Title: Neuropsychiatric symptoms as predictors of conversion from MCI to dementia: a machine learning approach

Running title: NPS as predictors of dementia: A ML approach.

Sabela C. Mallo, M.Sc. [1]; Sonia Valladares-Rodriguez, Ph.D [2]; David Facal, Ph.D [1];Cristina Lojo-Seoane, Ph.D ; Manuel J. Fernández-Iglesias, Ph.D [2]; Arturo X. Pereiro, Ph.D [1]

[1]Department of Developmental Psychology, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain.

[2]Department of Telematics Engineering, University of Vigo, Vigo, Spain.

Addresses:

Sabela C. Mallo, David Facal,, Arturo X. Pereiro, Onésimo Juncos-Rabadán: Development Psychology, University of Santiago de Compostela, Xosé María Suárez Núñez Street, Campus Sur, Santiago de Compostela, ES–15782, Spain.

Sonia Valladares-Rodriguez, Manuel J. Fernández-Iglesias: Department of Telematics Engineering, University of Vigo, Maxwell Street, Lagoas-Marcosende Campus, Vigo, 36310-ES, Spain.

Correpondence should be addressed to Sabela C. Mallo, Development Psychology, University of Santiago de Compostela, Xosé María Suárez Núnhez Street, Campus Sur, Santiago de Compostela, ES–15782, Spain; phone +34 881-813-949; fax +34 881-813-901; email: sabelacarme.mallo@usc.es

ABSTRACT

Objectives: To use a Machine Learning (ML) approach to compare Neuropsychiatric Symptoms (NPS) in participants of a longitudinal study who developed dementia and those who did not.

Design: Mann-Whitney U, chi-square and ML analysis. Nine ML algorithms were evaluated using a 10-fold stratified validation procedure. Performance metrics (accuracy, recall, F-1 score and Cohen's kappa) were computed for each algorithm, and graphic metrics (ROC and precision-Recall curves) and features analysis were computed for the best-performing algorithm.

Setting: Primary care health centres.

Participants: 128 participants: 78 cognitively unimpaired and 50 with MCI.

Measurements: Diagnosis at baseline, months from the baseline assessment until the 3rd follow-up or development of dementia (mean = 58.74 months), gender, age, Charlson Comorbidity Index, Neuropsychiatric Inventory-Questionnaire (NPI-Q) individual items, NPI-Q total severity and total stress score and Geriatric Depression Scale-15 items (GDS-15) total score.

Results: 30 participants developed dementia, while 98 did not. Most of the participants who developed dementia were diagnosed at baseline with amnestic multidomain MCI. The Random Forest Plot model provided the metrics that best predicted conversion to dementia. The algorithm indicated the importance of the metrics, in the following (decreasing) order: months from first assessment, age, diagnostic group at baseline, total NPI-Q severity score, total NPI-Q stress score and GDS-15 total score.

Conclusions: ML is a valuable technique for detecting the risk of conversion to dementia in MCI patients. NPS proxies were among the most important variables predicting

conversion, in accordance with previous findings suggesting that these symptoms are associated with neurocognitive disorders.

Key words: Neuropsychiatric Symptoms; Mild Cognitive Impairment; dementia; Behavioral and Psychological Symptoms of Dementia; diagnosis and classifications.

Introduction

Mild Cognitive Impairment (MCI) is considered a cognitive stage between no impairment and dementia (Jack *et al.*, 2018). Although people with MCI may remain stable or revert to being cognitively unimpaired (CU), the risk that they will develop dementia is relatively high, and the conversion rate increases with time from diagnosis and age (Facal *et al.*, 2015). Determining how cognitive, affective and clinical variables predict conversion is key for clinical and research purposes.

Neuropsychiatric Symptoms (NPS) are non-cognitive, behavioral or psychological symptoms associated with neurocognitive disorders (Lyketsos *et al.*, 2011). NPS are common in MCI (Monastero *et al.* 2009) and correlated with functional and cognitive impairment and increased caregiver stress (Feldman *et al.*, 2004).

NPS represent a Professional Interest Area (PIA) of the International Society to Advance Alzheimer's Research and Treatment (ISTAART), which has identified later life onset of sustained NPS as increasing the risk of cognitive decline and dementia (Ismail *et al.*, 2016). Recent research has established that NPS increase the likelihood that patients with MCI will develop dementia (Mortby *et al.*, 2017; Acosta *et al.*,2018) and that NPS scores are higher in participants who develop dementia than in those who do not, in both clinical (Rosenberg *et al.*, 2013) and population-based settings (Mortby *et al.*, 2017; Acosta, *et al.*, 2018). A previous longitudinal study of MCI patients concluded that NPS scores > 0 were associated with incident dementia and Alzheimer's disease (AD) (Rosenberg *et al.*, 2013). In a recent population-based study (Acosta *et al.*, 2018), the simultaneous presence of two symptoms corresponded to a 1.9-fold risk of dementia, whereas the presence of three to five symptoms increased the risk to 3-fold. Similarly, Mortby et al. established that NPS

were associated with a 3-fold risk of dementia (Mortby, *et al.*, 2017). Depression is one of the most commonly studied NPS (Ismail *et al.*, 2017) and its relation to increased risk of dementia has been demonstrated (Singh-Manoux *et al.*, 2017).

Techniques such as regression analysis (Mortby *et al.*, 2017; Acosta *et al.*, 2018) and Cox proportional hazard models (Rosenberg *et al.*, 2013) have been used to study the role of NPS in predicting progression from MCI to dementia. New scientific disciplines based on Machine Learning (ML) have emerged in recent years, thanks to the high processing capacities of state-of-the-art computers and the availability of massive amounts of data in electronic format. These approaches are based on exploiting the abilities of computers to learn from data (Abu-Mostafa *et al.*, 2012). In particular, ML provides innovative techniques that analyze data sets and automatically yield predictions or classifications. Furthermore, these methods have already demonstrated their potential to support diagnostics prediction or risk estimation by analyzing clinical and image features associated with dementia (Maroco *et al.*, 2011; Patel *et al.*, 2015; Facal *et al.*, 2019).

The purpose of the present study is to use an ML approach to explore the role of NPS in conversion from MCI to dementia. ML techniques were used to compare participants who developed dementia and those who did not, and the predictive value of NPS proxies according to the best performing ML model are discussed.

Methods

For this study, we selected 128 participants aged +50 years, belonging to the “Compostela Aging Study”, an on-going longitudinal study that began in 2008 (Juncos-Rabadán *et al.*, 2012), who completed baseline neurocognitive assessment and were followed up to 72

months later. Patients attendant Public Primary Care Health Centers in Santiago de Compostela and Vigo (Galicia, NW Spain) were referred to the study by their general practitioners (GPs). Inclusion criteria were subjective cognitive complaints, spontaneously reported by the patients. Exclusion criteria were: prior diagnosis of depression or other psychiatric disturbances according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria (American Psychiatric Association, 2013); prior diagnosis of neurological disease, including probable AD or other types of dementia, according to the NINCDS-ADRDA (McKhann *et al.*, 2011) and DSM-5 criteria (American Psychiatric Association, 2013); previous brain damage or brain surgery; undergoing chemotherapy; prior diagnosis of type II diabetes; sensory or motor disturbances; and consumption of substances that might affect normal performance of the tasks.

At baseline, the sample comprised 78 CU patients in the control group (60.9%) and 50 with MCI. MCI was diagnosed in accordance with Petersen's (Winblad *et al.*, 2004) criteria, as revised by Albert et al. (2011). The MCI group was also subdivided according to Petersen's criteria (Winblad *et al.*, 2004;Dubois *et al.,* 2007), as follows: multidomain amnestic-MCI (mda-MCI), 17 participants (13.3%), non-amnestic MCI (na-MCI), eight participants (6.3%) and single-domain amnestic MCI (sda-MCI), 25 participants (19.5%).

Participants underwent cognitive, affective and clinical examination conducted respectively by GPs, neurologists and psychogerontologists specialized in cognitive assessment. Diagnoses considered cognitive and clinical variables and were reached by consensus at special meetings held by the research team. Demographic, health and NPS data on the sample are shown in Table 1.

INSERT TABLE 1

Baseline evaluations were conducted between 2 January, 2008 and 11 November, 2012. A questionnaire on sociodemographic data was used to obtain information from the patients. The Charlson Comorbidity Index (CCI) (Charlson *et al.*, 1987) was obtained from the medical history to provide information about the health status of the patients. General cognitive functioning was evaluated with the Mini Mental State Examination (MMSE) (Folstein *et al.*, 1975). Cognitive impairment in several domains, such as memory, language, attention–calculation, praxis and executive functioning, were evaluated by the Cambridge Cognitive Assessment-Revised (CAMCOG-R) (Roth *et al.*, 1998). The California Verbal Learning Test (CVLT) (Delis *et al.*, 1987) was used to evaluate short-term and long-term memory. Functional assessment was conducted with the Lawton and Brody Index, to evaluate IADL (Lawton and Brody, 1969).

The cut-off point for diagnosis MCI was 1.5 standard deviations (SDs) below age and education norms on the corresponding tests. Participants who demonstrated normal cognitive functioning in the cognitive assessments, even those who reported subjective cognitive complaints, were included in the CU-control group. Amnestic MCI (aMCI) had a performance of 1.5 SDs below age norms for two measures of the CVLT (Short-Term Free Recall and Long-Term Free Recall). Mda-MCI was diagnosed when participants scored below age norms for two measures of the CVLT (Short-Term Free Recall and Long-Term Free Recall) and below age-related and education-related norms in the MMSE and on at least two subscales of the CAMCOG-R. Sda-MCI was diagnosed when participants scored below age norms for the CVLT, even though they performed normally in the MMSE and on the CAMCOG-R subscales. Na-MCI was diagnosed in participants who performed within

the normal range in the CVLT but 1.5 SDs below average in at least one of the other cognitive subscales of the CAMCOG-R.

The Neuropsychiatric Inventory-Questionnaire (NPI-Q) (Kaufer *et al.*, 2000), a short and reliable form of the Neuropsychiatric Inventory (NPI)(Cummings *et al.,* 1994), was administered. This measure is commonly used to assess NPS in both clinical and research settings (Ismail and Mortby, 2017). The NPI-Q evaluates, in a retrospective period of 4 weeks, the presence, frequency and severity of ten NPS (delusions, hallucinations, agitation/aggression, depression/dysphoria, anxiety, irritability, disinhibition, euphoria/elation, apathy, aberrant motor behavior) and two neurovegetative domains (changes in sleep and nightmare patterns, and change in appetite/eating). Hence, the informant assesses the level of distress experienced in relation to specific symptoms.

The Geriatric Depression Scale-15 items (GDS-15) (Yesavage and Sheikh, 1986), a self-reported screening scale specifically designed for the aging population, was used to assess depressive symptomatology.

Participants were informed at baseline about the longitudinal nature of the study and were contacted for follow-up assessments. They underwent the same cognitive assessment at follow-up and were re-diagnosed by the same research team around 48 and 72 months later (Facal, Guàrdia-olmos and Juncos-Rabadán, 2015).

The sample comprises 98 participants who completed the follow-up (mean = 58.74 months from the baseline assessment; S.D. = 8.55), and 30 who developed dementia during this period (mean = 41.90 months from the baseline assessment; S.D. = 23.70). Probable AD or other types of dementia were diagnosed according to the NINCDS-ADRDA (McKhann *et al.*, 2011) and DMS-5 (American Psychiatric Association, 2013) criteria, and progression

to dementia was confirmed by consultation of the medical history, recording the date of neurological diagnosis.

The study was approved by the Clinical Research Ethics Committee of the Xunta de Galicia (Galician Government) and was carried out in accordance with the provisions of the Declaration of Helsinki, as revised in Fortaleza 2009. Written informed consent was obtained from all participants before the study, and patient anonymity has been preserved.

Data were analyzed using SPSS v.20. The Mann-Whitney U test was used for descriptive analysis and group comparisons. Chi-square analysis was used to test for differences between diagnostic groups in converters and non-converters. Supervised ML techniques were used to test the value of NPS proxies in predicting conversion, with the dichotomous variable converter – non-converter considered the target variable. The data set consisted of 128 instances or triplets from participants and 20 input variables or features (including diagnosis at baseline assessment, months from the baseline assessment until the 3rd follow-up assessment or until conversion to dementia, gender, age, CCI, delusions, hallucinations, aggressiveness, depression/dysphoria, anxiety, euphoria/elation, apathy, disinhibition, irritability, aberrant motor behavior, sleep disturbances and eating disorders scores on the NPI-Q, total severity and total stress score on the NPI-Q and GDS-15 total score) to predict the target variable. Data analysis was performed with the Scikit-Learn (Pedregosa *et al.*, 2011) ML library under a Python ecosystem.

The data set was trained with 9 ML classifiers widely used in health research (Maroco *et al.,* 2011; Patel *et al.,* 2015; Facal *et al.*, 2019): 1) a linear model (i.e. LR- Logistic regression, which makes predictions using a logistic function based on the probability of a

default class; 2) hyperplanes (i.e. Support Vector Machines, which make classifications using a hyperplane that maximizes the margin or distance from the line only to the closest points); 3) similar instances or neighbors (i.e. Nearest Neighbors to compute the target-output as the class with the highest frequency from the K-most similar instances); (4) a Gaussian distribution (i.e. Gaussian Naïve Bayes, an extension of Naïve Bayes to obtain classification using real-valued attributes using the Gaussian Probability Density Function); 5) trees, used to split attributes based on values that minimize a loss function, such as the sum of squared errors (i.e. Random Forest Classifier and Extra Trees Classifier) and tree-based boosting algorithms, which are based on the creation and sequential addition of decision trees (i.e. Gradient Boosting Classifier and Ada Boost Classifier); and finally (6) neural networks (i.e. MLP - Multi-Layer Perceptron, a neural model that performs the classification by means of a linear function that maps the weighted inputs to the output of each neuron). To evaluate these ML algorithms, we used k-fold cross validation with k=10. This method is more effective for preventing biased or highly optimistic estimates than other methods such as simple training/test split, popularly known as overfitting (Schaffer, 1993).

For each of the aforementioned algorithms, we computed widely used performance metrics to evaluate the robustness of the ML prediction models, including score-based and graph metrics. We use the following score-based metrics: 1) Accuracy, i.e. the number of correct predictions divided by the total number of predictions; 2) Precision, i.e. the number of correct positive predictions divided by the total number of positive class values predicted (i.e. the sum of correct and incorrect positive predictions); 3) Recall, i.e. the number of positive predictions divided by the number of positive class values in the test data, in other words, the True Positive Rate; 4) F1-score, which expresses the balance between Precision and Recall; and 5) Cohen’s kappa, a more robust measure than the simple

percent agreement calculation, as it prevents the inclusion of positive values that occur by chance. We used the following graph metrics: 1) the ROC curve, a plot of True Positive Rate versus False Positive Rate; and 2) the Precision-Recall curve area, which shows the graphical relation between precision and recall. Finally, we used features analysis to explore the relevance of the input variables, i.e. we identified the input variables that contributed most to predicting converters and non-converter participants.

3. Results

Thirty participants developed dementia, while 98 did not. Comparisons between converters and non-converters for age, CCI and NPS information are summarized in Table 2.

INSERT TABLE 2.

Significant differences in converters and no converters were observed in relation to age, NPI-Q total severity, NPI-Q total stress, hallucinations, aggression and agitation, anxiety, apathy, disinhibition, irritability, aberrant motor behavior and eating disorders. Converters were older and their scores for these NPS measures were higher than in non-converters. Nevertheless, no significant differences were found for comorbidity, GDS-15 total score, delusions, depression/dysphoria, euphoria/elation or sleep disturbance.

Among the converters, at baseline 12 were diagnosed with mda-MCI (40%), two with na-MCI (6.6%), eight with sda-MCI (26.7%) and eight were CU (26.7%). Cross-tabulated data comparing diagnostic groups in converters and non-converters showed significant differences between groups (Fisher's exact test =26.75; $p < .001$); significant adjusted standardized residuals indicated that the number of patients diagnosed with mda-MCI who

developed dementia (12) was significantly higher than expected due to chance, and the number of CU individuals who developed dementia (8) was significantly lower than expected by chance (Table 3).

The dataset was processed using the nine ML algorithms outlined above. The classification abilities of the different algorithms, ordered from highest to lowest accuracy, are shown in Table 4.

INSERT TABLE 4.

The mean accuracy value acquired for all algorithms was higher than .79. The Random Forest performed best, with an accuracy greater than .86 (.00), and a Cohen's kappa, a measure of the validity of the classification without random influence, of .52 (.00) for this algorithm. The F1-score, which weights the harmonic average of precision and recall, reached a value of .71 (.16) in the Random Forest.

These results indicated that the Random Forest algorithm provided the best metrics for predicting the probability of conversion from MCI to dementia. The precision-recall curves and ROC curves obtained for this model indicated the high quality of classification (Figure 1 in the Supplement). The ROC curve was computed to assess the performance of the classifier, with an average AUC of .85±.16 (Figure 2 in the Supplement).

Finally, the Random Forest Plot algorithm was used to estimate the importance of the variables in the predictive models. The following six variables were the most important:

months from the first assessment, age, diagnostic group at baseline, total NPI-Q severity score, total NPI-Q stress score and GDS-15 total score (Table 5).

INSERT TABLE 5.

Discussion

To the best of our knowledge, this is the first study that uses an ML approach to explore the role of NPS in conversion from MCI to dementia. ML algorithms were used to compare a sample of 128 participants to predict the risk of dementia. Thirty participants developed dementia, and a higher proportion of mda-MCI participants than of CU participants developed dementia. The use of the innovate ML approach, based on the application of ML algorithms to socio-demographic data, basic health status and NPS proxies, enabled prediction of conversion to dementia with good precision and accuracy, both with a minimal standard deviation, specially the second one. The classification results from the nine ML algorithms were analyzed using a 10-fold stratified cross-validation procedure, which facilitate a robust classification score avoiding overfitting. The Random Forest Plot Classifier produced the best results in terms of accuracy and Cohen's kappa. This algorithm correctly predicted which participants would and would not develop dementia, with an accuracy rate of 86% ± .00 and precision of 89% ± .04. According to the individual values of the metrics used, precision was slightly higher than recall, which indicates that the true positive rate is slightly higher. The slightly low recall score of .71 (.28) may indicate a large number of false negatives in this sample. This result is possible, as progression from MCI to dementia occurs for reasons other than NPS, such as cognitive performance (Michaud *et al.,* 2017). Moreover, the Cohen's kappa score was fairly low.

However, in general the results showed that the Random Forest Plot Classifier is a good model for predicting conversion to dementia in MCI.

The variables were ranked according to their importance by means of the Random Forest Plot Classifier. For good prediction, the following were identified as the most important variables (in decreasing order): 1) months from the first assessment; 2) age; 3) diagnostic group at baseline; 4) total severity score on the NPI-Q; 5) total stress score on the NPI-Q; 6) GDS-15 total score. Accordingly, two diagnostic variables (months from the first assessment and diagnostic group at baseline), chronological age and three NPS proxies were the most important for predicting dementia. Individual items of the NPI-Q and comorbidity did not contribute as much predictive value to the model as the aforementioned variables.

The role of NPS in predicting conversion to dementia has traditionally been studied by regression analysis (Mortby *et al.,* 2017; Acosta *et al.,* 2018) and Cox proportional hazard models (Rosenberg *et al.,* 2013), and we used ML analysis as an innovative research approach.

In our study, global measures of NPS indicated a higher risk of conversion to dementia, with converters characterized by higher NPS scores than non-converters (Rosenberg, *et al.,* 2013; Forrester *et al.,* 2016; Mortby *et al.*, 2017; Acosta, *et al.,* 2018). These findings suggest that NPS should be considered when assessing the risk of dementia. Future studies should determine whether treatment of NPS helps to reduce the risk of dementia.

Our results are consistent with previously reported findings, confirming that the conversion rate increases with time since diagnosis and age of the participants (Facal *et al.,* 2015).

These results also indicated an association between caregiver stress related to NPS (measured with the NPI-Q) and an increased risk of dementia. Management of NPS, particularly of those most closely associated with caregiver distress, such as psychotic symptoms, are needed to reduce caregiver stress (Fischer, *et al.,* 2012). Further research is required to develop effective treatments, assistance and education programs (Fischer *et al.,* 2012; Ryan and Persad, 2012).

Considering depressive symptomatology, the GDS-15 played a more important role in classifying risk of conversion to dementia than the individual depression/dysphoria item of the NPI-Q. Depression is the most common behavioral symptom in MCI (Ismail *et al.*, 2017), and it is associated with a higher risk of dementia (Singh-Manoux *et al.*, 2017). Individuals with subsyndromal symptoms of depression may represent a subgroup of MCI that is highly vulnerable to accelerated cognitive decline (Gonzales *et al.*, 2017). The individual depression item of the NPI-Q is a single measure that simply collects information such as whether patients are sad, suffering from low mood or cry often (Kaufer *et al.*, 2000), while the GDS-15 is a screening scale including 15 questions (Yesavage and Sheikh, 1986). The GDS-15 provides much more extensive information, which may explain why the ML analysis indicated that it significantly increased the risk of dementia while the individual item of the NPI-Q did not. The GDS-15 is a self-reported measure while the individual item of the NPI-Q is informant-based. Therefore, the information may be qualitatively and significantly different (Sanchez-Villegas *et al.*, 2008). Importantly, research has suggested that the emergence of sustained depressive symptoms increases the risk of dementia, while established depressive symptoms do not (Singh-Manoux *et al.*, 2017).

In this respect, differentiating between the emergence of later life NPS that represents a change in a person's personality relative to the presence of longstanding symptoms is fundamental when predicting dementia (Tapiainen *et al.,* 2017). With the objective of improving the identification of patients at risk of dementia, the NPS PIA of the ISTAART published research diagnostic criteria for Mild Behavioral Impairment (MBI) (Ismail *et al.*, 2016). The Mild Behavioral Impairment-Checklist (MBI-C) was developed to assess MBI (Ismail *et al.*, 2017). Future studies using ML approaches to predict conversion from MCI to dementia must include MBI stages and would benefit from using information included in the MBI-C.

It should be recognized the relatively new application of ML techniques to the fields of aging and neurocognitive disorders. Nonetheless, the performance metrics obtained were robust and the algorithms used have been well tested in other similar areas.

In conclusion, the study findings suggest that ML techniques are helpful in detecting people with MCI and NPS at risk of dementia. Once the system was trained with the nine algorithms selected, the Random Forest Plot Classifier provided the best classification results. NPS proxies were among the most important variables for predicting conversion, adding further support to the hypothesis that NPS are associated with a higher risk of dementia in MCI. Future studies should focus on this topic with different psychiatric categories, such as MBI. Hence, studies are needed to determine the underlying mechanism through which NPS increase the chances of evolving into dementia. Finally, further research is required to develop a robust ML-based procedure approach for predicting diagnostic transitions along the spectrum of cognitive/behavioral impairment.

Conflict of interest

None.

Description of author's roles

Sabela C. Mallo, study concept and design, acquisition of data, manuscript writing, critical revision of the manuscript.

Sonia Valladares-Rodriguez, study concept and design, analysis and interpretation of data, critical revision of the manuscript.

David Facal, study concept and design, study supervision, critical revision of the manuscript.

Cristina Lojo-Seoane, study supervision, critical revision of the manuscript.

Manuel J. Fernández-Iglesias, analysis and interpretation of data, critical revision of the manuscript.

Arturo X. Pereiro, study supervision, critical revision of the manuscript.

Acknowledgments

The study was funded by the Spanish Ministry of Economy and Competitiveness (ref. PSI2014-55316-C3-1-R), Xunta de Galicia (grant ED431C2017/27), and the Galician Dementia Research Network (GAIN, Xunta de Galicia) (grant IN607C-2017/02). The first author is funded by a fellowship from the Spanish Ministry of Economy and Competitiveness (ref. BES-2015-071253).

References

Abu-Mostafa, Y. S., Magdon-Ismail, M. and Lin, H.-T. (2012) Learning from data. AML Book. New York, NY.

Acosta, I. *et al.* (2018) Neuropsychiatric symptoms as risk factors of dementia in a Mexican population: A 10/66 Dementia Research Group study, Alzheimers and Dementia, 14(3), pp. 271–279. doi: 10.1016/j.jalz.2017.08.015.

Albert, M. S. *et al.* (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimer's and Dementia, 7(3), pp. 270–279. doi: 10.1016/j.jalz.2011.03.008.

American Psychiatric Association (2013) DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th. American P. Washington, DC.

Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. R. (1987) A new method of classifying prognostic in longitudinal studies: development and validation, Journal of Chronic Diseases, pp. 373–383. doi: 0021-9681/87.

Cummings, J. *et al.* (1994) The Neuropsychiatric Inventory: Comprehensive Assessment of Psychopatlhology in Dementia, Neurology, 44, pp. 2308–2314. doi: 10.1212/WNL.44.12.2308.

Delis D. C., Kramer J. H. and Kaplan E., O. B. A. (1987) California Verbal Learning Test: Adult version. The Psychological Corporation. San Antonio, Texas.

Dubois, B. et al. (2007) Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS--ADRDA criteria, The Lancet Neurology. Elsevier, 6(8), pp. 734–746. doi: 10.1016/S1474-4422(07)70178-3.

Facal, D. *et al.* (2019) Machine learning approaches to studying the role of cognitive reserve in conversion from mild cognitive impairment to dementia, International Journal of Geriatric Psychiatry, pp. 0–1. doi: 10.1002/gps.5090.

Facal, D., Guàrdia-olmos, J. and Juncos-Rabadán, O. (2015) Diagnostic transitions in mild cognitive impairment by use of simple Markov models, 1, pp. 669–676. doi: 10.1002/gps.4197.

Feldman, H. *et al.* (2004) Behavioral symptoms in mild cognitive impairment.," Neurology, 63(4), p. 1199–1201. doi: 10.1212/01.WNL.0000118301.92105.EE.

Fischer, C. E., Ismail, Z. and Schweizer, T. A. (2012) Impact of neuropsychiatric symptoms on caregiver burden in patients with Alzheimers disease, Neurodegenerative Disease Management, 2(3), pp. 269–277. doi: 10.2217/nmt.12.19.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975) 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician,Journal of Psychiatric Research, 12(3), pp. 189–198. doi: 10.1016/0022-3956(75)90026-6.

Gonzales, M. M. *et al.* (2017) Cortical Atrophy is Associated with Accelerated Cognitive Decline in Mild Cognitive Impairment with Subsyndromal Depression, American Journal of Geriatric Psychiatry, 25(9), pp. 980–991. doi: 10.1016/j.jagp.2017.04.011.

Ismail, Z. *et al.* (2016) Neuropsychiatric symptoms as early manifestations of emergent dementia: Provisional diagnostic criteria for mild behavioral impairment, Alzheimers Dement., 12(2), pp. 195–202. doi: 10.1016/j.jalz.2015.05.017.

Ismail, Z., Elbayoumi, H., *et al.* (2017) Prevalence of Depression in Patients With Mild Cognitive Impairment: A Systematic Review and Meta-analysis, JAMA Psychiatry, 74(1), pp. 58–67. doi: 10.1001/jamapsychiatry.2016.3162.

Ismail, Z., Agüera-Ortiz, L., *et al.* (2017) The Mild Behavioral Impairment Checklist (MBI-C): A Rating Scale for Neuropsychiatric Symptoms in Pre-Dementia Populations, Journal of Alzheimers Disease, 56(3), pp. 929–938. doi: 10.3233/JAD-160979.

Ismail, Z. and Mortby, M. E. (2017) Cognitive and neuropsychiatric screening tests in older adults, in Chiu, H. and Shulman, K. (eds.) Mental Health and Illness of the Elderly. 1st ed. Singapoure: Springer Singapore, pp. 343–368. doi: 10.1007/978-981-10-0370-7_16-1.

Jack, C. R. *et al.* (2018) NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease, Alzheimers and Dementia. Elsevier Inc., 14(4), pp. 535–562. doi: 10.1016/j.jalz.2018.02.018.

Juncos-Rabadán, O. et *al.* (2012) "Prevalence and correlates of cognitive impairment in adults with subjective memory complaints in primary care centres," Dementia and Geriatric Cognitive Disorders, 33(4), pp. 226–232. doi: 10.1159/000338607.

Kaufer, D. *et al.* (2000) Validation of the NPI-Q, a Brief Clinical Form of the Neuropsychiatric Inventory, Journal of Neuropsychiatry, 12(2), pp. 233–239. doi: 10.1176/appi.neuropsych.12.2.233.

Lanctôt, K. L. *et al.* (2017) Apathy associated with neurocognitive disorders: Recent progress and future directions, Alzheimer's & Dementia, 13(1), pp. 84–100. doi: 10.1016/j.jalz.2016.05.008.

Lawton, M. P. and Brody, E. M. (1969) Assessment of Older People : Self-Maintaining and Instrumental Activities of Daily Living 1.

Lyketsos, C. G. *et al.* (2011) Neuropsychiatric symptoms in Alzheimer's disease," Alzheimers Dement. Elsevier Ltd, 7(5), pp. 532–539. doi: 10.1016/j.jalz.2011.05.2410.

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., and de Mendonça, A. (2011) Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests, BMC research notes. BioMed Central Ltd, 4(1), p. 299. doi: 1756-0500/4/299.

McKhann, G. M. *et al.* (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease,Alzheimers Dement, 7(3), pp. 263–269. doi: 10.1016/j.jalz.2011.03.005.

Michaud, T. L., Su, D., Siahpush, M., and Murman, D. L. (2017) The risk of incident mild cognitive impairment and progression to dementia considering mild cognitive impairment subtypes, Dementia and Geriatric Cognitive Disorders Extra, 7(1), pp. 15–29. doi: 10.1159/000452486.

Monastero, R., Mangialasche, F., Camarda, C., Ercolani, S., and Camarda, R. (2009) A systematic review of neuropsychiatric symptoms in mild cognitive impairment, Journal of Alzheimers Disease, 18(1), pp. 11–30. doi: 10.3233/JAD-2009-1120.

Mortby, M. E., Burns, R., Eramudugolla, R., Ismail, Z., and Anstey, K. J. (2017) Neuropsychiatric Symptoms and Cognitive Impairment: Understanding the Importance of Co-Morbid Symptoms, Journal of Alzheimers Disease, 59(1), pp. 141–153. doi: 10.3233/JAD-170050.

Patel, M. J., Andreescu, C., Price, J. C., Edelman, K. L., Reynolds III, C. F., and Aizenstein, H. J. (2015) Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction, International journal of geriatric psychiatry. Wiley Online Library, 30(10), pp. 1056–1067. doi: 10.1002/gps.4262.

Pedregosa, F. et al. (2011) "Scikit-learn: Machine learning in Python," Journal of machine learning research, 12, pp. 2825–2830.

Rosenberg, P. B., Mielke, M. M., Appleby, B. S., Oh, E. S., Geda, Y. E., & Lyketsos, C. G. (2013) The Association of Neuropsychiatric Symptoms in MCI With Incident Dementia and Alzheimer Disease, Am J Geriatr Psychiatry, 21(7), pp. 685–695. doi: 10.1016/j.jagp.2013.01.006.

Roth, M., Huppert, F. and Mountjoy, C. and Tym, E. (1998) The Cambridge Examination for Mental Disorder of the Elderly - revised, British Journal of Psychiatry. Cambridge: Cambridge University Pres, 149, pp. 698–709.

Ryan, K. A. and Persad, C. (2012) Neuropsychiatric Symptoms and Executive Functioning in Patients with Mild Cognitive Impairment: Relationship to Caregiver Burden, Dementia and Geriatric Cognitive Disorders, 34(3–4), pp. 206–215. doi: 10.1159/000339955.

Sanchez-Villegas, A. *et al.* (2008) Validity of a self-reported diagnosis of depression among participants in a cohort study using the Structured Clinical Interview for DSM-IV (SCID-I), BMC psychiatry. BioMed Central, 8(1), p. 43. doi: 10.1186/1471-244X-8-43.

Schaffer, C. (1993) “Overfitting avoidance as bias,” Machine Learning, 10(2), pp. 153–178. doi: 10.1007/BF00993504.

Singh-Manoux, A. *et al.* (2017) Trajectories of depressive symptoms before diagnosis of dementia: A 28-year follow-up study, JAMA Psychiatry, 74(7), pp. 712–718. doi: 10.1001/jamapsychiatry.2017.0660.

Tapiainen, V., Hartikainen, S. and Taipale, H. et al. (2017) Hospital-treated mental and behavioral disorders and risk of Alzheimer’s disease: A nationwide nested case-control study, European Psychiatry. Elsevier Masson SAS, 43, pp. 92–98. doi: 10.1016/j.eurpsy.2017.02.486.

Winblad, B. *et al.* (2004) Mild cognitive impairment - beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment, Journal of Internal Medicine, 256(3), pp. 240–246. doi: 10.1111/j.1365-2796.2004.01383.x.

Yesavage, J. A. and Sheikh, J. I. (1986) Geriatric depression scale (GDS): Recent evidence and development of a shorter version, Clinical Gerontologist, 5(1–2), pp. 165–173. doi: 10.1300/J018v05n01_09.

Figure 1.

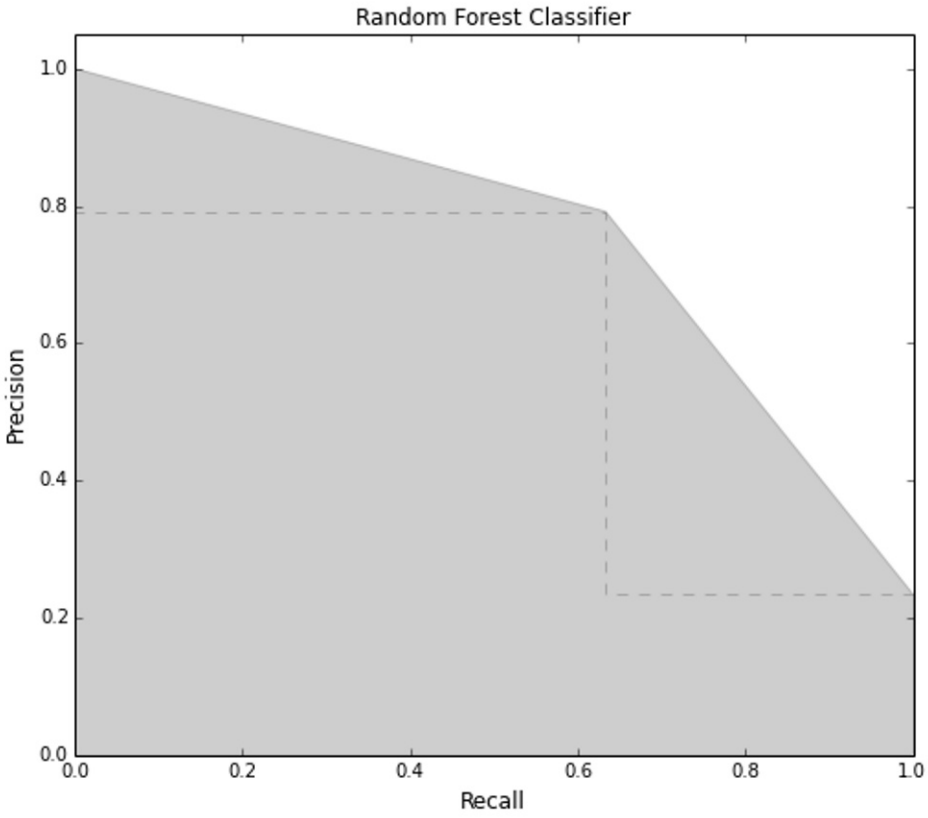
Random Forest Classifier
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Figure 2.

Table 1. Age, health status (Charlson Comorbidity Index) and NPS proxies for the four diagnostic groups. Mean, standard deviations (between brackets) and range are included.

	mda-MCI n=17	na-MCI n=8	sda-MCI n=25	CU n=78	Total n=128
Gender (woman)	11(8.6%)	7(5.5%)	7(5.5%)	49(38.3%)	74(57.8%)
Age (years)	72.41 (7.21) Range:60-87	68.5 (6.60) Range: 59-78	70 (7.63) Range:54-87	65.19 (7.81) Range:50-82	67.29 (8.04) Range: 50-87
CCI	1.12 (1.11) Range:0-4	1.50 (.75) Range:1-3	.60 (.82) Range:0-2	.72 (.77) Range:0-3	.80 (.85) Range: 0-4
Total NPI-Q severity score	5.58 (4.14) Range:0-14	5.50 (5.45) Range:0-16	3.16 (4.43) Range: 0-19	2.37 (2.97) Range:0-11	3.14 (3.79) Range: 0-19
Total NPI-Q stress score	6.12 (6.84) Range:0-23	6.25 (8.61) Range:0-25	3.88 (7.69) Range:0-36	2.47 (3.64) Range:0-15	3.46 (5.59) Range: 0-36
GDS-15	2.76 (2.28) Range:0-9	5.87 (3.27) Range: 0-9	2.00 (1.68) Range: 0-7	2.76 (2.06) Range: 0-9	2.84 (2.25) Range:0-12
Delusions	.06 (.24) Range:0-1	.00 (.00) Range:0-0	.04 (.20) Range: 0-1	.02 (.22) Range: 0-2	.03 (.21) Range: 0-2

Hallucinations	.06 (.24)	.00 (.00)	.08 (.40)	.00 (.00)	.02 (.20)
	Range:0-1	Range:0-0	Range:0-2	Range:0-0	Range: 0-2
Agitation/	.47 (.94)	.00 (.00)	.20 (.64)	.02 (.22)	.11 (.49)
aggressiveness	Range:0-3	Range:0-0	Range:0-3	Range:0-2	Range: 0-3
Depression	.59 (.87)	.62 (92)	.56 (.96)	.29 (.54)	.41 (.71)
	Range:0-2	Range:0-2	Range:0-3	Range:0-2	Range: 0-3
Anxiety	1.23 (1.14)	.87 (.99)	.52 (.92)	.52 (.97)	.64 (1.01)
	Range:0-3	Range:0-2	Range:0-3	Range:0-3	Range: 0-3
Euphoria/	.00 (.00)	.00 (.00)	00 (.00)	.06 (.24)	.04 (.19)
elation	Range:0-0	Range:0-0	Range:0-0	Range:0-1	Range: 0-1
Apathy	.59 (.79)	.87 (.83)	.32 (.75)	.18 (.53)	.30 (.65)
	Range:0-2	Range:0-2	Range:0-3	Range:0-3	Range: 0-3
Disinhibition	.18 (.39)	.25 (.71)	.32 (.80)	.06 (.33)	.14 (.50)
	Range:0-1	Range:0-2	Range:0-3	Range:0-2	Range: 0-3
Irritability	1 (1.12)	.50 (.76)	.64 (.86)	.38 (.74)	.52 (.84)
	Range:0-3	Range:0-2	Range:0-3	Range:0-3	Range: 0-3
Aberrant motor	.53 (.80)	.37 (.74)	.20 (.64)	.06 (.29)	.17 (.52)
behavior	Range:0-2	Range:0-2	Range:0-3	Range:0-2	Range: 0-3
Sleep	.47 (.87)	1.37 (1.06)	.24 (.60)	.59 (.90)	.55 (.88)
disturbances	Range:0-3	Range:0-3	Range:0-2	Range:0-3	Range: 0-3
Eating disorders	.41 (.79)	.62 (.91)	.04 (.20)	.14 (.44)	.19 (.53)
	Range:0-3	Range:0-2	Range:0-1	Range:0-2	Range: 0-3

Notes: Mda-MCI: multidomain amnestic-Mild Cognitive Impairment; na-MCI: non-amnestic Mild Cognitive Impairment; sda-MCI: single-domain amnestic Mild Cognitive Impairment; CU: cognitively unimpaired; CCI: Charlson Comorbidity Index; NPI-Q: Neuropscyhiatric Inventory-Questionnaire; GDS-15: Geriatric Depression Scale-15.

Table 2. Age, health status (Charlson Comorbidity Index) and NPS proxies for the converter – non-converter groups at baseline. Mean, standard deviations (between brackets), range and group comparisons (Mann-Whitney U test) are included.

	Converters (n=30)	Non-converters (n=98)	Group comparisons
Age (years)	74.06 (5.24) Range: 63-87	65.22 (7.61) Range: 50-87	Z=-5.54**
CCI	.7 (.79) Range: 0-2	.82 (.87) Range: 0-4	Z=-.63
Total NPI-Q severity score	5.47 (4.91) Range: 0-19	2.43 (3.07) Range: 0-16	Z=-3.25**
Total NPI-Q stress score	6.63 (8.45) Range: 0-36	2.50 (3.94) Range: 0-25	Z=-2.66**
GDS-15	2.73 (2.27) Range: 0-9	2.82 (2.26) Range: 0-9	Z=-.19
Delusions	.07 (.25) Range: 0-1	.02 (.20) Range: 0-2	Z=-1.76
Hallucinations	.10 (.40) Range: 0-2	.00 (.00) Range: 0-0	Z=2.56*
Agitation/ aggressiveness	.40 (.89) Range: 0-3	.03 (.22) Range: 0-2	Z=-3.57**

Depression	.57 (.86)	.35 (.66)	Z=-1.21
	Range: 0-3	Range: 0-3	
Anxiety	1.03 (1.09)	.52 (.95)	Z=-2.86**
	Range: 0-3	Range: 0-3	
Euphoria/	.07 (.25)	.03 (.17)	Z=-.89
elation	Range: 0-1	Range: 0-1	
Apathy	.57 (.90)	.22 (.54)	Z=-2.40*
	Range: 0-3	Range: 0-2	
Disinhibition	.33 (.71)	.08 (.39)	Z=-3.18**
	Range: 0-3	Range: 0-2	
Irritability	.90(1.09)	.41 (.71)	Z=-2.43*
	Range: 0-3	Range: 0-3	
Aberrant motor	.50 (.82)	.71(.32)	Z=-4.19**
behavior	Range: 0-3	Range: 0-2	
Sleep disturbances	.63 (.96)	.53 (.86)	Z=-.48
	Range: 0-3	Range: 0-3	
Eating disorders	.30 (.53)	.15(.52)	Z=-5.96**
	Range: 0-2	Range: 0-3	

Notes: ** < .05*;* * p < .01; CCI: Charlson Comorbidity Index; NPI-Q: Neuropsychiatric Inventory-Questionnaire; GDS-15: Geriatric Depression Scale-15.

Table 3. Cross-tabulation results (n and adjusted residuals *e*) for Converters and non-converter in each of the four diagnostic groups.

		Converters (n=30)		Non-converters (n=98)
		n	Adjusted standardized residuals	n
Diagnostic groups	mda-MCI	12	4.9*	5
	na-MCI	2	.1	6
	sda-MCI	8	1.1	17
	CU	8	-4.4*	70
Total		30		98

Notes: * $p < .01$; mda-MCI: multidomain amnestic-Mild Cognitive Impairment; na-MCI: non-amnestic Mild Cognitive Impairment; sda-MCI: single-domain amnestic Mild Cognitive Impairment; CU: cognitively unimpaired

Table 4. Metrics about prediction of converters and non-converters patients. Mean and standard deviations (between brackets).

ML classifier algorithms	Accuracy	Precision	F1score	Recall	Kappa
Random Forest	.86 (.00)	.89 (.04)	.75 (.16)	.71 (.28)	.52 (.00)
Extra Trees	.86 (.00)	.85 (.02)	.77 (.14)	.73 (.23)	.55 (.00)
Support Vector Machine	.85 (.00)	.83 (.04)	.76 (.15)	.73 (.23)	.53 (.00)
Logistic Regression	.84 (.00)	.85 (.01)	.73 (.18)	.69 (.29)	.47 (.00)
Artificial Neural Network	.84 (.00)	.80 (.07)	.76 (.14)	.74 (.20)	.52 (.00)
Gradient Boosting	.84 (.00)	.77 (.11)	.76 (.13)	.75 (.15)	.53 (.00)
Ada Boost	.83 (.00)	.76 (.14)	.77 (.12)	.77 (.11)	.53 (.00)
K-Nearest Neighbours	.80 (.00)	.75 (.05)	.61 (.26)	.60 (.37)	.26 (.00)
Gaussian Navy Bayes	.79 (.00)	.70 (.12)	.66 (.21)	.64 (.28)	.33 (.00)

Table 5. Measures of features importance with the Random Forest Plot algorithm.

Features	Magnitude of importance	Magnitude of importance (standardized to unit)
Months	.33	1.00
Age	.16	.48
Diagnosis at baseline	.09	.27
NPI-Q total severity	.07	.21
NPI-Q total stress	.06	.18
GDS-15 total	.05	.15
NPI-Q anxiety	.03	.09
NPI-Q apathy	.03	.09
CCI	.03	.09
NPI-Q sleep disturbances	.02	.06
NPI-Q aberrant motor behaviour	.02	.06
NPI-Q agitation and aggression	.02	.06
NPI-Q irritability	.02	.06
NPI-Q depression	.02	.06
NPI-Q disinhibition	.02	.06
NPI-Q eating disorders	.01	.03
NPI-Q hallucinations	.01	.03
Gender	.01	.03

NPI-Q euphoria	0.01	.03
NPI-Q delusions	0.00	.00

Notes: NPI-Q: Neuropscyhiatric Inventory-Questionnaire; GDS-15: Geriatric Depression Scale-15; CCI: Charlson Comorbidity Index.