Deloitte. *Dementia Economic Impact Report 2016*. Alzheimers New Zealand; 2017. https://www.alzheimers.org.nz/news/dementia-economic-impact-rpcort-2016

[6] Suzuki E. *Renewing priority for dementia: Where do we stand?* OECD; 2018. http://www.oecd.org/health/health-systems/Renewing-priority-for-dementia-Where-do-we-stand-2018.pdf

[7] NHS. *Is there a cure for dementia?* 2018. https://www.nhs.uk/conditions/dementia/cure/

[8] Ministry of Health. *New Zealand framework for dementia care*; 2013. https://bit.ly/3p2nfth

[9] Alzheimer's Association. (n.d.). *Medical tests*. https://www.alz.org/alzheimers-dementia/diagnosis/medical_tests
[10] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders 5th ed;* 2013.
[11] Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL & Chertkow H. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. Journa*l of the American Geriatrics Society;* 2005; *53(*4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x
[12] Hughes CP, Berg L, Danziger WL, Coben LA & Martin RL. A new clinical scale for the staging of dementia. *British Journal of Psychiatry;* 1982; *140*(6), 566–572. https://doi.org/10.1192/bjp.140.6.566
[13] Folstein M, Folstein S & McHugh P. Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research;* 1975; *12*, 189–198.
[14] Mohs RC, Rosen WG & Davis KL. The Alzheimer's disease assessment scale: An instrument for assessing treatment efficacy. *Psychopharmacology Bulletin;* 1983; *19*(3), 448–450. http://ci.nii.ac.jp/naid/10029841229/en/
[15] Pfeffer RI., Kurosaki TT, Harrah, CH, Chance JM & Filos S. Measurement of functional activities in older adults in the community. *Journals of Gerontology;* 1982; *37*(3), 323–329. https://doi.org/10.1093/geronj/37.3.323
[16] Balsis S, Benge JF, Lowe DA, Geraci L & Doody, R. S. How do scores on the ADAS-Cog, MMSE, and CDR-SOB correspond? *The Clinical Neuropsychologist;* 2015; *29*(7), 1002–1009. https://doi.org/10.1080/13854046.2015.1119312
[17] Sinha A & Sinha A. Mild cognitive impairment and its diagnosis to progression to dementia with several screening measures. *The Open Psychology Journal;* 2018; *11*(1), 142–147. https://doi.org/10.2174/1874350101811010142
[18] Mukadam N. Improving the diagnosis and prediction of progression in mild cognitive impairment. *International Psychogeriatrics;* 2018; *30*(10), 1419–1421. https://doi.org/10.1017/s1041610218001692
[19] Kavitha C, Mani V, Srividhya, SR, Khalaf, OI & Tavera Romero, CA. Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. Frontiers in public health; 2022; 10, 853294. https://doi.org/10.3389/fpubh.2022.853294
[20] James C, Ranson JM, Everson R, Llewellyn DJ. Performance of Machine Learning Algorithms for Predicting Progression to Dementia in Memory Clinic Patients. JAMA Netw Open. 2021;4(12):e2136553. doi:10.1001/jamanetworkopen.2021.36553
[21] Janiesch C, Zschech P & Heinrich K. Machine learning and deep learning. Electron Markets; 2021; 31, 685–695. https://doi.org/10.1007/s12525-021-00475-2
[22] Thabtah F, Peebles D, Retzler J, Hathurusingha C. 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,
[23] Clemmensen FK, Hoffmann K, Siersma V, Sobol N, Andersen BB, Vogel A, Lolk A, Gottrup H, Høgh P, Waldemar G, Hasselbalch SG & Frederiksen KS. 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; 2020. https://doi.org/10.1186/s12877-020-01926-9
[24] Kemp EC, Ebner MK, Ramanan S, Godek TA, Pugh EA, Bartlett HH, McDonald JW, Mecca MC, van Dyck CH & Mecca AP. Statin use and risk of cognitive decline in the ADNI cohort. The American Journal of Geriatric Psychiatry; 2020; 28(5), 507–517. https://doi.org/10.1016/j.jagp.2019.11.003
[25] Thabtah F, Spencer R & Ye Y. The correlation of everyday cognition test scores and the progression of Alzheimer's disease: a data analytics study. Health information science and systems; 2020; 8(1), 24. https://doi.org/10.1007/s13755-020-00114-8

[26]Marshall GA, Zoller AS, Lorius N, Amariglio RE, Locascio JJ, Johnson KA, Sperling RA & Rentz DM. Functional Activities Questionnaire items that best discriminate and predict progression from clinically normal to mild cognitive impairment. *Current Alzheimer Research;* 2015; *12*(5), 493-502. https://doi.org/10.2174/1567205012051505261150 03
[27]Mayo A. *Use of the Functional Activities Questionnaire in older adults with dementia.* Alzheimer's Association, D13; 2016; https://www.alz.org/careplanning/downloads/functional-activities-questionnaire.pdf
[28]Massachusetts Alzheimer's Disease Research Center (MADRC). (n.d.). Retrieved December 4, 2020. From https://www.madrc.org/
[29]Teng E, Becker BW, Woo E, Knopman DS, Cummings JL & Lu PH. Utility of the functional activities questionnaire for distinguishing mild cognitive impairment from very mild Alzheimer's disease. *Alzheimer Disease & Associated Disorders;* 2010; *24*(4), 348-353. https://doi.org/10.1097/WAD.0b013e3181e2fc84
[30]Beekly D, Ramos E, Lee W, Deitrich W, Jacka M, Wu J, Hubbard J, Koepsell T, Morris J & Kukull, W. The National Alzheimer's Coordinating Center (NACC) database: The uniform data set. *Alzheimer Disease and Associated Disorders;* 2007; *21*(3), 249–258.
[31]Hosmer D, Lemeshow S & Sturdivant R. *Applied logistic regression* (3rd ed., 398). John Wiley & Sons; 2013. https://doi.org/10.1002/9781118548387
[32]Lemoine B, Rayburn S & Benton R. Data fusion and feature selection for Alzheimer's Diagnosis. *International Conference on Brain Informatics;* 2010; *6334*, pp. 320-327. Toronto, ON, Canada: Springer-Verlag. https://doi.org/10.1007/978-3-642-15314-3_30
[33]Vapnik V. An overview of statistical learning theory. *IEEE Transactions on Neural Networks;* 1999; *10*(5).
[34]Battista P, Salvatore C & Castiglioni I. Optimizing neuropsychological assessments for cognitive, behavioral, and functional impairment classification: A machine learning study. *Behavioural Neurology;* 2017; 19. https://doi.org/10.1155/2017/1850909
[35]Cortes C & Vapnik V. Support-vector networks. *Machine Learning;* 1995; *20*(3), 273–297.
[36]Choi H & Jin KH. Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behavioural Brain Research;* 2018; *344,* 103–109. https://doi.org/10.1016/j.bbr.2018.02.017
[37]LeCun Y, Bengio Y & Hinton G. Deep learning. *Nature;* 2015; *521*, 436–44. https://doi.org/10.1038/nature14539
[38]So A, Hooshyar D, Park K & Lim H. Early diagnosis of dementia from clinical data by machine learning techniques. *Applied Sciences;* 2017; *7*(7), 651. https://doi.org/10.3390/app7070651
[39]Rish I. *An empirical study of the naive Bayes classifier*. Proceedings of IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence; 2001. New York, USA: IBM.
[40]Friedman N, Geiger D & Goldszmidt M. Bayesian network Ccassifiers. *Machine Learning;* 1997; *29*, 131–163. https://doi.org/10.1023/A:1007465528199
[41]Quinlan J. Bagging, boosting, and C4.5. *AAAI/IAAI;* 1996; *1,* 725-730.
[42]Liaw A & Wiener M. Classification and regression by randomForest. *R News;* 2002; *2*(3), 18–22.
[43]Kotsiantis S. Supervised machine learning: A review of classification techniques. *Informatica;* 2007; *31*(3), 249-268. https://doi.org/10.1007/s10462-007-9052-3
[44]Gardner M. & Dorling S. Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment;* 1998; *32*(14-15), 2627–2636. 1https://doi.org/0.1016/S1352-2310(97)00447-0
[45]Liu H & Setiono R. Chi2: feature selection and discretization of numeric attributes. Proceedings of the IEEE 7th International Conference on Tools with Artificial Intelligence; 1995; (pp. 388-391). IEEE.
[46]Mitchell T. *Machine learning*. McGraw-Hill; 1997.

[47]Chawla N, Bowyer K, Hall L & Kegelmeyer W. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence;* 2002; *16*, 321–357. https://doi.org/10.1613/jair.953
[48]Hall M. *Correlation-based feature selection for discrete and numeric class machine learning;* 2000; University of Waikato, New Zealand.
[49]Benesty J, Chen J, Huan Y & Cohen I. *Noise reduction in speech processing, 2.* Springer; 2009; https://doi.org/10.1007/978-3-642-00296-0_5
[50]Trambaiolli L, Spolaor N, Lorena A, Anghinah R & Sato J. Feature selection before EEG classification supports the diagnosis of Alzheimer's disease. *Clinical Neurophysiology;* 2017. http://dx.doi.org/10.1016/j.clinph.2017.06.251
[51]Zhou K, He W, Xu Y, Xiong G & Cai J. *Feature selection and transfer learning for Alzheimer's disease clinical diagnosis*. Applied Sciences; 2018. https://doi.org/10.3390/app808137
[52]Brand L, Nichols K, Wang H, Huang H & Shen L. Predicting longitudinal outcomes of Alzheimer's disease via a tensor-based joint classification and regression model. *Pacific Symposium on Biocomputing;* 2020; *25*, 7–18.
[53]World Dementia Council. *Defeating Dementia: The Road to 2025;* 2018. https://worlddementiacouncil.org/sites/default/files/2018-12/Defeating 20Dementia 20Report.pdf
[54]Alzheimer's Disease Neuroimaging Initiative (ADNI). *About ADNI;* 2017. http://adni.loni.usc.edu/about/
[55]Alzheimer's Disease Neuroimaging Initiative (ADNI). (n.d.). *ADNI procedures manual.* http://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI_GeneralProceduresManual.pdf
[56]Witten I, Hall M, Frank E, Holmes G, Pfahringer B & Reutemann P. The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter;* 2009; *11*(1), 10-18. https://doi.org/10.1145/1656274.1656278
[57]Kohavi R. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. International Joint Conference on Artificial Intelligence (IJCAI); 1995.
[58]Hoens TR & Chawla N. Imbalanced datasets: From sampling to classifiers. In *Imbalanced learning: Foundations, algorithms, and applications;* 2013; (pp. 43–59). https://doi.org/10.1002/9781118646106.ch3
[59]Fix E & Hodges J. *Discriminatory analysis: Nonparametric discrimination, consistency properties*. Int Stat Rev; 1989.
[60]Blagus R & Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics;* 2013; *14*, 106. https://doi.org/10.1186/1471-2105-14-106
[61]Harris Jr E. *Information gain versus gain ratio: A study of split method biases*. The MITRE Corporation; 2001.
[62]Quinlan J. *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers; 1993.

[3] Acknowledgment

[3] Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.