



Published in final edited form as:

*Lancet Reg Health Am.* 2023 January ; 17: . doi:10.1016/j.lana.2022.100387.

## Classification of Alzheimer's disease and frontotemporal dementia using routine clinical and cognitive measures across multicentric underrepresented samples: A cross sectional observational study

Marcelo Adrián Maito<sup>a,ad</sup>, Hernando Santamaría-García<sup>b,c,d,ad</sup>, Sebastián Moguilner<sup>a</sup>, Katherine L. Possin<sup>b,e</sup>, María E. Godoy<sup>a</sup>, José Alberto Avila-Funes<sup>f,g,h,ae</sup>, María I. Behrens<sup>i,ae</sup>, Ignacio L. Brusco<sup>j,ae</sup>, Martín A. Bruno<sup>k,ae</sup>, Juan F. Cardona<sup>l,ae</sup>, Nilton Custodio<sup>m,ae</sup>, Adolfo M. García<sup>a,b,n,o,ae</sup>, Shireen Javandel<sup>b,e,ae</sup>, Francisco Lopera<sup>p,ae</sup>, Diana L. Matallana<sup>q,ae</sup>, Bruce Miller<sup>b,e,ae</sup>, Maira Okada de Oliveira<sup>b,r,s,ae</sup>, Stefanie D. Pina-Escudero<sup>b,e,ae</sup>, Andrea Slachevsky<sup>t,u,v,ae</sup>, Ana L. Sosa Ortiz<sup>w,ae</sup>, Leonel T. Takada<sup>x,ae</sup>, Enzo Tagliazuchi<sup>y,z,ae</sup>, Victor Valcour<sup>b,aa,ae</sup>, Jennifer S. Yokoyama<sup>b,e,ae</sup>, Agustín Ibañez<sup>y,ab,ac,\*</sup>

<sup>a</sup>Cognitive Neuroscience Center (CNC), Universidad de San Andrés, Buenos Aires, Argentina

<sup>b</sup>Global Brain Health Institute, University of California, San Francisco, CA, USA

<sup>c</sup>Ph.D Program of Neuroscience, Psychiatry Department, Pontificia Universidad Javeriana, Bogotá, Colombia

<sup>d</sup>Center for Memory and Cognition Intellectus, Hospital San Ignacio, Bogotá, Colombia

<sup>e</sup>Memory and Aging Center, Department of Neurology, University of California, San Francisco, CA, USA

<sup>f</sup>Geriatrics Department, Instituto Nacional de Ciencias médicas y nutrición Salvador Zubirán, Mexico City, Mexico

<sup>g</sup>Centre de Recherche Inserm, U897, Brodeaux, France

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\*Corresponding author. Latin American Brain Health Institute (BrainLat), Universidad Adolfo Ibañez, Santiago de Chile, Chile. [agustin.ibanez@gbhi.org](mailto:agustin.ibanez@gbhi.org) (A. Ibañez).

<sup>ad</sup>Equal contribution.

<sup>ae</sup>Order of authorship in alphabetical order.

Contributors

Conceptualization: MM, HS-G, AI.

Methodology: MM, HS-G, SM, AI.

Investigation: MM, HS-G, AI, KP, MG, AAF, MB, IB, MB, JC, NC, AG, SJ, FL, DM, BM, MO, SP-E, AS, AS, LT, ET, VV, JY.

Visualization: MM, HS-G, SM, AI.

Supervision: AI.

Writing—original draft: HS-G, MM, AI.

Writing—review & editing: HS-G, MM, AI, KP, MG, JAAF, MB, IB, MB, JC, NC, AG, SJ, FL, DM, BM, MO, SP-E, AS, AS, LT, ET, VV, JY.

Access the data: MM and HS-G.

Decision to submit: AI, MM, HS-G.

Editor note

The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.lana.2022.100387>.

<sup>h</sup>University Victor Segalen Bourdeaux 2, Bordeaux, France

<sup>i</sup>Centro de Investigación Clínica Avanzada (CICA) Hospital Clínico Universidad de Chile, Departamento de Neurología y Neurocirugía, Hospital Clínico Universidad de Chile, Departamento de Neurociencia, Facultad de medicina Universidad de Chile and Departamento de Neurología y Psiquiatría, Clínica Alemana-Universidad del Desarrollo, Santiago, Chile

<sup>j</sup>Universidad Buenos Aires & Consejo Nacional de Investigaciones Científicas y técnicas (CONICET), Argentina

<sup>k</sup>Instituto de Ciencias Biomédicas de la Universidad Católica de Cuyo & Consejo Nacional de Investigaciones Científicas y técnicas (CONICET), Argentina

<sup>l</sup>Universidad del Valle, Cali, Colombia

<sup>m</sup>Unit Cognitive Impairment and Dementia Prevention, Peruvian Institute of Neurosciences, Lima, Peru

<sup>n</sup>National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina

<sup>o</sup>Departamento de Lingüística y Literatura, Facultad de Humanidades, Universidad de Santiago de Chile, Santiago, Chile

<sup>p</sup>Neuroscience Research Group, Universidad de Antioquia, Medellín, Colombia

<sup>q</sup>PhD Program of Neuroscience, Aging Institute, Pontificia Universidad Javeriana, Bogotá, Colombia

<sup>r</sup>Hospital Santa Marcelina, São Paulo, SP, Brazil

<sup>s</sup>University of São Paulo, São Paulo, SP, Brazil

<sup>t</sup>Neurology Department, Geroscience Center for Brain Health and Metabolism, Santiago, Chile

<sup>u</sup>Laboratory of Neuropsychology and Clinical Neuroscience (LANNEC), Physiopathology Program ICBM, East Neurologic and Neurosciences Departments, Faculty of Medicine, Hospital del Salvador and Faculty of Medicine, University of Chile, Santiago, Chile

<sup>v</sup>Servicio de Neurología, Departamento de Medicina, Clínica Alemana, Universidad del Desarrollo, University of Chile, Neuropsychiatry and Memory Disorders clinic (CMYN), Santiago, Chile

<sup>w</sup>Instituto Nacional de Neurología y neurocirugía, Ciudad de México, Mexico

<sup>x</sup>Hospital de Clinicas, University of Sao Paulo Medical School, Brazil

<sup>y</sup>Latin American Brain Health Institute (BrainLat), Universidad Adolfo Ibañez, Santiago de Chile, Chile

<sup>z</sup>Departamento de Física, Universidad de Buenos Aires & Instituto de Física de Buenos Aires (FIBA – CONICET), Buenos Aires, Argentina

<sup>aa</sup>Ph.D Program of Neuroscience, Psychiatry Department; Memory and Aging Center, Department of Neurology, University of California, San Francisco, USA

<sup>ab</sup>Universidad de San Andrés & Consejo Nacional de Investigaciones Científicas y técnicas (CONICET), Argentina

<sup>ac</sup>Global Brain Health Institute (GBHI), Trinity College Dublin, (TCD), Ireland

## Summary

**Background**—Global brain health initiatives call for improving methods for the diagnosis of Alzheimer’s disease (AD) and frontotemporal dementia (FTD) in underrepresented populations. However, diagnostic procedures in upper-middle-income countries (UMICs) and lower-middle income countries (LMICs), such as Latin American countries (LAC), face multiple challenges. These include the heterogeneity in diagnostic methods, lack of clinical harmonisation, and limited access to biomarkers.

**Methods**—This cross-sectional observational study aimed to identify the best combination of predictors to discriminate between AD and FTD using demographic, clinical and cognitive data among 1794 participants [904 diagnosed with AD, 282 diagnosed with FTD, and 606 healthy controls (HCs)] collected in 11 clinical centres across five LAC (ReDLat cohort).

**Findings**—A fully automated computational approach included classical statistical methods, support vector machine procedures, and machine learning techniques (random forest and sequential feature selection procedures). Results demonstrated an accurate classification of patients with AD and FTD and HCs. A machine learning model produced the best values to differentiate AD from FTD patients with an accuracy = 0.91. The top features included social cognition, neuropsychiatric symptoms, executive functioning performance, and cognitive screening; with secondary contributions from age, educational attainment, and sex.

**Interpretation**—Results demonstrate that data-driven techniques applied in archival clinical datasets could enhance diagnostic procedures in regions with limited resources. These results also suggest specific fine-grained cognitive and behavioural measures may aid in the diagnosis of AD and FTD in LAC. Moreover, our results highlight an opportunity for harmonisation of clinical tools for dementia diagnosis in the region.

**Funding**—This work was supported by the Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat), funded by NIA/NIH (R01AG057234), Alzheimer’s Association (SG-20-725707-ReDLat), Rainwater Foundation, Takeda (CW2680521), Global Brain Health Institute; as well as CONICET; FONCYT-PICT (2017-1818, 2017-1820); PIIIECC, Facultad de Humanidades, Usach; Sistema General de Regalías de Colombia (BPIN2018000100059), Universidad del Valle (CI 5316); ANID/FONDECYT Regular (1210195, 1210176, 1210176); ANID/FONDAP (15150012); ANID/PIA/ANILLOS ACT210096; and Alzheimer’s Association GBHI ALZ UK-22-865742.

## Keywords

Alzheimer’s Disease; Frontotemporal dementia; Underrepresented samples; Machine learning

---

## Introduction

Accurate diagnosis of Alzheimer's disease (AD) and frontotemporal dementia (FTD) remains a global brain health challenge.<sup>1-4</sup> Dementia prevalence will increase around 75% in high-income countries (HICs) and around 200% in lower-middle-income countries (LMICs) by 2050.<sup>5,6</sup> Moreover, the current scenario of dementia prevalence shows that two-thirds of individuals with dementia live in LMICs.<sup>6,7</sup> Diagnostic difficulties are even more problematic in both upper-middle-income countries (UMICs) and LMICs, as in the case in many Latin American countries (LAC)<sup>1,8</sup> in comparison with HICs. Improvement of diagnostic accuracy in a region with dramatic and progressive growth in the number of dementia cases is critical to provide tailored interventions.<sup>9</sup>

In LAC, there are numerous challenges impacting diagnostic accuracy for AD and FTD, including (a) the diversity and heterogeneity of instruments for assessing clinical and cognitive status; (b) the absence of standardized procedures to incorporate socio-demographic factors in diagnosis; (c) a dearth of harmonized diagnostic procedures across countries; (d) low dementia awareness among general practitioners and (e) under-developed training in dementia diagnosis.<sup>1,8,9</sup> Additionally, although methods that have been effective in high income countries (amyloid and tau PET imaging, and fluid-based biomarker assessment) could be effective solutions in the future, they currently are not widely available in LAC for both financial and logistic reasons, limiting their utility in informing clinical decisions.<sup>1,10</sup> Thus, developing methods to harmonise clinical, cognitive, and functional assessments is the most promising scalable approach available to diagnose AD and FTD in LAC.<sup>9</sup>

Although clinical and cognitive measures are widely used in LAC, several caveats limit their use for multicentric characterisation of dementia.<sup>1</sup> The lack of appropriate regional normative data in the clinical assessment hinders direct comparisons across countries.<sup>9</sup> Differences in sociocultural settings, genetic admixture, and clinical expertise in the region also increase clinical heterogeneity. The absence of pre-harmonized procedures across centres and countries brings additional barriers.<sup>7</sup> Failure to account for these limitations may lead to results that cannot be generalised to other settings, unjustified extrapolation of local patterns onto regional tendencies, and an inability to ascertain which data points prove to be the most robust drivers of findings. For these reasons, many studies in LAC rely on small datasets from cohorts from restricted geographical regions that may lead to results that are not generalisable. Thus, any direct multicentric comparison across LAC is challenging and potentially biased.

Here, we present a novel computational framework (Fig. 1) for classification of AD, FTD, and healthy controls (HCs) in multicentric heterogeneous samples from LAC using unharmonized clinical, demographic, and cognitive assessments. We used archival clinical datasets collected in heterogeneous populations and protocols from 11 centres in five LAC. Each centre that contributed data is an enrolment site for the Multipartner consortium to expand dementia research in dementia in Latin America (ReDLat).<sup>8</sup>

Our approach combined classical statistical methods (logistic regression models) and machine learning procedures<sup>11</sup> (support vector machine procedures, random forest, and sequential feature selection procedures) to identify the best factors to discriminate between AD, FTD, and HCs. Previous studies analysing dementia diagnosis using databases with high multidimensionality have revealed better accuracies with machine learning methods than with classical statistical models.<sup>12</sup> Thus, we anticipated higher accuracy with machine learning models than with logistic regression models. Studies in HICs have revealed a high capacity of cognitive tests to distinguish HCs from dementia patients and discriminate between different types of dementia.<sup>13,14</sup> In LAC, we expected more heterogeneity in the cognitive assessment batteries applied,<sup>13,15</sup> however in most countries, basic batteries are used to track cognition, neuropsychiatric symptoms, and functionality level in aging populations.<sup>1,8</sup> Against this background, we expected to achieve moderate discrimination between subtypes of dementia and HCs despite the data heterogeneity. Moreover, we predicted that general cognitive measures would help discriminate AD and FTD patients from HCs. Finally, considering that patients living with FTD are characterised by behavioural and frontal executive disturbances,<sup>16,17</sup> we predicted that instruments that track frontal functioning and behavioural changes (including IFS, NPI, and Mini-SEA) would best discriminate FTD from AD.

## Methods

### Study design

Cross-sectional observational study.

### Sample size calculation and derivation

We used a convenience sample that represented the total of AD and FTD cases recruited by the centres included in this study.

### Setting

The samples were recruited between January 2015 and October 2021 in the different sites of the Multi-Partner Consortium to Expand Dementia Research (ReDLat).

### Participants

Participants were recruited from eleven Latin American centres that participate in the Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat<sup>9,18</sup>) [Argentina (3 centres), Colombia (3 centres), Chile (2 centres), Mexico (2 centres), and Peru (1 centre)]. The total sample (n = 1792) included 904 participants living with AD, 282 with FTD and 606 HCs (full demographic information is provided in Table 2). All participants provided informed consent. The Institutional Review Boards and the Executive Committee of the ReDLat consortium reviewed and approved the current study.

### Clinical assessment across centres

Clinical diagnoses were established following the standard procedures employed at each research centre. In Colombia, centres diagnose patients through a consensus

conference held by a multidisciplinary team that includes psychiatrists, neurologists, neuropsychologists, and geriatricians. In Chile, Peru, Mexico, and Argentina, patients were diagnosed by experienced behavioural neurologists and geriatricians with input from the evaluating neuropsychologists. Each centre applied a heterogeneous set of neuropsychological measures to assess cognitive screening, frontal functioning, social cognition, neuropsychiatric symptoms, and functional status. Each country's neuropsychological battery was selected to align with their respective standard clinical procedures and the availability of instruments. The specific measures used in each centre are described in Table 3, and Supplementary Information S1. As expected, we detected a high number of not-at-random missing values among the neuropsychological measures (Supplementary Information S1) that were addressed by using conversion tables following recommended procedures.<sup>19,20</sup> Irrespective of the specific battery employed in each centre, all followed the international NINCDS-ADRDA criteria to diagnose AD<sup>21</sup> and published criteria to diagnose behavioural<sup>22</sup> and linguistic variants<sup>23</sup> of FTD. Finally, we normalised and standardised neuropsychological scores for each centre using its respective group of HCs for both train and test sets sampled randomly.

### Neuropsychological measures

Using previously published methods, the measures used across centres were harmonized,<sup>19,20</sup> standardised<sup>24</sup> and normalised<sup>25</sup> (section 2.3).

### Cognitive screening

Each centre tracked general cognitive functioning by using one of three types of scales comprising the Mini-Mental State Examination (MMSE),<sup>26</sup> the Montreal Cognitive Assessment (MoCA)<sup>27</sup> or the Addenbrooke's cognitive examination (ACE)<sup>28</sup> (Table 3). The MMSE is a classical instrument for assessing cognitive domains, such as verbal memory, working memory, language, and visuospatial functions. A score below 24 points has a sensitivity above 88.3% and a specificity close to 87% for detecting cognitive impairment in patients with dementia.<sup>27</sup> The MMSE was used in eight centres. The MoCA is a widely used cognitive screening tool, composed of 19 items that evaluate eight cognitive domains, including executive skills, naming, memory, attention, language, abstraction, deferred memory and orientation. It has a cut-off point of 26, a sensitivity of 87%, and a specificity of 87%.<sup>27</sup> This instrument was used in six centres. The ACE III is a cognitive screening instrument that evaluates the cognitive functions of attention, orientation, memory, language, visual perception and visuospatial skills. The ACE III was used in five centres.

### Executive function

Four countries assessed executive functioning using the INECO frontal screening (IFS).<sup>29</sup> The IFS encompasses verbal fluency, inhibitory control, processing speed, working memory and cognitive flexibility, and has shown to be effective in detecting executive dysfunction in patients with dementia.<sup>29,30</sup> The maximum possible score on the IFS is 30 points.

## Functionality

All countries assessed basic and instrumental activities of daily living using either the Pfeffer Functional Activity Questionnaire (FAQ)<sup>31</sup> or the Barthel index.<sup>32</sup> The FAQ assesses functioning in instrumental activities such as writing checks, paying bills, shopping and driving. It consists of 10 questions and is completed by an informant familiar with the patient's functioning. The Barthel index weighs difficulties in basic activities of daily living such as dressing, bathing, grooming, toilet use, bowels and bladder continency, and mobility. The FAQ was used in five centres, while the Barthel index was used in three centres.

## Neuropsychiatric symptoms

The Neuropsychiatric Inventory<sup>33</sup> (NPI) is used to track neuropsychiatric symptoms including delusions, hallucinations, conduct and sleep problems, depression, anxiety, and changes in eating patterns in dementia. Seven centres employed the NPI.

## Social cognition

The Social cognition and Emotional Assessment (SEA) in its short form (Mini-SEA)<sup>34</sup> was used in four centres. The Mini-SEA is composed of two segments: theory of mind (ToM) and emotion recognition.<sup>34</sup> ToM is assessed via the Faux-pas test which uses ten vignettes to track ability to detect social appropriateness. The emotion recognition test assesses ability to identify basic emotions using the Ekman pictures.

## Group differences in neuropsychological assessment

We performed an exploratory data analysis using basic statistical methods and approaches. One-way and two-way ANOVA was used for all Neuropsychological measurements, as well as Shapiro–Wilk normality tests and outlier detection using standard deviation, interquartile range, and percentile methods. We also tested for heteroskedasticity using Levene's test and performed T-tests for equal and unequal variances (Welch's t-test) and non-parametric T-tests (Mann–Whitney U test) when required. Classical statistical models were run for model selection, and we retained the best set by model significance,  $R^2$ , AIC, BIC, variables significance, and predictive power (Supplementary Information S3).

## Harmonisation across countries

Given the heterogeneity of clinical assessments across countries, there was a substantial amount of missing data. To harmonise the available data and increase the number of individuals with homogeneous cognitive measures, we applied the following three procedures.

**Harmonisation 1**—We harmonised the brief global cognitive assessments using equivalence tables<sup>19,20</sup> for MMSE-MoCA and MMSE-ACE-III. This procedure allows for estimation of MoCA and ACE-III scores using MMSE scores. We also used the equivalence tables to estimate MMSE scores using MoCA and ACE II scores.<sup>19,20</sup> Following this approach, we added a total of seven new converted-harmonised variables (Supplementary Information S2) and decreased the number of missing MMSE values by 325. The mean,



median and standard deviation for the original MMSE and converted-harmonised scores are provided in Table 4.

**Harmonisation 2**—To mitigate the impact of missing values and preserve the maximum number of individuals, we calculated z-scores for all neuropsychological measurements, resulting in a unified cognitive score.<sup>24</sup> Z-scores were calculated using normative data from each centre (Eq (1)).

$$x_z = (x - \mu) / s \quad \text{Eq (1)}$$

where:

$x_z$  is our new value.

$x$  is the original raw score

$\mu$  is the mean score for HCs from the site to which the patient belongs

$s$  is the standard deviation for HCs from the site to which the patient belongs.

By using these standardised neuropsychological scores, the number of available individuals for subsequent analysis was increased. We also calculated z-scores for demographic variables such as age and years of education, as well as the variables obtained in Harmonisation 1.

**Harmonisation 3**—A Min-Max scaling method<sup>25</sup> was implemented using Scikit-learn.<sup>35</sup> Each record was transformed from the original variable into a new one with a range between 0 and 1 (Eq (2)).

$$x_m = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad \text{Eq (2)}$$

where:

$x_m$  is our new value

$x$  is the original cell value

$x_{\min}$  is the minimum value of the  $x$  variable

$x_{\max}$  is the maximum value of the  $x$  variable

These methods have their strengths and weaknesses and are often required or recommended for Support Vector Machines.<sup>25</sup> This scaling method was implemented with the neuropsychological measurements, demographic variables and the variables obtained from Harmonisation 1.

After these procedures, the final dataset included the original variables, the converted variables (Harmonisation 1), the standardised variables (Harmonisation 2), and the normalised variables (Harmonisation 3).



## Data analysis

### Statistical and machine learning approaches

**Classical statistical approaches - logistic regression.:** We first tested logistic regression (LR) models, as they are efficient, easily interpretable, and widely used. An exploratory analysis revealed overlapping classes between many of the variables. As the data is not linearly separable, the regression models would not yield good classification results. In the following sections (2.5.2 and 2.5.3) we present the machine learning models used to surpass the LR limitations. For the LR models, Scikit-learn LR algorithms were used, optimizing the solver, the algorithm, and regularization both with and without an intercept term, using the scheme described in section 2.4.1.2 and 2.4.1.3.

**Machine learning approaches - random forest.:** We used the Scikit-learn Random Forest (RF) algorithm without bagging (bootstrap aggregating or resampling with replacement). RF better avoids overfitting<sup>36</sup> in contrast to LR models.

**Support vector machine models.:** We used non-linear Support Vector Machine (SVM) models to generate a predictive binary classification model. SVM transforms the feature space to establish a linear decision frontier with wide margins that result in less generalisation and a lower propensity to overfitting.<sup>37</sup> For the Scikit-learn RBF kernel, we optimized C and gamma hyperparameters. For the Scikit-learn Polynomial kernel, we optimised C, the degree of the polynomial, and the coefficient. C and gamma are both hyperparameters used for regularization; they stand for L2 squared and Kernel Coefficient, respectively.

**Cross-validation for hyperparameter tuning and training models.:** We performed cross-validation analyses to control hyperparameter tuning for each ML model. This allowed us to recursively search and test the best hyperparameters for each model to obtain the best generalization results. To this end, we ran a Grid Search for hyperparameter tuning. For each model, we searched for hyperparameters applying repeated stratified cross-validation, where the training sample was divided in  $k = 5$  folds at random, and each subset was used for training  $k-1$  times and 1 time for testing; all this was repeated  $n = 5$  times. The final training score for this procedure was obtained through the mean of the  $k * n = 25$  test scores. Next, we set aside 20% of the data as a validation set to apply the learned models on unseen data. Finally, we trained all models with the best hyperparameters given by our grid search using the same cross-validation strategy, but with  $k = 10$  and  $n = 100$ , and we then predicted on test data (the same 20% used for validation on the grid search).

**Sequential feature selection.:** Sequential Feature Selection algorithms were used as reported elsewhere<sup>11</sup> to reduce the space of  $d$  dimensional features to a subspace  $k$ , where  $k < d$ , to select the subset of features that were most predictive. This allowed us to remove irrelevant features, reducing generalization error and improving computational efficiency. We applied Sequential Backward Selection from Mlxtend.<sup>38</sup>

**Performance metrics.:** Seven performance metrics (accuracy, precision, sensitivity, specificity, F1 score, ROC AUC, confusion matrix) were used across all ML models.

## Data analysis and model selection procedures

### **Model selection with regression models and identification of best features to**

**discriminate AD from FTD.:** Our primary goal was to discriminate between AD patients and FTD patients. While our exploratory classification for AD vs HCs, and for FTD vs HCs was robust (Supplementary Information S3), our target classification was more challenging than modelling AD vs. HCs, and FTD vs. HCs because these two forms of dementia overlap in cognitive, social, and functional deficits.

**Methods to control imbalance in the sample size across classes.:** To obtain the train and test sets, stratified random sampling by site and diagnosis was carried out in all cases. This procedure ensured that the same proportion of diagnoses per site were assessed in both sets. Additionally, all classifiers were configured to account for class imbalance, adjusting class weights to be inversely proportional to their frequencies in the input data. This standard procedure can be described as follows:

$$W = n_S / (n_C * \text{freq}_C) \quad \text{Eq (3)}$$

where:

$W$  is an array of new weights with equal length to the number of classes.

$n_S$  is the sample size

$n_C$  is the number of classes.

$\text{freq}_C$  is the frequency of classes for the sample.

**Classification models comparing dementia cases and HCs.:** Our primary goal was to discriminate between AD patients and FTD patients. These two forms of dementia partially overlap in cognitive, social, and functional deficits resulting in greater challenges for classification compared to models used to differentiate AD vs. HCs; or FTD vs. HCs (patients vs controls). Nonetheless, we performed exploratory classification models for the latter comparisons and these results are reported in Supplementary Information S3.

## Country-level analyses

Furthermore, we ran a group of analyses per country to assess the multi-country variability of RF models. RF models ( $n = 5000$ ) were trained and evaluated using the same parameters for each country, as described in 3.2.

## Role of the funding source

Funders did not have any role in study design, data collection, data analysis, interpretation, or writing of this report.

## Results

The distribution of the number of AD, FTD, and controls participants are listed in Table 1. Demographic information is provided in Table 2. Ethnicity and gender identity were not

available. The role of sex as a predictor of AD, FTD, and controls was assessed in all models described below.

### **Model selection**

First, we classified AD vs FTD following several methods (LR, RF, and SVM). The best model for discriminating between AD and FTD patients was the RF model and included (in order of importance), social cognition (Mini-SEA), neuropsychiatric symptoms (NPI), cognitive screening (MMSE normalised following the Van Steenoven method), age, executive function (IFS), years of education, and sex. This model showed strong discriminative power (accuracy = 0.932, sensitivity = 0.75, specificity = 0.972, ROC AUC = 0.965). The discriminative power of this model was confirmed with SVM analyses (as revealed by the accuracy = 0.9, sensitivity = 0.62, specificity = 0.972, and AUC = 0.8). An LR model with the same variables was less effective (accuracy = 0.82, sensitivity = 0.62, specificity = 1, and AUC = 0.5) (Table 5).

### **Assessing the discriminative power of the model in a test sample**

As a confirmatory analysis, we assessed the discriminative power of our model and its statistical significance assuming a p-value = .0001. We trained and evaluated a total of 5000 RF classifiers scrambling the class labels and counting how many times the AUC of the RF classifier applied on the scrambled labels was greater than that of the original classifier (Fig. 2).

### **AD vs. FTD classification controlling for demographic factors**

Secondary analyses aimed to determine whether demographic variables influenced the results. A sequential feature selection (SFS) algorithm<sup>38</sup> was implemented over the entire sample. This algorithm eliminates one-by-one the variables with less predictive power. Cross-validation is run until an optimal model is reached, with fewer variables and a higher average score. The best model scaled by the Min-Max normalisation method was reached on step five. The features with the greatest predictive capacity were cognitive screening (MMSE - Van Steenoven), social cognition (Mini-SEA), and neuropsychiatric symptoms (NPI) (Table 6). The demographic variables were among the least powerful features in the ranking.

### **Complementary results (machine learning model for discriminating HCs vs. patients)**

A RF model was optimized for discriminating AD vs HCs and FTD vs HCs and these models reached high scores (both models reached an accuracy = +0.9). The best discriminators of AD vs. HCs were measures that track social cognition (Mini-SEA), neuropsychiatric symptoms (NPI), cognitive screening (MMSE – Van Steenoven) and executive functioning (IFS) and, to a lesser degree, age, education and sex. By contrast, the best discriminators of FTD vs. HCs were the measures of neuropsychiatric symptoms (NPI), social cognition (Mini-SEA), executive functioning (IFS), age, education, cognitive screening (MMSE – Van Steenoven) and sex (Supplementary Information S7-S9).

To control for the possible effects of missing data in the AD and FTD classification, we ran a new model using only the data from two sites where data was complete. The model achieved

almost identical results to those reported when including all centres (accuracy = 0.875, sensitivity = 0.714, specificity = 0.91, ROC AUC = 0.91, see Supplementary information 10 and Supplementary Fig. 4).

### Country-level results

RF models were assessed at the country level to evaluate the variability across LACs. Our results revealed adequate feature scores and similar predictors in each country's discrimination (Supplementary Information S11-S12). An expected variability at the country level was observed regarding power and order of features.

### Discussion

This study implemented classical statistical and machine learning procedures to discriminate between HCs, AD, and FTD patients using archival clinical, cognitive, and demographic data collected in clinical settings across multiple LAC. Despite the data heterogeneity, machine learning approaches were highly effective in differentiating groups. A Random Forest model proved most successful in discriminating between AD and FTD (accuracy = 0.91 and ROC AUC = 0.96). The most significant factors in discriminating between AD and FTD in LAC were social cognition, neuropsychiatric symptoms, cognitive screening, and to a lesser extent, demographics (age, education, and sex).

To the best of our knowledge, this is the first study using machine learning models to assess the capacity of clinical, cognitive, and demographic data collected from multicentric clinical settings to discriminate between HCs, AD, and FTD in LAC. Currently, there is an unmet need to study dementia diagnosis across LAC, as highlighted by multiple international associations.<sup>5,39</sup> The data-driven techniques employed here provide a reliable approach to accurately discriminate between conditions, in a region known for exhibiting high clinical heterogeneity, having no systematic procedures for clinical assessment of dementia, and presenting limited access to biomarkers.<sup>1,8,10,40-43</sup> Additionally, the results add new evidence supporting complementary automatized methods for clinical discrimination of dementia. Such methods highlight the discriminative potential of measures of executive functioning, neuropsychiatric symptoms, and social cognition as compared to classical cognitive measures.

Our findings are consistent with those from studies in HICs,<sup>13,14</sup> showing that in underrepresented populations such as those found in LAC, machine learning analysis of clinical and cognitive tests can have high accuracy in discriminating AD from other dementias. Particularly, a ranking of conventional cognitive tests (social cognition > executive functioning > neuropsychiatric symptoms > cognitive screening) discriminated between AD and FTD patients with high accuracy in our sample. Results confirm that social cognition,<sup>34</sup> executive functioning and neuropsychiatric symptoms<sup>44-46</sup> are useful to discriminate between FTD and AD patients.<sup>46</sup>

The functionality scales did not have high discriminative values in predicting AD and FTD in our study, as previously reported results.<sup>44,47</sup> High dispersion of functionality values and the inability to control for other potential confounders such as disease severity could explain

the reduced discriminative power and similar impairments in instrumental activities of daily living.

Although FTD patients have an earlier average age of disease onset<sup>48,49</sup> and tend to be more highly educated than AD patients,<sup>48</sup> the classical demographic factors (age, sex, and education) had lower predictive scores than the cognitive variables in this study. The sample used in this study exhibited a high dispersion in age, sex, and education levels in AD and FTD patients. That heterogeneity could explain why cognitive measures performed better than demographic factors in discriminating between AD and FTD. Nonetheless, our results coincided with past studies showing that demographic factors are less effective predictors than cognitive measures in differentiating AD from FTD.<sup>50</sup>

Our results also revealed that social cognition (Mini-SEA), neuropsychiatric symptoms (NPI), executive functioning (IFS), and cognitive screening tests (MMSE normalised following Van Steenoven) are the best predictors to differentiate HCs vs. AD and HCs vs. FTD. The best discriminators for AD vs. HCs were the measures tracking social cognition, neuropsychiatric symptoms, cognitive screening and executive functioning and, to a lesser degree, age, education and sex. By contrast, social cognition and neuropsychiatric symptoms were more useful in differentiating FTD from HCs (Supplementary Information S7-S9). Results are consistent with past studies showing that in comparison with HCs, patients with AD and FTD tend to more frequently exhibit impairment in cognitive processes,<sup>51</sup> social cognition<sup>16</sup> and neuropsychiatric symptoms (comparing HCs vs. AD<sup>45,52</sup> and HCs vs. FTD<sup>44,45</sup>). However, in our study, the measures that track social cognition and executive functioning held more weight as discriminators of HCs vs. AD and FTD than cognitive screening tools (MMSE) or demographic factors. Although heterogeneity issues could explain this pattern of results, it could also suggest a high sensitivity of measures tracking specific frontal functioning<sup>16,41,53,54</sup> in discriminating healthy and pathologic aging.

The country-level analyses also revealed similar prediction scores of AD and FTD as those run with data gathered from all countries. Not surprisingly, the accuracy level decreased in particular countries, and the order of features showed country-level variability. These differences can be explained by disparate sample sizes, heterogeneity in data collection, clinical assessment procedures, and other unknown country-level differences. Beyond these expected differences, our results reached correspondence in predictive values and relevant features of AD and FTD, allowing a regional clinical dementia standardization.

To the best of our knowledge, no other studies have aimed to develop an approach to distinguish between AD and FTD using multicentric, heterogeneous, archival clinical datasets from underrepresented, low-resource settings. The combination of classical statistical approaches and machine learning methods resulted in an effective technique.<sup>55</sup> Logistic regression is sensitive to capturing outliers and correlated features, modulating the estimation of coefficients when data is sparse, and assuming a linear relationship between the inputs and the output.<sup>25</sup> However, complex datasets usually require non-linear decision boundaries to improve classification outcomes. Particularly, Support Vector Machines and Random Forest models can learn non-linear decision boundaries and thus, achieve better classification results than logistic regression.<sup>37</sup>

Our study obtained the most accurate results by using machine learning models confirmed with standard cross-validation procedures with training and testing sets. These methods are more robust to account for heterogeneity of data and complex interaction between features. Our results support a better performance of machine learning methods than classical statistical approaches in predicting dementia outcomes when databases combine multiple variables and interactions.<sup>12</sup> Accuracy levels achieved by machine learning procedures in our study are comparable with those observed in machine learning studies analysing neuroimaging and other biomarkers classification of AD and FTD.<sup>56</sup>

Furthermore, results revealed that random forest was superior in classifying AD and FTD cases. The lower classification power of the logistic regressions and SVM (compared to Random Forest) could be explained by their reduced accuracy with multidimensional datasets. In fact, variance across groups can impact the SVM hyperplane to discriminate groups.<sup>57,58</sup> The heterogeneity across groups in the present research suggest that this is the case. By contrast, our results confirm that Random Forest is one of the most accurate procedures for datasets with multidimensional information and larger interactions between variables.<sup>59,60</sup>

Together, results reveal important insights for the study of dementia diagnosis in LAC's underrepresented populations, including (a) conventional data used for clinical assessment of dementia is a reliable source of information to discriminate AD, FTD, and controls across different LACs, partially compensating the limited regional access to biomarkers; (b) similar discrimination scores for AD and FTD were observed at the regional level, as found in other studies using biomarkers [1–3]; (c) the regional reproducibility in discriminating groups supported the efficacy of clinical procedures for dementia diagnosis in Latin America; (d) combined neuropsychiatric symptoms, executive functions, and social cognition are relevant measures to characterize dementia in LACs.

Our work does have important limitations. Differences in protocols between centres led to missing values, which resulted in a reduction in the number of samples used for modelling. The data collection process across centres was not harmonised, only global scores were used, and severity scores were not available. These are typical challenges when using multicentric data. We addressed this problem by harmonising and reducing missing data following previously published procedures.<sup>19,20,24,25</sup> Moreover, we tested and verified the replicability of our results in a complementary model using only the information from two centres with a high proportion of complete data. Results from the complementary models produced similar predictive scores and similar predictors as the models that included data from all sites. Although these results suggest that missing data did not ostensibly affect the classification processes in our study, future studies should test the replicability of our data-driven models in large datasets with more balanced data across different centres and compare those results using missing data imputation methods.

Our study was also limited by a smaller than optimal sample size for the statistical methods used. However, the sample size was comparable to, and often larger than, other multicentric studies of dementia.<sup>15,61-63</sup> Furthermore, our applied methodology allows us to integrate futures samples to our current data and test our models on new unseen data. Another

potential limitation of our study is the imbalance in sample size of the two diagnostic groups. To account for this, we performed additional procedures to control for the imbalance in the classifications (section 2.4.2.2). While this imbalance is expected given previous studies in LAC which have described a higher prevalence of AD (representing 55% of dementia cases seen in clinics) than FTD (representing between 2.8% and 1.5% of dementia cases [1–3]), future studies may benefit from increasing the number of FTD cases included.

Specific data describing neuropsychiatric symptoms beyond the total NPI score was not available in the current study. Future investigation assessing predictors of AD and FTD should track neuropsychiatric symptoms with more granularity. Furthermore, our study goals are focused on LAC where differences in dementia presentation and varied prevalence in dementia risk factors across countries are determined by varied social, medical, and genetic risks. A comparison of data from LAC with other regions is not within the scope of this study. However, the present framework can be used in future studies to assess data from other regions across the world using similar approaches.

Finally, dementia diagnoses in our study were based on clinical expertise without biomarker confirmation as access to traditional biomarkers is limited in LAC due to cost. While it would be ideal for future studies to test the classification power against biomarker confirmed cases, our results indicate that the usage of clinical and neuropsychological data could help support dementia diagnosis, particularly in underrepresented populations where access to biomarkers is limited. Looking forward, we expect to add more underrepresented samples from other LMIC and UMICs to further test the generalizability of our approach and develop new models.

## Conclusions

A high classification accuracy for AD, FTD, and HCs was obtained by combining classical statistical and machine learning procedures applied to demographic, cognitive, and behavioural data from clinical settings across LAC. Results highlight the importance of combining conventional clinical assessment with fine-grained cognitive tests tracking social cognition, executive functioning, behavioural symptoms, and cognitive screening to diagnose AD and FTD in LAC. We developed a robust methodological approach to deal with highly heterogeneous data based on the measurements described above from archival clinical datasets to classify dementia subtypes and HCs. These findings support the use of combined statistical analysis, statistical modelling, and machine learning methods for multicentric studies involving low-resource regions with restricted access to biomarkers for underrepresented cohorts. Dementia services in LAC could benefit from improving diagnosis accuracy by applying the current methodological approach.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.



## Acknowledgments

The authors thank the clinical teams and researchers from the ReDLat consortium, as well as Lucas Neufeld for the help with the elaboration of the procedure figure.

### Declaration of interests

Bruce Miller: support for the present Manuscript (National Institutes of Health (R01AG057234); Grants or contracts (NIH 1R01AG070883, NIH/NIA R35AG072362), Royalties or licenses (Cambridge University Press, Guilford Publications, Inc.; Johns Hopkins Press, Oxford University Press, Taylor & Francis Group, Elsevier, Inc.), payment or honoraria for lectures (UCLA- Grand Rounds, Korean Dementia Society Talks, UCI MIND Seminar Series, Taipei Medical University, Dementia Center; Support for attending meetings and/or travel (UCLA, CalArts, GBHD); Participation on Data Safety Monitoring Board or Advisory Board (Stanford University ADRC—External Advisor; The Buck Institute for Research on Aging—Scientific Advisor; Arizona Alzheimer’s Disease Center—External Advisor); Leadership of fiduciary role in other board, society, committee or advocacy group (Institute for Neurodegenerative Diseases – Affiliated Faculty, The Bluefield Project to Cure FTD– Director and Internal Advisor, Tau Consortium—Co-Director, and Scientific Advisor, Global Brain Health Institute GBHI—Co-Director). Jennifer Yokoyama: All support for the present Manuscript: (NIH-NIA R01 AG062588, R01 AG057234, P30 AG062422; NIH-NINDS U54 NS123985); Grants or contracts from any entity (Rainwater Charitable Foundation – Institution; Alzheimer’s Association – Institution; Global Brain Health Institute – Institution; Mary Oakley Foundation – Institution); Payment or honoraria for lectures, presentations, speakers, bureaus, manuscript writing or educational events (Washington University in St. Louis School of Medicine NeuroGenomics and Informatics Center); Participation on Data Safety Monitoring Board or Advisory Board (Epstein Family Alzheimer’s Collaboration - Scientific Advisory Board member—payment made to institution). Ana Luisa Sosa: Grants or contracts from any entity (FINGERS, ReDLat, DIAN-Tu, DIAN-obs); Payment or honoraria for lectures, presentations, speakers, bureaus, manuscript writing or educational events (Asopharma, Lundbeck, Carnot, Apopharma, Biogen and Roche); Support for attending meetings and/or travel (ADI meeting 2022, by Carnot); Participation on Data Safety Monitoring Board or Advisory Board (Carnot and Lundbeck - Advisory Board). All other authors declare they have no competing interests.

### Data sharing statement

All data, code, and materials used in the analyses are available from the corresponding author upon reasonable request.

### References

1. Parra MA, Baez S, Sedeño L, et al. Dementia in Latin America: paving the way toward a regional action plan. *Alzheimer’s Dement*. 2021;17(2):295–313. [PubMed: 33634602]
2. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 2020;396(10248):413–446. [PubMed: 32738937]
3. Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. *Lancet*. 2017;390(10113):2673–2734. [PubMed: 28735855]
4. Dawson WD, Bobrow K, Ibanez A, et al. The necessity of diplomacy in brain health. *Lancet Neurol*. 2020;19(12):972–974.
5. Nichols E, Steinmetz JD, Vollset SE, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health*. 2022;7(2):e105–e125. [PubMed: 34998485]
6. Ferri CP, Jacob KS. Dementia in low-income and middle-income countries: different realities mandate tailored solutions. *PLoS Med*. 2017;14(3):e1002271. [PubMed: 28350797]
7. Custodio N, Wheelock A, Thumala D, Slachevsky A. Dementia in Latin America: epidemiological evidence and implications for public policy. *Front Aging Neurosci*. 2017;9.
8. Parra MA, Baez S, Allegri R, et al. Dementia in Latin America: assessing the present and envisioning the future. *Neurology*. 2018;90(5):222–231. [PubMed: 29305437]
9. Ibanez A, Yokoyama JS, Possin KL, et al. The multi-partner consortium to expand dementia research in Latin America (ReDLat): driving multicentric research and implementation science. *Front Neurol*. 2021;12:303.

10. Parra MA, Orellana P, Leon T, et al. Biomarkers for dementia in Latin American countries: gaps and opportunities. *Alzheimer's Dement.* 2022.
11. Moguilner S, Birba A, Fittipaldi S, et al. Multi-feature computational framework for combined signatures of dementia in underrepresented settings. *J Neural Eng.* 2022;19(4).
12. Spooner A, Chen E, Sowmya A, et al. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Sci Rep.* 2020;10(1):1–10. [PubMed: 31913322]
13. Garcia-Gutierrez F, Delgado-Alvarez A, Delgado-Alonso C, et al. Diagnosis of Alzheimer's disease and behavioural variant frontotemporal dementia with machine learning-aided neuropsychological assessment using feature engineering and genetic algorithms. *Int J Geriatr Psychiatr.* 2021;37(2).
14. Gregory CA, Orrell M, Sahakian B, Hodges JR. Can frontotemporal dementia and Alzheimer's disease be differentiated using a brief battery of tests? *Int J Geriatr Psychiatr.* 1997;12(3).
15. Battista P, Salvatore C, Castiglioni I. Optimizing neuropsychological assessments for cognitive, behavioral, and functional impairment classification: a machine learning study. *Behav Neurol.* 2017;2017:1850909. [PubMed: 28255200]
16. Baez S, Manes F, Huepe D, et al. Primary empathy deficits in frontotemporal dementia. *Front Aging Neurosci.* 2014;6:262. [PubMed: 25346685]
17. Migeot JA, Duran-Aniotz CA, Signorelli CM, Piguet O, Ibanez A. A predictive coding framework of allostatic-interoceptive overload in frontotemporal dementia. *Trends Neurosci.* 2022;45:838–853. [PubMed: 36057473]
18. Ibanez A, Parra MA, Butler C. The Latin America and the Caribbean consortium on dementia (LAC-CD): from networking to research to implementation science. *J Alzheimers Dis.* 2021;82(s1):S379–S394. [PubMed: 33492297]
19. Kim R, Kim H-J, Kim A, Jang M-H, Kim HJ, Jeon B. Validation of the conversion between the mini-mental state examination and montreal cognitive assessment in Korean patients with Parkinson's disease. *J Mov Disord.* 2018;11(1):30–34. [PubMed: 29316782]
20. Matías-Guiu JA, Pytel V, Cortés-Martínez A, et al. Conversion between Addenbrooke's cognitive examination III and mini-mental state examination. *Int Psychogeriatr.* 2017;30(8):1227–1233. [PubMed: 29223183]
21. 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. [PubMed: 17616482]
22. Rascovsky K, Hodges JR, Knopman D, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain.* 2011;134(Pt 9):2456–2477. [PubMed: 21810890]
23. Gorno-Tempini ML, Hillis AE, Weintraub S, et al. Classification of primary progressive aphasia and its variants. *Neurology.* 2011;76(11):1006–1014. [PubMed: 21325651]
24. Donnelly-Kehoe PA, Pascariello GO, García AM, et al. Robust automated computational approach for classifying frontotemporal neurodegeneration: multimodal/multicenter neuroimaging. *Alzheimer's Dement.* 2019;11:588–598.
25. Larose CD. *LDLDT. Data Mining and Predictive Analytics.* 2nd ed. 2015.
26. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12(3):189–198. [PubMed: 1202204]
27. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* 2005;53(4):695–699. [PubMed: 15817019]
28. Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, Hodges JR. *Addenbrooke's Cognitive Examination.* American Psychological Association (APA); 2000.
29. Torralva T, Roca M, Gleichgerrcht E, López P, Manes F. INECO Frontal Screening (IFS): a brief, sensitive, and specific tool to assess executive functions in dementia. *J Int Neuropsychol Soc.* 2009;15(5):777–786. [PubMed: 19635178]

30. Gleichgerrcht E, Roca M, Manes F, Torralva T. Comparing the clinical usefulness of the Institute of cognitive neurology (INECO) frontal screening (IFS) and the frontal assessment battery (FAB) in frontotemporal dementia. *J Clin Exp Neuropsychol*. 2011;33(9):997–1004. [PubMed: 21923634]
31. Pfeffer RI, Kurosaki TT, Harrah C Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. *J Gerontol*. 1982;37(3):323–329. [PubMed: 7069156]
32. Mahoney F, Barthel DW. Functional evaluation ; the Barthel index. A simple index of the independence useful in scoring improvement in the rehabilitation of the chronically ill. *Md State Med J*. 1965;14:61–66.
33. Cummings JL. The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. *Neurology*. 1997;48(5):s10S–6.
34. Funkiewiez A, Bertoux M, de Souza LC, Lévy R, Dubois B. The SEA (Social Cognition and Emotional Assessment): a clinical neuropsychological tool for early diagnosis of frontal variant of frontotemporal lobar degeneration. *Neuropsychology*. 2012;26(1):81–90. [PubMed: 21895376]
35. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.
36. Breiman L. *Random Forests*. SpringerLink; 2018.
37. Schölkopf B, Smola AJ. *Learning with Kernels*. The MIT Press; 2018.
38. Raschka S. *MLxtend: providing machine learning and data science utilities and extensions to Python's scientific computing stack*. *J Open Source Softw*. 2018;3(24):638.
39. Miranda JJ, Barrientos-Gutiérrez T, Corvalan C, et al. Understanding the rise of cardiometabolic diseases in low-and middle-income countries. *Nat Med*. 2019;25(11):1667–1679. [PubMed: 31700182]
40. Ibañez K, Polke J, Hagelstrom RT, et al. Whole genome sequencing for the diagnosis of neurological repeat expansion disorders in the UK: a retrospective diagnostic accuracy and prospective clinical validation study. *Lancet Neurol*. 2022;21(3):234–245. [PubMed: 35182509]
41. Baez S, García AM, Ibanez A. The social context network model in psychiatric and neurological diseases. *Curr Top Behav Neurosci*. 2016;30:379–396.
42. Duran-Aniotz C, Orellana P, Leon Rodriguez T, et al. Systematic review: genetic, neuroimaging, and fluids biomarkers for frontotemporal dementia across Latin America countries. *Front Neurol*. 2021;12:663407. [PubMed: 34248820]
43. Duran-Aniotz C, Sanhueza J, Grinberg LT, et al. *The Latin American Brain Health Institute, A Regional Initiative To Reduce the Scale and impact of Dementia*. Wiley Online Library; 2022.
44. Santacruz Escudero JM, Beltrán J, Palacios Á, et al. Neuropsychiatric symptoms as predictors of clinical course in neurodegeneration. A longitudinal study. *Front Aging Neurosci*. 2019;11:176. [PubMed: 31396074]
45. Ismail Z, Smith EE, Geda Y, et al. Neuropsychiatric symptoms as early manifestations of emergent dementia: provisional diagnostic criteria for mild behavioral impairment. *Alzheimers Dement*. 2016;12(2):195–202. [PubMed: 26096665]
46. Santamaria-Garcia H, Reyes P, Garcia A, et al. First symptoms and neurocognitive correlates of behavioral variant frontotemporal dementia. *J Alzheimers Dis*. 2016;54(3):957–970. [PubMed: 27567867]
47. Park LQ, Harvey D, Johnson J, Farias ST. Deficits in everyday function differ in AD and FTD. *Alzheimer Dis Assoc Disord*. 2018;29(4):301–306.
48. Borroni B, Alberici A, Agosti C, Premi E, Padovani A. Education plays a different role in Frontotemporal Dementia and Alzheimer's disease. *Int J Geriatr Psychiatr*. 2008;23(8):796–800.
49. Hodges JR, Piguet O. Progress and challenges in frontotemporal dementia research: a 20-year review. *J Alzheimers Dis*. 2018;62(3):1467–1480. [PubMed: 29504536]
50. Hutchinson AD, Mathias JL. Neuropsychological deficits in frontotemporal dementia and Alzheimer's disease: a meta-analytic review. *J Neurol Neurosurg Psychiatry*. 2007;78(9):917–928. [PubMed: 17371908]
51. Arevalo-Rodriguez I, Smailagic N, Figuls MRI, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev*. 2015;3.

52. Ng KP, Pascoal TA, Mathotaarachchi S, et al. Neuropsychiatric symptoms predict hypometabolism in preclinical Alzheimer disease. *Neurology*. 2017;88(19):1814–1821. [PubMed: 28404803]
53. Eslinger PJ, Moore P, Anderson C, Grossman M. Social cognition, executive functioning, and neuroimaging correlates of empathic deficits in frontotemporal dementia. *J Neuropsychiatry Clin Neurosci*. 2011;23(1):74–82. [PubMed: 21304142]
54. Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci*. 2006;7(4):268–277. [PubMed: 16552413]
55. Bari Antor M, Shafayet Jamil AHM, Mamtaz M, et al. A comparative analysis of machine learning algorithms to predict Alzheimer's disease. *J Healthc Eng*. 2021:9917919. [PubMed: 34336171]
56. Kim JP, Kim J, Park YH, et al. Machine learning based hierarchical classification of frontotemporal dementia and Alzheimer's disease. *Neuroimage Clin*. 2019;23.
57. Bari Antor M, Jamil A, Mamtaz M, et al. A comparative analysis of machine learning algorithms to predict Alzheimer's disease. *J Healthc Eng*. 2021;2021:9917919. [PubMed: 34336171]
58. Kharoubi R, Oualkacha K, Mkhadri A. The cluster correlation-network support vector machine for high-dimensional binary classification. *J Stat Comput Simulat*. 2019;89(6):1020–1043.
59. Altman N, Krzywinski M. Ensemble methods: bagging and random forests. *Nat Methods*. 2017;14(10):933–934.
60. Breiman L. Random forest. *Mach Learn*. 2001;45(1):5–32.
61. Weakley A, Williams JA, Schmitter-Edgecombe M, Cook DJ. Neuropsychological test selection for cognitive impairment classification: a machine learning approach. *J Clin Exp Neuropsychol*. 2015;37(9):899–916. [PubMed: 26332171]
62. Grassi M, Perna G, Caldirola D, Schruers K, Loewenstein DA. A clinically-translatable machine learning algorithm for the prediction of Alzheimer's disease conversion in individuals with mild and premild cognitive impairment. *J Alzheim Dis*. 2018;61:1555–1573.
63. Gurevich P, Stuke H, Kastrup A, Stuke H, Hildebrandt H. Neuropsychological testing and machine learning distinguish Alzheimer's disease from other causes for cognitive impairment. *Front Aging Neurosci*. 2017;9.

## Research in context

### Evidence before this study

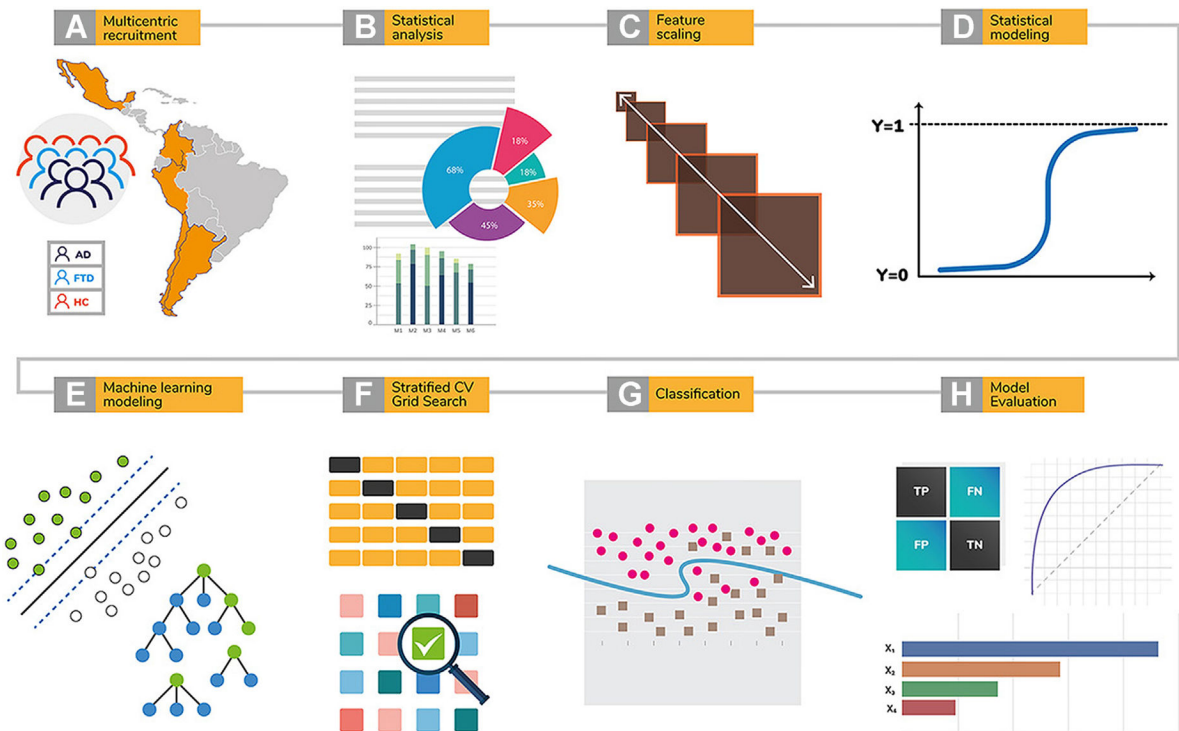
We developed a methodology to assess the best predictors of Alzheimer's disease and frontotemporal dementia using routine clinical and cognitive measures collected from underrepresented samples across multiple centres in Latin America (ReDLat cohort). Diagnostic procedures in this region are challenging due to a lack of harmonised diagnostic methods and clinical procedures, and limited access to biomarkers. Dementia studies in Latin America often rely on small datasets, meaning results are often not generalisable. This work represents a unique attempt to present a comprehensive approach to exploring archival clinical datasets from a large sample of 1794 individuals (904 diagnosed with Alzheimer's disease, 282 diagnosed with frontotemporal dementia and, 606 healthy controls) from eleven centres across five countries.

### Added value of this study

Based on a comprehensive review of the literature, this is the first attempt to develop an approach to differentiate and classify Alzheimer's disease and frontotemporal dementia patients with multicentric, heterogeneous, underrepresented, archival clinical datasets in low-resources settings. Here we applied data harmonization techniques, classical statistical approaches and machine learning methods to classify and differentiate Alzheimer's disease from frontotemporal dementia, and both from healthy controls.

### Implications of all the available evidence

Our results are comparable to findings from single-centre studies with homogeneous cohorts, machine learning studies using neuroimaging and other biomarker-based classification of Alzheimer's disease and frontotemporal dementia patients. We found a high classification accuracy of the dementia and healthy controls cases by combining classical statistical and machine learning procedures applied to demographic, cognitive and behavioural data from clinical settings across Latin American countries. Importantly, some cognitive and behavioural measures including executive function, social cognition and cognitive screening measures produced higher predictive values than demographic factors such as education, age or sex. We describe a methodological approach to deal with highly heterogeneous data from underrepresented populations.

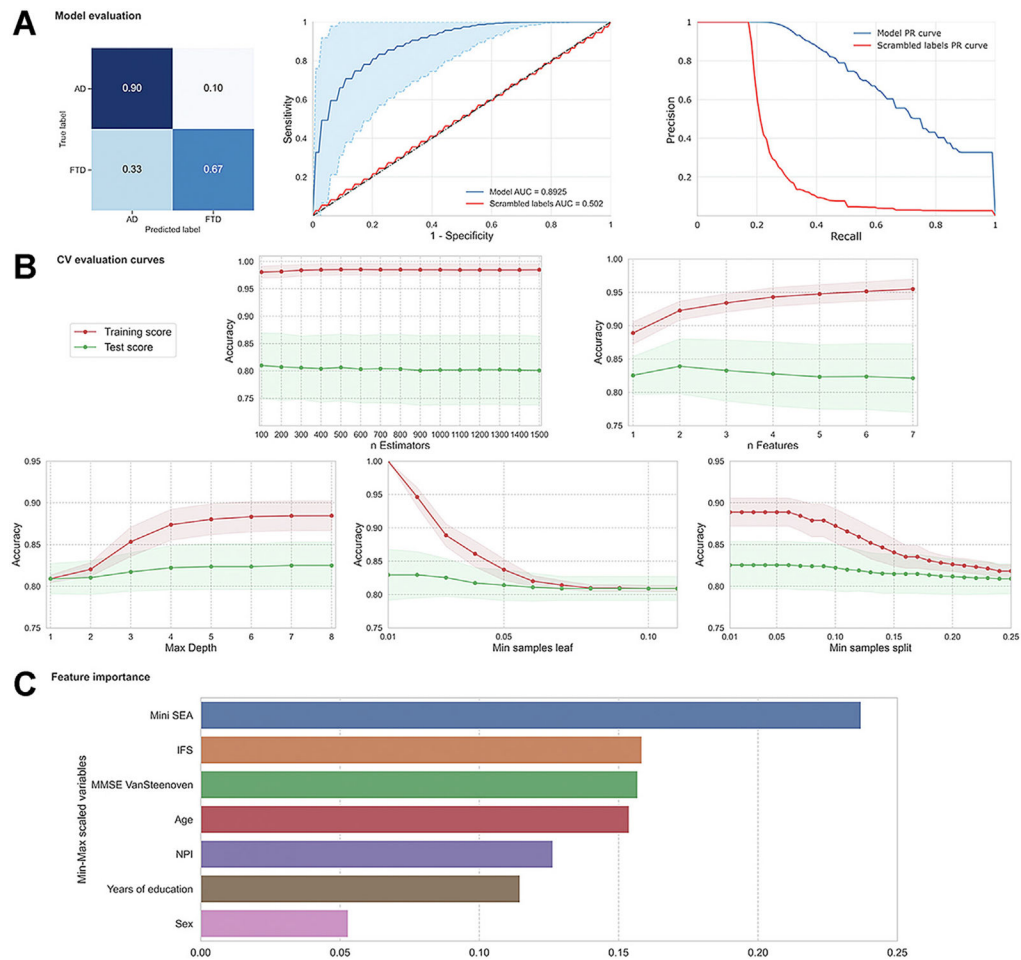


**Fig. 1: Approach to dementia diagnosis with standard and non-standard methodologies.**

(A) Data was combined from 1792 participants seen in eleven centres in five LAC, including 904 patients with AD, 282 patients with FTD, and 606 HCs. Data comprised cognitive screening (MMSE, MoCA, and ACE-III), executive function (IFS), social cognition (Mini-SEA), functional status (Pfeffer FAQ and Barthel index), and neuropsychiatric symptoms (NPI). Demographic data included year of birth, country, years of education, and sex.

(B) Basic statistical comparisons included Welch's t-test (unequal variances t-test), Mann-Whitney U test, paired t-test, one-way and two-way ANOVA, Shapiro-Wilk test of normality, Levene test for equal variances, outliers' detection by standard deviation, percentiles and IQR methods. (C) Neuropsychological measurements and demographic variables were standardised and normalised. Cognitive screening total scores were converted (detailed in section 2.3). (D) Standard logistic regression models were applied. (E) Baseline machine learning models: support vector machines, and random forest were applied. (F) The stratified k-fold cross-validation scheme with standard grid-search approach was used to test different model's regularization. (G) Classification was obtained using out of sample testing sets. (H) ROC AUC, Precision-Recall curve, Accuracy, Sensitivity, specificity and F1-score were used for model evaluation. AD patients were set as the "negative" and FTD patients as the "positive" class. LAC, Latin American Countries; AD, Alzheimer's Disease; FTD, Frontotemporal Dementia; MMSE, Mini Mental State Examination; MoCA, Montreal Cognitive Assessment; ACE-III, Addenbrooke's Cognitive Examination; IFS, Ineco Frontal Screening; Mini-SEA, Mini-Social Cognition & Emotional Assessment; Pfeffer FAQ, Pfeffer Functional Activity Questionnaire; NPI, Neuropsychiatric Inventory; ANOVA, Analysis of Variance; IQR, Interquartile Range; ROC AUC, Receiver Operating Characteristics Area Under Curve.





**Fig. 2: Mean Main results, including random forest model evaluation, CV validation curves, and feature importance test sets.**

(A) The upper left graph shows the confusion matrix in relative magnitudes. The upper central graph shows the ROC curves with confidence intervals and its AUC, the upper right graph shows the Precision-Recall curves for the original classifier (blue) and the scrambled one (red). (B) The central and lower graphs show the validation curves on five RF hyperparameters (regularization) on Stratified Cross Validation with  $k = 5$ , repeats = 20. The dots represent the average score for  $k * n = 100$  scores for each range of possible values shown on the x axis. The red and green areas around the lines represent the standard deviation for the results of training and testing for each of the 100 scores. (C) The bottom graph shows the importance of every feature of the original model. The most important ones were social cognition (Mini-SEA), executive functions (IFS), cognitive screening (measured with MMSE normalized via Van Steenoven method), age, neuropsychiatric symptoms (NPI), years of education and sex.



**Table 1:**

Distribution of type of individuals by country.

Country	AD	HCS	FTD	Total
Argentina	257	192	53	502
Chile	197	145	59	401
Colombia	320	232	155	707
Mexico	30	21	7	58
Peru	100	16	8	124
Total	904	606	282	1792

AD, Alzheimer Disease; FTD, Frontotemporal Dementia; HCs, Healthy Controls.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:**

Demographic information.

	AD	FTD	HCs	Statistic	Pairwise comparisons using a one-way ANOVA P values
Sex (%) <sup>a</sup>	M = 310 (17.3%) F = 594 (33.15%)	M = 141 (7.87%) F = 141 (7.87%)	M = 187 (10.43%) F = 419 (23.38%)	32.123	(P < 0.0001)
Age <sup>b</sup>	81.58 (9.95%)	72.33 (9.14%)	73.65 (10.9%)	146.27	AD vs CN (P < 0.001) AD vs FTD (P < 0.001) CN vs FTD (P > 0.05)
Years Education <sup>b</sup>	10.29 (4.88%)	12.77 (5.10%)	13.58 (4.7%)	81.133	AD vs CN (P < 0.001) AD vs FTD (P < 0.001) CN vs FTD (P > 0.05)

AD, Alzheimer Disease; FTD, Frontotemporal Dementia; HCs, Healthy Controls.

<sup>a</sup>Chi-Square Test (Critical value).<sup>b</sup>Anova test. Mean (standard deviation).

**Table 3:**

Assessments by test type and center.

Center	MMSE	MoCA	ACE-III	IFS	Mini-SEA	Barthel	Pfeiffer	NPI	NPIc	Sample(n)
Arg 1	0	3	31	34	0	0	0	34	0	54
4Arg 2	248	0	0	0	0	0	0	0	0	249
Arg 3	0	63	101	181	114	0	38	23	22	199
Chi 1	1	111	1	0	0	0	0	0	0	127
Chi 2	271	177	176	173	159	0	155	229	223	274
Col 1	59	29	59	59	29	30	0	0	0	59
Col 2	324	35	0	34	0	343	0	13	11	454
Col 3	187	188	0	99	0	0	94	94	0	194
Mex 1	19	17	0	0	0	12	0	0	0	21
Mex 2	37	0	0	0	0	0	37	37	37	37
Per 1	124	0	124	124	124	0	124	124	0	124
Total	1270	623	492	704	426	385	448	554	293	1792

MMSE, Mini Mental State Examination; MoCA, Montreal Cognitive Assessment; ACE-III, Addenbrooke's Cognitive Examination; IFS, Ineco Frontal Screening; Mini-SEA, Mini-Social Cognition & Emotional Assessment; Barthel, Barthel scale; Pfeiffer FAQ, Pfeiffer Functional Activity Questionnaire; NPI, Neuropsychiatric Inventory; NPIc, Neuropsychiatric Inventory Caregiver. Arg, Argentina; Chi, Chile; Col, Colombia; Mex, Mexico; Per, Peru.

**Table 4:**

Mean, median and standard deviation for MMSE scores.

	MMSE conversion method											
	MMSE original			Van Steenoven			Lawton			Matias-Guiu		
	Median	Mean	Std	Median	Mean	Std	Median	Mean	Std	Median	Mean	Std
AD	22	20.74	5.45	22	20.78	5.36	22	20.77	5.34	22	20.86	5.4
FTD	23.5	22.24	6.23	23	22.35	6.1	23	22.33	6.08	24	22.47	6.04

MMSE, Mini Mental State Examination; AD, Alzheimer Disease, FID, Frontotemporal Dementia, Std, Standard deviation.

**Table 5:**

Best RF model compared with SVM and LR.

Type	Train		Test						
	Accuracy		Accuracy	Precision	Specificity	Sensitivity	F1 Score	ROC AUC	C.Matrix
Logistic Regression	0.81		0.82	0	1	0	0	0.5	[1, 0, 1, 0]
Support Vector Machine	0.99		0.9	0.83	0.97	0.62	0.71	0.8	[0.97, 0.03, 0.375, 0.625]
Random Forest	0.98		0.93	0.86	0.97	0.75	0.8	0.97	[0.97, 0.03, 0.25, 0.75]

**Table 6:**

Sequential Feature Selection (SFS) algorithm for best Random Forest model.

Step	Min Max scaled feature names	AVG score	CI bound	Std dev	Std error
1	Mint-SEA total, NPI total, MMSE Van Steenoven total, Years of education, Sex, IFS total, Age	0.8322	0.0282	0.0982	0.0140
2	Mint-SEA total, NPI total, MMSE Van Steenoven total, Years of education, Sex, IFS total	0.8322	0.0279	0.0973	0.0139
3	Mint-SEA total, NPI total, MMSE Van Steenoven total, Years of education, Sex	0.8393	0.0309	0.1077	0.0154
4	Mint-SEA total, NPI total, MMSE Van Steenoven total, Years of education	0.8489	0.0295	0.1028	0.0147
5	Mint-SEA total, NPI total, MMSE Van Steenoven total	0.8500	0.0291	0.1014	0.0145
6	Mint-SEA total, NPI total	0.8093	0.0344	0.1201	0.0172
7	Mint-SEA total	0.7211	0.0286	0.0996	0.0142