A Study on Psychometric Assessment Data for Autonomous Dementia Detection

Chloe M. Barnes
c.barnes1@aston.ac.uk
Aston University
Birmingham, West Midlands, UK

ABSTRACT

Dementia and Alzheimer's disease are characterised by cognitive decline, and diagnoses are predicted to rise due to an ageing population. Psychometric assessments are widely used by clinicians to inform the diagnosis of dementia, however these may not be as accurate or accessible for patients with lower levels of literacy or socioeconomic status. This study explores how machine learning models can detect dementia when trained on combinations of attributes from multi-modal datasets containing psychometric and Magnetic Resonance Imaging (MRI) data. When psychometric testing is not available, results show that the Random Forest classifier achieves a balanced accuracy, sensitivity and specificity of 84.75%, 79.10%, and 90.41% respectively before the dataset was standardised, and 84.34%, 78.27%, and 90.41% after – outperforming identical models trained on data from a single psychometric test.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning approaches**; **Supervised learning by classification**.

KEYWORDS

Dementia Classification, Machine Learning, Multi-Modal Data

ACM Reference Format:
Chloe M. Barnes. 2022. A Study on Psychometric Assessment Data for Autonomous Dementia Detection. In *The 15th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '22), June 29-July 1, 2022, Corfu, Greece.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3529190.3534726

1 INTRODUCTION

Dementia is characterised by degeneration of cognitive ability that impacts a person's independence and daily life, with Alzheimer's disease, a neurodegenerative disorder, being the most common cause [20, 23]. Recent projections indicate that the number of diagnosed cases of dementia across the world will triple by 2050 [20]; tools that facilitate the detection of early signs of dementia and cognitive decline may therefore become increasingly important to prevent and diagnose dementia as the ageing population grows [21].

PETRA '22, June 29-July 1, 2022, Corfu, Greece

ACM ISBN 978-1-4503-9631-8/22/06.
https://doi.org/10.1145/3529190.3534726

While many psychometric and interview-based instruments exist to assess cognitive ability and the presence of dementia – such as the General Practitioner Assessment of Cognition (GPCOG), Mini-Cog, and Brief Cognitive Rating Scale (BCRS) – the Clinical Dementia Rating (CDR) [13] and Mini-Mental State Examination (MMSE) [6] are two of the most commonly used [15, 19]. However, these measures each have notable limitations, including: potentially poor test-retest reliability [22] due to the subjective nature of these interview-based assessments; inaccurate results when used in certain cultures due to the abstract questions in the assessments [10]; the potential for the MMSE to be negatively influenced or require adjustment when the patient is from a low socioeconomic background, or has a low level of education or literacy [9, 19], or even if the patient has physical disabilities such as arthritis or paralysis [16]; and CDR being an inaccessible method of assessment, if not adjusted, for those living in community-living situations that do not have an informant that can be present during the assessment [15]. In scenarios where these otherwise reliable tools for assessing cognitive ability are inaccessible or potentially inaccurate, a dementia diagnosis may therefore rely more heavily on data obtained by non-psychometric methods – else, the risk of a false negative diagnosis may be greater.

The study presented in this paper is therefore designed to explore how machine learning methods may be utilised to detect dementia when psychometric assessments may or may not be possible or accurate. This could be another tool for use by clinicians to inform their assessments and diagnoses of dementia, which could lead to more accessible detection of dementia for those that may otherwise receive inaccurate psychometric results – for example, in communities with a low socioeconomic status [19], or those that live alone or in community-living scenarios [15]. A variety of machine learning models are trained in this study using: (1) data obtained purely from psychometric assessments (CDR and MMSE); (2) data obtained purely from MRI scans and clinical data (age, sex, etc.); and (3) a combination of both psychometric assessment, MRI scan and clinical data. A further contribution of this study is an exploration into the role of standardisation of the data when training machine learning models to detect dementia, since the attributes can vary in scale by orders of magnitude.

2 BACKGROUND AND RELATED WORK

As the population ages, methods to detect cognitive decline and dementia will be increasingly important for prevention and treatment, to reduce the load on healthcare services [21]. Machine learning techniques have been widely explored in elderly care, for example: the monitoring of stroke patients [11]; fall detection [25]; activity recognition by assistive robots [1]; dementia classification in

community-living arrangements [24]; or predicting the future decline in cognitive ability in the elderly [14]. Dementia has also been detected in EEG signals using machine learning [7]; transfer learning between different biological signals such as EEG and EMG is also possible [4], which could be another way to assist dementia prediction, monitoring, or diagnosis in the future. One of the limitations of the CDR psychometric assessment is that it can be inaccessible or inaccurate if not adjusted for patients without an informant or those in community-living scenarios [15]; machine learning thus has potential to also help inform clinicians in the future when diagnosing dementia, for those living in shared or solitary environments without an informant.

To explore whether machine learning can be used to detect dementia with or without the presence of psychometric assessment data such as that which is obtained from the CDR test, this study uses the OASIS-2 dataset [12], consisting of longitudinal MRI data collected from 150 adults aged between 60 and 98, with and without dementia over the course of 373 total visits. An explanation of the attributes contained within the dataset is presented in Table 1. Briefly, the dataset consists of 15 attributes that: identify the subject, MRI session, visit number, and days since the previous visit; describe the subject in terms of sex, dominant hand, age, years of education (EDUC), and socioeconomic status (SES); provide psychometric assessment data (CDR and MMSE); are obtained from MRI or clinical assessments (eTIV, nWBV, ASF); and state the subject's dementia classification (Group). The subjects are classified as having dementia, not having dementia, or 'converted', meaning that the subject received a dementia diagnosis during the study.

Many experimental studies have trained machine learning models to analyse and detect dementia using this dataset. Battineni et al. [3] trained a Support Vector Machine (SVM) model to detect dementia in a multi-class classification problem (i.e. predict each of the three classes in the Group attribute); the model was trained on attributes such as Subject ID, MR Delay, Age, MMSE, CDR and nWBV achieves an accuracy and precision of 68.75% and 64.18% respectively. A limitation of this study is that some attributes are removed during the preprocessing stage which could potentially affect accuracy positively, while other attributes used as identifiers (like Subject ID) may negatively affect accuracy. In contrast, a more in-depth study compares how different models perform when viewing this as a binary classification problem: the subject has dementia or not. Khan and Zubair [8] explore the effect of removing or imputing missing values, showing that models trained only on data recorded from the first MRI visits of subjects, such as an SVM, Naive Bayes, and Random Forest achieve an accuracy of 69%, 71%, and 75.3% respectively when missing values are removed; the accuracy score however drops for these models with imputation of missing values. Other work proposes a deep neural network model to predict dementia in a binary classification problem [2] using all attributes except from the identifiers (Subject and MRI IDs) and the dominant hand (since all subjects are right-handed); the study explores how removing combinations of the age, gender, and socioeconomic status of a subject affects dementia detection, with age being found to affect the accuracy of the model more than the other two. A limitation here is that the classification problem is reduced from multi-class to binary by combining all samples with a subject classed as 'converted' into the 'dementia' class, which could negatively affect accuracy since some of these samples would be from subjects yet to receive a dementia diagnosis.

Table 1: Descriptions of the attributes contained in the OASIS-2 Longitudinal Dataset [12].

Attribute	Description
Subject ID	Identifier of the subject
MRI ID	Identifier of the MRI session
Group	Diagnosis of dementia, no dementia, or 'converted' (subject diagnosed with dementia during the project)
Visit	Identifier of the subject's repeat visits, each subject visited at least two and at most five times (1-5)
MR Delay	Days between the subject's visit and the previous visit; visits are spaced by at least a year
M/F	Male or Female (males = 62, females = 88)
Hand	Dominant hand used by the subject (all subjects in the study were right-handed)
Age	Age of the subject at the time of the visit (min = 60, max = 98)
EDUC	Number of years of education (min = 6, max = 23)
SES	Socioeconomic Status according to the Hollingshead Index of Social Position (lowest = 5, highest = 1)
MMSE	Mini-Mental State Examination score (min = 0, max = 30)
CDR	Clinical Dementia Rating (0 = no dementia, 0.5 = questionable, 1 = mild, 2 = moderate, 3 = severe [13])
eTIV	Estimated total intracranial volume (cm^3)
nWBV	Normalised whole-brain volume
ASF	Atlas Scaling Factor

To complement previous analyses of this dataset, the study presented in this paper explores how data obtained from psychometric assessments affect the accuracy of various machine learning models when detecting dementia. The aims of this study are therefore threefold: to train machine learning models to detect dementia when presented with a binary classification problem (predict the Group target attribute (Table 1), which labels a subject as having dementia or not); to investigate the effect that data obtained from non-psychometric and psychometric assessment, as well as a combination of the two, can affect dementia classification accuracy, in the case that psychometric testing is not available; and to explore whether standardisation of the data affects classification accuracy.

3 EXPERIMENTAL DESIGN

3.1 Data Preprocessing

The OASIS-2 dataset [12] presents an imbalanced, multi-class classification problem, with 190 samples classifying the subject as not having dementia, 146 as having dementia, and 37 as 'converted' – meaning the subject was not diagnosed with dementia at their first visit, but had a changed diagnosis during a subsequent visit. Since the 'converted' class applies to all samples of data taken from a subject who received a change in dementia diagnosis, the choice was made to remove this class for this study. Reducing the problem to a binary classification problem is intended to not only improve classification accuracy since the multi-class problem was imbalanced, but to provide a benchmark for future work that could then attempt to classify the 'converted' samples as well.

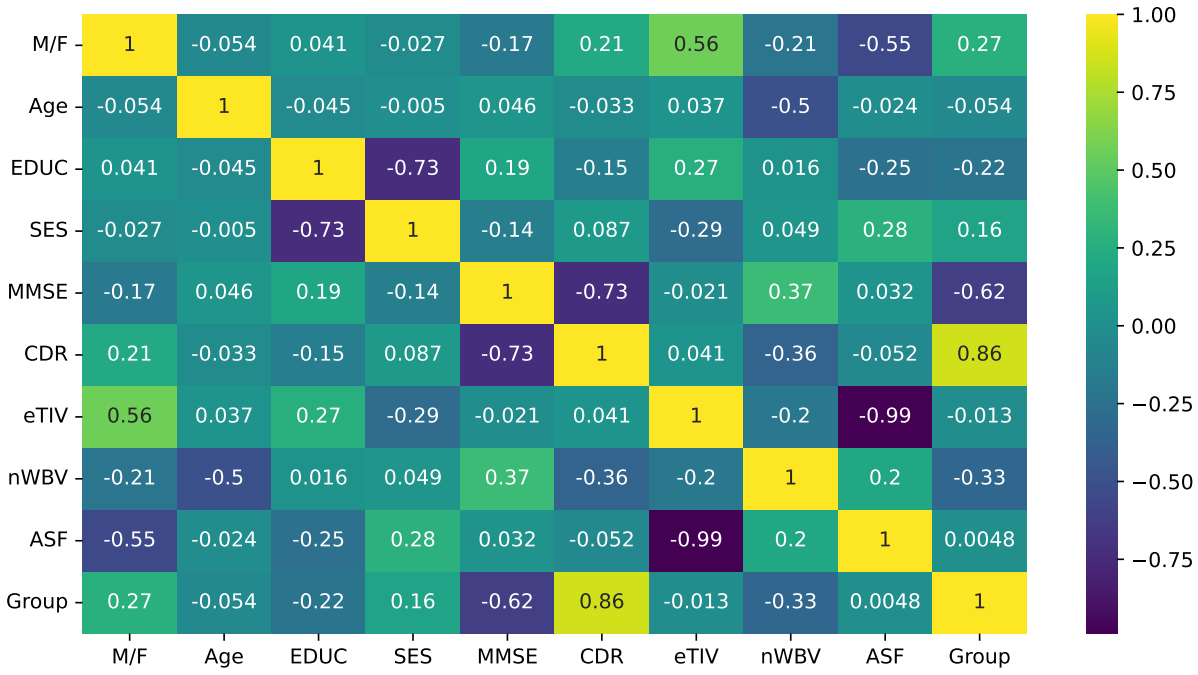
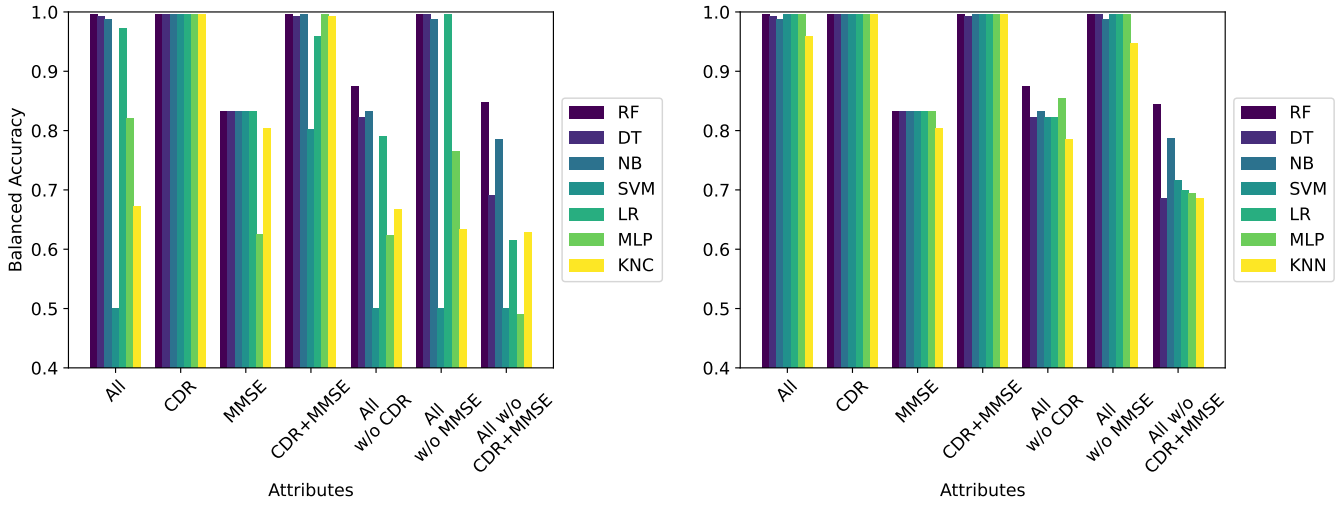


Figure 1: Attribute Correlation Matrix

The attributes selected to train models to predict whether a subject has dementia (Group) are: M/F, Age, EDUC, SES, MMSE, CDR, eTIV, nWBV, and ASF. Subject ID, MRI ID, Visit, and MR Delay are not used to train the models since these attributes identify the subject or MRI visit, and are not measures that can be used in cognitive assessment. The Hand attribute is also removed, as each subject is identified as right-handed. The correlation matrix shown in Figure 1 highlights the correlations between these nine attributes, based on Pearson's Correlation Coefficient (r). Most notably, eTIV and ASF have a strong negative correlation, as ASF is a scaling factor for the intracranial volume [5]. Men also tend to have a greater eTIV than women [5], shown by a correlation of $r = 0.56$ in Figure 1. There is also a negative correlation between the number of years of education a subject has (EDUC) and their socioeconomic status (SES), since more educated subjects are likely to have a lower SES score (lower score means higher socioeconomic status). CDR and MMSE are also negatively correlated, as a high CDR score and low MMSE score are indicative of dementia. The correlation matrix also highlights that the two most important attributes in this dataset for predicting the target (Group) are CDR ($r = 0.86$) and MMSE ($r = -0.62$); a negative correlation with MMSE indicates that a greater MMSE score correlates to subjects not having dementia, whereas for CDR, greater scores indicate dementia.

As MMSE has a notable correlation with determining whether a subject has dementia, imputing missing values for this attribute could potentially result in incorrect predictions being made. Thus, samples with missing values (in SES or MMSE) were instead removed, leaving 317 samples out of 336. Future work could explore the effect of different methods of imputation on classification using this dataset in more detail, as Khan and Zubair [8] found that accuracy could be positively or negatively affected by imputation depending on the classifier. After preprocessing, the dataset used in the rest of the study contains 190 samples where subjects do not have dementia, and 127 with dementia. Of these 317 samples, 180 were from female subjects and 137 from male subjects.

One major contribution of this study is an exploration of how the selected attributes affect dementia detection when using machine learning methods. The CDR and MMSE scores are the result of psychometric clinical assessments, whereas the other selected attributes are measured without interview. CDR > 0 indicates that a subject shows signs of dementia (only two subjects *without* dementia have CDR > 0 in the dataset), and MMSE scores closer to 0 are indicative of more severe dementia than those closer to 30. Both of these psychometric assessments are highly informative for the diagnosis of dementia, which can be seen in the correlation matrix presented in Figure 1. However, these measures can potentially be affected by a subject's education level or culture, or be less effective when an informant is not available (such as in community-living scenarios) [10, 15]. This study is therefore designed to ascertain the extent to which dementia can be successfully detected without CDR or MMSE scores if unavailable or unobtainable, with a set of seven experiments using the following sets of attributes: (1) All nine attributes identified above (M/F, Age, EDUC, SES, MMSE, CDR, eTIV, nWBV, and ASF); (2) CDR only; (3) MMSE only; (4) CDR and MMSE only; (5) All attributes except CDR; (6) All attributes except MMSE; and (7) All attributes except CDR and MMSE. By training models using each of these combinations of attributes, the effect of CDR and MMSE measures on the classification of dementia, as well as the viability of detecting dementia when either of these measures are unavailable can be observed.



(a) Balanced Accuracy with Non-Standardised Data

(b) Balanced Accuracy with Standardised Data

Figure 2: Mean Balanced Accuracy of models trained using different selections of attributes, before and after standardisation. Tables 2 and 3 present performance scores (including Balanced Accuracy) for each selection of attributes.

Finally, `skikit learn`'s [17] `StandardScaler` was used to standardise the dataset as the values for each attribute varied in terms of scale (e.g. eTIV values range from 1106 to 2004, and nWBV from 0.644 to 0.837). The experiments in this study were conducted using an Intel Core i7-10700K CPU (3.80GHz), used `scikit-learn` [17] to train and analyse the models, and were repeated with non-standardised and standardised data for comparison.

3.2 Training Machine Learning Models

Seven machine learning models were chosen to explore their effectiveness in detecting dementia in non-standardised and standardised data: (1) Random Forest (RF); (2) Decision Tree (DT); (3) Naive Bayes (NB); (4) Support Vector Machine (SVM) with Radial Basis Function kernel (RBF); (5) Logistic Regression (LR); (6) Multi-Layer Perceptron (MLP); and (7) K-Nearest Neighbours (KNN). Each model was trained using 10-fold cross-validation with the same seed for comparison. Accuracy, precision and the F1 score were recorded for each experiment; additionally, since the dataset is imbalanced (non-dementia = 190, dementia = 127), the balanced accuracy, sensitivity (also known as 'recall') and specificity were also recorded. Sensitivity and specificity measure the probability of 'true positive' and 'true negative' values respectively.

4 RESULTS

The results for the experimental study outlined above are presented in this section; Figure 2 presents a visualisation of the balanced accuracy for each of the experiments, when each model is trained using non-standardised (Figure 2a) and standardised data (Figure 2b), for each set of attributes. A detailed breakdown of the mean performance metrics obtained in each experiment across each fold in 10-fold cross-validation are recorded in Tables 2 and 3 respectively.

The first notable observation is that models trained using standardised data perform as well, if not better, than those trained using non-standardised data. The two exceptions to this are the Random Forest and Naive Bayes models, which perform identically when trained with non-standardised and standardised data, but classify one fewer true positive and true negative respectively when the dataset contains all attributes except CDR and MMSE. This indicates that these two classifiers are generally robust against the effects of standardisation of the data, but removing the psychometric attributes CDR and MMSE from the dataset can slightly degrade performance of the model when the dataset is standardised. Saying this, the Decision Tree model also performs identically regardless of standardisation, however excluding the aforementioned psychometric attributes CDR and MMSE from the dataset when standardised slightly improves the performance. In contrast, the accuracy of the Logistic Regression, Multi-Layer Perceptron, and K-Nearest Neighbours models is equivalent or improves when trained on standardised data compared to non-standardised data.

Despite SVMs being a popular choice of machine learning model for detecting dementia [18], standardisation of the dataset appears to impact the SVM model more than the others explored in this study. When training the model with non-standardised data, only the CDR and MMSE attributes, or a combination of the two demonstrated an increase in accuracy above classifying all data as the most common target class (non-dementia). However, with standardised data, the model's performance in terms of accuracy is generally on par with the others.

With standardised data, all models perform similarly when the dataset contains at least one of the psychometric assessment rating attributes (CDR and MMSE). However, when training the models on only non-psychometric assessment rating attributes (i.e. all except CDR and MMSE), the Random Forest model achieves the highest score in each recorded metric, with a balanced accuracy and accuracy of 84.75% and 85.46% respectively for non-standardised data, and 84.34% and 85.33% for standardised data.

Table 2: Mean performance scores across each fold in 10-fold cross-validation for experiments with non-standardised data, with the standard deviation in (brackets). Bold indicates rows with the greatest Balanced Accuracy (Bal. Accuracy) and Accuracy.

Classifier	Attributes	Bal. Accuracy	Accuracy	Precision	F1	Sensitivity	Specificity
Random Forest	All	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR	0.8745 (0.084)	0.8899 (0.069)	0.9032 (0.104)	0.8458 (0.115)	0.8005 (0.139)	0.9485 (0.043)
	All w/o MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR + MMSE	0.8475 (0.060)	0.8582 (0.053)	0.8546 (0.079)	0.8149 (0.069)	0.7910 (0.117)	0.9041 (0.053)
Decision Tree	All	0.9880 (0.015)	0.9874 (0.015)	0.9840 (0.032)	0.9842 (0.019)	0.9857 (0.029)	0.9902 (0.020)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR	0.8324 (0.061)	0.8456 (0.050)	0.8219 (0.065)	0.7960 (0.082)	0.7766 (0.107)	0.8883 (0.049)
	All w/o MMSE	0.9880 (0.015)	0.9874 (0.015)	0.9840 (0.032)	0.9842 (0.019)	0.9857 (0.029)	0.9902 (0.020)
	All w/o CDR + MMSE	0.7846 (0.100)	0.7884 (0.093)	0.7374 (0.077)	0.7370 (0.108)	0.7464 (0.158)	0.8227 (0.064)
Naive Bayes	All	0.9928 (0.011)	0.9906 (0.014)	0.9749 (0.038)	0.9869 (0.020)	1.0000 (0.000)	0.9857 (0.022)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	0.9928 (0.011)	0.9906 (0.014)	0.9749 (0.038)	0.9869 (0.020)	1.0000 (0.000)	0.9857 (0.022)
	All w/o CDR	0.8227 (0.069)	0.8458 (0.056)	0.8829 (0.135)	0.7771 (0.103)	0.7048 (0.118)	0.9407 (0.058)
	All w/o MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR + MMSE	0.6904 (0.111)	0.7167 (0.083)	0.6478 (0.201)	0.5974 (0.190)	0.5726 (0.209)	0.8083 (0.100)
Support Vector Machine (RBF)	All	0.5000 (0.000)	0.5992 (0.054)	0.0000 (0.000)	0.0000 (0.000)	0.0000 (0.000)	1.0000 (0.000)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	0.8023 (0.077)	0.8427 (0.060)	0.9800 (0.060)	0.7426 (0.123)	0.6092 (0.148)	0.9955 (0.014)
	All w/o CDR	0.5000 (0.000)	0.5992 (0.054)	0.0000 (0.000)	0.0000 (0.000)	0.0000 (0.000)	1.0000 (0.000)
	All w/o MMSE	0.5000 (0.000)	0.5992 (0.054)	0.0000 (0.000)	0.0000 (0.000)	0.0000 (0.000)	1.0000 (0.000)
	All w/o CDR + MMSE	0.5000 (0.000)	0.5992 (0.054)	0.0000 (0.000)	0.0000 (0.000)	0.0000 (0.000)	1.0000 (0.000)
Logistic Regression	All	0.9719 (0.037)	0.9781 (0.028)	0.9826 (0.035)	0.9667 (0.043)	0.9535 (0.065)	0.9902 (0.020)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	0.9592 (0.030)	0.9654 (0.026)	0.9826 (0.035)	0.9536 (0.035)	0.9282 (0.054)	0.9902 (0.020)
	All w/o CDR	0.7906 (0.080)	0.8204 (0.057)	0.8475 (0.141)	0.7314 (0.121)	0.6574 (0.150)	0.9238 (0.065)
	All w/o MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR + MMSE	0.6150 (0.082)	0.6564 (0.054)	0.6152 (0.220)	0.4694 (0.148)	0.4021 (0.155)	0.8278 (0.091)
Multi-Layer Perceptron	All	0.8207 (0.104)	0.8392 (0.079)	0.8615 (0.139)	0.7585 (0.166)	0.7451 (0.251)	0.8964 (0.125)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.6251 (0.067)	0.7003 (0.066)	1.0000 (0.000)	0.3821 (0.169)	0.2502 (0.135)	1.0000 (0.000)
	Only CDR + MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR	0.6235 (0.072)	0.6845 (0.067)	0.8138 (0.236)	0.4223 (0.142)	0.3007 (0.120)	0.9463 (0.067)
	All w/o MMSE	0.7647 (0.155)	0.7821 (0.135)	0.7855 (0.213)	0.6766 (0.242)	0.6804 (0.312)	0.8490 (0.167)
	All w/o CDR + MMSE	0.4907 (0.041)	0.5164 (0.122)	0.2488 (0.307)	0.1966 (0.218)	0.2895 (0.377)	0.6919 (0.417)
K-Nearest Neighbours	All	0.6716 (0.088)	0.7003 (0.082)	0.6842 (0.196)	0.5743 (0.122)	0.5108 (0.109)	0.8323 (0.105)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8045 (0.079)	0.8427 (0.060)	0.9646 (0.072)	0.7454 (0.125)	0.6246 (0.163)	0.9843 (0.035)
	Only CDR + MMSE	0.9928 (0.011)	0.9906 (0.014)	0.9749 (0.038)	0.9869 (0.020)	1.0000 (0.000)	0.9857 (0.022)
	All w/o CDR	0.6680 (0.083)	0.6972 (0.079)	0.6827 (0.195)	0.5694 (0.115)	0.5037 (0.097)	0.8323 (0.105)
	All w/o MMSE	0.6335 (0.097)	0.6532 (0.094)	0.5882 (0.161)	0.5436 (0.125)	0.5199 (0.121)	0.7471 (0.119)
	All w/o CDR + MMSE	0.6277 (0.095)	0.6470 (0.095)	0.5832 (0.162)	0.5366 (0.122)	0.5128 (0.116)	0.7426 (0.126)

5 DISCUSSION AND IMPLICATIONS

Generally speaking, the balanced accuracy of the trained models in this study is positively affected if the model is trained on standardised data – as can be seen in Figure 2. The trained models with the best performance have balanced accuracy and accuracy scores of 99.51% and 99.38% respectively (Tables 2 and 3); a sensitivity score of 100% for these best-performing models means that all true positives are accurately classified, but a specificity of 99.02% indicates that two true negatives are misclassified as positives. Looking at the raw data, these two misclassified samples of non-dementia subjects have CDR = 0.5, which otherwise indicates the subject has received a dementia diagnosis (usually CDR > 0 is an indicator of dementia). These best-performing scores are only observed to be achievable if the model is trained on a dataset including the CDR attribute (this is not to say it is *always* observed to be achieved).

The results show that models trained using at least one of the two psychometric attributes (CDR and MMSE) outperform models that are trained without psychometric data, with CDR being the most reliable indicator for dementia in this study; this shows that, if available, psychometric testing still provides valuable information for the detection of dementia, regardless of the limitations. The exception to this is the Random Forest model, which, when trained on all non-psychometric attributes, achieves a greater balanced accuracy and accuracy than when trained on just the MMSE attribute. Without the non-psychometric assessment rating attributes (i.e. all

Table 3: Mean performance scores across each fold in 10-fold cross-validation for experiments with standardised data, with the standard deviation in (brackets). Bold indicates rows with the greatest Balanced Accuracy (Bal. Accuracy) and Accuracy.

Classifier	Attributes	Bal. Accuracy	Accuracy	Precision	F1	Sensitivity	Specificity
Random Forest	All	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR	0.8745 (0.084)	0.8899 (0.069)	0.9032 (0.104)	0.8458 (0.115)	0.8005 (0.139)	0.9485 (0.043)
	All w/o MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR + MMSE	0.8434 (0.057)	0.8550 (0.051)	0.8533 (0.080)	0.8102 (0.066)	0.7827 (0.111)	0.9041 (0.053)
Decision Tree	All	0.9880 (0.015)	0.9874 (0.015)	0.9840 (0.032)	0.9842 (0.019)	0.9857 (0.029)	0.9902 (0.020)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR	0.8324 (0.061)	0.8456 (0.050)	0.8219 (0.065)	0.7960 (0.082)	0.7766 (0.107)	0.8883 (0.049)
	All w/o MMSE	0.9880 (0.015)	0.9874 (0.015)	0.9840 (0.032)	0.9842 (0.019)	0.9857 (0.029)	0.9902 (0.020)
	All w/o CDR + MMSE	0.7870 (0.100)	0.7915 (0.094)	0.7451 (0.077)	0.7404 (0.106)	0.7464 (0.158)	0.8275 (0.068)
Naive Bayes	All	0.9928 (0.011)	0.9906 (0.014)	0.9749 (0.038)	0.9869 (0.020)	1.0000 (0.000)	0.9857 (0.022)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	0.9928 (0.011)	0.9906 (0.014)	0.9749 (0.038)	0.9869 (0.020)	1.0000 (0.000)	0.9857 (0.022)
	All w/o CDR	0.8227 (0.069)	0.8458 (0.056)	0.8829 (0.135)	0.7771 (0.103)	0.7048 (0.118)	0.9407 (0.058)
	All w/o MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR + MMSE	0.6862 (0.111)	0.7136 (0.082)	0.6445 (0.201)	0.5909 (0.190)	0.5642 (0.211)	0.8083 (0.100)
Support Vector Machine (RBF)	All	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR	0.8223 (0.105)	0.8492 (0.084)	0.8891 (0.154)	0.7724 (0.151)	0.6937 (0.172)	0.9509 (0.059)
	All w/o MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR + MMSE	0.7160 (0.118)	0.7420 (0.102)	0.7051 (0.184)	0.6327 (0.175)	0.5880 (0.192)	0.8441 (0.094)
Logistic Regression	All	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR	0.8230 (0.086)	0.8426 (0.073)	0.8514 (0.144)	0.7783 (0.121)	0.7265 (0.135)	0.9195 (0.072)
	All w/o MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR + MMSE	0.6991 (0.127)	0.7260 (0.098)	0.6671 (0.224)	0.6037 (0.205)	0.5701 (0.219)	0.8280 (0.091)
Multi-Layer Perceptron	All	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8318 (0.044)	0.8551 (0.034)	0.9051 (0.088)	0.7916 (0.067)	0.7148 (0.103)	0.9489 (0.046)
	Only CDR + MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR	0.8547 (0.078)	0.8712 (0.068)	0.8785 (0.130)	0.8200 (0.109)	0.7737 (0.108)	0.9357 (0.064)
	All w/o MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR + MMSE	0.6950 (0.091)	0.7165 (0.078)	0.6618 (0.147)	0.6117 (0.135)	0.5807 (0.150)	0.8093 (0.076)
K-Nearest Neighbours	All	0.9584 (0.038)	0.9654 (0.033)	1.0000 (0.000)	0.9549 (0.041)	0.9168 (0.076)	1.0000 (0.000)
	Only CDR	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	Only MMSE	0.8045 (0.079)	0.8427 (0.060)	0.9646 (0.072)	0.7454 (0.125)	0.6246 (0.163)	0.9843 (0.035)
	Only CDR + MMSE	**0.9951** (0.010)	**0.9938** (0.013)	**0.9840** (0.032)	**0.9917** (0.017)	**1.0000** (0.000)	**0.9902** (0.020)
	All w/o CDR	0.7859 (0.106)	0.8209 (0.077)	0.8650 (0.094)	0.7143 (0.176)	0.6300 (0.208)	0.9419 (0.045)
	All w/o MMSE	0.9471 (0.045)	0.9557 (0.041)	1.0000 (0.000)	0.9416 (0.052)	0.8942 (0.091)	1.0000 (0.000)
	All w/o CDR + MMSE	0.6854 (0.106)	0.7166 (0.107)	0.7085 (0.135)	0.5902 (0.142)	0.5266 (0.177)	0.8441 (0.120)

except CDR and MMSE), the two best-performing models regardless of whether the dataset is standardised or not are tree-based: the Random Forest and the Decision Tree. Therefore, in this study, the Random Forest classifier is found to be the most effective for classifying whether subjects have received a dementia diagnosis both when psychometric assessment data is available and unavailable – even when the dataset may or may not be standardised.

6 CONCLUSION AND FUTURE WORK

The presented experimental study explores whether dementia can be detected in longitudinal data using various machine learning methods – without relying on psychometric assessment rating data; this type of assessment may be inaccurate for unavailable for those with lower socioeconomic backgrounds or levels of education [9, 19], those in community-living situations [15], or those with physical disabilities [16]. Further, the effect of the standardisation of the dataset on classification accuracy is also investigated, by repeating the experiments before and after standardisation.

By training machine learning models using the OASIS-2 dataset [12], the Random Forest classifier was found to detect dementia with a higher balanced accuracy and accuracy when psychometric data was unavailable compared to the other models included in the study, achieving 84.75% and 85.46% respectively pre-standardisation, and 84.34% and 85.33% post-standardisation. The results indicate that dementia can be detected with greater accuracy if the dataset contains some form of psychometric assessment data (in this case,

CDR or MMSE, with CDR being the most reliable indicator). However, if this is not possible or accessible, a Random Forest trained without data from psychometric assessments, and only MRI and clinical data, can be a viable alternative.

A limitation of this study is that samples with missing data were removed during preprocessing, with all belonging to the minority class. This imbalance could potentially affect classification accuracy and model performance, therefore future work will explore the effect of imputing missing values in the dataset instead of removing them, as well as methods to balance the dataset. Parameter tuning for the best models will also be conducted as future work to improve the results. Another avenue of future work could investigate the study as a multi-class classification problem instead of binary; the third, 'converted' class was removed from the dataset during preprocessing due to the dataset being heavily imbalanced, but again, methods to balance the dataset could be explored in the future.

ACKNOWLEDGMENTS

The dataset used in this study (Open Access Series of Imaging Studies (OASIS) [12]): is publicly available at http://www.oasis-brains.org; has Principal Investigators D. Marcus, R, Buckner, J. Csernansky, and J. Morris; and is supported by grants P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, and U24 RR021382.

REFERENCES

[1] David Ada Adama, Ahmad Lotfi, Caroline Langensiepen, Kevin Lee, and Pedro Trindade. 2017. Learning Human Activities for Assisted Living Robotics. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '17)*. Association for Computing Machinery, New York, NY, USA, 286–292. https://doi.org/10.1145/3056540.3076197

[2] Shakila Basheer, Surbhi Bhatia, and Sapiah Binti Sakri. 2021. Computational Modeling of Dementia Prediction using Deep Neural Network: Analysis on OASIS Dataset. *IEEE Access* 9 (2021), 42449–42462. https://doi.org/10.1109/ACCESS.2021.3066213

[3] Gopi Battineni, Nalini Chintalapudi, and Francesco Amenta. 2019. Machine Learning in Medicine: Performance Calculation of Dementia Prediction by Support Vector Machines (SVM). *Informatics in Medicine Unlocked* 16 (2019), 100200. https://doi.org/10.1016/j.imu.2019.100200

[4] Jordan J Bird, Jhonatan Kobylarz, Diego R Faria, Anikó Ekárt, and Eduardo P Ribeiro. 2020. Cross-Domain MLP and CNN Transfer Learning for Biological Signal Processing: EEG and EMG. *IEEE Access* 8 (2020), 54789–54801. https://doi.org/10.1109/ACCESS.2020.2979074

[5] Randy L Buckner, Denise Head, Jamie Parker, Anthony F Fotenos, Daniel Marcus, John C Morris, and Abraham Z Snyder. 2004. A Unified Approach for Morphometric and Functional Data Analysis in Young, Old, and Demented Adults using Automated Atlas-Based Head Size Normalization: Reliability and Validation Against Manual Measurement of Total Intracranial Volume. *Neuroimage* 23, 2 (2004), 724–738. https://doi.org/10.1016/j.neuroimage.2004.06.018

[6] Marshal F Folstein, Susan E Folstein, and Paul R McHugh. 1975. "Mini-Mental State": A Practical Method for Grading the Cognitive State of Patients for the Clinician. *Journal of Psychiatric Research* 12, 3 (1975), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

[7] Cosimo Ieracitano, Nadia Mammone, Amir Hussain, and Francesco C Morabito. 2020. A Novel Multi-Modal Machine Learning Based Approach for Automatic Classification of EEG Recordings in Dementia. *Neural Networks* 123 (2020), 176–190. https://doi.org/10.1016/j.neunet.2019.12.006

[8] Afreen Khan and Swaleha Zubair. 2020. An Improved Multi-Modal Based Machine Learning Approach for the Prognosis of Alzheimer's Disease. (2020). https://doi.org/10.1016/j.jksuci.2020.04.004

[9] Renata Kochhann, Juliana Santos Varela, Carolina Saraiva de Macedo Lisboa, and Márcia Lorena Fagundes Chaves. 2010. The Mini Mental State Examination: Review of Cutoff Points Adjusted for Schooling in a Large Southern Brazilian Sample. *Dementia & Neuropsychologia* 4 (2010), 35–41. https://doi.org/10.1590%2FS1980-57642010DN40100006

[10] Wee Shiong Lim, Mei Sian Chong, and Suresh Sahadevan. 2007. Utility of the Clinical Dementia Rating in Asian Populations. *Clinical Medicine & Research* 5, 1 (2007), 61–70. https://doi.org/10.3121/cmr.2007.693

[11] Andrea Mannini, Diana Trojaniello, Andrea Cereatti, and Angelo M Sabatini. 2016. A Machine Learning Framework for Gait Classification using Inertial Sensors: Application to Elderly, Post-Stroke and Huntington's Disease Patients. *Sensors* 16, 1 (2016), 134. https://doi.org/10.3390/s16010134

[12] Daniel S Marcus, Anthony F Fotenos, John G Csernansky, John C Morris, and Randy L Buckner. 2010. Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults. *Journal of Cognitive Neuroscience* 22, 12 (12 2010), 2677–2684. https://doi.org/10.1162/jocn.2009.21407

[13] John C Morris. 1993. The Clinical Dementia Rating (CDR): Current Version and Scoring Rules. *Neurology* 43 (1993), 2412–2414. https://doi.org/10.1212/WNL.43.11.2412-a

[14] Kyoung-Sae Na. 2019. Prediction of Future Cognitive Impairment Among the Community Elderly: A Machine-Learning Based Approach. *Scientific Reports* 9, 1 (2019), 1–9. https://doi.org/10.1038/s41598-019-39478-7

[15] Ma Shwe Zin Nyunt, Mei Sian Chong, Wee Shiong Lim, Tih Shih Lee, Philip Yap, and Tze Pin Ng. 2013. Reliability and Validity of the Clinical Dementia Rating for Community-Living Elderly Subjects without an Informant. *Dementia and Geriatric Cognitive Disorders Extra* 3, 1 (2013), 407–416. https://doi.org/10.1159/000355122

[16] Verna C Pangman, Jeff Sloan, and Lorna Guse. 2000. An Examination of Psychometric Properties of the Mini-Mental State Examination and the Standardized Mini-Mental State Examination: Implications for Clinical Practice. *Applied Nursing Research* 13, 4 (2000), 209–213. https://doi.org/10.1053/apnr.2000.9231

[17] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[18] Enrico Pellegrini, Lucia Ballerini, Maria del C Valdes Hernandez, Francesca M Chappell, Victor González-Castro, Devasuda Anblagan, Samuel Danso, Susana Muñoz-Maniega, Dominic Job, Cyril Pernet, Grant Mair, Tom J MacGillivray, Emanuele Trucco, and Joanna M Wardlaw. 2018. Machine Learning of Neuroimaging for Assisted Diagnosis of Cognitive Impairment and Dementia: A Systematic Review. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 10 (2018), 519–535. https://doi.org/10.1016/j.dadm.2018.07.004

[19] Marcia Scazufca, Osvaldo P Almeida, Homero P Vallada, Wernestty A Tasse, and Paulo R Menezes. 2009. Limitations of the Mini-Mental State Examination for Screening Dementia in a Community with Low Socioeconomic Status. *European Archives of Psychiatry and Clinical Neuroscience* 259, 1 (2009), 8–15. https://doi.org/10.1007/s00406-008-0827-6

[20] Philip Scheltens, Bart De Strooper, Miia Kivipelto, Henne Holstege, Gael Chételat, Charlotte E Teunissen, Jeffrey Cummings, and Wiesje M van der Flier. 2021. Alzheimer's Disease. *The Lancet* 397, 10284 (2021), 1577–1590. https://doi.org/10.1016/S0140-6736(20)32205-4

[21] Jennifer A Smith, Minjung Kho, Wei Zhao, Miao Yu, Colter Mitchell, and Jessica D Faul. 2021. Genetic Effects and Gene-by-Education Interactions on Episodic Memory Performance and Decline in an Aging Population. *Social Science & Medicine* 271 (2021), 112039. https://doi.org/10.1016/j.socscimed.2018.11.019 Social and Behavioral Science at the Forefront of Genomics: Discovery, Translation, and Health Equity.

[22] Robert J Spencer, Carrington R Wendell, Paul P Giggey, Leslie I Katzel, David M Lefkowitz, Eliot L Siegel, and Shari R Waldstein. 2013. Psychometric Limitations of the Mini-Mental State Examination Among Nondemented Older Adults: An Evaluation of Neurocognitive and Magnetic Resonance Imaging Correlates. *Experimental Aging Research* 39, 4 (2013), 382–397. https://doi.org/10.1080/0361073X.2013.808109

[23] Sukriti Srivastava, Razi Ahmad, and Sunil Kumar Khare. 2021. Alzheimer's Disease and its Treatment by Different Approaches: A Review. *European Journal of Medicinal Chemistry* 216 (2021), 113320. https://doi.org/10.1016/j.ejmech.2021.113320

[24] Prabitha Urwyler, Reto Stucki, Luca Rampa, René Müri, Urs P Mosimann, and Tobias Nef. 2017. Cognitive Impairment Categorized in Community-Dwelling Older Adults with and without Dementia using In-Home Sensors that Recognise Activities of Daily Living. *Scientific Reports* 7, 1 (2017), 1–9. https://doi.org/10.1038/srep42084

[25] Diana Yacchirema, Jara Suárez de Puga, Carlos Palau, and Manuel Esteve. 2019. Fall Detection System for Elderly People using IoT and Ensemble Machine Learning Algorithm. *Personal and Ubiquitous Computing* 23, 5 (2019), 801–817. https://doi.org/10.1007/s00779-018-01196-8