Original Paper

Debbie Rankin[1] PhD, Corresponding author, d.rankin1@ulster.ac.uk, +4428 7167 5841
Michaela Black[1] PhD, mm.black@ulster.ac.uk
Bronac Flanagan[1] PhD, b.flanagan@ulster.ac.uk
Catherine Hughes[2] PhD, c.hughes@ulster.ac.uk
Adrian Moore[3] PhD, a.moore@ulster.ac.uk
Leane Hoey[2] PhD, l.hoey@ulster.ac.uk
Jonathan Wallace[4] MSc, jg.wallace@ulster.ac.uk
Chris Gill[2] PhD, c.gill@ulster.ac.uk
Paul Carlin[5] MPA, p.carlin@open.ac.uk
Anne M Molloy[6] PhD, amolloy@tcd.ie
Conal Cunningham[7] MD, ccunningham@stjames.ie
Helene McNulty[2] PhD, h.mcnulty@ulster.ac.uk

[1] School of Computing, Engineering and Intelligent Systems, Ulster University, Derry~Londonderry, Northern Ireland, United Kingdom
[2] Nutrition Innovation Centre for Food and Health, School of Biomedical Sciences, Ulster University, Coleraine, Northern Ireland, United Kingdom
[3] School of Geography and Environmental Sciences, Ulster University, Coleraine, Northern Ireland, United Kingdom
[4] School of Computing, Ulster University, Jordanstown, Northern Ireland, United Kingdom
[5] Faculty of Wellbeing, Education and Social Care, The Open University, Belfast, Northern Ireland, United Kingdom
[6]School of Medicine, Trinity College Dublin, Dublin 2, Ireland (AMM)
[7]Mercers Institute for Research on Ageing, St James's Hospital, Dublin, Ireland (CC)

Identifying Key Predictors of Cognitive Dysfunction in Older People Using Supervised Machine Learning Techniques

Abstract

Background:
Machine learning techniques, specifically classification algorithms, may be effective to assist in understanding key health, nutritional and environmental factors associated with cognitive function in ageing populations.

Objective:
The objective of this study was to use classification techniques to identify the key patient predictors considered most important in the classification of poorer cognitive performance, which is an early risk factor for dementia.

Methods:

Data was utilised from the Trinity-Ulster and Department of Agriculture (TUDA) study, which included detailed information on sociodemographic, clinical, biochemical, nutritional, and lifestyle factors on 5,186 older adults recruited from the Republic of Ireland and Northern Ireland, a proportion of whom (20%) were followed up 5-7 years later for reassessment. Cognitive function at both time points was assessed using a battery of tests, including the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) with a score <70 classed as poorer cognitive performance. This paper trained three classifiers; decision trees, Naïve Bayes and random forests, to classify the RBANS score and to identify key health, nutritional and environmental predictors of cognitive performance and cognitive decline over the follow-up period, and assessed their performance, taking note of the variables these optimised classifiers deemed as of key importance for their computational diagnostics.

Results:
In the classification of a 'low' RBANS score (<70), our classification models performed well (range F1-score 0.73 to 0.93), all highlighting the individual's score from the Timed Up and Go (TUG) test, the age the participant left education and whether or not the participant's family reported memory concerns as of key importance. The classification models performed well in classifying a greater rate of decline in the RBANS score (range F1-score 0.66-0.85), also indicating the TUG score as of key importance, followed by blood indicators: plasma homocysteine (tHcy), vitamin B6 biomarker (plasma pyridoxal-5-phosphate; PLP) and glycated haemoglobin (HbA1c).

Conclusions:
The results presented here would suggest that it may be possible for a healthcare professional to make an initial evaluation, with a high level of confidence, of the potential for cognitive dysfunction using only a few short, non-invasive questions, thus providing a quick, efficient and non-invasive way to help them decide whether or not a patient requires a full cognitive evaluation. This approach has the potential benefits of making time and cost savings for health service providers and avoiding stress created through unnecessary cognitive assessment in low risk patients.

Trial Registration: Clinical Trials.gov Identifier: NCT02664584

Keywords: Classification; Supervised Machine Learning; Cognition; Nutrition; Timed Up and Go

Introduction

Background

Globally, populations are ageing; by 2050 it is estimated that more than 2 billion people will be aged over 60 years [1]. Cognitive function generally declines with age and ranges in severity from mild cognitive impairment (MCI) to dementia. MCI can be defined as cognitive decline greater than expected for an individual's age and

education level but that does not interfere with activities of daily living, whereas dementia impacts profoundly on normal functioning [2,3]. Dementia currently affects 50 million people worldwide and it is estimated that this will increase to 152 million by 2050. The annual cost of dementia is estimated at US $1 trillion and is expected to more than double by 2030 [4]. Therefore, strategies that promote better brain health and well-being in older age are an urgent public health priority.

Alzheimer's disease is the most common form of dementia, with other forms including vascular dementia, dementia with Lewy bodies, frontotemporal dementia and mixed dementia. Risk factors for dementia are disease dependent, but commonly include age, genetics, medical conditions including cardiovascular disease and diabetes, diet, lifestyle and environmental factors [5]. A major recent report highlighted the complexity of dementia and the potential to prevent or delay the onset of the disease through interventions targeted at modifiable risk factors [6]. In particular, nutrition has been identified as a key area of interest and emerging evidence links lower status of certain vitamins with cognitive dysfunction in older adults, whilst certain dietary patterns and components appear to have protective roles in maintaining cognitive health [7].

The application of data mining within healthcare has become increasingly popular, driven particularly by the large amount of complex data available that tests the capabilities of traditional statistical approaches [8]. In healthcare, as in other areas, data mining has provided a means of accessing and analysing large volumes of data in order to better inform and to drive change. Classification models, in particular, have been utilised extensively in the understanding of MCI. These models can help us to understand patterns in the behaviour of data in terms of diagnosing MCI, specifically in the consideration of key features pertaining to a diagnosis of impairment [9,10] or predicting the progression of the impairment [11]. Furthermore, models have been developed in order to apply a more objective approach to the MCI diagnosis [12], not to undermine but rather to support a clinician's analysis [13]. Na et al [14] have investigated the use of non-invasive, easy-to-collect variables that are commonly collected in community health care settings such as sociodemographic, health, functional and interpersonal variables, for the prediction of cognitive impairment amongst community dwelling older adults, using the Korean Longitudinal Study of Ageing (KLoSA) dataset [15] and a Gradient Boosting Machine classifier.

Many studies apply machine learning approaches to the popular Open Access Series of Imaging Studies (OASIS) [16], Alzheimer's Disease Neuroimaging Initiative (ADNI) [17] and Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing (AIBL) [18] datasets consisting of neuroimaging data (e.g. MRI and PET scan data) from participants ranging from no cognitive impairment to MCI to Alzheimer's disease. These datasets also include a range of demographic, biomarker, clinical and cognitive assessment data. Ding et al [19] used a Bayesian network approach for the classification of Alzheimer's disease with heterogeneous features from the AIBL dataset and demonstrated that machine learning could be used to select features

and their appropriate combinations that are relevant for Alzheimer's disease severity classification with high accuracy. Korolev et al [20] used a kernel-based classifier and the ADNI data set to develop a prognostic model for predicting MCI-to-dementia progression over a three-year period.

The aim of our study was to compare a selection of data analytics techniques to identify determinants of cognitive health in community dwelling older adults using existing data from the TUDA study. The TUDA study was designed to investigate nutritional, health and lifestyle factors in the development of diseases of ageing including dementia. A range of analytical models on the data were developed to determine factors that may predict poorer cognitive performance and cognitive decline over time, assessed using an in-depth neuropsychiatric test.

Methods

CRISP-DM Methodology

Within this study the widely used Cross-Industry Process for Data Mining (CRISP-DM) research methodology was adopted [21]. CRISP-DM has 6 main steps; Business understanding, Data understanding, Data preparation, Modelling, Evaluation and Deployment. In the business understanding phase, the objective of this study was defined: to use classification techniques to identify the key patient predictors considered most important in the classification of cognitive dysfunction, itself a predictor of dementia. In the data understanding phase, the data quality was examined to understand data collection methods and the features contained within the TUDA dataset, as described in the next section (The Data). In the data preparation phase, the TUDA dataset was pre-processed to cleanse the dataset and select features relevant to the modelling phase. Feature selection methods and the results of feature selection are described in subsequent sections (Methods > The Data and Statistical Analysis Techniques, and Results > Feature Selection). In the modelling phase a number of machine learning modelling techniques were selected and applied to the prepared data and their parameters were calibrated to optimal values in order to increase the knowledge extracted from the data (described in Methods > Machine Learning Techniques, and Results). Upon building the models that produced the highest quality knowledge from the data analysis perspective, the models were thoroughly evaluated to ensure robustness and achievement of the business objectives. The knowledge gained from the models was then presented to clinical experts in a way that could be used and understood.

The Data

The TUDA cohort provides detailed nutrition and health data, along with related lifestyle, clinical, and biochemical details, on a total of 5186 community-dwelling older adults aged 60-102 years, making this cohort one of the most comprehensively characterised cohorts of its kind for ageing research internationally. With an overall goal to address the prevention of age-related disease, the TUDA study is aimed at investigating nutrition and related factors in the development of common diseases of ageing. TUDA study participants were recruited

between 2008 and 2012 from hospital outpatient or General Practice clinics in the Republic of Ireland or Northern Ireland, via standardised protocols for participant sampling, assessment and data recording and with centralised laboratory analysis. In brief, the inclusion criteria for the TUDA study were being born on the island of Ireland, aged >60 years, and without an existing diagnosis of dementia. Non-fasting blood samples were collected from all participants and a wide range of parameters, including routine biochemistry and haematological profiles, along with biomarkers of micronutrient status, were measured. A comprehensive health and lifestyle questionnaire was administered as part of the 90-minute interview to capture medical and demographic details, along with comprehensive information on medication and vitamin supplement usage. Physiological function tests, blood pressure, bone health (DXA scans) and cognitive function tests were also measured. A subset of approximately 20% of participants (n=987) were reassessed 5-7 years after their initial assessment in order to investigate progression of risk factors and disease over time.

A summary of the characteristics of the subset of the TUDA cohort (n=2869) analysed in this work is shown in (Table 1). Pre-processing and feature selection performed on the original dataset to reach this subset of data are described in subsequent Methods and Results sections.

Table 1. General Characteristics of TUDA Study Participants[a]

	Characteristic	**Males (n = 1191)**	**Females (n = 1678)**
	Age, mean (SD) (year)	72.1 (7.8)	72.2 (7.8)
	Education, mean (SD) (years)[b]	16.3 (3.3)	16.1 (2.8)
Health and Lifestyle			
	BMI (kg/m^2)	28.9 (4.3)	28.7 (5.7)
	Waist to Hip Ratio	0.97 (0.07)	0.88 (0.07)
	Instrumental Activities of Daily Living	25.0 (4.1)	24.9 (3.5)
	Physical Self Maintenance Score	23.3 (1.6)	23.1 (1.7)
	Timed Up and Go (seconds)	12.9 (9.1)	13.0 (8.0)
	Living alone % (n)	21.8 (260)	37.7 (632)
	Current smoker % (n)	10.2 (122)	11.6 (194)
	Alcohol (units/week)	8.8 (14.6)	2.9 (6.7)
	Socio-economically most deprived % (n)	24.4 (291)	25.4 (426)
Neuropsychiatric assessment			
	MMSE score	27.8 (1.4)	27.9 (1.4)
	RBANS score	87.3 (14.5)	88.9 (15.2)
	RBANS class = 'low' (target) % (n)[c]	11.2 (133)	10.0 (168)
	RBANS class = 'high' (target) % (n)[c]	88.8 (1058)	90.0 (1510)
	FAB score	15.7 (2.2)	15.9 (2.1)
	Depression CES-D score	4.8 (6.2)	6.1 (7.7)
	Anxiety (HAD score)	2.6 (3.2)	3.5 (3.8)
Clinical Measures			
	White cell count (10^9/L)	7.1 (3.6)	6.9 (3.3)
	Haemoglobin (g/DL)	14.2 (1.5)	13.0 (1.3)

	Mean corpuscular volume (FI)	90.7 (5.5)	90.6 (5.1)
	Platelet count (10^9/L)	229 (59.0)	265 (66.9)
	Urea (mmol/L)	7.2 (2.9)	6.7 (2.3)
	Creatinine (μmol/L)	98 (31.0)	79 (22.4)
	Albumin (g/L)	42 (3.7)	42 (3.4)
	Gamma GT (U/L)	43 (47.5)	34 (36.0)
	Sodium (mmol/L)	140 (5.1)	139 (3.2)
	Potassium (mmol/L)	4.3 (0.5)	4.2 (0.4)
	Calcium (mmol/L)	2.3 (0.1)	2.3 (0.1)
	Phosphate (mmol/L)	1.0 (0.2)	1.1 (0.2)
	Alkaline Phosphatase (U/L)	82 (34.2)	82 (25.7)
	Low-Density Lipoprotein (LDL; mmol/L)	2.23 (0.8)	2.58 (0.9)
	High-Density Lipoprotein (HDL; mmol/L)	1.23 (0.4)	1.55 (0.4)
	Triglycerides (mmol/L)	1.78 (1.0)	1.62 (1.0)
	C-reactive protein (mg/L)	6.1 (11.1)	5.5 (11.9)
	Glycated haemoglobin (HbA1c; %)	6.0 (1.0)	5.9 (0.7)
	Parathyroid Hormone (pg/ml)	45.2 (30.8)	47.2 (31.9)
	Glomerular filtration rate (ml/min)	77.2 (25.3)	67.8 (22.6)
Nutritional Biomarkers			
	Red blood cell folate (nmol/L)	1053 (591.1)	1100 (582.7)
	Serum vitamin B12 (pmol/L)	267 (191.0)	296 (277.3)
	Plasma vitamin B6 (PLP; nmol/L)	74.1 (53.2)	81.5 (69.7)
	Riboflavin (EGRac)	1.35 (0.2)	1.34 (0.2)
	Total plasma homocysteine (μmol/L)	15.1 (5.9)	14.1 (5.1)
	Total Vitamin D (nmol/L)	51.6 (25.9)	56.0 (30.1)

[a]Abbreviations: MMSE: Mini Mental State Examination; RBANS: Repeatable Battery for the Assessment of Neuropsychological Assessment; FAB: Frontal Assessment Battery; CES-D: Centre for Epidemiological Studies Depression; HADs: Hospital Anxiety and Depression Scale; PLP; pyridoxal-5'-phosphate; EGRac: erythrocyte glutathione reductase activation coefficient with a higher EGRac value indicating poorer riboflavin status.
[b]Education refers to the age of leaving formal education
[c]RBANS score <70 is assigned class 'low' and an RBANS score >=70 is assigned class 'high'

Cognitive function was assessed at both timepoints using three assessment tools, the Mini Mental State Examination (MMSE), the Frontal Assessment Battery (FAB) and RBANS, and the rate of cognitive decline was calculated over the 5-7 year follow-up period. For the purposes of this paper, the cognitive function outcome indicator is categorised based on RBANS. RBANS is an age-adjusted and sensitive neuropsychiatric battery for assessing global cognitive function [22]. This tool has also been validated to assess specific cognitive domains within the brain, including immediate and delayed memory, visual-spatial, language, and attention, which are combined to provide a total score, with lower scores generally indicative of poorer cognitive performance.

The rate of RBANS change over the 5-7 year period between initial assessment and follow-up assessment is computed as the difference between a participant's RBANS score at each sampling point, normalised to account for the time between each assessment, where this can differ by up to 2 years across participants (see Equation (1)).

Equation (1)

$$RBANS_{rate\ of\ change} = \frac{RBANS_{assessment\ 2} - RBANS_{assessment\ 1}}{Time\ between\ assessments\ 1\ and\ 2}$$

The dataset initially contained 525 variables. During pre-processing, the data was cleansed to detect and correct inaccurate values, identify missing values and ensure consistent coding of these, ensure consistent coding of categorical variables, identify spelling and coding inconsistencies and correct these, transform text variables into categorical variables where possible, ensure numeric values fell within an appropriate and accurate range, check for consistency amongst dependent variables and correct any errors, and finally check for duplicate data and remove any redundancy. Normalisation was carried out on the data table, including non-loss decomposition to decompose the large data table into smaller tables, transforming composite attributes into separate attributes, transforming multi-valued attributes and repeating columns into separate tables and re-coding text attributes to categorical attributes where possible. This process reduced the number of variables to 345 within the dataset. These variables were a combination of text, categorical and numerical variables.

Feature Selection

Dimension reduction is an important stage for understanding the information in a dataset. Typical dimension reduction techniques, such as principal component analysis (PCA) [23], describe all the numerical variables contained within a dataset in terms of a number of linear combinations (fewer than the original number of features) of these features. Although a widely used and appreciated method for reducing the number of dimensions within a dataset, PCA is only valid for numerical features. Additionally, often a more transparent feature selection method is required to remove redundant features of various types in order to reduce the size of the dataset without losing potentially valuable information. Whilst a range of feature selection techniques exist, due to the nature of the features in the TUDA dataset and the prior knowledge that a large number of variables were likely to be highly correlated, correlation analysis and clustering were used in this study to allow highly correlated features to be determined and redundant features to be removed. These methods also helped us to discuss, evaluate and agree on the features to be retained in collaboration with the data gatekeepers and expert clinicians who had an in-depth knowledge of the data. Further feature selection was not carried out as we elected to retain as many features as possible for use in training the classifiers. This section describes the feature selection techniques performed and the results of feature selection are described in the Results section.

Manual Feature Selection

Manual feature selection was performed to remove features containing large amounts of missing data and therefore considered not useful to the analysis. Free-text variables that could not be encoded were also removed. Based on expert clinical

knowledge, features deemed irrelevant to the study were removed, as well as a number of subjective features where a comparable objective laboratory-obtained feature existed in the dataset.

Correlation and Association

Correlation analysis is necessary before the development of classification models for two primary reasons; "Algorithms might "overfit" predictions to spurious correlations in the data; multicollinear, correlated predictors could produce unstable estimates" [24] and "Perfectly correlated variables are truly redundant in the sense that no additional information is gained by adding them" [25]. In other words, as many machine learning algorithms rely on linearly independent variables, strongly correlated variables must be evaluated and removed to avoid unreliable results. Moreover, two variables which follow the same behaviour add little to the information gained by the dataset and thus are considered redundant. Correlation analysis allows the determination of highly correlated variables, which may undermine the consequential data analysis results. Due to the difference in categorisation of the variables within the dataset, correlation coefficients would be calculated for numerical-numerical pairs, whereas the strength of association are necessary for categorical-categorical variables and categorical-numerical variables. Correlations between numerical variables were calculated using Spearman's non-parametric correlation coefficient [26], the strength of association between categorical variables calculated using Cramér's V statistic [27], and the coefficient of determination (R2) was calculated between categorical and numerical variables [28].

Clustering

Clustering is useful in feature selection [25] to analyse the data to find structural patterns. Clustering can be used together with correlation analysis to identify those variables that behave in a similar manner and thus the information offered by the variables may prove redundant. Clustering of variables can take one of two forms; hierarchical, which outputs an informative hierarchy, and non-hierarchical, which divides the data into clusters, within which the variables may behave similarly. Due to the nature of the information this study seeks to derive, the focus was placed on hierarchical clustering, illustrated specifically in the form of tree structures, or dendrograms.

Ascendant hierarchical clustering can use a mixture of both numeric and categorical variables, to arrange variables into homogenous clusters, that is, variables which are strongly related to each other [29]. The algorithm for finding these related clusters follows the concepts of PCA and Multiple Correspondence Analysis (MCA). In PCA and MCA, the dataset is analysed to find new linearly independent variables to describe the same set of data. In this hierarchical clustering, these new synthetic variables are used as the centre points of the clusters and each original variable is then grouped according to its similarity to the cluster centre, either using the sum of the correlation ratio, for numeric variables, and the squared correlation, for categorical variables.

Machine Learning Techniques

Machine learning techniques are regularly employed for detecting patterns and dependencies within data, such as within healthcare data. Specifically, machine learning algorithms can be used to look for combinations of variables and generate rules within data which can be used to reliably predict outcomes [24]. This style of problem relies on classification algorithms where predictor variables are used to predict an outcome, or class variable. These predictions are based on a training sample of the data, usually consisting of a random sample of about 70-80% of the available data. The developed model comprises rules based on this training data, and then tested against the remaining data (see Figure 1). The training procedure is repeated on a number of different subsets of the data to reduce the likelihood of overfitting the model. In this study 10-fold cross-validation was used to measure the performance of classifiers, Initially the data was split into a training set (75%) and an evaluation set (25%). Models were trained using the training set with 10-fold cross validation applied (with a 90%/10% train/test split at each fold). The modelling techniques of decision trees, random forests and Naive Bayes were selected for their ease of interpretability. It is crucial that the results of modelling in this study can be explained to clinical experts. The individual algorithms were developed using R's caret package, specifically using the train and predict functions. The evaluation dataset was used to evaluate the performance of the model found to be optimal during training for each of the three respective techniques considered.

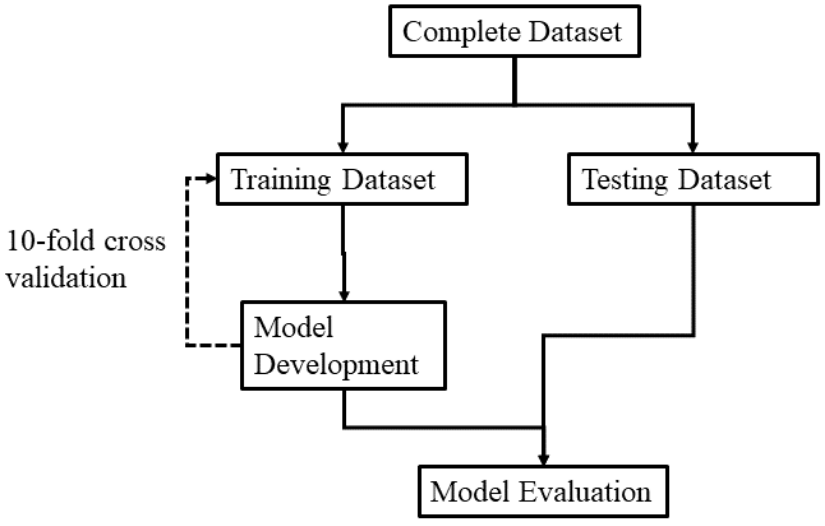

Figure 1. Model development and testing protocol

Decision Tree

Decision Trees are one of the most common machine learning algorithms when using a combination of continuous and categorical variables, chosen for their computational efficiency and readability. The Classification and Regression Tree (CART) [30] algorithm in particular lends itself well to explanatory knowledge discovery [31] due to its transparency. CART decision trees are developed using a top-down recursive algorithm, where the dataset is split into ever-smaller subsets

according to some pre-decided metric, most commonly using either the Gini impurity index or a permutation importance measure. The measures used are described below. The rpart implementation of the CART decision tree algorithm in R's caret package has been used in this study. This implementation automatically applies pruning, choosing a range of complexity parameters (cp) automatically selects the optimal model using the cp that provides the highest accuracy.

The resulting decision tree easily translates itself to a series of rules that can be used to classify the test data. The advantages of using decision tree classification lie in its ease of application, particularly as both numerical and categorical input variables require little to no pre-processing, its transparency for interpretation, as the resulting tree can be explained using Boolean logic, and its computational efficiency particularly with large datasets. In addition, decision tree classification does not require domain knowledge or parameter setting [31]. However, traditional decision trees are also the least robust of the machine learning classification methods as they are prone to overfitting and therefore rely very heavily on the training data. Often a small change in the training data can result in large changes in the developed tree. These shortcomings can be addressed using the Random Forest Algorithm.

Random Forest

The random forest algorithm [32] works in a similar manner to decision trees, but where the CART algorithm results in a single tree, the random forest algorithm results in a 'forest' of trees. Each of the maximal trees within the random forest will have been developed using a random subset of the predictor variables [33]. Each split within the tree is then calculated according to a given performance metric from only within this subset of variables. Typically, many trees are considered, thus reducing the prediction error, as the model prediction will reflect the average prediction across all the trees. As a result, the random forest algorithm is considered robust, flexible and highly suited to large datasets [34]. The random forest algorithm in R's caret package has been used in this study. This implementation chooses a range of mtry parameters, where mtry is the number of variables available for splitting at each tree node, which have a strong influence on predictor variable importance estimates [35]. The mtry parameter providing the highest accuracy is used to select the optimal model.

Naïve Bayes

The Naïve Bayes algorithm for classification is based on Bayes Theorem which describes the most likely outcome (Y) based on k number of observations ($X=\{x_1,x_2,\ldots x_k\}$). This can be written as P(Y|X) and, as the algorithm is 'naïve' and all variables are considered independent, is calculated using Equation (2).

Equation (2)

$$P(Y|\mathbf{X}) = \frac{P(\boldsymbol{X}|Y)P(\boldsymbol{X})}{P(Y)} = \frac{1}{P(Y)}\prod_{i}^{k} P(x_i|Y)P(x_i)$$

The probability of an outcome P(Y), the probability of an observation being described by X, P(X), and the probability of an observation being described by X, given that they can be classed by Y, P(X|Y), can all be estimated using the given dataset. For its use as a classifier, an observation is classified according to the most likely class based on the random variables the observation describes. The benefits of the Naïve Bayes classifier are its theoretical low error rate, however, based on its underlying independence of the variables in practice this may not be the case. The Naïve Bayes algorithm in R's caret package has been used in this study.

Importance and Accuracy Measures

Gini Impurity Index

The Gini impurity index describes the likelihood of an incorrect classification using a random variable (var) and is described mathematically as Equation (3).

Equation (3)

$$\text{Gini}(\text{var}) = 1 - \sum_{i}^{m} p_i$$

Where p_i is the probability of a correct classification, according to m classes. By considering the variables resulting in a minimal Gini impurity index, this metric will therefore determine the best (most pure) variables to use in order to split the training data until a convergence criterion is met.

Permutation Importance

Permutation variable importance [32] is calculated by using the effect the variable has for the overall prediction performance. This performance can be predicted using the out-of-bag (OOB) prediction error, calculated by taking a mean prediction error rate of those trees that did not include the specific variable [34].

Performance Evaluation

To compare the performance of each classification model, a variety of evaluation metrics were used. The accuracy, precision, recall and F1 scores were computed. Precision, recall and F1 scores take account of true and false positives and negatives, respectively, whereas accuracy considers only true positives and true negatives [36].

Results

Feature Selection

Manual selection

Initially 6 features deemed irrelevant for analysis were removed including participant identification numbers and cohort category (which described the clinic from which the participants were selected). Nine free text variables and 9 variables with inconsistent questioning were removed. Additionally, 94 subjective features

were removed in favour of more objective laboratory-obtained results. Several of the removed subjective features had high numbers of missing values and therefore removal of these in favour of subjective features assisted in handling missing data whilst ensuring that there was no information loss within the dataset and data duplication was also minimised. For example, nutritional status based on blood analysis (e.g. measure of key vitamin biomarkers) was retained over self-reported dietary intake (e.g. supplement and fortified food use).

Correlation/correspondence analysis

Initial investigation into cognitive function with the TUDA dataset, as measured using the RBANS score, highlights that as expected RBANS decreases with age (Figure 2).

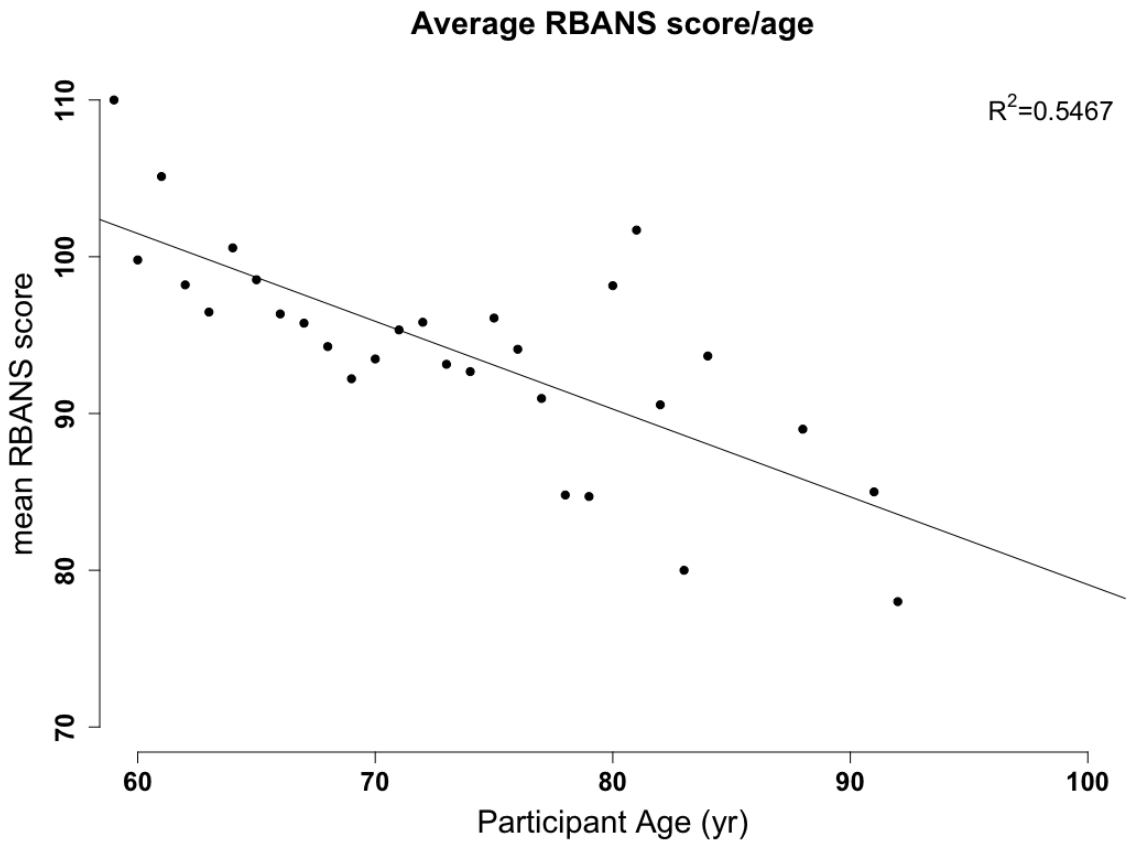

Figure 2. Mean RBANS score as a function of participant's age. The graph shows a general decrease in RBANS score as age increases.[a]

[a] RBANS scores have been averaged by age, thus each point represents the average score for any particular age. One outlier existed for age = 86. This has been removed and the R value recalculated accordingly.

Correlation analysis and association analysis was carried out. The key results of this analysis are shown in (Multimedia Appendix 1). We observe a relationship between variables concerning follow-up questions within the questionnaire (e.g. medication use and duration of use). Based on this 41 features related to follow-up questions

were removed. We also observe a high correlation between the use of specific medications (e.g. bisphosphonate medications: Risedronate, Ibandronic acid and Etidronate). These medications could be grouped into bone and hormone-related categories and therefore we amalgamated each respective subset into a new variable. Specifically, two new variables were added for bone and hormone-related medication, encompassing the many types of bone medications, including the bisphosphonates, and hormone–related medications from the original dataset, respectively. This resulted in the removal of 30 features and the addition of 2 new features. Furthermore, scores for each assessment element of RBANS were removed and only the total score was retained. The total RBANS score is later used as the target variable in classification.

We also removed the other neuropsychiatric test results (MMSE, FAB, Hospital anxiety and depression scale (HADS), Centre for Epidemiological Studies depression scale (CES-D), and functional test results (instrumental activities of daily living (IADL) and the physical self-maintenance scale (PSMS)) from the dataset, as they are clinical assessment tools as opposed to individual predictor variables. This resulted in the removal of a further 72 features. The correlation matrix between these scores is shown in Figure 3.

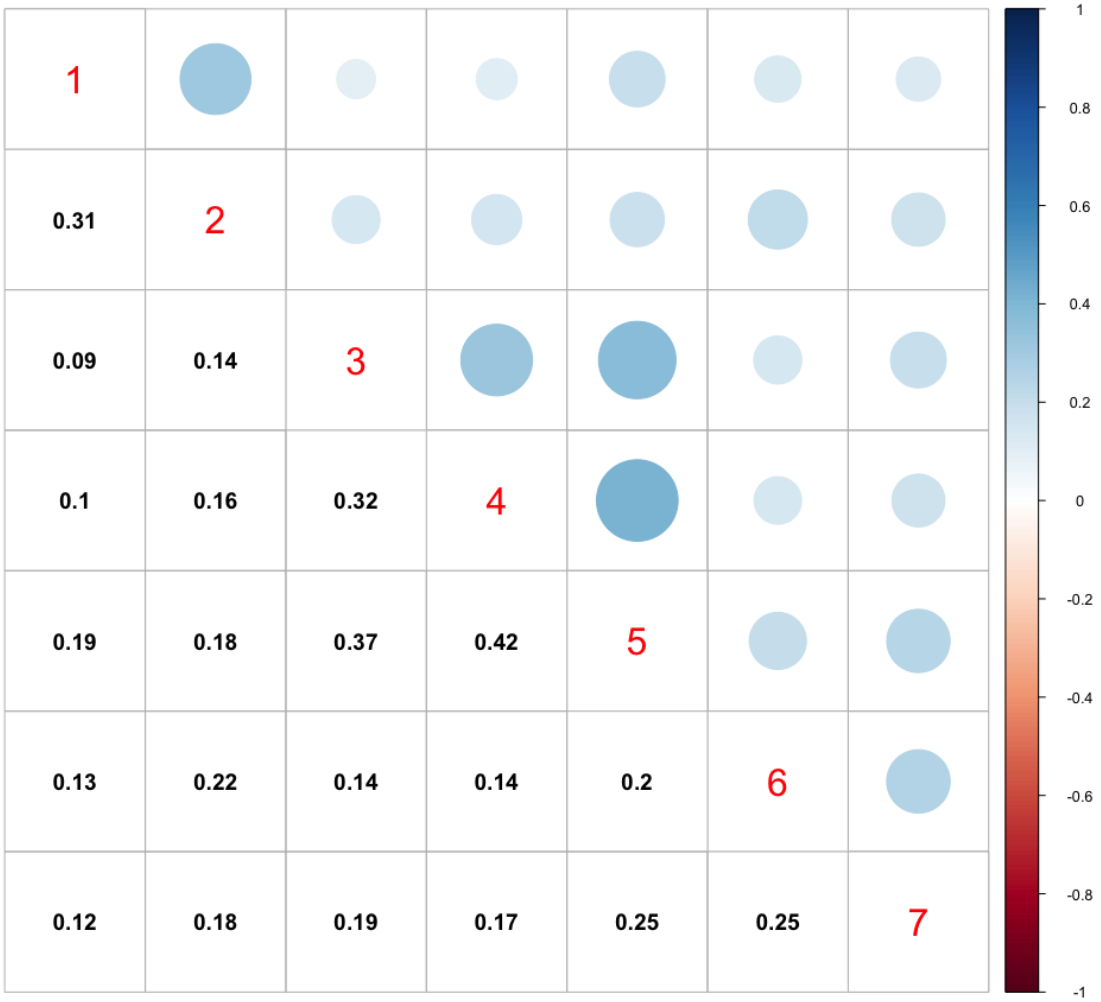

Figure 3. Correlation matrix using Spearman's (nonparametric) coefficient between participant test scores, ignoring observations with missing data[a]

[a] Variable descriptors are as follows: 1 = Hospital Anxiety and Depression Scale (HADS) total score; 2 = Depression questionnaire total score; 3 = Mini Mental State Examination (MMSE) total score; 4 = Frontal Assessment Battery (FAB) total score; 5 = Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) total score; 6 = Physical Maintenance Scale (PSMS) total score; 7 = Instrumental Activities of Daily Living (IADL) total score

The resulting subset of features following this stage of selection reduced the dataset from 345 variables to 69 plus the class variable (RBANS score) (see Multimedia Appendix 2).

Clustering

Clustering analysis was carried out using the ClustOfVar package within R Studio [29] to determine variable clusters and the strengths of their relationships. As expected, the scores from the clinical assessments, RBANS and its sub-component tests, FAB and MMSE, are closely related (Figure 4). The participant's age was closely related to kidney function, as indicated by the glomerular filtration rate (GFR), and together these form a cluster with the physical diagnostic tests of IADL, TUG and PSMS (Figure 5).

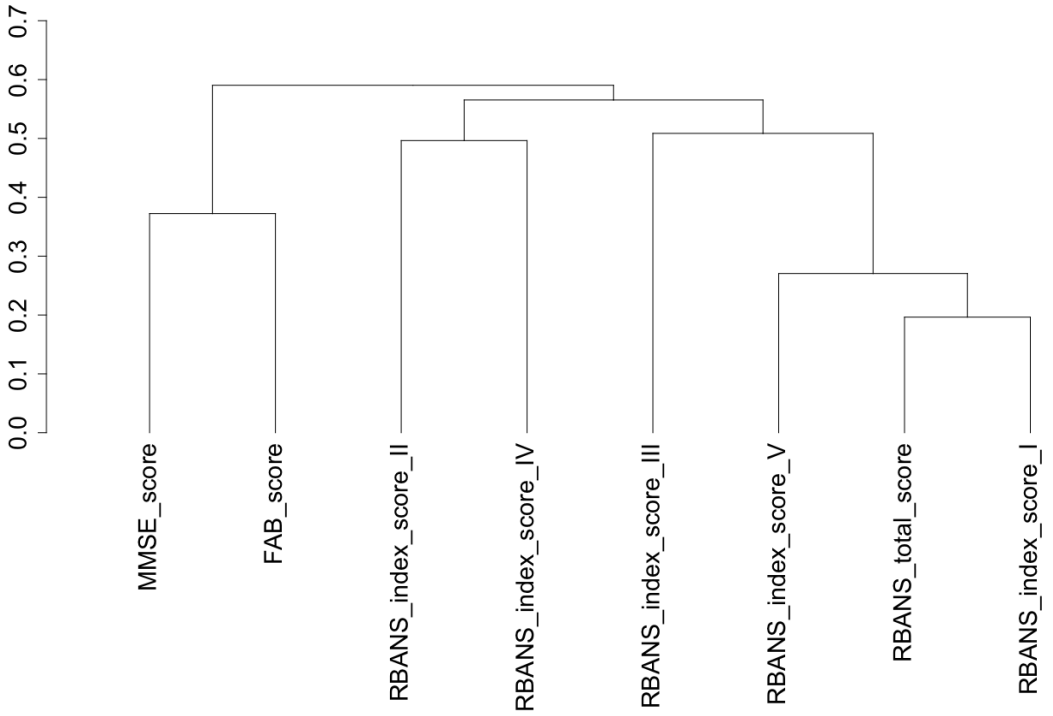

Figure 4. Hierarchical clustering of variables depicted as a dendrogram showing strong relationships between clinical assessment scores from the RBANS, FAB and MMSE assessments[a]

[a] Variable descriptors are as follows: MMSE_score = Mini Mental State Examination (MMSE) total score; FAB_score = Frontal Assessment Battery (FAB) total score; RBANS_index_score_I = Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) immediate memory score; RBANS_index_score_II = Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) visuospatial/constructional score; RBANS_index_score_III = Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) language score; RBANS_index_score_IV = Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) attention score; RBANS_index_score_V = Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) delayed memory score; RBANS_total_score = Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) total score

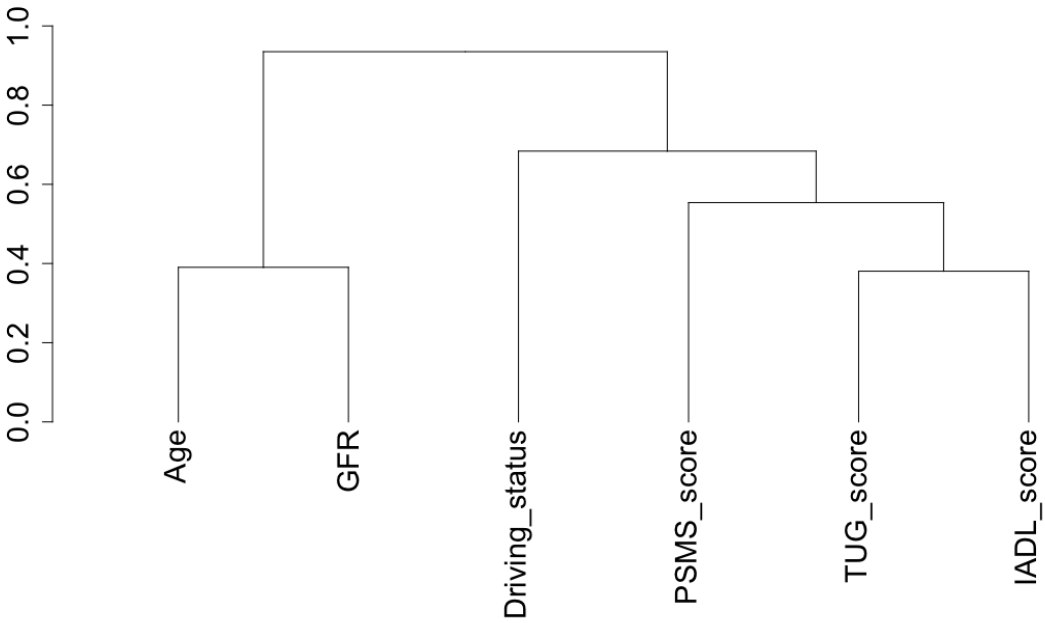

Figure 5. Hierarchical clustering of variables depicted as a dendrogram showing the close relation between a participant's age and kidney function (GFR), which together form a cluster with the physical diagnostic tests of IADL, TUG and PSMS[a]

[a] Variable descriptors are as follows: Age = participant's age; GFR = kidney function; Driving_status = driving status; PSMS_score = Physical Maintenance Scale (PSMS) total score; TUG score = Timed Up and Go score; IADL_score = Instrumental Activities of Daily Living (IADL) total score

Following feature selection, the dataset contained 69 features and 5186 observations, however missing data still remained. To retain as much data as possible whilst minimising the chance of statistical bias, participant records were imputed by replacing missing values with the average or expected value, in this case, according to the participant's age and gender. As in other studies of RBANS score [37], participants with visual (224 participants) or arthritic problems (1445 participants) were omitted as they would have been hindered from carrying out certain tasks within the test and thus their results may be unreliable, as were those displaying an MMSE score of below 24 (647 participants). Upon removing the relevant records, 2869 observations remained.

RBANS Classification

Classification models were utilised for two purposes: to discover if a model could be developed to predict a low RBANS score, representing poorer cognitive function, from the TUDA dataset; and to determine if the developed model could be used to identify key health, nutritional and environmental predictors of these low scores .

The target variable in this analysis is the RBANS total score. For this analysis, the RBANS score was categorised using a data-driven clustering approach to find 2 natural groupings within the data identifying those with poorer cognitive performance as having an RBANS score <70 (assigned class 'low') and an RBANS score >= 70 was indicative of normal cognitive performance (assigned class 'high').

Class imbalance [38] within the dataset was resolved using over-sampling, in which a random sample of the smaller class was replicated until class sizes were equal.

The supervised modelling techniques of decision trees, random forest and Naïve Bayes were applied with 69 predictor variables (listed in Multimedia Appendix 2). The dataset (n = 2869) was split into a training set (75%, n = 2152) and an evaluation set (25%, n = 717). The models were trained using the training set with 10-fold cross-validation applied and the results are shown in (Table 2). For the decision tree model the complexity parameter value of 0.020 for pruning was found to produce the highest accuracy. For the random forest model the mtry value of 58 was found to produce the highest accuracy.

Table 2. Classification of RBANS score performance measures when models trained with 10-fold cross validation (training set size = 2869)

Classification Technique	Accuracy mean (SD)	Precision mean (SD)	Recall mean (SD)	F1 mean (SD)
Decision Tree	0.737 (0.020)	0.795 (0.037)	0.643 (0.051)	0.709 (0.028)
Naïve Bayes	0.500 (0.000)	0.500 (0.000)	1.000 (0.000)	0.667 (0.000)
Random Forest	0.990 (0.006)	1.000 (0.000)	0.981 (0.011)	0.990 (0.006)

The models were then evaluated using the held out 25% evaluation dataset and the accuracy of these models ranged from 60.4% using the decision tree, to 87.7% using the random forest algorithm (Table 3). The random forest algorithm performed best in this comparison in terms of both accuracy and F1 score, with the decision tree algorithm performing worst. This is expected in terms of robustness, specifically pertaining to problems with overfitting by the decision tree algorithm, which has been rectified somewhat using multiple trees within the random forest.

Table 3. Classification of RBANS score performance measures when applied to the evaluation dataset (training set size = 2869, evaluation set size = 717)

Classification Technique	Overall Accuracy	Precision	Recall	F1
Decision Tree	0.604	0.926	0.596	0.725
Naïve Bayes	0.876	0.876	0.100	0.934
Random Forest	0.877	0.882	0.992	0.934

The key predictors of the RBANS total score in the decision tree were as follows: the participants scores from the TUG functional mobility test, representing the time it takes a participant to get out of a chair, walk three metres, turn around and walk back to return to their original seated position; the age the participant left education; whether any family members were concerned about the participant's memory, and the participant's GFR as shown in (Figure 6). This decision tree predicted that a person who took under 13 seconds to perform the TUG test and left education after 16 years old was classed as a 'high' RBANS scorer (i.e. indicative of normal cognitive performance). The decision tree classification model also

highlights the importance of the TUG test alone; if a participant took longer than 13 seconds to perform the test, they were most likely to be a low scorer, indicative of poorer cognitive performance.

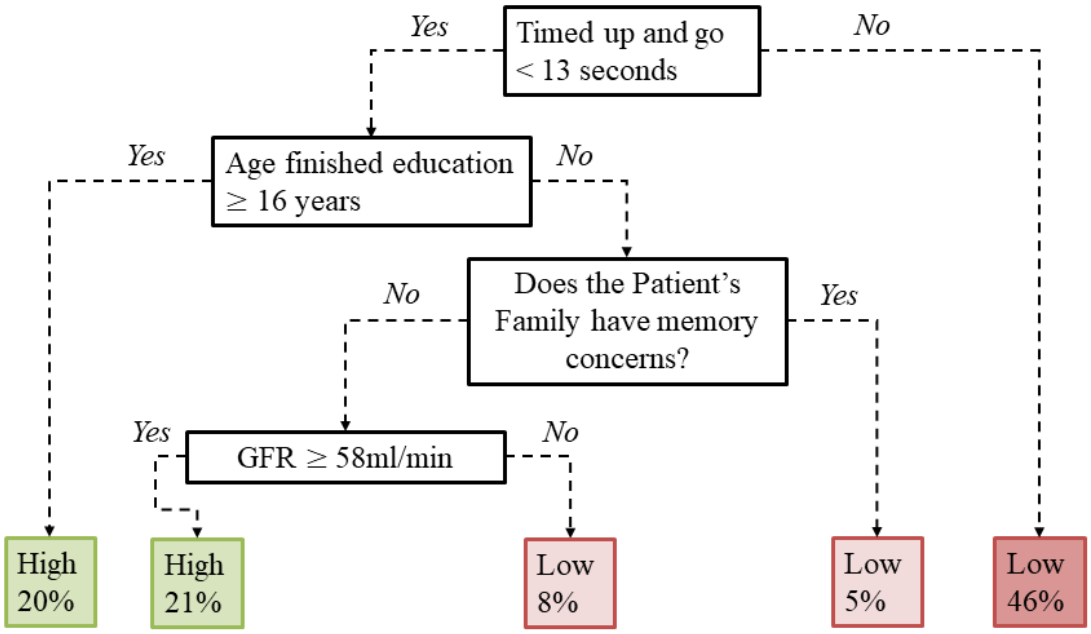

Figure 6. Decision tree classifier of RBANS score

Similarly, the Naïve Bayes and random forest algorithms also detect the TUG score, the age the participant left education and the participant's age as being highly informative features as shown in (Figure 7) and (Figure 8) for Naïve Bayes and random forest models, respectively, with the Naïve Bayes algorithm adding a participant's driving status, and the random forest algorithm adding GFR to form the top four informative variables within these respective algorithms.

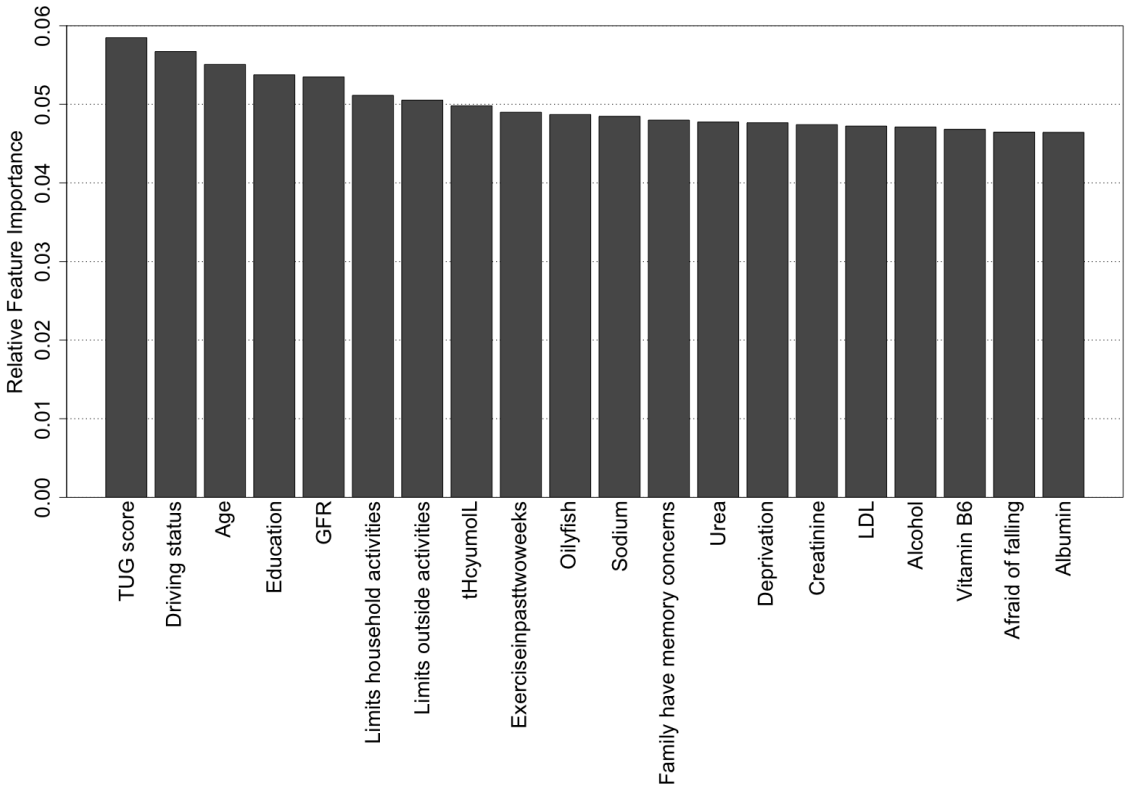

Figure 7. Twenty most important features for classification of RBANS score as detected using feature permutation using a Naïve Bayes classifier (see Multimedia Appendix 2 for feature descriptions)

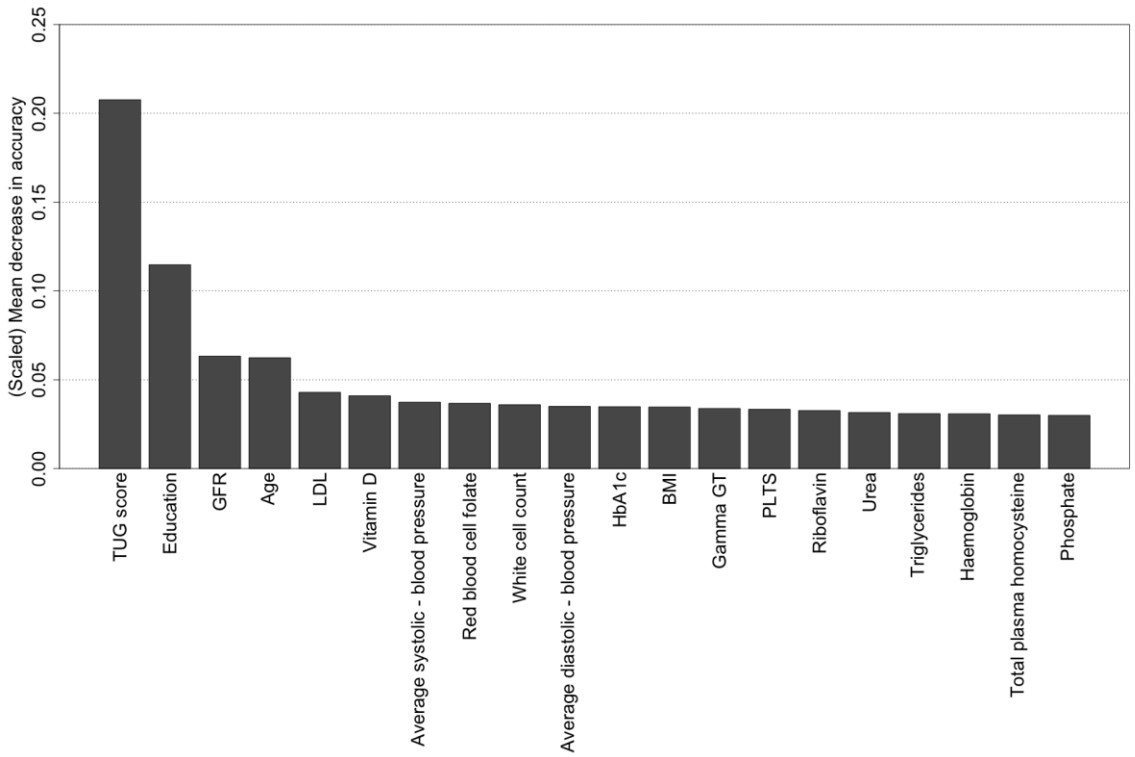

Figure 8. Twenty most important features for classification of RBANS score as detected using feature permutation using a random forest classifier (see Multimedia Appendix 2 for feature descriptions)

The informative nature of the four most important features determined by the most accurate classifier (random forest), as shown in (Figure 8) was confirmed when these algorithms were rerun using only this subset of four features. Again, 10-fold cross validation was applied to train the model on the training dataset (n = 2152) with results shown in (Table 4). For the decision tree model the complexity parameter value of 0.010 for pruning was found to produce the highest accuracy. For the random forest model the mtry value of 2 was found to produce the highest accuracy. The models were then evaluated using the held out 25% evaluation dataset. Training on the four most important features as determined by the random forest model resulted in a decrease in accuracy for the random forest model from 87.7% to 80.1% (Table 5). A larger reduction in accuracy was observed for the Naïve Bayes model, decreasing from 87.6% to 69.3%, whilst the decision tree model increased in accuracy from 60.4% to 72.5% when trained on this reduced dataset compared with training on the original dataset containing 69 variables.

Table 4. Classification of RBANS score performance measures when models trained with 10-fold cross validation (training set size = 2869) and four key variables: (1) age the participant left education, (2) TUG Score, (3) GFR measure and (4) participant's age

Classification Technique	Accuracy mean (SD)	Precision mean (SD)	Recall mean (SD)	F1 mean (SD)

Decision Tree	0.688 (0.020)	0.702 (0.026)	0.655 (0.045)	0.677 (0.020)
Naïve Bayes	0.693 (0.012)	0.775 (0.021)	0.545 (0.026)	0.640 (0.018)
Random Forest	0.929 (0.013)	1.000 (0.000)	0.857 (0.026)	0.923 (0.015)

Table 5. Classification of RBANS score performance measures when models trained using four key variables: (1) age the participant left education, (2) TUG Score, (3) GFR measure and (4) participant's age when applied to the evaluation dataset (training set size = 2869, evaluation set size = 717)

Classification Technique	Overall Accuracy	Precision	Recall	F1
Decision Tree	0.725	0.928	0.732	0.819
Naïve Bayes	0.598	0.946	0.557	0.701
Random Forest	0.801	0.878	0.889	0.883

Classifying cognitive decline using rate of change of RBANS score

A subset (n=987) of TUDA study participants were reassessed using an identical protocol 5-7 years after the initial assessment. A result of this follow-up assessment enabled the creation of a new variable to add to the original TUDA dataset for these 987 participants; the rate of change of RBANS score. This variable would act as a measure of predicted cognitive decline (or improvement) over the 5-7 year follow-up period. The same classification models of decision tree, Naïve Bayes and random forest were applied to the TUDA data (n=987), using the new 'Rate of RBANS change' as the classification variable. This variable was categorised using the scheme that one half standard deviation below the mean was 'Acute Decline', and above this boundary was 'Normal', or expected change. The variable was normalised to adjust for differing periods of time between first and second RBANS assessments (between 5 and 7 years) amongst participants. The dataset (n = 987) was split into a training set (75%, n = 740) and an evaluation set (25%, n = 247). The models were trained using the training set with 10-fold cross-validation applied and the results are shown in (Table 6). For the decision tree model the complexity parameter value of 0.035 for pruning was found to produce the highest accuracy. For the random forest model the mtry value of 2 was found to produce the highest accuracy.

Table 6. Classification of RBANS score performance measures when models trained with 10-fold cross validation (training set size = 987)

Classification Technique	Accuracy mean (SD)	Precision mean (SD)	Recall mean (SD)	F1 mean (SD)
Decision Tree	0.603 (0.045)	0.613 (0.053)	0.571 (0.151)	0.582 (0.083)
Naïve Bayes	0.499 (0.008)	0.499 (0.008)	0.997 (0.009)	0.665 (0.007)
Random Forest	0.962 (0.026)	0.978 (0.035)	0.946 (0.031)	0.962 (0.028)

The models were then evaluated using the held out 25% evaluation dataset and the results are shown in (Table 7). Although the accuracy from these classification models is lower than that reported for the classification of the RBANS score, ~70% vs 90% for random forest classifiers, it nevertheless indicates the possibility of

using our existing variables for predicting a perhaps pathological rate of cognitive decline to a reasonable level of accuracy. The decision tree performed poorest, however the information it provides (Figure 9) indicates that TUG test score is again the most informative attribute, followed by the participant's blood measures of total plasma homocysteine (tHcy), vitamin B6 biomarker (PLP) and HbA1c.

Table 7. Classification performance for rate of change of RBANS score when applied to the evaluation dataset (training set size = 740, evaluation set size = 287)

Classification Technique	Overall Accuracy	Precision	Recall	F1
Decision Tree	0.547	0.735	0.605	0.664
Naïve Bayes	0.739	0.739	1.000	0.850
Random Forest	0.702	0.735	0.933	0.822

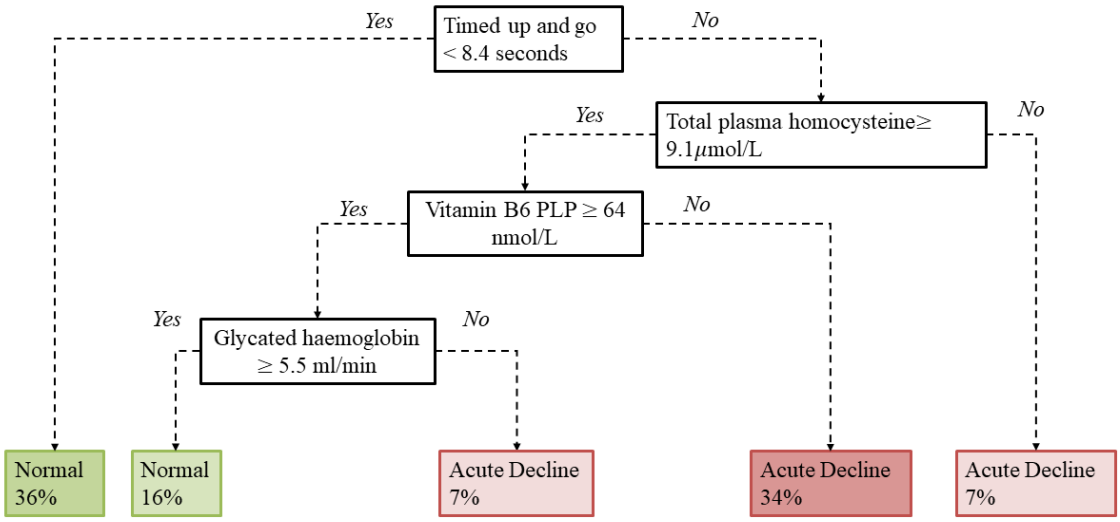

Figure 9. Decision tree classifier of rate of change of RBANS score

Furthermore, using permutation importance measures (Figures 10 and 11), it has been indicated that the same key variables for the classification of RBANS score are no longer of such importance for the classification of rate of RBANS score change. Instead, the blood measures of vitamin B6 PLP and Urea, coupled with the results of the TUG test and the participant's age are likely key predictors, particularly using the (best performing) Naïve Bayes algorithm (Figure 10).

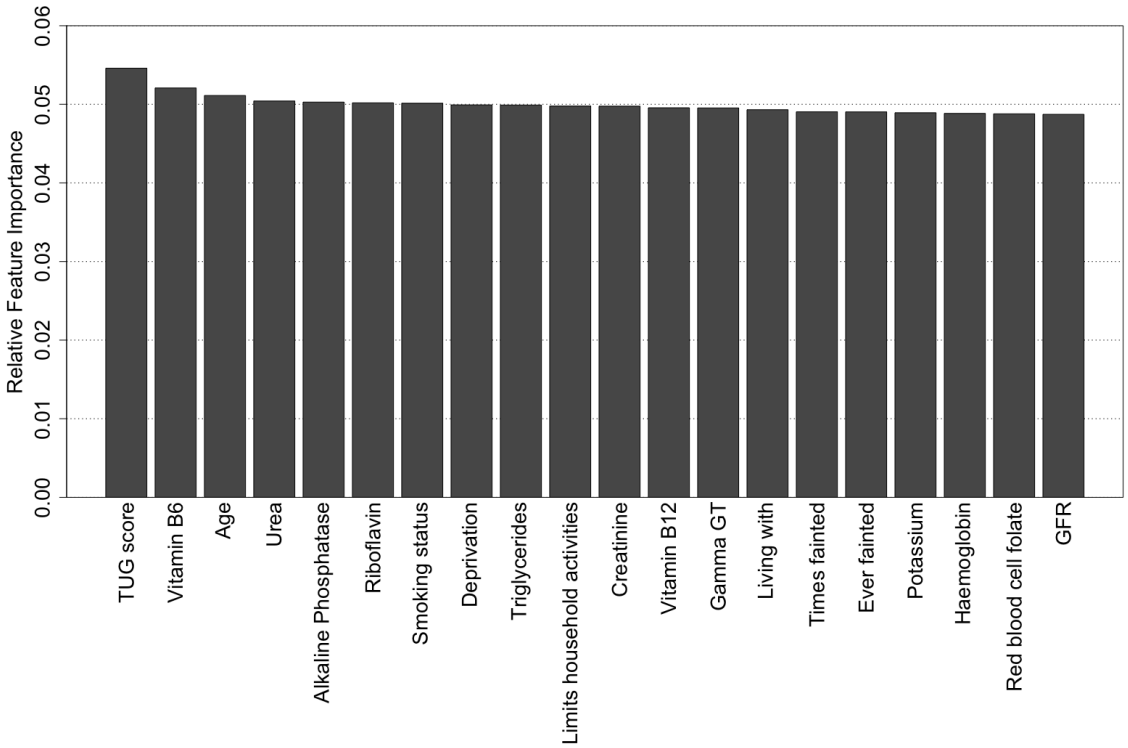

Figure 10 Twenty most important features for predicting rate of RBANS change as detected using feature permutation using a Naïve Bayes classifier (see Multimedia Appendix 2 for feature descriptions)

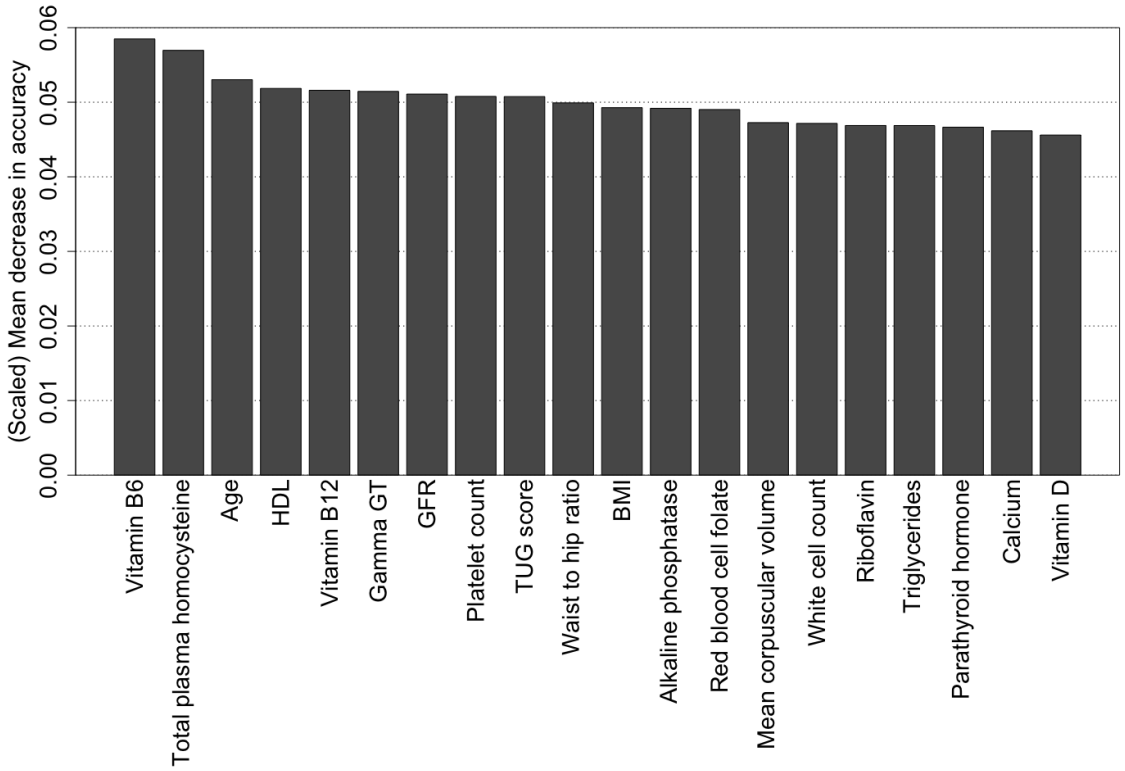

Figure 11 Twenty most important features for predicting rate of RBANS change as detected using feature permutation using a Random Forest classifier (see Multimedia Appendix 2 for feature descriptions)

Discussion

The results of this study indicate that modelling of a variety of clinical, lifestyle and socio-demographic factors using machine learning techniques may help predict poorer cognitive function in older people with a high level of accuracy (circa 90%) and using a small number of non-invasive indicators. The approach is also useful, although slightly less accurate (c.70%), in helping to predict the rate of cognitive decline over a 5-7 year time period with a small number of measures being the most influential health, nutritional and environmental predictors. The results are important for clinicians and health service providers, especially at the early stages of engagement and diagnosis of cognitive dysfunction in older patients, by identifying those patients most in need of more intensive investigation. Furthermore, these findings may be useful for informing nutritional and lifestyle interventions aimed at maintaining brain health in the adult population.

The results presented here suggest that it may well be possible for a healthcare professional to make an initial prediction (with a high level of confidence) of cognitive dysfunction using only a few short, non-invasive questions. Whilst the approach is not a diagnostic instrument for detecting the presence or absence of dementia, it has particular merit in that it could provide a very quick, efficient and

non-invasive screening method to help clinicians decide, at an early consultation stage, whether or not a patient should be investigated further using more in-depth cognitive assessment tools. Similarly, a recent study [14] using a machine learning approach to develop a Gradient Boosting Machine classifier with the KLoSA dataset [15], also identified sociodemographic, functional and health related factors, amongst others, as the most important predictors of cognitive impairment. The authors concluded that the model could be used to screen for cognitive impairment in a community healthcare setting. Using such an approach may therefore offer potential benefits to both health service providers and older patients. It may make time and cost savings for health service providers as cognitive tests are often laborious to administer (e.g. approximately 30 minutes for RBANS used in the current study) and avoid testing of low-risk patients. As a result, any unnecessary stress associated with cognitive testing may be reduced or avoided in older adults. The current results also suggest that some additional invasive clinical measures may be required to identify those individuals at greatest risk of future cognitive decline, providing valuable information that could help clinicians design the most appropriate intervention and treatment strategies for patients on a case by case basis.

It is interesting in the prediction of poorer cognitive performance that, in addition to participant age, the models identified non-invasive physical, behavioural and socio-economic variables over invasive clinical measures as the most influential predictors (with the exception of GFR) while the opposite was true for predicting rate of change (with TUG being the exception). This would suggest that non-clinical factors are much better in predicting poorer cognitive performance in older people, while clinical measures are needed to predict cognitive decline.

Machine learning methods produce the best classification models and predictive outcomes based on the quality and quantity (comprehensiveness) of the input variables. The potential for bias still remains, for example, when a key variable is missing from the data. Consequently, the results from the models need to be evaluated for theoretical and, in health outcome studies, clinical plausibility to determine their value and potential for real world application [39].

In this study, all three models identified TUG and the age a participant left education as the most important predictive variables. In terms of plausibility, this is encouraging as both these factors have been frequently identified and cited in the literature from large cohort studies as being important risk factors of cognitive dysfunction [6,40]. In support of these findings, we previously reported using geo-demographic analysis of the current cohort that socioeconomic status, namely area-based deprivation, was an important determinant of cognitive dysfunction alongside age, years of education, depression, and TUG test [41]. The emergence of age left education as the dominant variable from the socio-economic cluster is particularly interesting as it has consistently been found to be the most important individual socio-economic factor related to cognitive function across the lifecycle [42]. Furthermore, two recent population based longitudinal studies in the USA and UK

have indicated that higher educational attainment, particularly in early life could help protect against a decline in cognitive function as people age [43,44]. Reduced physical function, measured using tools like TUG, has also been associated with lower socio-economic status [45] and cognitive dysfunction [46]. The TUG test reflects an individual's strength and mobility, inherently assessing gait, balance and to a lesser degree cognition and vision. It is a screening tool routinely used to assist clinicians identifying patients at risk of falling [47]. A cut off of ≥12 seconds is commonly applied to identify individuals at high risk of falls, but these cut off levels are applied differently across various studies [48]. Within the current study a TUG score of >13 seconds was associated with poor cognitive performance and a score of >8 seconds predicted future risk of cognitive decline. These selected predictors, and their associated split points, from the machine learning analytics, are consistent with other studies, where poor functional performance was correlated with lower executive function in those with MCI and Alzheimer's disease [49,50] and is associated with future dementia occurrence [51]. Moreover, the TUG test can be considered, in a sense, a global measure of body function. Poor performance has been associated with increased CVD and mortality, as well as all-cause mortality in older adults [52-54] and in patients with chronic kidney disease [55]. Additional predictors beyond TUG selected in the decision trees as informative are also linked with poor cognitive performance, including a measure of kidney function, GFR. Low GFR is associated with poorer cognitive performance [56], with a recent study reporting individuals with impaired kidney function had lower cognitive performance compared to individuals with normal kidney function. Furthermore, in frail older adults with poor TUG scores, the severity of renal dysfunction is independently correlated with cognitive impairment [57]. Consequently, it is clear that the various machine learning approaches are identifying appropriate factors with known links to cognitive performance.

When the machine learning approaches were applied to identify the predictors of rate of cognitive decline in TUDA participants over a 5-7 year follow-up period, vitamin B6 biomarker status (as measured by blood concentrations of the active form of the vitamin, pyridoxal-5'-phosphate) at baseline emerged, after the TUG test, as one of the key predictors. High proportions of older adults in population-based surveys from the US and Europe, including the UK, are reported to have deficient or low B6 status [58]. Vitamin B6 has a number of important biological roles including immunomodulating effects, and in clinical and population-based studies, blood B6 concentrations are found to be inversely associated with inflammatory conditions, neurodegenerative diseases and depression, and to predict the risk of cardiovascular disease and certain cancers [59]. Of note, vitamin B6 and related B-vitamins (namely, folate, vitamin B12 and riboflavin) are required as cofactors in one-carbon metabolism, a series of essential reactions involving the transfer of one-carbon units for DNA synthesis and repair, homocysteine metabolism and in the methylation of phospholipids, proteins, DNA and neurotransmitters [60]. There is a growing body of evidence to indicate that one-carbon metabolism and related B-vitamins may be important for maintaining cognitive health in ageing. The majority of research to date has focused on folate and vitamin B12. Although vitamin B6 has

been less extensively investigated, the findings of the current paper are in agreement with other observational studies. Low vitamin B6 status has been associated with cognitive dysfunction [61,62] and cognitive decline [63,64] in older people. Low vitamin B6 status was associated with cognitive decline in the Veterans Affairs Normative Ageing Study [64]. More recently, low baseline status of vitamin B6 was also associated with a greater than expected rate of cognitive decline in a cohort of community dwelling older adults in Northern Ireland [63]. Of greater importance, a number of randomised controlled trials demonstrated that vitamin B6 supplementation in combination with other B-vitamins reduces the rate of cognitive decline in older people [65,66] and a reduced rate of brain atrophy as measured using MRI [67]. Furthermore, other evidence from the TUDA study indicates that vitamin B6 along with folate and riboflavin are associated with an increased risk of depression [7]. This machine learning approach has identified vitamin B6 as an important determinant of cognitive health in the TUDA study and whilst biologically plausible and supported by other scientific evidence the possible beneficial effects of vitamin B6 on cognitive health would need to be confirmed in randomised controlled trials.

What is very interesting from a clinical setting are the changes in the selected predictors within machine learning models when comparing the RBANS total score model versus the rate of change of the RBANS score model. The age a participant left education is a dominant predictor from the socio-economic cluster in the RBANS total score model however, it becomes an uninformative predictor for the rate of change of the RBANS score model and actually disappears from the models. This implies that while this socio-economic factor is an important predictor in cognitive dysfunction (diagnosis), it is not important when predicting rate of cognitive decline. So, while patients may start off on a different baseline due to some socio-economic predictors, their rate of cognitive decline is not influenced by these socio-economic predictors.

Whilst this paper focuses on key health, nutritional and environmental predictors of cognitive dysfunction and rate of change of cognitive function using machine learning techniques, as part of the project, the research team also sought input from Personal and Public Involvement (PPI): patients, carers, and clinicians. This engagement focused on causation of cognitive dysfunction, particularly in relation to age, activity and genetics, considered as measures of risk. This aspect of the work in terms of engagement with PPI, their expectations and how these align with the findings of this work will be the focus of a future research publication.

Limitations

The current study had a number of strengths but also limitations. The main limitation is that the TUDA study is observational in design and thus residual confounding and reverse causality cannot be ruled out in this analysis. Also, owing to the low instances of participants with poorer cognitive performance as indicated by an RBANS below 70 (target class = “low”), this class was underrepresented within the training dataset and therefore oversampling had to be performed to

allow for more balanced classifier training. This artificial approach of boosting the number of samples was necessary for the classifier, but, coupled with the imputation of missing data, no new information would have been attained. This led to an imbalance between the precision and recall accuracy metrics, although this was remedied with the use of the F1 score. Generally, the algorithms performed well in the classification of RBANS score. The decision trees performed poorest but, as explained in the text, they were still capable of drawing out key and transparent information. Whilst an extensive comparison of classification approaches was not the focus of this study, we recognise that alternative variations of the algorithms used in this study exist, for example, C4.5 and C5.0 for decision trees, as well as other learning algorithms such as neural networks and boosting algorithms. These alternative approaches may potentially yield better results, and we intend on investigating these in future whilst ensuring the interpretability of results remains a key objective. In addition, the performance of the classifiers could have been improved using a dimension reduction technique such as PCA, however this would have impacted the interpretability of the classifier, as was the objective of the paper.

The main strength of this study was the utilisation of data from the TUDA study, a large and comprehensively characterised cohort of older community-dwelling adults. Furthermore, a subset of the TUDA study cohort was re-examined 5-7 years later using standardised protocols at both timepoints. This enabled changes in cognition to be tracked over time and the rate of cognitive decline to be calculated compared to most observational studies that measure cognition at one time point only. The primary outcome of the study was based on the RBANS total test, a sensitive neuropsychiatric battery for global cognitive assessment. As comprehensive data was available, this permitted objective laboratory measures over subjective measures of nutritional status to be included in the analytical models, thus providing more robust data on predictors of cognitive function.

Conclusions

In conclusion, the derived classification models were able to identify a small number of key non-invasive predictors that are able to predict cognitive dysfunction and then rate of change of cognitive function with a high level of accuracy in the TUDA study. The TUG score, the age the participant left education and whether or not a participant's family reported memory concerns emerged as key predictors that could potentially be incorporated into a screening tool for cognitive dysfunction for healthcare professionals to identify individuals in need of further in-depth cognitive evaluation. Given the burden on healthcare resources, this could result in improvements in the efficiency of dementia screening and present cost and time savings for the relevant health professions. Furthermore, the results provide evidence to identify key targets that could be included in public health strategies aimed at dementia prevention. Further investigation is necessary to test the accuracy of the identified predictors in other large cohorts and using other cognitive assessment tools. The TUDA data enables extensive opportunities of future investigations of the aging population.

Acknowledgements
We acknowledge the support of the eHealth & Data Analytics Dementia Pathfinder Programme and HSCB eHealth Directorate in this work, under award 15F-1701.

Authors' Contributions
DR, BF and MB contributed to the design, model development, analysis and interpretation of the study. CH, LH, AM, CG, HMN, PC and JW contributed to the design and interpretation of data and models. BF and DR drafted the manuscript. CH, LH, AM, CG and HMN contributed to the clinical aspects of the manuscript. All authors reviewed the manuscript critically for scientific and technical content, and all authors gave final approval of this version for publication.

Conflicts of Interest
None declared.

Abbreviations
AIBL: Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing
ADNI: Alzheimer's Disease Neuroimaging Initiative
CART: Classification and regression tree
CES-D: Centre for Epidemiological Studies depression scale
CRISP-DM: Cross-Industry Process for Data Mining
CVD: Cardiovascular Disease
DXA: Dual-energy X-ray Absorptiometry
FAB: Frontal assessment battery
GFR: Glomerular filtration rate
HADS: Hospital anxiety and depression scale
HbA1c: Glycated haemoglobin
IADL: Instrumental activities of daily living
KLoSA: Korean Longitudinal Study of Ageing
MCA: Multiple correspondence analysis
MCI: Mild cognitive impairment
MMSE: Mini mental state examination
OASIS: Open Access Series of Imaging Studies
OOB: Out-of-bag
PCA: Principal component analysis
PLP: Vitamin B6 marker Pyridoxal 5-phosphate
PSMS: Physical maintenance scale
RBANS: Repeatable battery for the assessment of neuropsychological status
tHcy: Total plasma homocysteine
TUDA: Trinity-Ulster and Department of Agriculture study
TUG: Timed Up and Go

References
1. P. D. United Nations, Department of Economic and Social Affairs, World Population Prospects: the 2015 revision, key findings and advance tables. Popul. Dev. Rev. 2015; 41(3):557–561

2. Gauthier, S, Reisberg, B, Zaudig, M, Petersen, R, Ritchie, K, Broich, K, Belleville, S, Brodaty, H, Bennett, D, Chertkow, H, Cummings, J, de Leon, M, Feldman, H, Ganguli, M, Hampel, H, Scheltens, P, Tierney, M, Whitehouse, P, Winblad, B. Mild cognitive impairment. Lancet 2006; 367(9518):1262-70 [doi: https://doi.org/10.1016/S0140-6736(06)68542-5].
3. World Health Organisation. 2015. International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) Mental and behavioural disorders. Available at: http://apps.who.int/classifications/icd10/browse/2016/en#/
4. Alzheimer's Disease International. 2019. World Alzheimer Report 2019: Attitudes to dementia. Alzheimer's Disease International, London. Available at: https://www.alz.co.uk/research/WorldAlzheimerReport2019.pdf
5. Alzheimer's Disease International. 2014. World Alzheimer Report 2014: Dementia and risk reduction. An analysis of protective and modifiable factors, London. Available at: https://www.alz.co.uk/research/WorldAlzheimerReport2014.pdf
6. Livingston, G, Sommerlad, A, Orgeta, V, Costafreda, S, Huntley, J, Ames, D, Ballard, C, Banerjee, S, Burns, A, Cohen-Mansfield, J, Cooper, C, Fox, N, Gitlin, L, Howard, R, Kales, H, Larson, E, Ritchie, K, Rockwood, K, Sampson, E, Samus, Q, Schneider, L, Selbaek, G, Teri, L, Mukadam, N. Dementia prevention, intervention, and care. Lancet 2017; 390(10113):2673-2734 [doi: https://doi.org/10.1016/S0140-6736(17)31363-6]
7. Moore, K, Hughes, C, Ward, M, Hoey, L, McNulty, H. Diet, nutrition and the ageing brain; current evidence and new directions. Proceedings of the Nutrition Society 2018; 77(2):152-163 [doi: https://doi.org/10.1017/S0029665117004177]
8. Koh, H, Tan, G. Data Mining Applications in Health care. J. Healthc. Inf. Manag. 2011; 19(2):64–72. PMID: 15869215
9. Saeys, Y, Inza, I, Larrañaga, P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007; 23(19):2507–2517 [doi: https://doi.org/10.1093/bioinformatics/btm344]
10. Bedner, P, Steinhäuser, C. Crucial role for astrocytes in epilepsy. Pathological Potential of Neuroglia: Possible New Targets for Medical Intervention 2014. ISBN: 978-1-4939-0974-2
11. Albert, M, Zhu, Y, Moghekar, A, Mori, S, Miller, MI, Soldan, A, Pettigrew, C, Selnes, O, Li, S, Wang, MC. Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years. Brain 2018; 141(3):877-887 [doi: https://doi.org/10.1093/brain/awx365]
12. Linz, N, Tröger, J, Alexandersson, J, Wolters, M, König, A, Robert, P. Predicting dementia screening and staging scores from semantic verbal fluency performance. IEEE International Conference on Data Mining Workshops, ICDMW; 2017 Nov 18-21; New Orleans, LA, USA. IEEE; 2017 [doi: https://doi.org/10.1109/ICDMW.2017.100]
13. Petersen, R. Mild cognitive impairment as a diagnostic entity. J. Intern. Med. 2004; 256(3):183–194 [doi: https://doi.org/10.1111/j.1365-2796.2004.01388.x]

14. Na, K. Prediction of future cognitive impairment among the community elderly: A machine-learning based approach. Scientific Reports 2019; 9(3335) [doi: https://doi.org/10.1038/s41598-019-39478-7]
15. Korea Employment Information Service. 2015. Korean Longitudinal Study of Ageing (KLoSA). https://survey.keis.or.kr/eng/klosa/klosa01.jsp
16. Marcus, D, Fotenos, A, Csernansky, J, Morris, J, Buckner, R. Open Access Series of Imaging Studies (OASIS): Longitudinal MRI Data in Nondemented and Demented Older Adults. J Cogn Neurosci. 2010; 22(12):2677-2684 [doi: https://doi.org/10.1162/jocn.2009.21407]
17. Petersen, R, Aisen, P, Beckett, L, Donohue, M, Gamst, A, Harvey, D, Jack, C, Jagust, W, Shaw, L, Toga, A, Trojanowski, J, Weiner, M. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology. 2010; 74(3):201-209. [doi: https://doi.org/10.1212/WNL.0b013e3181cb3e25]
18. Ellis, K, Bush, A, Darby, D, De Fazio, D, Foster, J, Hudson, P, Lautenschlager, N, Lenzo, N, Martins, R, Maruff, P, Masters, C, Milner, A, Pike, K, Rowe, C, Savage, G, Szoeke, C, Taddei, K, Villemagne, V, Woodward, M, Ames, D, AIBL Research Group. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr. 2009; 21(4):672-687. [doi: https://doi.org/10.1017/S1041610209009405]
19. Ding, X, Bucholc, M, Wang, H, Glass, D, Wang, H, Clarke, D, Bjourson, A, Dowey, L, O'Kane, M, Prasad, G, Maguire, L, Wong-Lin, K. A hybrid computational approach for efficient Alzheimer's disease classification based on heterogeneous data. Scientific Reports 2018; 8(9774) [doi: https://doi.org/10.1038/s41598-018-27997-8]
20. Korolev, I, Symonds, L, Bozoki, A, Alzheimer's Disease Neuroimaging Initiative. Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification. PLoS ONE 2016; 11(2):e0138866 [doi: https://doi.org/10.1371/journal.pone.0138866]
21. Wirth, R, Hipp, J. CRISP-DM: Towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining; 2000 Apr 11-13; Manchester UK. Lancashire: Practical Application Company; 2000. ISBN: 1902426088
22. Randolph, C, Tierney, M, Mohr, E, Chase, T. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary Clinical Validity. J Clin Exp Neuropsychol 1998; 20(3):310-9 [doi: https://doi.org/10.1076/jcen.20.3.310.823]
23. Wold, S, Esbensen, K, Geladi, P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems 1987; 2(1-3):37-52 [doi: https://doi.org/10.1016/0169-7439(87)80084-9]
24. Obermeyer, Z, Emanuel, E. Predicting the future-big data, machine learning, and clinical medicine. New England Journal of Medicine 2016; 375(13):1216-1219 [doi: https://doi.org/10.1056/NEJMp1606181]

25. Guyon, I, Elisseeff, A. An introduction to variable and feature selection. J. Mach. Learn. Res. 2003; 3:1107-1135 [doi: https://doi.org/10.5555/944919.944968]
26. Campbell, M, Swinscow, T. Statistics at Square One. 11th edition. John Wiley & Sons; 2011. ISBN: 1444357808
27. Cramér, H. Mathematical Methods of Statistics (PMS-9). Princeton: Princeton University Press; 1999. ISBN: 0691080046
28. Glantz, S, Slinker, B, Neilands, T. Primer of Applied Regression & Analysis of Variance. 3rd edition. USA: McGraw-Hill Education; 2016. ISBN: 0071824111
29. Chavent, M, Kuentz-Simonet, V, Liquet, B, Saracco, J. ClustOfVar: An R package for the clustering of variables. J. Stat. Softw. 2012; 50(13):1-16 [doi: https://doi.org/10.18637/jss.v050.i13]
30. Gordon, A, Breiman, L, Friedman, J, Olshen, R, Stone, C. Classification and Regression Trees. Biometrics 1984; 40(3): 874
31. Han, J, Kamber, M, Pei, J. Classification: Basic Concepts. Data Mining. Concepts and Techniques. 3rd Edition. Morgan Kaufmann Publishers; 2011. p. 327-391. ISBN: 9780123814791
32. Breiman, L. Random Forests. Mach. Learn. 2001; 45:5-32 [doi: https://doi.org/10.1023/A:1010933404324]
33. Genuer, R, Poggi, JM, Tuleau-Malot, C. Variable selection using random forests. Pattern Recognit. Lett. 2010; 31(14):2225-2236 [doi: https://doi.org/10.1016/j.patrec.2010.03.014]
34. Nembrini, S, Konig, IR, Wright, MN. The revival of the Gini importance?. Bioinformatics 2018; 34(21):3711-3718 [doi: https://doi.org/10.1093/bioinformatics/bty373]
35. Strobl, C, Boulesteix, A, Kneib, T, Augustin, T, Zeileis, A. Conditional variable importance for random forests. BMC Bioinformatics 2008; 9(307) [doi: https://doi.org/10.1186/1471-2105-9-307]
36. Davis, J, Goadrich, M. The relationship between precision-recall and ROC curves. Proceedings of the 23rd international conference on Machine learning; 2006 Jun 25-29; Pittsburgh, Pennsylvania, USA. ACM; 2006 [doi: https://doi.org/10.1145/1143844.1143874]
37. Hook, J, Marquine, M, Hoelzle, J. Repeatable battery for the assessment of neuropsychological status effort index performance in a medically ill geriatric sample. Arch. Clin. Neuropsychol. 2009; 24(3):231-235 [doi: https://doi.org/10.1093/arclin/acp026]
38. Longadge, R, Dongre, S. Class Imbalance Problem in Data Mining: Review. International Journal of Computer Science and Network 2013; 2(1)
39. Crown, W. Potential application of machine learning in health out-comes research and some statistical cautions. Value Health 2015; 18(2):137–140 [doi: https://doi.org/10.1016/j.jval.2014.12.005]
40. Kivimaki, M, Batty, G, Pentti, J, Shipley, M, Sipila, P, Nyberg, S, Suominen, S, Oksanen, T, Stenholm, S, Virtanen, M, Marmot, M, Singh-Manoux, A, Brunner, E, Lindbohm, J, Ferrie, J, Vahtera J. Association between socioeconomic status and the development of mental and physical health conditions in adulthood:

a multi-cohort study. Lancet Public Health 2020; 5(3):140-149 [doi: https://doi.org/10.1016/S2468-2667(19)30248-8]
41. McCann, A, McNulty, H, Rigby, J, Hughes, C, Hoey, L, Molloy, A, Cunningham, C, Casey, M, Tracey, F, O'Kane, M, McCarroll, K, Ward, M, Moore, K, Strain, J, Moore A. Effect of Area-Level Socioeconomic Deprivation on Risk of Cognitive Dysfunction in Older Adults. J Am Geriatr Soc 2018; 66(7):1269-1275 [doi: https://doi.org/10.1111/jgs.15258]
42. Norton, S, Matthews, F, Barnes, D, Yaffe, K, Brayne, C. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. Lancet Neurol 2014; 13(8): 788-94 [doi: https://doi.org/10.1016/S1474-4422(14)70136-X]
43. Langa, K, Larson, E, Crimmins, E, Faul, J, Levine, D, Kabeto, M, Weir, D. A comparison of the prevalence of dementia in the United States in 2000 and 2012. JAMA Intern Med 2017; 177(1):51-58 [doi: https://doi.org/10.1001/jamainternmed.2016.6807]
44. Wu, C, Odden, M, Fisher, G, Stawski, R. The association of retirement age with mortality: a population-based longitudinal study among older adults in the United States. J Epidemiol Community Health 2016; 70(9):917-923 [doi: https://doi.org/10.1136/jech-2015-207097]
45. Stringhini, S, Carmeli, C, Jokela, M, Avendaño, M, McCrory, C, d'Errico, A, Bochud, M, Barros, H, Costa, G, Chadeau-Hyam, M, Delpierre, C, Gandini, M, Fraga, S, Goldberg, M, Giles, G, Lassale, C, Kenny, R, Kelly-Irving, M, Paccaud, F, Layte, R, Muennig, P, Marmot, M, Ribeiro, AI, Severi, G, Steptoe, A, Shipley, M, Zins, M, Mackenbach, J, Vineis, P, Kivimäki, M. Socioeconomic status, non-communicable disease risk factors, and walking speed in older adults: multi-cohort population based study. BMJ 2018; 360:k1046 [doi: https://doi.org/10.1136/bmj.k1046]
46. Donoghue, O, Frances Horgan, N, Savva, G, Cronin, H, O'Regan, C, Kenny, R. Association between timed up-and-go and memory, executive function, and processing speed. J Am Geriatr Soc 2012; 60(9):1681-6 [doi: https://doi.org/10.1111/j.1532-5415.2012.04120.x]
47. Barry, E, Galvin, R, Keogh, C, Horgan, F, Fahey, T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta-analysis. BMC Geriatr. 2014; 14:14 [doi: https://doi.org/10.1186/1471-2318-14-14]
48. Lusardi, M, Fritz, S, Middleton, A, Allison, L, Wingood, M, Phillips, E, Criss, M, Verma, S, Osborne, J, Chui, K. Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability. J Geriatr Phys Ther 2017; 40(1):1–36 [doi: https://doi.org/10.1519/JPT.0000000000000099]
49. Blackwood, J, Shubert, T, Forgarty, K, Chase, C. Relationships Between Performance on Assessments of Executive Function and Fall Risk Screening Measures in Community-Dwelling Older Adults. J Geriatr Phys Ther 2016; 39(2):89-96 [doi: https://doi.org/10.1519/JPT.0000000000000056]
50. Ansai, J, Andrade, L, Nakagawa, T, Vale, F, Caetano, M, Lord, S, Rebelatto, J. Cognitive Correlates of Timed Up and Go Subtasks in Older People With

Preserved Cognition, Mild Cognitive Impairment, and Alzheimer's Disease. Am J Phys Med Rehabil 2017; 96(10):700-705 [doi: https://doi.org/10.1097/phm.0000000000000722]
51. Lee, J, Shin, D, Jeong, S, Son, K, Cho, B, Yoon, J, Park, B, Kwon, I, Lee, J, Kim, S. Association Between Timed Up and Go Test and Future Dementia Onset. J Gerontol A Biol Sci Med Sci 2018; 73(9):1238-1243 [doi: https://doi.org/10.1093/gerona/glx261]
52. Bergland, A, Jorgensen, L, Emaus, N, Strand, B. Mobility as a predictor of all-cause mortality in older men and women: 11.8 year follow-up in the Tromso study. BMC Health Serv Res 2017; 17(1):22 [doi: https://doi.org/10.1186/s12913-016-1950-0]
53. Son, K, Shin, D, Lee, J, Kim, S, Yun, J, Cho, B. Association of timed up and go test outcomes with future incidence of cardiovascular disease and mortality in adults aged 66 years: Korean national representative longitudinal study over 5.7 years. BMC Geriatr 2020; 20: 111 [doi: https://doi.org/10.1186/s12877-020-01509-8]
54. Chua, K, Lim, W, Lin, X, Yuan, J, Koh, W. Handgrip Strength and Timed Up-and-Go (TUG) Test are Predictors of Short-Term Mortality among Elderly in a Population-Based Cohort in Singapore. J Nutr Health Aging 2020; 24(4):371-378 [doi: https://doi.org/10.1007/s12603-020-1337-0]
55. Roshanravan, B, Robinson-Cohen, C, Patel, K, Ayers, E, Littman, A, De Boer, I, Alp Ikizler, T, Himmelfarb, J, Katzel, L, Kestenbaum, B, Seliger, S. Association between physical performance and all-cause mortality in CKD. J Am Soc Nephrol 2013; 24:822–830 [doi: https://doi.org/10.1681/ASN.2012070702]
56. Martens, R, Kooman, J, Stehouwer, C, Dagnelie, P, van der Kallen, C, Koster, A, Kroon, A, Leunissen, K, Nijpels, G, van der Sande, F, Schaper, N, Sep, S, van Boxtel, M, Schram, M, Henry, R. Estimated GFR, Albuminuria, and Cognitive Performance: The Maastricht Study. Am J Kidney Dis 2017; 69(2):179-191 [doi: https://doi.org/10.1053/j.ajkd.2016.04.017]
57. Coppolino, G, Bolignano, D, Gareri, P, Ruberto, C, Andreucci, M, Ruotolo, G, Rocca, M, Castagna, A. Kidney function and cognitive decline in frail elderly: two faces of the same coin? Int Urol Nephrol 2018; 50(8):1505-1510 [doi: https://doi.org/10.1007/s11255-018-1900-3]
58. Bates, C, Pentieva, K, Prentice, A. An Appraisal of Vitamin B6 Status Indices and Associated Confounders, in Young People Aged 4-18 Years and in People Aged 65 Years and Over, in Two National British Surveys. Public Health Nutr 1999; 2(4):529-35 [doi: https://doi.org/10.1017/s1368980099000713]
59. Ueland, PM, McCann, A, Midttun, O, Ulvik, A. Inflammation, Vitamin B6 and Related Pathways. Mol Aspects Med 2017; 53:10-27 [doi: https://doi.org/10.1016/j.mam.2016.08.001]
60. McNulty, H, Ward, M, Hoey, L, Hughes, C, Pentieva, K. Addressing Optimal Folate and Related B-vitamin Status Through the Lifecycle: Health Impacts and Challenges. Proc Nutr Soc 2019; 78)3):449-62 [doi: https://doi.org/10.1017/S0029665119000661]
61. Riggs, K, Spiro, A, Tucker, K, Rush, D. Relations of vitamin B-12, vitamin B-6, folate, and homocysteine to cognitive performance in the normative aging

study. American Journal of Clinical Nutrition 1996; 63(3):306-14 [doi: https://doi.org/10.1093/ajcn/63.3.306]
62. Kim, H, Kim, G, Jang, W, Kim, S, Chang, N. Association between intake of B vitamins and cognitive function in elderly Koreans with cognitive impairment. Nutrition Journal 2014; 13(1):118 [doi: https://doi.org/10.1186/1475-2891-13-118]
63. Hughes, C, Ward, M, Tracey, F, Hoey, L, Molloy, A, Pentieva, K, McNulty, H. B-Vitamin Intake and Biomarker Status in Relation to Cognitive Decline in Healthy Older Adults in a 4-Year Follow-Up Study. Nutrients. 2017; 9(1):53 [doi: https://doi.org/10.3390/nu9010053]
64. Tucker, K, Qiao, N, Scott, T, Rosenberg, I, Spiro, A. High homocysteine and low B vitamins predict cognitive decline in aging men: the Veterans Affairs Normative Aging Study. American Journal of Clinical Nutrition 2005; 82(3):627-635 [doi: https://doi.org/10.1093/ajcn.82.3.627]
65. de Jager, C, Oulhaj, A, Jacoby, R, Refsum, H, Smith, A. Cognitive and clinical outcomes of homocysteine-lowering B-vitamin treatment in mild cognitive impairment: a randomized controlled trial. International Journal of Geriatric Psychiatry 2012; 27(6):592-600 [doi: https://doi.org/10.1002/gps.2758]
66. Cheng, D, Kong, H, Pang, W, Yang, H, Lu, H, Huang, C, Jiang, Y. B Vitamin Supplementation Improves Cognitive Function in the Middle Aged and Elderly With Hyperhomocysteinemia. Nutr Neorosci 2016; 19(10):461-466 [doi: https://doi.org/10.1179/1476830514Y.0000000136]
67. Smith, A, Smith, S, de Jager, C, Whitbread, P, Johnston, C, Agacinski, G, Oulhaj, A, Bradley, K, Jacoby, R, Refsum, H. Homocysteine-lowering by b vitamins slows the rate of accelerated brain atrophy in mild cognitive impairment: a randomized controlled trial. PLoS ONE 2010; 5(9):e12244 [doi: https://doi.org/10.1371/journal.pone.0012244]
68. Pellegrini, E, Ballerini, L, Valdes Hernandez, M, Chappell, F, Gonzalez-Castro, V, Anblagan, D, Danso, S, Munoz-Maniega, S, Job, D, Pernet, C, Mair, G, MacGillivray, T, Trucco, E, Wardlaw, J. Machine learning for neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. Alzheimers Dement (Amst). 2018; 10:519-535 [doi: https://doi.org/10.1016/j.dadm.2018.07.004]