International Journal of Medical Informatics 170 (2023) 104966

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Personalized screening and risk profiles for Mild Cognitive Impairment via a Machine Learning Framework: Implications for general practice

Maria Basta [a], Nicholas John Simos [b,*], Maria Zioga [a], Ioannis Zaganas [a], Simeon Panagiotakis [c], Christos Lionis [a], Alexandros N Vgontzas [a]

[a] School of Medicine, University of Crete, Heraklion, Crete, Greece
[b] Foundation of Research and Technology, Heraklion, Crete, Greece
[c] Internal Medicine Department, Heraklion University Hospital, Heraklion, Crete, Greece

ARTICLE INFO

Keywords:
Random Forest
Age-related cognitive impairment
Dementia
Mini-mental state examination
Model-agnostic analysis
Model explainability

ABSTRACT

Objectives: Diagnosis of Mild Cognitive Impairment (MCI) requires lengthy diagnostic procedures, typically available at tertiary Health Care Centers (HCC). This prospective study evaluated a flexible Machine Learning (ML) framework toward identifying persons with MCI or dementia based on information that can be readily available in a primary HC setting.

Methods: Demographic and clinical data, informant ratings of recent behavioral changes, self-reported anxiety and depression symptoms, subjective cognitive complaints, and Mini Mental State Examination (MMSE) scores were pooled from two aging cohorts from the island of Crete, Greece (N = 763 aged 60–93 years) comprising persons diagnosed with MCI (n = 277) or dementia (n = 153), and cognitively non-impaired persons (CNI, n = 333). A Balanced Random Forest Classifier was used for classification and variable importance-based feature selection in nested cross-validation schemes (CNI vs MCI, CNI vs Dementia, MCI vs Dementia). Global-level model-agnostic analyses identified predictors displaying nonlinear behavior. Local level agnostic analyses pinpointed key predictor variables for a given classification result after statistically controlling for all other predictors in the model.

Results: Classification of MCI vs CNI was achieved with improved sensitivity (74 %) and comparable specificity (73 %) compared to MMSE alone (37.2 % and 94.3 %, respectively). Additional high-ranking features included age, education, behavioral changes, multicomorbidity and polypharmacy. Higher classification accuracy was achieved for MCI vs Dementia (sensitivity/specificity = 87 %) and CNI vs Dementia (sensitivity/specificity = 94 %) using the same set of variables. Model agnostic analyses revealed notable individual variability in the contribution of specific variables toward a given classification result.

Conclusions: Improved capacity to identify elderly with MCI can be achieved by combining demographic and medical information readily available at the PHC setting with MMSE scores, and informant ratings of behavioral changes. Explainability at the patient level may help clinicians identify specific predictor variables and patient scores to a given prediction outcome toward personalized risk assessment.

1. Background and significance

Whereas dementia is the most common of the age-related degenerative diseases in modern societies, milder forms of neurocognitive decline, such as Mild Cognitive Impairment (MCI), are more prevalent and have a significant impact on the wellbeing of affected persons and their families [1]. A comprehensive neuropsychiatric and neuropsychological evaluation, available at tertiary health care facilities, is typically required to set the diagnosis of MCI [2]. Consequently, MCI is highly underdiagnosed especially among persons with low education and living in rural areas [3]. In this setting, primary healthcare centers (PHCs) are the first point of contact for most elderly. General practitioners, PHC nurses and other personnel could be trained to recognize early signs and symptoms and administer cognitive screening

* Corresponding author at: Computational Bio-Medicine Laboratory, Institute of Computer Science, Foundation for Research and Technology–Hellas (FORTH), Nikolaou Plastira 100, P.O Box 1385, Vassilika Vouton, 70013 Heraklion, Crete, Greece.
E-mail addresses: nicholasjohnsimos@gmail.com (N. John Simos), lionis@med.uoc.gr (C. Lionis), avgontzas@pennstatehealth.psu.edu (A.N. Vgontzas).

https://doi.org/10.1016/j.ijmedinf.2022.104966
Received 27 September 2022; Received in revised form 7 December 2022; Accepted 12 December 2022
Available online 16 December 2022

instruments, such as the Mini Mental State Examination (MMSE), which has adequate diagnostic accuracy in identifying persons with dementia but poor performance in discriminating persons with MCI from cognitively non-impaired elderly [4]. In clinical practice additional variables are usually considered toward MCI diagnosis, such as physical and psychiatric comorbidities (more importantly late-onset depression and overall emotional status), polypharmacy and drug interactions. Changes in cognition are usually accompanied by behavioral symptoms (such as depressed mood, irritability, or subtle changes in personality) [5]. Cognitive and behavioral difficulties are often under- or overestimated by persons with MCI highlighting the importance of informant reports as diagnostic aids [6].

Machine Learning models (ML) and Artificial Intelligence (AI) are inherently suitable to address the challenges posed by multi-modal datasets, including: (i) scalability, (ii) high dimensionality, (iii) heterogeneity and complexity and (iv) distribution of the data [7,8]. A key advantage of these methods is their ability to automate the process of hypothesis generation and evaluation, in comparison to conventional statistical approaches. Ensemble models have been used extensively for feature selection and optimal classification of patients into MCI and cognitively non-impaired groups, as well as for identifying persons with MCI at higher risk for developing dementia. Recent *meta*-analyses report average accuracies ranging between 60 and 98 % based on combinations of neuropsychological test results [9] and 70–80 % on the basis of neuroimaging biomarkers [10]. However very few studies have combined clinical (e.g., physical comorbidities, polypharmacy), informant-based neuropsychiatric manifestations, patient self-reported cognitive and emotional symptoms, and patient scores on brief cognitive screening tools (e.g., MMSE) in order to support health care professionals in the PHC setting to achieve early detection of MCI.

Although ML and AI are becoming well-established tools in various areas of healthcare research, explainability, transparency and, most importantly, accountability and responsibility are often overlooked. Especially in healthcare-related topics, the ability to understand why and/or how a particular model reached specific predictions is of paramount importance. Without attempting to investigate models further, the possibility of erroneous results, overfitting or fitting using spurious and unimportant features and characteristics is increased. Explainable AI (XAI) provides computational tools to improve understanding of underlying mechanisms driving the results of ML-based classification [11–14]. These tools are often applied in the form of model-agnostic analyses conducted on already developed and tested models, of varying types of underlying estimators, to produce explainability/interpretability profiles on a global or subject level. Particularly in rural regions where access to relevant healthcare specialties (e.g., neurologists, psychiatrists, neuropsychologists) is limited, decision-support algorithms that could be used by specially trained PHC practitioners to estimate risk for neurocognitive disorders may be valuable in the context of early detection efforts. The present work addresses this problem with data from two elderly cohorts from the island of Crete, Greece where our group has previously reported very high rates of underdiagnosis of neurocognitive impairment among community-dwelling persons over 60 years of age [3] despite local prevalence rates in the upper limits of the range found in other European countries (i.e., 10.8 % for dementia of any type and 32.4 % for MCI).

2. Objectives

The first aim of this study was to assess the overall accuracy of a flexible ML framework toward differentiating persons with MCI from cognitively non-impaired elderly and from patients with dementia, based on information that can be readily available in a PHC setting. This information includes sociodemographic, clinical, self-reported symptoms of anxiety and depression, and informant-rated behavioral changes in daily life.

A secondary aim was to apply model-agnostic explainability/interpretability analyses to aid interpretation of prediction results in evaluating the classification process and to further investigate the contribution of specific predictor variables to the classification results. Specifically, model agnostic analyses were applied: (i) at the global (population-specific) level to help clarify which features are most significant for this comparison and how they contribute toward model decisions and, (ii) at the local (i.e., person-specific) level to identify predictor variables of primary importance for a particular clinical prediction.

3. Methods

3.1. Participants

Data for the ML modeling were derived from two cohorts of community-dwelling adults aged > 60 years from the island of Crete, Greece: (a) The Cretan Aging Cohort (CAC, n = 506 [3]) recruited from 13 PHC centers mainly in rural regions, and (b) the SKEPSI cohort (n = 257) of self-referred urban-dwelling participants [15]. Both cohorts were initially tested during 2013–2014 as part of prospective studies on aging. Participants from CAC were initially screened using MMSE and referred for comprehensive neuropsychological and neuropsychiatric examination if they scored < 24. To ensure that a representative sample of cognitively non-impaired elders were examined, 181 participants who scored in the low-risk range on MMSE (≥24points) were also included. Participants from the SKEPSI cohort either responded to advertisements in local media inviting persons aged 50 years or older to be tested for "memory and other cognitive difficulties they may be experiencing" or were referred for neuropsychological testing by local physicians. Assessments were performed, and clinical diagnosis was reached, using identical instruments and procedures in both cohorts by the same group of experts (neurologists, gerontologists, psychiatrists, and neuropsychologists) [3]. The final sample of 763 persons included 277 persons meeting formal clinical criteria for MCI, 153 persons with dementia, and 333 cognitively non-impaired persons (CNI group). Sociodemographic characteristics of the total sample and each of the three groups are presented in Table 1. Additional details on participant recruitment and testing procedures is available in the Supplementary Material.

3.2. Predictor variables

The following 80 variables were included as predictors in all ML models (Model Set 1 and 2):

- Sociodemographic: Age and education in years, living alone, family status, residence (urban, rural), (former) occupation type (sedentary vs manual).
- Medical: Number of physical illnesses, number of major operations, polypharmacy (defined by > 4 medications).
- MMSE total score
- Informant scales: Cambridge Behavioral Inventory [16], a 45-item questionnaire providing carer ratings on the following domains: Memory/orientation/attention, Challenging Behaviors, Self-care, Motivation, Mood, Eating Behavior, Abnormal beliefs, Stereotypic Behaviors, Sleep; Mayo Fluctuations Scale [17] providing scores on four common manifestations of Lewy Body Dementia (daytime sleepiness/lethargy, excessive daytime sleep, disorganized ideas, staring into space). Item-level scores were entered in the models.
- Self-reported mental health symptoms: Total scores on the Center for Epidemiological Studies Depression Scale (CESD [18]) and State Trait Anxiety Inventory Form Y (STAI [19]), assessing symptoms of depression and anxiety, respectively.
- Presence of at least one type of cognitive complaint (episodic memory, word finding, name recall).

Table 1
Sample sociodemographic and clinical characteristics.

	Total sample	NI	MCI	Dementia
n	763	333	277	153
Age (years)	72.4 ± 9.0	67.5 ± 8.1[#]	74.6 ± 7.7[#§]	79.2 ± 6.6[$§]
Range	60–93	60–90	60–93	60–92
Education (years)	7.4 ± 4.6	9.3 ± 4.6[#]	6.1 ± 4.1[#§]	5.7 ± 3.8[$§]
Range	0–23	0–23	0–20	0–17
Women (%)	63.3	67.6[#$]	62.5[#]	55.6[$]
Marital status (%)				
Single	2.1	2.1	1.7	2.6
Married	69.5	71.3	68.8	66.4
Widowed	24.1	19.3[#$]	27.4[#]	29.6[$]
Divorced	4.4	7.3[#$]	2.1[#]	1.3[$]
Occupation (%)				
Professional	28.0	36.9[#$]	22.6[#]	15.8[$]
Agricultural/skilled worker	42.3	33.0[#$]	46.2[#]	58.6[$]
Small Business/mixed	29.7	30.1	31.3	25.7
Geographic origin (%)				
Urban	46.6	60.4[#$]	37.5[#§]	33.1[$§]
Small town	10.3	10.5	9.8	10.6
Rural	43.1	29.1[#$]	52.7[#§]	56.3[$§]
Number of physical diseases	3.4 ± 1.8	3.2 ± 1.8[#]	3.6 ± 1.8[#]	3.3 ± 1.9
Number of operations	1.5 ± 1.5	1.5 ± 1.5[$]	1.6 ± 1.7[§]	1.1 ± 1.3[$§]
Polypharmacy (%)	40.1	31.9[#$]	46.9[#]	48.0[$]

Values depict means ± 1 standard deviation, unless otherwise specified. Statistically significant differences ($p < .05$) are noted with [#] for comparisons between NI and MCI, [$] for comparisons between NI and Dementia, [§] for comparisons between MCI and AD groups. Polypharmacy: >4 medications.

Additional variables, which were included only in model Set 2 included 13 age- and education-adjusted z scores representing performance on tests of memory, visuoconstructive ability, language, attention/processing speed, and executive functions.

3.3. Machine learning pipeline

The analysis pipeline adopted to address the main objective of the study entailed preprocessing steps, feature selection, model training and testing.

3.3.1. Handling of missing values

Univariate imputation via replacement with mean was used on missing values. Missing values were never over 20 % per variable. The balanced random forest model was utilized in the present study, an improvement on the classic RF with the addition of internal random bootstrap sample undersampling in order to achieve better class balancing [20]. To avoid bias, imputation was performed inside the cross-validation loop separately for training and testing subsets.

3.3.2. ML classifier

The Balanced Random Forest Classifier (RF) was used for classification and consensus feature importance-based feature selection in a nested cross-validation scheme. RF models use bagging, an ensemble technique employed to enhance prediction accuracy and control model overfitting [21]. Ensemble or averaging ML methods combine multiple "weak" learners, often decision trees and average their predictions in order to produce the final class estimate. Individual decision trees are characterized by high variance in their estimate and do not perform well, by combining multiple complimentary simple estimators a substantially better prediction machine is created. Decision trees get their name from the tree-like structure of the set of rules that defines them. They consist of multiple splits, nodes and leaves (terminal nodes), which produce the output variables.

In the present work cross validation utilized 80/20 stratified split, externally repeated 400 times to stabilize classification metrics. Internally, a stratified 6-fold split was repeated 60 times in order to derive the top 15 most important features across all internal iterations (thus "consensus" features). This approach was utilized in previous work of our team [22,23] in order to avoid overfitting and obtain stable and reliably reproducible final features of increased significance to the particular clinical comparison [24,25]. The number of selected features in the final models was kept relatively low in order to avoid model overfitting and to aid interpretability. Furthermore, adding more features did not aid in model performance or produce noteworthy results in terms of local and global level explainability. Externally, the number of selected features (via consensus feature-importance ranking) was set to 15. In the nested CV internal loop, more features were selected to allow for the calculation of more accurate feature prevalence statistics across iterations.

3.3.3. Models tested

ML models were built to perform binary classification problems (CNI vs MCI, CNI vs Dementia, MCI vs Dementia). A multi-class classification model was also tested. Primary models only considered variables that may be available to PHC practitioners (sociodemographic and clinical information, MMSE score, severity of anxiety and depression symptoms as measured by self-reported instruments, informant-based ratings of everyday cognitive difficulties and other behavioral manifestations of age-related neurodegenerative conditions;Model Set 1). For comparison purposes, a second set of models were examined which in addition to all variables included in Model Set 1, also considered scores on the standardized neuropsychological tests routinely used in MCI and dementia diagnosis (age- and education-adjusted z scores derived from Greek population norms. Specificity, sensitivity, accuracy, precision, F-score, and AUC were used to evaluate the performance of the cross-validated model on the test set.

3.3.4. Model agnostic analysis

Model-agnostic analysis was applied to the final cross-validated models, which were trained only with the sets of variables that emerged as significant features (Model Sets 1 and 2, separately).

Global-level analyses aimed to identify variables that display nonlinear behavior as predictors (partial dependence plots [26,27]), and potentially determine clinically useful cutoff scores to aid interpretation of the results of the neuropsychological evaluation. The term "nonlinear' here refers to variable responses that demonstrated abrupt, step-type changes in predicted class membership, i.e., a limited increase in a variable's value leads to a significant change in estimated prediction value, potentially causing a change in the person's predicted class.

At the patient/local level we sought to identify predictor variables that emerge as key contributors to a given classification result after statistically controlling for all other predictors in the model [28,29]. This was possible with the use of ceteris paribus profiles [30] as well as break down plots [31], both created for individual subject predictions while utilizing the model trained on the remaining subjects.

The ceteris paribus profiles and break-down plots (local level) as well as partial dependence plots (global level) were developed using the *dalex* Python package [29], the default values in the arguments of the main function were applied. Ceteris paribus profiles (subject-specific) indicate for each valuable separately, the estimated change in prediction (continuous value that determines class membership) regarding variable value fluctuations. The Partial Dependence Plots (PDP) (and Accumulated Local Effects (ALE)) are created as group averages of individual ceteris paribus profiles, offering the same potential interpretation on a more global level.

4. Results

Results comparing the three study groups on the predictor variables using univariate analyses are presented in the Supplementary Material.

4.1. Classification results

Table 2 presents the performance indices of the two sets of the ML models. Model Set 1 displayed fair classification accuracy (74 %) on the main comparison of interest (CNI vs MCI) with balanced sensitivity and specificity estimates (74 and 73 %, respectively). Corresponding values based on MMSE alone are listed in the Supplementary Material. In Fig. 1 (upper panel) the 15 top features/predictors are presented for the CNI vs MCI comparison according to their impact on model output (Shap values), i.e., their relative contribution to the final predictive model's performance, separately for each class (blue: CNI, red: MCI). MMSE and age had the highest contribution to model performance, with additional sociodemographic, behavioral, emotional and clinical variables featuring in the top 15 list. Specifically, older age, lower education, rural residence, memory problems (forgetting recent events, repetitive questioning, loss of objects, forgetting what day it is), behavioral manifestations (rigid thinking), emotional difficulties (self-reported symptoms of anxiety and depression), multicomorbidities and polypharmacy emerged as important features in correctly identifying persons with MCI in comparison to cognitively non-impaired persons.

Classification accuracy was considerably higher for the secondary comparison (MCI vs Dementia: accuracy = 85 %, sensitivity/specificity = 87 %; CNI vs Dementia: accuracy = 94 %, sensitivity/specificity = 94 %). Several of the highest-ranking features for the former contrast (see Fig. 1, lower panel) were common to those supporting the CNI vs MCI contrast, presented in Fig. 1-upper panel (MMSE, age, education, anxiety and depression symptoms, memory problems) with some additional features (difficulties performing everyday tasks, confusion, disorganized ideas). Please note that in the upper panel of Fig. 1 (CNI vs MCI model), CNI represents class 0 and MCI class 1, while in the lower panel (MCI vs Dementia model), MCI represents class 0 while Dementia is class 1.

As expected, the second set of (comparison) models performed considerably better for the CNI vs MCI contrast (82 %, 83 %, and 81 % for overall accuracy, sensitivity, and specificity, respectively). Adding neuropsychological test scores as predictors did not improve model performance for the secondary comparisons (MCI vs Dementia, CNI vs Dementia).

4.2. Model agnostic analyses: Global level

The prominent role of some of the highest-ranking predictors (MMSE, age, education) toward discriminating CNI from MCI participants according to Model 1 is supported by the abrupt response in estimated model prediction driven by the shift of variable values in the partial dependence plots (Fig. 2).

The partial dependence plots displayed in Figs. 2 and 3 refer to all participants in each of the two groups of interest (CNI and MCI) and are averages of individual ceteris paribus plots. In this type of explainability diagram, all variables are kept fixed except for the one displayed. The final plot shows the expected change in prediction value in the 0 to 1 range (y axis) as a function of a specific variable's values. Prediction values < 0.5 indicate an increased probability of 'low' class (class ''0') membership (CNI in this case), while prediction values ≥ 0.5 indicate an increased probability of 'high' class (class '1') membership (MCI in this

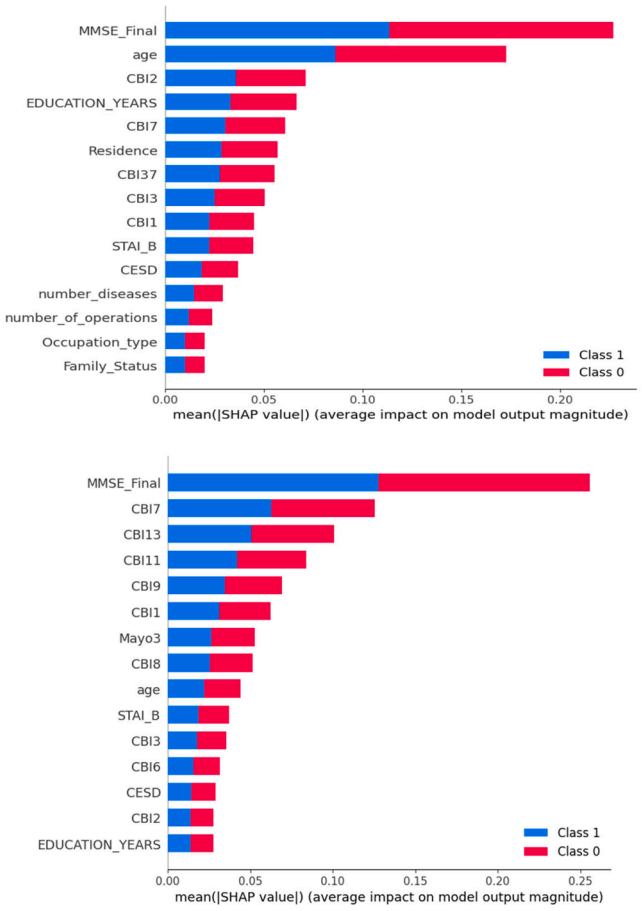

Fig. 1. Most significant features that emerged from Model type 1. Upper panel: Model discriminating between CNI (class 0) vs MCI (class 1) participants. Lower panel: Model discriminating between MCI (class 0) vs Dementia groups (class 1). Abbreviations; MMSE: total Mini Mental Status Examination score, Mayo3: Disorganized thoughts, CBI1: Forgets events that took place in the previous days (e.g., conversations, trips, etc), CBI2: Asks the same questions over and over again, CBI3: Loses things or does not remember where he/she placed them, CBI6: Has difficulty concentrating when reading or watching television, CBI7: Forgets what day it is, CBI8: Appears to be confused or "lost" in unfamiliar surroundings, CBI9: Has difficulty using electrical appliances (e.g., television, radio, stove, washing machine), CBI11: Has difficulty using the telephone, CBI13: Has difficulty handling money or paying bills, CBI37: Remains fixed in his/her ideas (even when s/he is clearly wrong), STAI_B: State Trait Anxiety Inventory Form Y (Trait Anxiety), CESD: Center for Epidemiological Studies Depression Scale total score.

Table 2
Results of the Balance RF classifier in differentiating between participants with MCI or dementia from cognitively non-impaired controls (CNI) according to two types of models (Model sets 1 and 2).

Contrast	Model set	Accuracy	Precision	Sensitivity	Specificity	F1	ROC-AUC
CNI vs MCI	1	74 ± 3	68 ± 4	74 ± 5	73 ± 5	71 ± 4	74 ± 3
CNI vs MCI	2	82 ± 3	77 ± 4	83 ± 5	81 ± 4	80 ± 3	82 ± 3
MCI vs Dementia	1	85 ± 3	90 ± 3	87 ± 5	87 ± 5	88 ± 3	84 ± 3
MCI vs Dementia	2	84 ± 4	90 ± 4	86 ± 5	86 ± 5	88 ± 3	83 ± 4
CNI vs Dementia	1	94 ± 2	86 ± 6	94 ± 5	94 ± 3	90 ± 4	94 ± 2
CNI vs Dementia	2	94 ± 2	89 ± 5	91 ± 8	96 ± 3	90 ± 4	93 ± 4

Abbreviations; MCI: Mild Cognitive Impairment, CNI: Cognitively Non-Impaired.

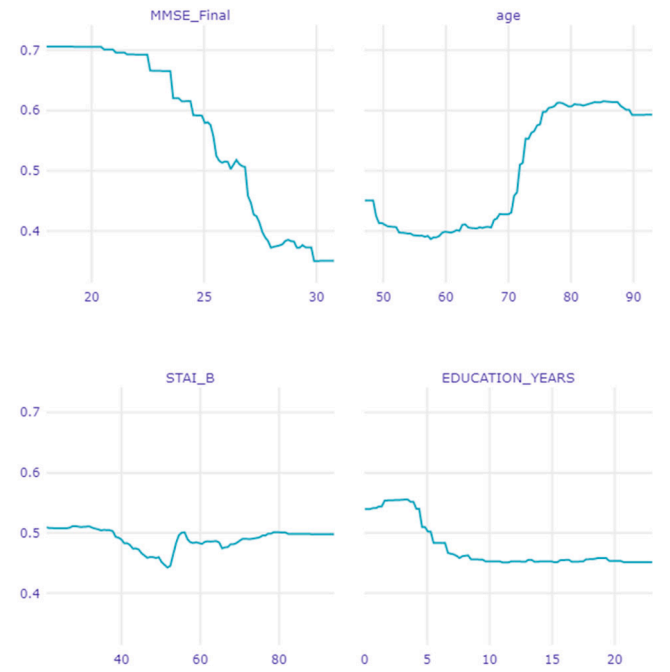

Fig. 2. Partial dependence plots at the global (group) level of the four highest-ranking predictors: CNI vs MCI (Model 1). Abbreviations; MMSE: total Mini Mental Status Examination score, STAI_B: State Trait Anxiety Inventory Form Y (Trait Anxiety).

case).

As shown in Fig. 2, MMSE scores < 23 points 'mandate' that a participant be categorized as MCI (class '1', prediction > 0.6), whereas persons with values > 24 are significantly more likely to be classified in the CNI group by the model (prediction values < 0.5). Regarding age, persons aged 75 years or older are given an increased probability for an MCI diagnosis, albeit not significant enough to drive a definitive diagnosis according to the developed model. Furthermore, more years of education potentially reduce the probability of MCI class membership. Specifically, individuals with >7 or 8 years of education have a significantly higher chance of not being characterized as suffering MCI than subjects with <4 years of education. Importantly, the partial dependence plots in Fig. 3 indicate that informant ratings as low as 1 on the 4-point CBI scale indicating even occasional presence of a given memory or behavioral manifestation is sufficient to drive the model's prediction slightly over the 0.5 class membership threshold.

4.3. Model agnostic analyses: Local level

Model agnostic analyses at the local (person-specific) level are designed to aid interpretation of a given classification decision by the model and to identify potential risk factors for a specific participant. The first example presented here (Fig. 4) is of a CNI participant, with a prediction probability of 0.03, indicating a very low probability of membership to the MCI class. Small increases in the values of informant-ratings of memory problems (CBI questions 1, 3, 7) and rigid thinking (CBI question 37) each have the potential of increasing the projected prediction by as much as 0.3 points. Considering that each individual ceteris paribus profile is calculated while keeping all other values fixed, if these values are all increased simultaneously, the model would likely produce a probability value exceeding the threshold for membership to the MCI class.

Fig. 5 illustrates a ceteris paribus plot of a person correctly classified to the MCI class with a very high probability (0.85). A change in MMSE score from (the observed) 21 to 26 points, would shift model prediction below 0.5 and thus toward the CNI class.

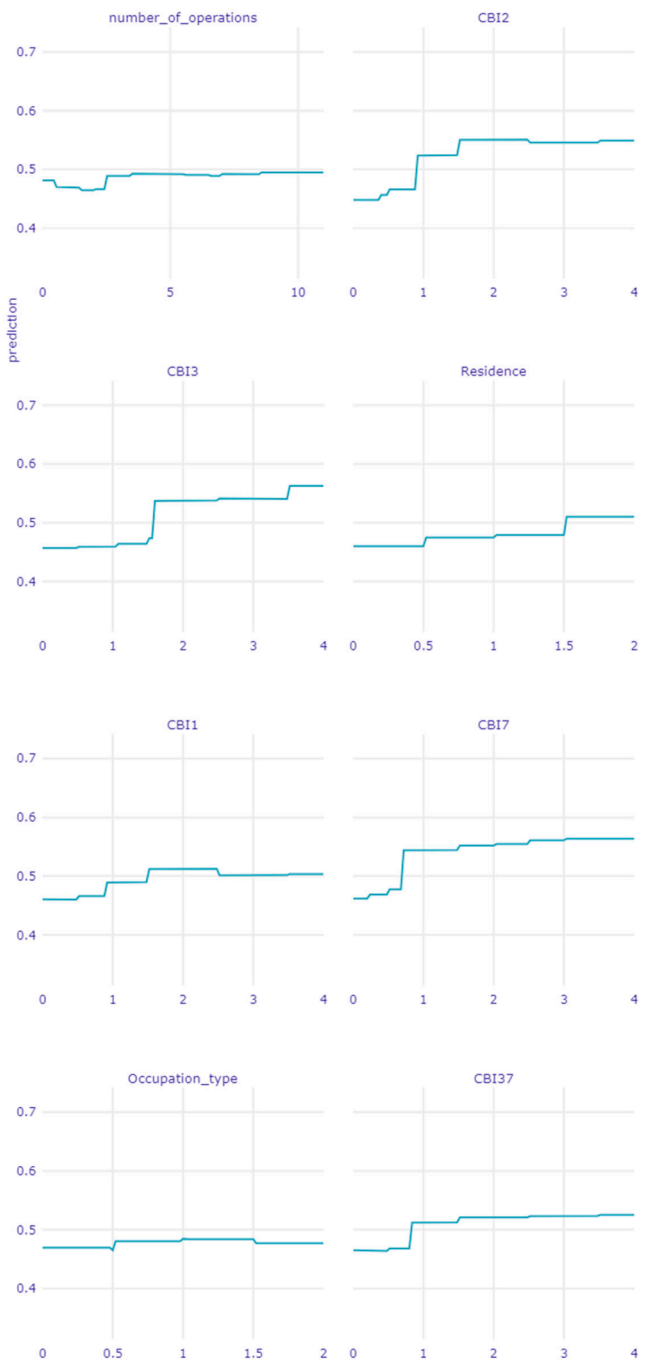

Fig. 3. Partial dependence plots at the global (group) level for Model 1 discriminating between CNI vs MCI participants: variables 5–15 among the highest-ranking predictors. Abbreviations; CBI1: Forgets events that took place in the previous days (e.g., conversations, trips, etc), CBI2: Asks the same questions over and over again, CBI3: Loses things or does not remember where he/she placed them, CBI7: Forgets what day it is, CBI37: Remains fixed in his/her ideas (even when s/he is clearly wrong).

5. Discussion

The key finding of the present study is that by considering a relatively small set of sociodemographic, clinical, brief cognitive screening, and behavioral variables it is possible to achieve fair sensitivity and specificity (74 %) in identifying persons who actually meet criteria for MCI from cognitively non-impaired elderly (Model Set 1). This was the primary contrast of interest in the present study given that diagnosis of MCI typically requires extensive cognitive/neuropsychological testing

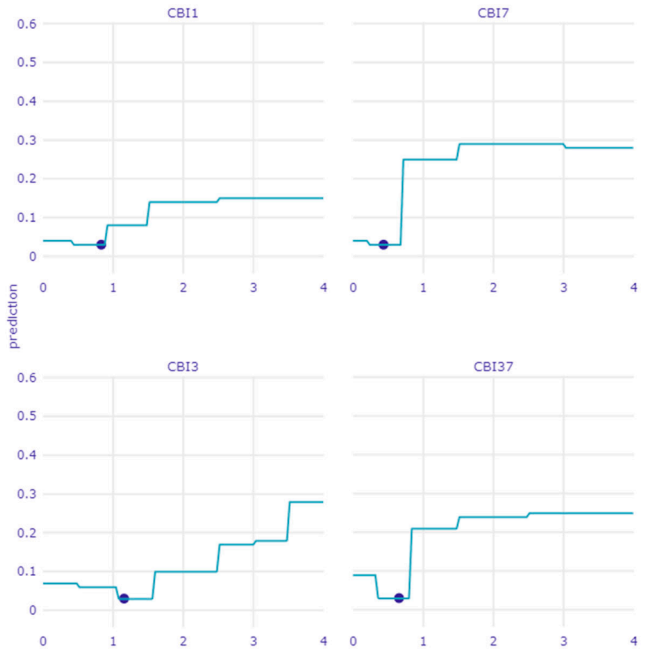

Fig. 4. Ceteris paribus profile of a CNI participant, derived from Model type 1 contrasting CNI vs MCI groups. Observed values are marked by dots. Abbreviations; CBI1: Forgets events that took place in the previous days (e.g., conversations, trips, etc), CBI3: Loses things or does not remember where he/she placed them, CBI7: Forgets what day it is, CBI37: Remains fixed in his/her ideas (even when s/he is clearly wrong).

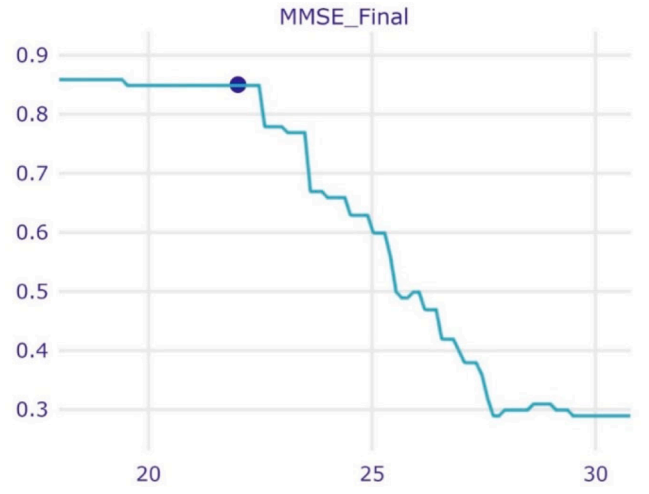

Fig. 5. Ceteris paribus profile of a MCI participant, derived from Model type 1 contrasting CNI vs MCI groups. The observed value is marked by a dot.

which is available only at tertiary HC centers. While sociodemographic and clinical variables can be readily extracted from the patient's medical record, the remaining predictor variables can be obtained from health care personnel following minimal training. Importantly, by combining informant reports on a handful of relevant questions and obtaining brief self-reported ratings of anxiety and depression it is possible to complement the results of brief cognitive screening tools (such as the MMSE) and significantly improve its sensitivity (37.2 to 41.6 % depending on the cutoff). These findings are consistent with previous literature, reporting a wide range of MMSE performance in identifying persons with MCI from cognitively non-impaired elderly (sensitivity: 18–86 %, specificity: 48–100 %) [4].

The most significant factors among those entered in the model to predict MCI vs CNI, were MMSE score (<24 points), age (>75 years) and education (<4years). Education is a well-known protective factor for dementia as indicated by large epidemiological studies has been shown to contribute in decrease of dementia rate [32]. Consistent with previous reports, this study confirmed the significance of education in predicting MCI among those with <4 years of education or CNI in those with seven or more years of education. It should be noted, however, that education may be a risk factor for neurocognitive impairment indirectly through socioeconomic status by limiting access to high quality healthcare services. While the two groups did not differ significantly on self-rated symptoms of anxiety and depression both scores were among the 15 top-ranking features in the model classifying CNI from MCI persons. Such results in the context of ML modeling often indicate that these variables are involved in more complex interactions with other variables in determining class probability. Conversely, despite the small (non-significant) tendency of persons with MCI to report cognitive complaints as compared to persons in the CNI group, this variable did not rank among the strongest predictors in the corresponding models. This finding is consistent with the notion that the probability of reporting cognitive difficulties in elderly is primarily determined by non-cognitive factors (such as emotional status and physical health [33]).

As expected, classification performance was considerably higher when applying Model type 1 to the discrimination of persons with dementia from persons diagnosed with MCI (approximately 85 %). This performance level is notable for two reasons: Firstly, for not taking into account informant-ratings of daily functionality which weighted considerably toward establishing clinical diagnosis of dementia in the present cohort. Secondly, given the relatively mild severity of dementia characterizing the present cohort. However, this performance is somewhat lower than that achieved by MMSE alone.

In view of the rapidly growing application of ML to address clinical questions and problems, it is important to stress the dangers and pitfalls of ill-designed ML models. Conceptually, the most crucial issues concern poor evaluation strategies, ineffective feature selection, and choices that can easily lead to overfitted, biased or seemingly "over-performing" models. For these reasons, we have selected a nested cross-validation strategy paired with an importance-based consensus feature selection. In addition, we carefully examined individual feature importance and prediction response over the entire range of feature values on a global level (sample-wide).

Whereas explainability at the group level may help future users to interpret classification results, routine clinical use could benefit from information pertaining to the importance of specific predictor variables and individual values to a given prediction outcome. This approach is meaningful given the high level of individual variability in relevant profiles, i.e., ceteris paribus and breakdown profiles. Aside from facilitating the in-depth study of each predictor's effect on the estimator's outcome, individualized predictor-prediction and feature importance profiles allow for personalized risk assessment and can direct treatments such as cognitive conditioning or guide the production of personalized recommendations based on the test performance and estimated outcome of each individual.

5.1. Limitations

The most notable limitations of the present results relate to the cross-sectional nature of the dataset used for model training and cross-validation and to certain sample characteristics. Regarding the former issue, MCI diagnosis, achieved following comprehensive neuropsychological and neuropsychiatric evaluation of all participants, could be considered as tentative given the likelihood of reverting to normal status at a later assessment especially in non-clinical settings [34]. Sample characteristics are also notable for the high percentage of persons residing in rural areas (43.1 %) and having <6 years of formal education (26.3 %).

6. Conclusions

The findings of this study suggest that machine learning techniques could contribute to better, faster, and simpler diagnostic procedures within PHC. These results extend previous research indicating that machine learning techniques could help optimize algorithms to improve detection of dementia and/or progression from MCI to dementia based on health records [35]. Future research that integrates additional, easy to obtain biomarkers, such as voice-derived indices of neurocognitive decline, via ensemble or artificial neural networks, are forthcoming to improve identification of persons likely to suffer from MCI and identify early markers of progression to dementia.

Clinical Relevance Statement

The current results stress the need to train GPs and PHC personnel to recognize specific risk factors and manifestations of MCI. The current results show that clinical data obtained at the PHC level could help differentiate MCI from normal cognition and dementia.

Ethical Approval

This study was approved by the institutional review board of the University Hospital of Heraklion, Crete, Greece, all participants provided written informed consent.

Funding

This study was supported by the following sources: A grant from the National Strategic Reference Framework (ESPA) 2007-2013, Thales Program, entitled “A multi-disciplinary network for the study of Alzheimer’s Disease”, and a grant from the Cross-border Cooperation Programme “Greece-Cyprus 2007-2013”, entitled: “Advanced Age: Designing a protocol for the Evaluation of Cognitive Functions and Quality of Life and Evidence-Based Interventions”, Project Acronym “SKEPSI”.

CRediT authorship contribution statement

Maria Basta: Conceptualization, Resources, Writing - original draft. **Nicholas John Simos:** Methodology, Software, Visualization, Writing - original draft. **Maria Zioga:** Methodology, Writing - review & editing. **Ioannis Zaganas:** Resources, Writing - review & editing. **Simeon Panagiotakis:** Resources, Writing - review & editing. **Christos Lionis:** Resources, Writing - review & editing. **Alexandros N Vgontzas:** Conceptualization, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2022.104966.

References

[1] K. Dean, G. Wilcock, Living with mild cognitive impairment: the patient’s and carer’s experience, Int. Psychogeriatr. 24 (6) (2012) 871–881.
[2] B. Winblad, K. Palmer, M. Kivipelto, Mild cognitive impairment–beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment, J. Intern. Med. 256 (3) (2004) 240–246.
[3] I. Zaganas, P. Simos, M. Basta, S. Kapetanaki, S. Panagiotakis, I. Koutentaki, et al., The Cretan aging cohort: cohort description and burden of dementia and mild cognitive impairment, Am. J. Alzheimer’s Dis. Other Dementias. 34 (1) (2019) 23–33.
[4] A. Mitchell, A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment, J. Psychiatr. Res. 43 (4) (2009) 411–431.
[5] E. Martin, L. Velayudhan, Neuropsychiatric symptoms in mild cognitive impairment: a literature review, Dement. Geriatr. Cogn. Disord. 49 (2) (2020) 146–155.
[6] R. Tsang, K. Diamond, L. Mowszowski, S. Lewis, S. Naismith, Using informant reports to detect cognitive decline in mild cognitive impairment, Int. Psychogeriatr. 24 (6) (2012).
[7] P. Tan, M. Steinbach, V. Kumar, Introduction to data mining, Pearson Educ India, 2016.
[8] C. Bishop, Pattern recognition and machine learning, Springer, 2006.
[9] P. Battista, C. Salvatore, M. Berlingeri, A. Cerasa, I. Castiglioni, Artificial intelligence and neuropsychological measures: The case of Alzheimer’s disease, Neurosci. Biobehav. Rev. 114 (2020) 211–228.
[10] C. Salvatore, P. Battista, I. Castiglioni, Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines, Curr. Alzheimer Res. 13 (5) (2016) 509–533.
[11] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf GZY. XAI—Explainable artificial intelligence David. Sci Robot [Internet]. 2019; (December):1. Available from: http://www.darpa.mil/program/explainable-artificial-intelligence.
[12] F.K. Dosilovic, M. Brcic, N. Hlupic, Explainable artificial intelligence: A survey. 2018 41st Int Conv Inf Commun Technol Electron Microelectron MIPRO 2018 - Proc. 2018;210–5.
[13] A. Adadi, M. Berrada, Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), IEEE Access 6 (2018) 52138–52160.
[14] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2021.
[15] E. Karademas, P. Simos, I. Zaganas, S. Tziraki, S. Panagiotakis, M. Basta, et al., The impact of mild cognitive impairment on the self-regulation process: A comparison study of persons with mild cognitive impairment and cognitively healthy older adults, J. Health Psychol. 24 (3) (2019) 351–361.
[16] G. Wear, C. Wedderburn, E. Mioshi, C. Williams-Gray, S. Mason, R. Barker, et al., The Cambridge Behavioural Inventory revised, Dement Neuropsychol. 2 (2) (2008) 102–107.
[17] T. Ferman, G. Smith, B. Boeve, R. Ivnik, R. Petersen, D. Knopman, et al., DLB fluctuations: specific features that reliably differentiate DLB from AD and normal aging, Neurology 62 (2) (2004) 181–187.
[18] K. Fountoulakis, A. Iacovides, S. Kleanthous, S. Samolis, S.G. Kaprinis, K. Sitzoglou, et al., Reliability, validity and psychometric properties of the Greek translation of the Center for Epidemiological Studies-Depression (CES-D) Scale, BMC Psychiatry 1 (2001) 1–10.
[19] C.D. Spielberger, State-Trait Anxiety Inventory, Palo Alto Consult Psychol Press, 1983.
[20] https://imbalancedlearn.org/stable/references/generated/imblearn.ensemble.BalancedRandomForestClassifier.html.
[21] L. Breiman, Random Forests, Mach. Learn. 45 (2001) 5–32.
[22] N.J. Simos, E. Kavroulakis, G.C. Manikis, G. Bertsias, E. Papadaki, K. Marias, Machine learning classification of neuropsychiatric systemic lupus erythematosus patients using resting-state fmri functional connectivity. IST 2019 - IEEE Int Conf Imaging Syst Tech Proc. 2019;(Ml):8–13.
[23] N.J. Simos, S.I. Dimitriadis, E. Kavroulakis, G.C. Manikis, G. Bertsias, P. Simos, et al., Quantitative identification of functional connectivity disturbances in neuropsychiatric lupus based on resting-state fMRI: A robust machine learning approach, Brain Sci. 10 (11) (2020) 1–18.
[24] S. Parvandeh, H.W. Yeh, M.P. Paulus, B.A. McKinney, Consensus features nested cross-validation, Bioinformatics 36 (10) (2020) 3093–3098.
[25] Y. Zhong, P. Chalise, J. He, Nested cross-validation with ensemble feature selection and classification model for high-dimensional biological data, Commun. Stat. Simul. Comput. [Internet]. 2020; 0(0): 1–18. Available from: https://doi.org/10.1080/03610918.2020.1850790.
[26] M.G. Brandon, pdp: An R package for constructing partial dependence plots, R J. 9 (1) (2017) 421.
[27] W.A. Daniel, Z. Jingyu, Visualizing the effects of predictor variables in black box supervised learning models, R. Stat. Soc. Ser. B (Statistical Methodol.) 82 (4) (2020) 1059–1086.
[28] B.P. Law, DALEX: Explainers for Complex Predictive Models in R, Available from, J. Mach. Learn. Res. [Internet]. 19 (2018) 1–5, https://pbiecek.github.io/DALEX.
[29] H. Baniecki, W. Kretowicz, P. Piatyszek, J. Wisniewski, P. Biecek, dalex: Responsible machine learning with interactive explainability and fairness in python, J. Mach. Learn. Res. 22 (2021) 1–7.
[30] A. Fisher, C. Rudin, F. Dominici, All models are wrong, but many are useful: Learning a variable’s importance by studying an entire class of prediction models simultaneously, J. Mach. Learn. Res. 20 (Vi) (2019).
[31] S. Mateusz, B. Przemyslaw, Explanations of model predictions with live and breakDown packages, arXiv [Internet]. 2018; Available from: arxiv:180401955.
[32] C. Satizabal, A. Beiser, V. Chouraki, G. Chêne, C. Dufouil, S. Seshadri, Incidence of dementia over three decades in the Framingham Heart Study, N. Engl. J. Med. 374 (6) (2016) 523–532.

[33] A. Koyanagi, L. Smith, S.J. Il, H. Oh, K. Kostev, L. Jacob, et al., Multimorbidity and Subjective Cognitive Complaints: Findings from 48 Low-and Middle-Income Countries of the World Health Survey, J. Alzheimers Dis. 81 (4) (2021) 1737–1747.
[34] H. Kaduszkiewicz, M. Eisele, B. Wiese, J. Prokein, M. Luppa, T. Luck, et al., Prognosis of mild cognitive impairment in general practice: results of the German AgeCoDe study, Ann. Fam. Med. 12 (2) (2014) 158–165.
[35] R.C. Yates, J.S. Julia, K. Leah, E.I. Abisola, J.L. Sei, F. Sharon, et al., External Validation of the eRADAR Risk Score for Detecting Undiagnosed Dementia in Two Real-World Healthcare Systems, J. Gen Int. Med. 2022.