Received: 4 September 2018 | Accepted: 5 March 2019

DOI: 10.1002/gps.5090

RESEARCH ARTICLE

WILEY International Journal of Geriatric Psychiatry

Machine learning approaches to studying the role of cognitive reserve in conversion from mild cognitive impairment to dementia

David Facal[1] | Sonia Valladares-Rodriguez[2] | Cristina Lojo-Seoane[1] | Arturo X. Pereiro[1] | Luis Anido-Rifon[2] | Onésimo Juncos-Rabadán[1]

[1] Department of Developmental and Educational Psychology, University of Santiago de Compostela, Santiago de Compostela, Spain

[2] School of Telecommunication Engineering, University of Vigo, Vigo, Spain

Correspondence
D. Facal, Facultade de Psicoloxía, Universidade de Santiago de Compostela, Rúa Xosé María Suárez Núñez, s/n, Campus Vida, 15782 Santiago de Compostela, Spain.
Email: david.facal@usc.es

Funding information
Axencia Galega de Innovación, Grant/Award Number: Rede Galega de Investigación en Demencias REGIDEM; Consellería de Cultura, Educación e Ordenación Universitaria, Grant/Award Number: ED431C-2017/27; Spanish Directorate General of Scientific and Technical Research, Grant/Award Numbers: PSI2014-55316-C3-1-R and PSI2017-89389-C2-1-R

Objectives: The overall aim of the present study was to explore the role of cognitive reserve (CR) in the conversion from mild cognitive impairment (MCI) to dementia. We used traditional and machine learning (ML) techniques to compare converter and nonconverter participants. We also discuss the predictive value of CR proxies in relation to the ML model performance.

Methods: In total, 169 participants completed the longitudinal study. Participants were divided into a control group and three MCI subgroups, according to the Petersen criteria for diagnosis. Information about the participants was compared using nine ML classification techniques. Seven relevant performance metrics were computed in order to evaluate the accuracy of prediction regarding converter and nonconverter participants.

Results: ML algorithms applied to socio-demographic, basic health, and CR proxy data enabled prediction of conversion to dementia. The best performing models were the gradient boosting classifier (accuracy (ACC) = 0.93; F1 = 0.86, and Cohen κ = 0.82) and random forest classifier (ACC = 0.92; F1 = 0.79, and Cohen κ = 0.71). Use of ML techniques corroborated the protective role of CR as a mediator of conversion to dementia, whereby participants with more years of education and higher vocabulary scores survived longer without developing dementia.

Conclusions: We used ML approaches to explore the role of CR in conversion from MCI to dementia. The findings indicate the potential value of ML algorithms for detecting risk of conversion to dementia in cognitive aging and CR studies. Further research is required to develop an ML-based procedure that can be used to make robust predictions.

KEYWORDS

cognitive reserve, dementia, diagnostic transitions, educational level, gradient boosting classifier, machine learning, mild cognitive impairment, random forest classifier, supervised learning, vocabulary

1 | INTRODUCTION

Mild cognitive impairment (MCI) is a condition characterized by cognitive impairment with minimal impairment of instrumental activities of daily living. MCI is considered a cognitive stage of the cognitive continuum traditionally divided into three categories: cognitively unimpaired (CU), MCI, and dementia.[1] Although MCI sufferers have an increased risk of progressing to dementia than their age-matched controls, most remain stable or return to normality, and only a proportion convert to dementia.[2-4] Prediction of the risk of MCI individuals converting to Alzheimer disease (AD) and identification of how demographic, clinical, and lifestyle characteristics influence this risk are increasingly relevant in relation to the prevention of cognitive impairment.

Years of schooling, occupational complexity, crystallized intelligence (vocabulary level), engagement in leisure activities, participation in social and cultural events, and integrity of social networks have been shown to have protective effects against cognitive decline.[5-9] All of these factors contribute to increasing cognitive reserve (CR), defined as an active process whereby the brain adapts to a situation of deterioration by using cognitive processing resources to compensate for the deficits.[10] The CR construct has been formalized by considering two latent variables: level of education (including proxies such as years of formal schooling, occupational attainment, reading habits, and vocabulary knowledge) and lifestyle (including participation in social and cultural activities).[9] Longitudinal studies have focused on confirming the protective effects of CR on general cognitive performance and memory over time[11,12] and on the role of CR in the transition between diagnostic states.[13]

Diagnostic transitions in MCI and eventual conversion to dementia have been studied by survival analysis,[14] polynomial regression analysis,[15] multinomial logistic regression (LR) analysis,[16] and Markov models.[3] learning models using cognitive reserve proxies as predictors yielded learning models using cognitive reserve proxies as predictors yielded learning (ML), a scientific discipline that focuses on how computers learn from data,[17] can build classifiers by automatically learning the inherent structure of a data set. In this study, we focused on the supervised learning category, which starts with the goal of predicting a known output or target. This innovative technique has shown enormous potential for supporting diagnosis prediction, risk estimation, and classification learning for diagnosis.[18-21]

The aim of the study was to use ML approaches to explore the role of CR in conversion of MCI to dementia. For this purpose, we used traditional and ML techniques to compare converter and nonconverter participants. We also discuss the predictive value of CR proxies in relation to the most accurate ML models.

Key points

- Machine learning models using cognitive reserve proxies as predictors yielded high predictive values regarding the risk of conversion to dementia.
- The Gradient Boosting Classifier predicted conversion to dementia with an accuracy of 0.93 and a precision of 0.88.
- The Random Forest Classifier predicted conversion to dementia with an accuracy of 0.92 and a precision of 0.86.
- The six most important predictors using both algorithms were months from baseline, age, diagnostic group, WAIS and Peabody vocabulary scores and number of years of education.

2 | METHODS

2.1 | Participants

The study included 169 participants aged over 50 years (range, 50-87 y) who completed a baseline assessment and attended a follow-up evaluation for about 54 months within an ongoing longitudinal study, which began in 2008.[22] All participants were referred to us by general practitioners according to the study criteria of subjective cognitive complaints (participants spontaneously reported that their cognition was not as good as before) and the absence of prior diagnosis of cognitive impairment, clinical stroke, traumatic brain injury, motor-sensory defects, alcohol or drug abuse/dependence, or diagnosis of any significant medical or psychiatric illnesses.

At baseline, the sample comprised 101 CU participants as the control group (49.8%) and 68 participants who fulfilled the core NIA-AA 2011 MCI criteria.[23] The MCI group was distributed according to the Petersen criteria,[24] as follows: 22 participants were included in the multidomain amnestic MCI (ma-MCI) group (13% of the total sample), 14 participants in the nonamnestic MCI (na-MCI) group (8.3%), and 32 participants in the single-domain amnestic MCI (sda-MCI) group (18.9%). The diagnoses were made by considering clinical and cognitive assessments and were reached by consensus at special diagnostic meetings held by the research team. Demographic, health, and CR-related data for the different groups of participants are shown in Table 1.

2.2 | Baseline assessment and diagnosis

The baseline assessments were conducted between 2 January 2008 and 11 November 2012 by psychologists specializing in psychogerontology and with training in neuropsychological assessment. Participants underwent a demographic and health interview and completed extensive neuropsychological and cognitive assessment, including the California Verbal Learning Test (CVLT)[25,26] and the Spanish, revised version of the Cambridge Cognitive Examination (CAMCOG-R), which comprises subscales for specific domains such as language, attention-calculation, praxis, and executive functioning and is sensitive to MCI detection.[27] In relation to health status, the Charlson Comorbidity Index was obtained from the medical history, and history of memory disorders in the family was considered separately (yes/no

TABLE 1 Baseline demographic, health status, and cognitive reserve data for each diagnostic group[a]

	ma-MCI	na-MCI	sda-MCI	Control	Total
Age (y)	71.36 (7.58) Range, 54-87	66.64 (7.72) Range, 50-78	71.03 (8.33) Range, 54-87	64.99 (8.06) Range, 50-82	67.10 (8.44) Range, 50-87
Education (y)	10.54 (4.96) Range, 3-25	9 (3.59) Range, 3-18	9.56 (3.72) Range, 5-20	9.44 (4.89) Range, 1-22	9.57 (4.58) Range, 1-25
Occupational status	Worked, 4 (18.2%) Retired, 18 (81.8%)	Worked, 5 (35.7%) Retired, 9 (64.3%)	Worked, 6 (18.8%) Retired, 26 (81.3%)	Worked, 42 (41.6%) Unemployed, 3 (3%) Retired, 56 (55.4%)	Worked, 57 (33.7%) Unemployed, 3 (1.8%) Retired, 109 (64.5%)
Co-morbidity	1.04 (1.09) Range, 0-4	1 (0.88) Range, 0-3	0.62 (0.87) Range, 0.3	0.72 (0.78) Range, 0-3	0.77 (0.85) Range, 0-4
History of memory disorders	Yes, 11 (50%) No, 11 (50%)	Yes, 5 (35.7%) No, 9 (64.3%)	Yes, 11 (34.4%) No, 21 (65.6%)	Yes, 55 (54.5%) No, 46 (45.5%)	Yes, 82 (48.5%) No, 87 (51.5%)
Occupational attainment	Unqualified worker, 11 (50%); housewife, 1 (4.5%); qualified worker, 6 (27.3%): middle ranking manager, 4 (18.2%)	Unqualified worker, 8 (57.1%); housewife, 3 (21.4%); qualified worker, 3 (21.4%)	Unqualified worker, 14 (43.8%); housewife, 7 (21.9%); qualified worker, 9 (28.1%): middle-ranking manager, 2 (6.3%)	Unqualified worker, 42 (41.6%); housewife, 17 (16.8%); qualified worker, 28 (27.7.8%): middle-ranking manager, 12 (11.9%); high ranking or director, 2 (2%)	Unqualified worker, 75 (44.4%); housewife, 28 (16.6%); qualified worker, 46 (27.7%): middle-ranking manager, 18 (10.7%); high ranking or director, 2 (1.2%)
Reading habits	Never, 3 (13.6%); once a week, 5 (22.7%); twice a week, 2 (9.1%); every day, 12 (54.5%)	Never, 3 (21.4%); once a week, 2 (14.3%); twice a week, 5 (35.7%); every day, 4 (28.6%)	Never, 5 (15.6%); once a week, 5 (15.6%); twice a week, 3 (9.4%); every day, 19 (59.4%)	Never, 11 (10.9%); once a week, 13 (12.9%); twice a week, 17 (16.8%); every day, 60 (59.4%)	Never, 22 (13%); once a week, 25 (14.8%); twice a week, 27 (16%); every day, 95 (56.2%)
Frequency of social activities	Never, 8 (36.4%); rarely, 6 (27.3%); often, 6 (27.3%); always, 2 (9.1%)	Never, 5 (35.7%); rarely, 2 (14.3%); often, 5 (35.7%); always, 2 (14.3%)	Never, 10 (31.3%); rarely, 7 (21.9%); often, 8 (25%); always, 7 (21.9%)	Never, 33 (32.7%); rarely, 27 (26.7%); often, 20 (19.8%); always, 21 (20.8%)	Never, 56 (33.1%); rarely, 42 (24.9%); often, 39 (23.1%); always, 32 (18.9%)
Frequency of cultural activities	Never, 17 (77.3%); often, 3 (13.6%); always, 2 (9.1%)	Never, 12 (85.7%); rarely, 1 (7.1%); always, 1 (7.1%)	Never, 18 (56.3%); rarely, 2 (6.3%); often, 4 (12.5%); always, 8 (25%)	Never, 71 (70.3%); rarely, 11 (10.9%); often, 7 (6.9%); always, 12 (11.9%)	Never, 118 (69.8%); rarely, 14 (8.3%); often, 14 (8.3%); always, 23 (13.6%)
WAIS vocabulary score	43.68 (13.23) Range, 15-71	40.21 (14.56) Range, 18-66	48.25 (11.43) Range, 22-70	51.29 (12.83) Range, 23-75	48.80 (13.18) Range, 15-75
Peabody test score	56.32 (14.23) Range, 23-87	54.61 (18.64) Range, 16-88	58.56 (17.55) Range, 19-84	64.39 (17.72) Range, 7-94	61.43 (17.60) Range, 7-94

Abbreviations: ma-MCI, multidomain amnestic mild cognitive impairment; na-MCI, nonamnestic mild cognitive impairment; sda-MCI, single-domain amnestic mild cognitive impairment; WAIS, Wechsler Adult Intelligence Scale.

[a]For numerical variables, the values shown are means (and standard deviations). For categorical variables, the values are frequencies (and percentages).

$^{**}P < 0.01$.

$^{*}P < 0.05$.

response). Demographic and health changes were recorded during follow-up assessments; all participants underwent the same cognitive assessment as before and were rediagnosed by the same research team, again by consensus at specific meetings.

For all the diagnoses, the cut-off was 1.5 standard deviations (SDs) below age and education norms on the corresponding tests. All participants who displayed normal cognitive performance in the cognitive assessment, even though they reported subjective cognitive complaints, were included in the CU-control group. For the MCI participants, the general criteria outlined by Albert et al[24] were applied: (i) presence of concerns regarding a change in cognition corroborated by an informant; (ii) impairment in one or more cognitive domains; (iii) preservation of independence in functional abilities with minimal aid or assistance; and (iv) no dementia according to the criteria considered by National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and

Related Disorders Association (NINCDS-ADRDA) and in the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*), Fourth Edition. For amnestic MCI (aMCI), the criteria proposed by Petersen and colleagues[24,28] were applied, and performance of 1.5 SDs below age norms for two measures of the Spanish version of the CVLT (Short-Term Free Recall and Long-Term Free Recall) was considered to indicate the condition. ma-MCI was diagnosed in those participants who scored below age norms for two measures of the Spanish version of the CVLT (Short-Term Free Recall and Long-Term Free Recall) and below age-related and education-related norms in the Mini-Mental State Examination (MMSE) and on at least two subscales of the CAMCOG-R. Participants were diagnosed as sda-MCI when they scored below age norms for the Spanish version of the CVLT but performed normally in the MMSE and in the CAMCOG-R subscales. Finally, na-MCI was diagnosed in participants who performed within the normal range in the Spanish version of the CVLT but 1.5 SDs below average in at least one of the other cognitive subscales of the CAMCOG-R.

The study received approval from the Ethics in Clinical Research Committee of the Galician Government and was conducted in accordance with the provisions of the Declaration of Helsinki as revised in Seoul in 2008. Written informed consent was obtained from all participants prior to the study.

2.3 | Cognitive reserve proxies

Proxy variables have been used to operationalize the theoretical construct of CR.[29] In this study, an ad hoc questionnaire including the following measures was administered to the participants: (a) total number of years of formal schooling; (b) occupational attainment, which evaluates the complexity of the profession according to the protocol outlined in the Network for Efficiency and Standardization of Dementia Diagnosis (NEST-DD) project,[7] on a scale of 1 to 6 (1, no occupation; 2, unqualified worker; 3, housewife; 4, qualified worker, small business employee, office worker, or sales person; 5, middle-ranking civil servant or manager, small business owner, teacher, or specialist in subordinate position; 6, high-ranking civil servant or director, university lecturer, or self-employed person with high responsibility); (c) reading habits, which evaluates the frequency of reading during the last 3 years on a scale of 0 to 4 (0, never; 1, occasionally; 2, once a week; 3, twice a week; and 4, every day); (d) frequency of social and cultural activities, which evaluates participation in these activities during the last 3 years on a scale of 0 to 4 (0, never; 1, rarely; 2, occasionally; 3, often; and 4, always). The level of vocabulary knowledge was evaluated by the vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS)[30] and the Peabody picture vocabulary test.[31]

2.4 | Follow-up assessments

All participants were informed about the longitudinal nature of the study at baseline, and they were invited to attend follow-up assessment around 18 months after the baseline assessment (first follow-up), around 36 months after the baseline assessment (second follow-up), and around 54 months after the baseline assessment (third follow-up). In the current study, the sample comprises those 131 participants who completed the third follow-up assessment about 5 years after the baseline assessment without conversion to dementia (mean = 58.78 mo from the baseline assessment; SD = 8.22) and 38 participants who converted to dementia during this period (mean = 41.79 mo from the baseline assessment; SD = 24.35). Diagnosis of probable AD or other types of dementia was made according to the criteria used by NINCDS-ADRDA and *DSM-IV-TR* criteria. Conversion to dementia was confirmed by consultation of the medical history and considering the date of neurological diagnosis.

2.5 | Data analysis

Mann-Whitney *U* and Pearson chi-squared tests were run in SPSS for descriptive analysis and group comparisons.[32] Supervised ML techniques were also used to test the value of CR proxies for predicting conversion to dementia, with the dichotomous variable converter-nonconverter as the target variable. For this purpose, we started with a data set consisting of 169 instances of the total number of participants and 14 input variables or features (including diagnosis at baseline assessment, months from the baseline assessment until the third follow-up assessment or until conversion to dementia, gender, age, occupational status, Charlson Comorbidity Index, history of memory disorders, years of formal schooling, occupational attainment, reading habits, social activities, cultural activities, and Peabody test and WAIS vocabulary subtest scores), which enabled us to predict the target variable. The data were analysed using the Scikit-Learn[33] ML library, written in Python programming language.

The ML algorithms selected for training the system are all widely used in medical research [18-20] (a) LR is based on a multinomial LR model with a ridge estimator; (b) support vector machines (SVMs) construct a hyper-plane or set of hyper-planes in a high or infinite dimensional space used for classification; (c) the nearest neighbours method assigns unclassified sample data to the nearest of a set of previously classified data; (d) Gaussian naive Bayes is based on classification mechanisms that use Gaussian processes; (e) ensemble methods combined the predictions of several base estimators built by averaging or boosting methods, including the random forest, extra trees, gradient boosting, and gradient boosting classifiers; and finally, (f) artificial neural network implements a multi-layer perceptron (MLP) algorithm that classifies by using the cross-entropy loss function. Finally, we used *k*-fold cross-validation ($k = 10$) to test and evaluate the aforementioned algorithms. In this process, the data set is divided into *k* subsets. For each analysis, one of the *k* subsets is used as the test set, and the other $k - 1$ subsets form the training set. Performance statistics are calculated across all *k* trials. This is an effective means of preventing overfitting, and it also provides a good indication of how well the classifier will perform with unseen data (for more information, see Appendix S1).

For each of the above algorithms, performance metrics were computed in order to assess the prediction obtained. Again, the following metrics are widely used in ML approaches: (a) Accuracy is calculated as

the proportion of true-positive and true-negative predictions in all cases evaluated; (b) precision is a measure that estimates the probability that a positive prediction is correct; (c) recall is the proportion of instances belonging to the positive class that are correctly predicted as positive; (d) the F1 score indicates a weighted harmonic average of precision and recall; (e) Cohen kappa is a measure of the validity of the classification when confronted with a random result; (f) the receiver operating characteristic (ROC) curve is a plot of true-positive rate versus false-positive rate; and finally (g) the precision-recall curve area indicates the graphical relationship between precision and recall. Finally, we computed the ranking of the importance of the features in order to determine which input variables contribute most to predicting converter and nonconverter participants. For this purpose, we applied the feature selection functions supported by the ensemble ML algorithms (for further information, see Appendix S1).

3 | RESULTS

3.1 | Differences in demographic, health, and CR differences between converters and nonconverters

Comparisons between converters and nonconverters in relation to demographic, health status, and CR proxies are shown in Table 2. Significant differences were found for age, occupational status, WAIS vocabulary subtest and Peabody scores, and diagnostic group at baseline. The converters were older, and most of them were retired. In addition, their vocabulary scores were lower than those of nonconverters, and a higher percentage of converters was diagnosed with sda-MCI and ma-MCI.

3.2 | Machine learning classification of converters and nonconverters

The data were processed using the nine ML algorithms outlined above. The metrics included in Table 3, ordered by the highest to lowest accuracy, show the classification abilities exploring the role of demographic, health, and CR proxies in the conversion from MCI to dementia. The mean accuracy obtained for all algorithms exceeded 0.79. The gradient boosting and random forest performed best, with an accuracy greater than 0.90. The F1 score weighted harmonic average of precision and recall exceeded 0.79 for the gradient boosting, random forest, and AdaBoost classifiers. Finally, the Cohen kappa measure of the validity of the classification without random influence was greater than 0.71 for gradient boosting, AdaBoost, and random forest classifiers.

Following accuracy and precision outcomes, as the best predictions of conversion or nonconversion were obtained from the ensemble of gradient boosting and random forest, we selected both algorithms. Accordingly, we developed precision-recall curves and ROC curves for these algorithms, as an indication of classification quality. We computed the precision-recall curve, in which a larger area below the curve indicates both greater precision and higher recall (see Figure 1). We also computed the ROC curve to assess the performance of the classifier. Both have an excellent average area under the curve (AUC) (mean ROC AUC 0.93 for gradient boosting and 0.92. for random forest) (see Figure 2).

Finally, both algorithms were used to compute the importance of the features in the predictive models. In both methods, the following six features were the most important: months from the first assessment, age, diagnostic group at baseline, WAIS vocabulary score, Peabody test score, and number of years of formal schooling (Table 4).

4 | DISCUSSION

Innovative approaches based on ML techniques were used to investigate the role of CR in conversion from MCI to dementia. In order to predict conversion and nonconversion to dementia, we applied ML algorithms to data from a sample of 169 participants. The clinical procedure identified a total of 38 participants who converted from MCI to dementia. In parallel, we used a 10-fold stratified cross-validation procedure to analyse the classification results from the nine ML algorithms. The best results were obtained with ensemble methods (gradient boosting classifier, random forest classifier) using accuracy as the index parameter, as

TABLE 2 Demographic, health status, and cognitive reserve information at baseline according to converter or nonconverter groups, and group comparisons (Mann-Whitney *U* test and Pearson chi-squared test, where applicable)

	Converters	Nonconverters	Group Comparisons
Age (y)	74.03 (5.88) Range, 62-87	65.09 (8.01) Range, 50-87	$Z = -5.96^{**}$
Education (y)	9.92 (4.39) Range, 4-25	9.47 (4.65) Range, 1-22	$Z = -0.76$
Occupational status	Employed, 54 (41.2%) Unemployed, 3 (2.3%) Retired, 74 (56.5%)	Employed, 3 (7.9%); Retired, 35 (92.1%)	$\chi^2 = 16.36^{**}$
Co-morbidity	0.80 (0.87) Range, 0-4	0.66 (0.78) Range, 0-2	$Z = -0.840$
History of memory disorders	Yes, 67 (51.1%) No, 64 (48.9%)	Yes, 15 (39.5%) No, 23 (60.5%)	$\chi^2 = 1.61$

(Continues)

TABLE 2 (Continued)

	Converters	Nonconverters	Group Comparisons
Occupational attainment	Unqualified worker, 54 (41.2%); housewife, 23 (17.6%); qualified worker, 39 (29.8%); middle-ranking manager, 13 (9.9%); high ranking or manager, 2 (1.5%)	Unqualified worker, 21 (55.3%); housewife, 5 (13.2%); qualified worker, 7 (18.4%): middle-ranking manager, 5 (13.2%)	$\chi^2 = 3.92$
Reading habits	Never, 16 (12.2%); once a week, 21 (16%); twice a week, 20 (15.3%); every day, 74 (56.5%)	Never, 6 (15.8%); once a week, 4 (10.5%); twice a week, 7 (18.4%); every day, 21 (55.3%)	$\chi^2 = 1.08$
Frequency of social activities	Never, 42 (32.1%); rarely, 31 (23.7%); often, 29 (22.1%); always, 29 (22.1%)	Never, 14 (36.8%); rarely, 11 (28.9%); often, 10 (26.3%); always, 3 (7.9%)	$\chi^2 = 3.91$
Frequency of cultural activities	Never, 92 (70.2%); rarely, 10 (7.6%); often, 11 (8.4%); always, 18 (13.7%)	Never, 26 (68.4%); rarely, 4 (10.5%); often, 3 (7.9%), always; 5 (13.2%)	$\chi^2 = 0.328$
WAIS vocabulary score	44.18 (13.01) Range, 15-71	50.14 (12.97) Range, 18-75	$Z = -2.20$*
Peabody test score	56.39 (17.52) Range, 23-87	62.89 (17.42) Range, 7-94	$Z = -2.17$*
Diagnostic group at baseline	Controls, 93 (71%); sda-MCI, 20 (15.3%); na-MCI, 11 (8.4%); ma-MCI, 7 (5.3%)	Controls, 8 (21%); sda-MCI, 12 (31.6%); na-MCI, 3 (7.9%); ma-MCI, 15 (39.5%)	$\chi^2 = 46.04$**

Abbreviations: ma-MCI, multidomain amnestic mild cognitive impairment; na-MCI, non amnestic mild cognitive impairment; sda-MCI, single-domain amnestic mild cognitive impairment; WAIS, Wechsler Adult Intelligence Scale.

**$P < 0.01$.

*$P < 0.05$.

TABLE 3 Metrics for prediction of converter and nonconverter participants

Algorithms	Accuracy ↓	Precision	Recall	F1 Score	κ (Cohen Kappa)
Gradient boosting classifier	0.93	0.88	0.88	0.86	0.82
Random forest classifier	0.92	0.86	0.73	0.79	0.71
AdaBoost classifier	0.90	0.82	0.78	0.79	0.73
Extra trees classifier	0.86	0.79	0.55	0.64	0.52
Support vector machine	0.85	0.73	0.60	0.64	0.54
Logistic regression	0.84	0.74	0.39	0.49	0.42
Gaussian naive Bayes	0.84	0.65	0.79	0.69	0.58
Artificial neural network	0.82	0.71	0.48	0.50	0.46
Nearest neighbour classifier	0.79	0.56	0.41	0.43	0.33

this indicates how close the value is to the true value of the quantity being measured.

The present innovative approach, based on the application of ML algorithms to socio-demographic data, basic health status, and CR proxies, enabled prediction of conversion to dementia, with good precision and sensitivity (ie, F1 close to the maximum value). The gradient boosting classifier and random forest classifier algorithms yielded a high rate of correct prediction of converters and nonconverters (accuracy higher than 91% and precision higher than 86%). According to the individual values of the metrics used in this study, the precision was slightly higher than recall, which indicates that the true-positive rate is slightly higher, taking into account only the predictions of conversion from MCI to dementia. Moreover, these results are robust with respect to a random classification according to the Cohen κ obtained, considered almost perfect, with a value above 0.81.[34] Furthermore, both graphical curves computed (precision-recall curve and ROC curve) show the high accuracy of classification.

As gradient boosting and random forest algorithms provided the best prediction, we computed the ranking of the importance of different variables for both. In both methods, a set of six variables was the most important for achieving a good prediction; ie, months from the first assessment, age, diagnostic group at baseline, WAIS vocabulary score, Peabody test score, and number of years of formal schooling were the most important input variables for detecting

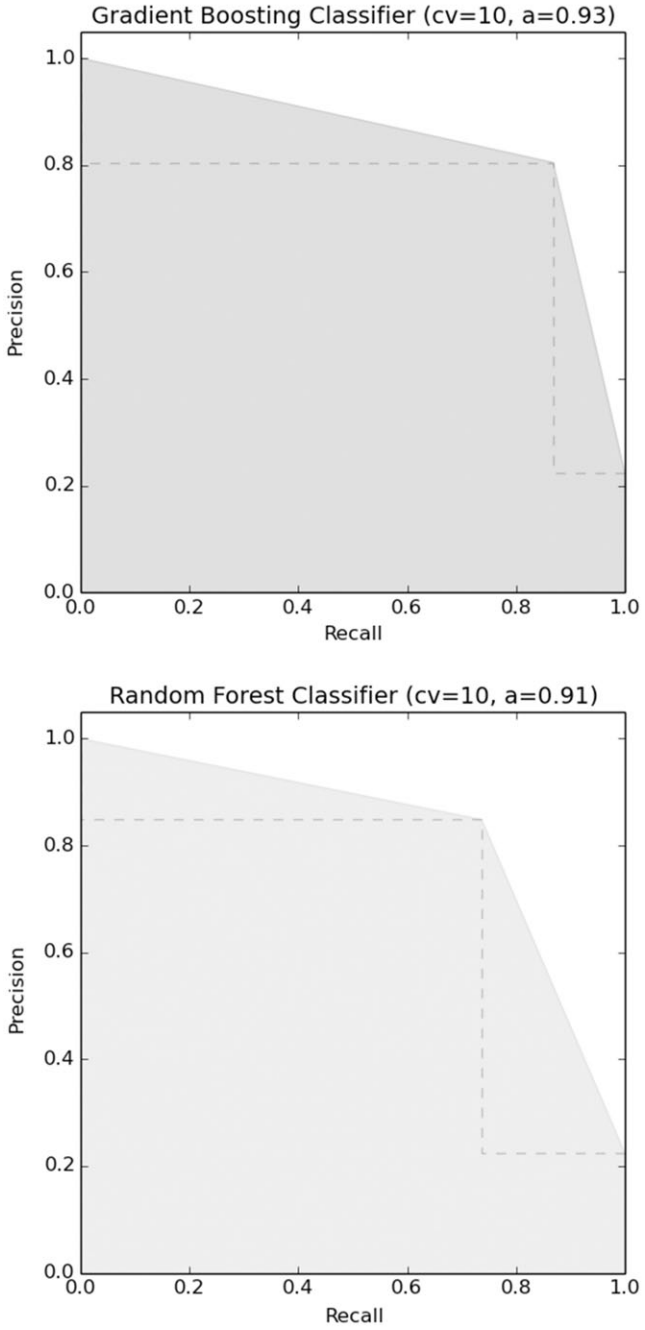

FIGURE 1 Comparison of precision-recall curve metrics for evaluating classifier output for the gradient boosting and random forest algorithms

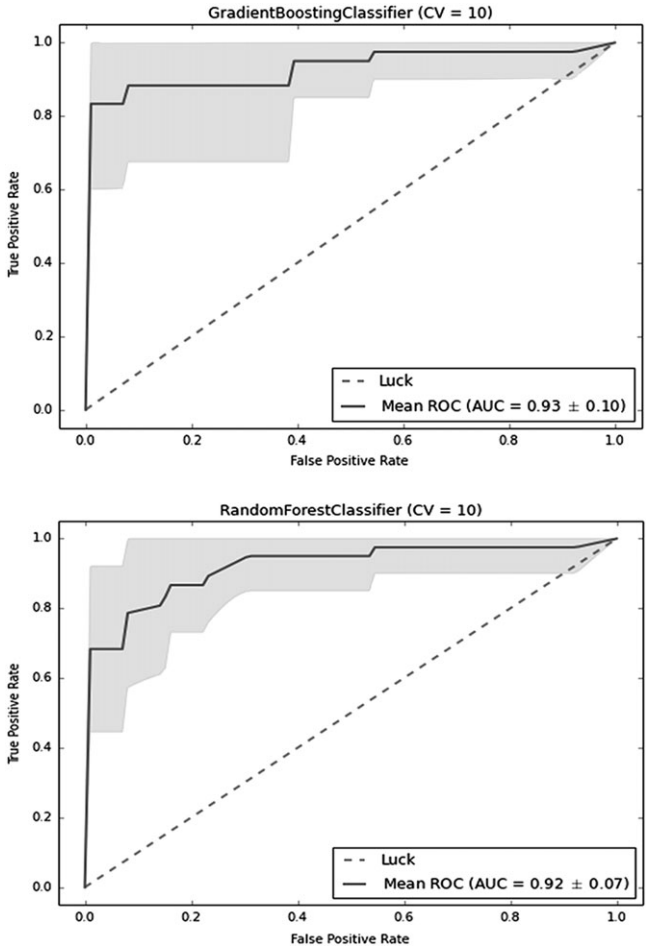

FIGURE 2 Comparison of ROC curve for evaluating the classification performance of the best two algorithms, using cv-fold cross-validation (cv = 10) and a binary classification (ie, converters and nonconverters). AUC indicates area under the curve; cv, cross-validation; ROC, receiver operating characteristic

conversion from MCI and AD. Accordingly, two diagnostic variables (the diagnosis itself and the months from the diagnosis to the final follow-up or conversion to dementia), chronological age, and three CR proxies were the most important for predicting conversion to dementia using both gradient boosting and random forest classifiers. The variables years of formal schooling and vocabulary knowledge are both related to the level of education. Other proxies related to the levels of education, such as occupational attainment and reading habits (the latter only in the random forest classifier), also add some predictive value to the models, although they are lower in the ranking of importance. Variables related to lifestyle, such as social activities and cultural activities, are also lower in the ranking, and cultural activities are only important for the random forest classifier. These findings are consistent with previously reported findings, as proxies related to educational level, such as years of schooling and level of vocabulary, have been shown to have greater loading in the CR construct and greater predictive value of cognitive impairment than have proxies related to lifestyle, such as social or cultural activities. Complexity in the accurate collection of data on lifestyle CR proxies has been cited as a potential improvement in research on CR.[9,35-37]

The findings obtained using ML techniques corroborated the protective role of CR previously reported in relation to the aforementioned proxies.[6,10,11,13] In a follow-up study with a time lapse of around 18 months, Lojo-Seoane et al[11] observed that CR had direct effects on episodic memory and general cognitive performance, and that it also influenced episodic memory and general cognitive performance, at both baseline and follow-up, through the availability of processing resource measures in working memory tasks. Serra et al[13] examined how CR modulates the rate of conversion from aMCI to AD in a sample

TABLE 4 Measuring the importance of different parameters with ensemble machine learning algorithms[a]

Random Forest Classifier		Gradient Boosting Classifier	
Months	0.31	Months	0.46
Age	0.20	Age	0.19
Diagnosis at baseline	0.12	WAIS vocabulary score	0.15
WAIS vocabulary score	0.07	Diagnosis at baseline	0.08
Peabody score	0.07	Peabody score	0.07
Education (y)	0.06	Education (y)	0.02
Occupational status	0.03	Social activities	0.01
Social activities	0.03	Occupational attainment	0.01
Co-morbidity	0.02	Co-morbidity	0.01
Occupational attainment	0.02	History of memory disorders	0.00
Cultural activities	0.02	Cultural activities	0.00
Reading habits	0.02	Gender	0.00
History of memory disorders	0.01	Reading habits	0.00
Gender	0.01	Occupational status	0.00

Abbreviation: WAIS, Wechsler Adult Intelligence Scale.

[a]Ranking of features, sorted by the highest to lowest weight.

of 42 participants followed up for 2 years. These researchers obtained a global enrichment index as a total score proxy of CR. They observed that participants with high CR survived longer without developing AD. These findings are consistent with the CR hypothesis[10] and support the role of CR as a mediator of conversion to dementia.

This study has some limitations. Although ML techniques are widely used in different health fields, their use in studies of cognitive aging and cognitive impairment is relatively new.[18-20] Further investigation into the validity of such algorithms in this field is therefore required, using bigger samples and novel diagnostic approaches. Likewise, our study uses a well-constructed longitudinal data set, considering the complexity of MCI diagnosis and transitions between diagnostic categories in this field.[3,38,39] Therefore, the performance metrics obtained are robust, and the algorithms used have been well tested in other similar fields.

5 | CONCLUSION

We used ML approaches to explore the role of CR in conversion from MCI to dementia. The findings indicate the potential value of ML algorithms for detecting risk of conversion to dementia in cognitive aging and CR studies. Accordingly, we validated a data set composed of 14 variables from baseline assessment variables to CR proxies. After training the system with the nine algorithms selected, the best classification results were achieved by ensemble methods (gradient boosting plus random forest classifiers). Moreover, some CR proxies were among the most important variables for predicting conversion to dementia, adding evidence to the increasing literature about the CR hypothesis. Further research is required to develop a robust procedure ML-based approach for making robust predictions about diagnostic transitions in cognitive impairment.

ACKNOWLEDGEMENTS

This work was financially supported by the Spanish Directorate General of Scientific and Technical Research (Project PSI2014-55316-C3-1-R and PSI2017-89389-C2-1-R) and by the Galician Government (Consellería de Cultura, Educación e Ordenación Universitaria, Ref: ED431C-2017/27; Axencia Galega de Innovación Rede Galega de Investigación en Demencias REGIDEM, http://regidem.imaisd.es).

CONFLICT OF INTEREST

None declared.

ORCID

David Facal https://orcid.org/0000-0001-6230-0090

REFERENCES

1. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14(4):535-562.
2. Cui Q, Zhang Q, Takahashi H. The neural mechanism of encountering misjudgment by the justice system. *PLoS ONE*. 2013;8(9):e75434. https://doi.org/10.1371/journal.pone.0075434
3. Facal D, Guàrdia-Olmos J, Juncos-Rabadán O. Diagnostic transitions in mild cognitive impairment by use of simple Markov models. *Int J Geriatr Psychiatry*. 2015;30(7):669-676. https://doi.org/10.1002/gps.4197
4. Michaud TL, Su D, Siahpush M, Murman DL. The risk of incident mild cognitive impairment and progression to dementia considering mild cognitive impairment subtypes. *Dement Geriatr Cogn Dis Extra*. 2017;7(1):15-29. https://doi.org/10.1159/000452486
5. Chapko D, McCormack R, Black C, Staff R, Murray A. Life-course determinants of cognitive reserve (CR) in cognitive aging and dementia—a systematic literature review. *Aging Ment Health*. 2017;0(0):1-12. https://doi.org/10.1080/13607863.2017.1348471
6. Ferreira D, Bartrés-Faz D, Nygren L, et al. Different reserve proxies confer overlapping and unique endurance to cortical thinning in healthy middle-aged adults. *Behav Brain Res*. 2016; 311:375-383.
7. Garibotto V, Borroni B, Kalbe E, et al. Education and occupation as proxies for reserve in aMCI converters and AD: FDG-PET evidence. *Neurol*. 2008;71(17):1342-1349. https://doi.org/10.1212/01.wnl.0000327670.62378.c0
8. Giogkaraki E, Michaelides MP, Constantinidou F. The role of cognitive reserve in cognitive aging: results from the neurocognitive study on aging. *J Clin Exp Neuropsychol*. 2013;35(10):1024-1035. https://doi.org/10.1080/13803395.2013.847906
9. Lojo-Seoane C, Facal D, Guardia-Olmos J, Juncos-Rabadan O. Structural model for estimating the influence of cognitive reserve on cognitive performance in adults with subjective memory complaints. *Arch Clin Neuropsychol*. 2014;29(3):245-255. https://doi.org/10.1093/arclin/acu007
10. Stern Y. Cognitive reserve. *Neuropsychologia*. 2009;47(10):2015-2028. https://doi.org/10.1016/j.neuropsychologia.2009.03.004
11. Lojo-Seoane C, Facal D, Guàrdia-Olmos J, Pereiro A, Juncos-Rabadán O. Effects of cognitive reserve on cognitive performance in a

follow-up study in older adults with subjective cognitive complaints. The role of working memory. *Front Aging Neurosci*. 2018;10:189.

12. Sumowski J, Rocca M, Leavitt V, et al. Brain reserve and cognitive reserve protect against cognitive decline over 4.5 years in. *Am Acad Neurol*. 2014;(2):4-5.
13. Serra L, Musicco M, Cercignani M, et al. Cognitive reserve and the risk for Alzheimer's disease: a longitudinal study. *Neurobiol Aging*. 2015;36(2):592-600. https://doi.org/10.1016/j.neurobiolaging.2014.10.010
14. Ding D, Zhao Q, Guo Q, et al. Progression and predictors of mild cognitive impairment in Chinese elderly: a prospective follow-up in the Shanghai Aging Study. *Alzheimer's Dement Diagnosis. Assess Dis Monit*. 2016;4:28-36.
15. Cloutier S, Chertkow H, Kergoat M-J, Gauthier S, Belleville S. Patterns of cognitive decline prior to dementia in persons with mild cognitive impairment. *J Alzheimers Dis*. 2015;47(4):901-913.
16. Han JW, Kim TH, Lee SB, et al. Predictive validity and diagnostic stability of mild cognitive impairment subtypes. *Alzheimers Dement*. 2012;8(6):553-559.
17. Abu-Mostafa YS, Magdon-Ismail M, Lin H-T. *Learning from Data*. 4 NY, USA: AMLBook New York; 2012.
18. Lehmann C, Koenig T, Jelic V, et al. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). *J Neurosci Methods*. 2007;161(2):342-350. https://doi.org/10.1016/j.jneumeth.2006.10.023
19. Tripoliti EE, Fotiadis DI, Argyropoulou M, Manis G. A six stage approach for the diagnosis of the Alzheimer's disease based on fMRI data. *J Biomed Inform*. 2010;43(2):307-320.
20. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, de Mendonça A. 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 Res Notes*. 2011;4(1):299.
21. Patel MJ, Andreescu C, Price JC, Edelman KL, Reynolds CF, Aizenstein HJ. Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. *Int J Geriatr Psychiatry*. 2015;30(10):1056-1067. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4683603/
22. Juncos-Rabadán O, Pereiro AX, Facal D, et al. Prevalence and correlates of cognitive impairment in adults with subjective memory complaints in primary care centres. *Dement Geriatr Cogn Disord*. 2012;33(4):226-232.
23. Albert MS, DeKosky ST, Dickson D, et al. 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. *Alzheimers Dement*. 2011;7(3):270-279. https://doi.org/10.1016/j.jalz.2011.03.008
24. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med*. 2004;256(3):183-194. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2796.2004.01388.x
25. Delis DC, Kramer JH, Kaplan E, Thompkins BAO. In: Psychological Corporation 1987, ed. *CVLT: California Verbal Learning Test-Adult Version: Manual*. Psychologi. San Antonio, Texas: Psychological Corporation; 1987.
26. Benedet MJ, Alejandre MÁ. TAVEC: Test de Aprendizaje Verbal España-Complutense. Tea Madrid; 1998.
27. Conde-Sala JL, Garre-Olmo J, Vilalta-Franch J, et al. Predictors of cognitive decline in Alzheimer's disease and mild cognitive impairment using the CAMCOG: a five-year follow-up. *Int Psychogeriatr*. 2012;24(6):948-958. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/9C6D2CBAAE51305220F3EDBB0F167DF1/S1041610211002158a.pdf/predictors_of_cognitive_decline_in_alzheimers_disease_and_mild_cognitive_impairment_using_the_camcog_a_fiveyear_followup.pdf
28. Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol*. 2007;6(8):734-746.
29. Jones RN, Manly J, Glymour MM, Rentz DM, Jefferson AL, Stern Y. Conceptual and measurement challenges in research on cognitive reserve. *J Int Neuropsychol Soc*. 2011;17(4):593-601. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/C634271A2EEB94093997398182A4EE28/S1355617710001748a.pdf/div-class-title-conceptual-and-measurement-challenges-in-research-on-cognitive-reserve-div.pdf
30. Wechsler D. *WAIS-III, Wechsler Adult Intelligence Scale: Administration and Scoring Manual*. Circle Pines, MN: Psychological Corporation; 1997.
31. Dunn LM, Dunn LM. *Peabody Picture Vocabulary Test, Revised: Forms L and M*. Circle Pines, MN: American Guidance Service; 1981.
32. Sanchez-Villegas A, Schlatter J, Ortuno F, et al. 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*. 2008;8(1):43.
33. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
35. Fritsch T, Smyth KA, McClendon MJ, et al. Associations between dementia/mild cognitive impairment and cognitive performance and activity levels in youth. *J Am Geriatr Soc*. 2005;53(7):1191-1196.
36. Juncos-Rabadán O, Pereiro AX, Facal D, Reboredo A, Lojo-Seoane C. Do the Cambridge Neuropsychological Test Automated Battery episodic memory measures discriminate amnestic mild cognitive impairment? *Int J Geriatr Psychiatry*. 2014;29(6):602-609.
37. Solé-Padullés C, Bartrés-Faz D, Junqué C, et al. Brain structure and function related to cognitive reserve variables in normal aging, mild cognitive impairment and Alzheimer's disease. *Neurobiol Aging*. 2009;30(7):1114-1124.
38. del Carmen Díaz-Mardomingo M, García-Herranz S, Rodríguez-Fernández R, Venero C, Peraita H. Problems in classifying mild cognitive impairment (MCI): one or multiple syndromes? *Brain Sci*. 2017;7(9). https://doi.org/10.3390/brainsci7090111
39. Fang ML, Coatta K, Badger M, et al. Informing understandings of mild cognitive impairment for older adults: implications from a scoping review. *J Appl Gerontol*. 2017;36(7):808-839. https://doi.org/10.1177/0733464815589987

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Facal D, Valladares-Rodriguez S, Lojo-Seoane C, Pereiro AX, Anido-Rifon L, Juncos-Rabadán O. Machine learning approaches to studying the role of cognitive reserve in conversion from mild cognitive impairment to dementia. *Int J Geriatr Psychiatry*. 2019;1–10. https://doi.org/10.1002/gps.5090