Psychological Medicine

cambridge.org/psm

Prediction of depression cases, incidence, and chronicity in a large occupational cohort using machine learning techniques: an analysis of the ELSA-Brasil study

Diego Librenza-Garcia[1,2,3], Ives Cavalcante Passos[1,2], Jacson Gabriel Feiten[1,2], Paulo A. Lotufo[4,5], Alessandra C. Goulart[4,5], Itamar de Souza Santos[4,5], Maria Carmen Viana[6], Isabela M. Benseñor[4,5] and Andre Russowsky Brunoni[4,5,7]

[1]Laboratory of Molecular Psychiatry, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; [2]Programa de Pós-Graduação em Psiquiatria e Ciências do Comportamento, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; [3]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Canada; [4]Department of Internal Medicine, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil; [5]Hospital Universitário, Universidade de São Paulo, São Paulo, Brazil; [6]Department of Social Medicine, Postgraduate Program in Public Health, Center of Psychiatric Epidemiology (CEPEP), Federal University of Espírito Santo, Vitória, Brazil and [7]Laboratory of Neurosciences (LIM-27), Department and Institute of Psychiatry, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil

Original Article

Cite this article: Librenza-Garcia D, Passos IC, Feiten JG, Lotufo PA, Goulart AC, de Souza Santos I, Viana MC, Benseñor IM, Brunoni AR (2020). Prediction of depression cases, incidence, and chronicity in a large occupational cohort using machine learning techniques: an analysis of the ELSA-Brasil study. *Psychological Medicine* 1–9. https://doi.org/10.1017/S0033291720001579

Received: 7 October 2019
Revised: 26 April 2020
Accepted: 11 May 2020

Key words:
Incident depression; machine learning; major depressive disorder; prognosis

Author for correspondence:
Andre Russowsky Brunoni,
E-mail: brunoni@usp.br

Abstract

Background. Depression is highly prevalent and marked by a chronic and recurrent course. Despite being a major cause of disability worldwide, little is known regarding the determinants of its heterogeneous course. Machine learning techniques present an opportunity to develop tools to predict diagnosis and prognosis at an individual level.

Methods. We examined baseline (2008–2010) and follow-up (2012–2014) data of the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil), a large occupational cohort study. We implemented an elastic net regularization analysis with a 10-fold cross-validation procedure using socioeconomic and clinical factors as predictors to distinguish at follow-up: (1) depressed from non-depressed participants, (2) participants with incident depression from those who did not develop depression, and (3) participants with chronic (persistent or recurrent) depression from those without depression.

Results. We assessed 15 105 and 13 922 participants at waves 1 and 2, respectively. The elastic net regularization model distinguished outcome levels in the test dataset with an area under the curve of 0.79 (95% CI 0.76–0.82), 0.71 (95% CI 0.66–0.77), 0.90 (95% CI 0.86–0.95) for analyses 1, 2, and 3, respectively.

Conclusions. Diagnosis and prognosis related to depression can be predicted at an individual subject level by integrating low-cost variables, such as demographic and clinical data. Future studies should assess longer follow-up periods and combine biological predictors, such as genetics and blood biomarkers, to build more accurate tools to predict depression course.

Introduction

Mood disorders account for almost 50% of the burden of mental disorders (Kupfer, Frank, & Phillips, 2012). Among them, major depression has a chronic, recurrent course, and is highly prevalent. In fact, its chronicity is especially difficult to tackle, as research is usually focused on the management of acute depressive episodes (Andrews, 2008). In addition, depression has heterogeneous trajectories, varying from a single or a few episodes to an intermittent course that can persist over the lifespan (Musliner, Munk-Olsen, Eaton, & Zandi, 2016).

However, relatively little is known on the sociodemographic and clinical predictors associated with depression recurrence and incidence, as most available cohort data present limitations. For instance, several cohort studies investigated specific subgroup of patients (e.g. perinatal depression, geriatric depression, and depression in children and adolescents), enrolled only those already depressed at baseline, performed a short-term follow-up, or presented high attrition rates (Beard, Tracy, Vlahov, & Galea, 2008; Musliner et al., 2016; Skapinakis, Weich, Lewis, Singleton, & Araya, 2006; Spijker et al., 2004). An additional issue is that most studies have been conducted in developed countries (Musliner et al., 2016). It is reasonable to assume that the course of depression is different in low- and middle-income countries that present substantial economic disparity and low social support for the poorest people.

Furthermore, standard investigation has focused on traditional statistical approaches focused on group-level results (Bzdok, Altman, & Krzywinski, 2018). In this context, machine

learning approaches can be advantageous and have increasingly been used in prognostic psychiatry, as they can model complex datasets to produce accurate predictive tools based on learning algorithms. In addition, machine learning focuses on results at the individual patient level (Bzdok et al., 2018; Dwyer, Falkai, & Koutsouleris, 2018). For example, there is supporting evidence for the influence of socioeconomic inequality in the association between depression and gender (Rai, Zitko, Jones, Lynch, & Araya, 2013).

In this study, we used a machine learning approach to predict depression cases, incidence, and chronicity in a large Brazilian occupational cohort, using clinical and sociodemographic data. This study has the potential to aid in mental health policies development. Additionally, our findings could be employed in other low- and middle-income countries that present populations with similar characteristics.

Methods

Study design and participants

ELSA-Brasil is an occupational, prospective cohort study of 15 105 civil servants from six public institutions in major Brazilian cities (São Paulo, Rio de Janeiro, Salvador, Porto Alegre, Belo Horizonte, and Vitória) (Aquino et al., 2012). All active or retired employees of these institutions aged 35–74 years were eligible for the study. Exclusion criteria were current or recent pregnancy (4 months prior to the first interview), intention to quit working at the institution in the near future, severe cognitive or communication impairments, and, if retired, residing outside of a study center's corresponding metropolitan area. All local ethics committees approved the study and all participants provided written, informed consent prior to assessment. Since this was an observational cohort, subjects were informed of all their clinical and mental diagnoses and referred to an appropriate medical appointment, but no intervention was provided by the study.

The first wave ($n = 15\,105$ participants) of ELSA took place from August 2008 to December 2010 and the second wave ($n = 13\,922$ participants) took place from September 2012 to December 2014.

Predictor variables

As predictors, we selected variables that are easily accessible to clinicians and that can be collected in a single clinic visit. In consequence, the models created can be used in large populations without a significant increase in the cost of assistance.

The following baseline variables were investigated as predictors:

(a) For sociodemographic variables, information was collected regarding sex, age, educational level (presence or absence of a university degree), self-reported race (white *v.* non-white), marital status (married *v.* other), and familial monthly income.
(b) Regarding clinical variables, we assessed obesity (defined as a body mass index >30 kg/m^2 and obtained by measured weight and height) and smoking status (never a smoker *v.* past or present smoker). To evaluate general health status, participants were asked to judge their health according to a Likert scale (Chor et al., 2013). The answers were categorized into very good/good health status *v.* moderate/poor/very poor health. Finally, we used dietary information to identify those who presented a heavy alcohol consumption (Chor et al., 2013), defined as more than 210 (men) or 140 (women) grams of alcohol consumed per week (equivalent to 15 and 10 glasses of drink per week, respectively) (Piccinelli et al., 1997).
(c) Regarding mental disorders, we used the Portuguese version of the Clinical Interview Schedule-Revised (CIS-R) (Nunes et al., 2016), which is a structured interview for measurement and diagnosis of non-psychotic psychiatric morbidity in the community (Lewis, Pelosi, Araya, & Dunn, 1992). The questionnaire includes 14 sections covering common psychiatric symptoms and assessing the following ICD-10 diagnosis: general anxiety disorder (GAD, F41.1), panic disorder (PD, F41.0), social anxiety disorder (SAD, F40.1), and obsessive-compulsive disorder (F42). The sensitivity/specificity of the CIS-R is 74/98% for any mental disorder and 75/98% for depressive episodes based on the ICD-10 diagnostic criteria (Head et al., 2013).
(d) For psychotropic use, all participants were asked regarding the use of prescription and non-prescription medicines, and continuous and non-continuous use of medication taken in the past 2 weeks. All participants were instructed to bring to the study clinic all medications and prescription forms for examination. We assessed whether participants were using antidepressants and/or benzodiazepines as a categorical variable (yes or no). A complete review on psychotropic use in the ELSA-Brasil study can be found elsewhere (Brunoni et al., 2013).
(e) Finally, we assessed the presence or absence of at least one of the following negative life events in the past 12 months: being assaulted or robbed, being hospitalized, bereavement/mourning of a relative, severe financial problems, or ending an intimate relationship.

Outcome variables

At both waves, a shortened version of the CIS-R was applied to diagnose depression. Therefore, we could define the following clinical courses: no depression (absence of depression at both waves), incident depression (depression only at wave 2), and chronic depression (depression at both waves). Of note, the CIS-R limits detailed enquiry to the previous week of the assessment, since memory for psychological symptoms and the validity of the responses becomes poor when a longer period of enquiry is used in community samples (Das-Munshi, Castro-Costa, Dewey, Nazroo, & Prince, 2014). Since the CIS-R does not evaluate the entire period between baseline and follow-up, it is not possible to differentiate if a subject that was depressed at both waves had a persistent or a recurrent course. In addition, due to the low number of participants with depression at the baseline, we did not develop models including only depressed subjects.

Data analysis

Descriptive analyses were reported as means (with standard deviations) or absolute and relative frequencies. We divided participants into four groups based on the outcomes. We used χ^2 or Student *t* tests to analyze demographic and clinical variables among these groups.

The machine learning analysis was performed with R software (Version R 3.3.1) and R Studio (Version 0.99.902) using the R

package caret (Version 6. 0-73) (R Core Team, 2018). Machine learning approaches may outperform traditional multiple regression: (1) coefficients are unstable when high correlations exist among predictors, which leads to low replication of predictions in independent samples (Berk, 2016); and (2) traditional regression assumes additivity, whereas the predictors considered here might have non-additive effects.

Machine learning analysis

The elastic net is a machine learning method that uses regularization with an embedded feature selection procedure. Through a cost function composed of both L1 (Lasso regression) and L2 (Ridge regression) weight magnitude penalties, the method can remove predictors with low impact to the outcome while regularizing for improved generalization. The coefficients of the non-relevant features are shrunk toward zero, eliminating correlated variables, simplifying the model, and reducing overfitting. As our dataset is composed of several attributes, identifying the most important ones enables a wider applicability and more practical use of our predictive models. We performed bivariate elastic net regularization to explore the association of the predictive variables and the outcome. The training procedure was performed with 10-fold classification, feature selection, hyperparameter tuning, and sampling correction. To select the hyperparameters α and λ, we selected a sequence of values from 0.1 to 0.9 in intervals of 0.1 (α), and from 0.001 to 0.1 in intervals of 0.001 (λ) and tested all combinations of these values in cross-validation loop. The hyperparameters combination that maximized the AUC was selected for the final model.

Using both wave 1 and wave 2, we defined the following outcomes: (1) no depression (absence of depression at baseline and follow-up); (2) any depression (incident, chronic, or remitted depression); (3) incident depression (depression at wave 2 but not at wave 1); and (4) chronic depression (depression at both waves). Then, the following models were developed to distinguish: (A) no depression *v.* any type of depressive course (incident or chronic or remitted); (B) no depression *v.* incident depression; and (C) no depression *v.* chronic depression. All variables were normalized or standardized before being used in the elastic net equation. We performed the missing data imputation by using median for numeric variables and mode for categorical variables, using the training dataset (van Buuren, 2018).

Individual-level predicted probabilities based on the elastic net algorithm were created, as well as the receiver operating characteristic (ROC) curves, and the area under the curve (AUC) was calculated to evaluate the predictive performance. Additionally, we calculated sensitivity, specificity, balanced accuracy, positive predictive value (PPV), and negative predictive value (NPV). We used a cut-off of 0.5 as the boundary for class decision, i.e. the algorithm will classify probabilities above 50% as belonging to the positive outcome level and below to the negative outcome level. Finally, we plotted how PPV and NPV change *vis-à-vis* different cut-offs for class boundary decision.

Cross-validation

For each analysis, we randomly split our baseline data into training (75% of the whole sample) and test datasets (25%). We deployed a standard machine learning protocol with 10-fold cross-validation, feature selection, hyperparameter tuning, and class imbalance correction in the training dataset (Fig. 1). We repeated 10-fold cross-validation 10 times to improve tuning.

Class imbalance

Class imbalance introduces a bias toward classifying all the data as the majority class, which usually leads to poor detection of the infrequent class. The class imbalance problem was addressed through a resampling step, which entailed under-sampling the majority class in each analysis followed by algorithm training. Specifically, we used down-sampling, a procedure that randomly selects instances of the majority class to match class frequencies of the minority class. This process was repeated in 1000 iterations to allow us to use all instances in the training set. The algorithm-predicted probabilities were averaged over the resampling iterations. In unbalanced sets, the regular accuracy can be a misleading measure of performance, biased toward predicting the majority class. There is a possibility, for example, for the classifier to simply predict all cases of the majority class, leading to an artificial high performance that, on the other hand, is unable to predict most or all instances of the minority class, which is the event we were initially trying to detect (Brodersen, Ong, Stephan, & Buhmann, 2010).

The balanced accuracy can be defined as $1/2((TP/P) + (TN/N))$ based on a confusion matrix of predicted instances *v.* actual instances, with *TP* being true-positive cases, *TN* true-negative cases, *P* all positive cases, and *N* all negative cases. This way, if the algorithm is poor to detect either the true-positive or the true-negative cases, the balanced accuracy will be low and close to chance. Therefore, when dealing with unbalanced datasets, AUC and balanced accuracies are more proper performance metrics to evaluate a machine learning classifier than the traditional accuracy values (Buda, Maki, & Mazurowski, 2018; Luque, Carrasco, Martín, & de las Heras, 2019).

Supplementary analysis

In addition to the main analysis, eight more models are available in the online Supplementary material (online Supplementary Tables S4–S7), including models: (a) excluding subjects with missing data in the generalized anxiety disorder variable, (b) with no correction for class imbalance, (c) using the random forest algorithm, (d) with 100 random splits for train/test, (e) with sensitivity analyses excluding SAD, GAD, and OCD variables, one at a time, (f) using the least absolute shrinkage and selection operator (LASSO) algorithm. The selected α and λ parameters for the main models (online Supplementary Table S8) and the elastic net regression penalized β coefficients for the main models (online Supplementary Table S9) are also available in the online Supplementary material.

Results

Out of the 15 105 participants included at wave 1, 1180 (7.8%) did not complete the assessment at wave 2, the main reasons being death and moving outside of the metropolitan area of the study after retiring. We found that 499 (3.58%) participants presented with a new depressive episode, 426 (3.06%) remitted, 160 (1.15%) persisted or recurred in a depressive episode, and 12 837 (92.21%) presented no depression at both waves. Descriptive analyses of demographic and clinical variables are described in online Supplementary Table S1, and missing data frequency and distribution for each variable are presented in online Supplementary Fig. S1 and Table S2.

Table 1 shows model performance for each analysis when the cut-off for class boundary decision is set at a 0.5 probability, i.e. if the individual has a probability $\geqslant$0.5, it is classified as positive,

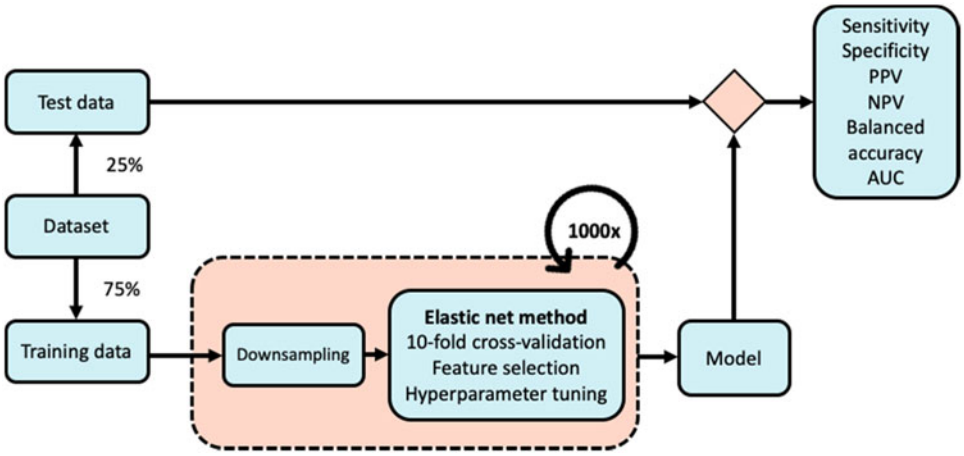

Fig. 1. Elastic net procedure for training and testing data.

Table 1. Performance metrics for the elastic net models to predict the three clinical outcomes

Model	Sensitivity	Specificity	PPV	NPV	Balanced accuracy	AUC
A	0.67	0.78	0.19	0.97	0.73	0.79 (0.76–0.82)
B	0.61	0.75	0.07	0.98	0.68	0.71 (0.66–0.77)
C	0.81	0.84	0.07	1.00	0.82	0.90 (0.86–0.95)

AUC, area under the curve; NPV, negative predictive value; PPV, predictive positive value.
(A) Depression *v.* non-depression; (B) incident depression *v.* non-depression; (C) chronic depression *v.* non-depression.

and when the probability is <0.5, as negative. Figure 2 shows the ROC curves and AUC values for the predictive models regarding outcomes 1–3, and the selected variables with their relative relevance weights to each model. Online Supplementary Fig. S2 and Table S3 show PPVs and NPVs for different cut-offs of class boundaries. Supplementary analysis can be found in online Supplementary Tables S4–S7. Online Supplementary Table S8 shows the selected values for hyperparameters α and λ parameters selected in the main analysis.

Classifying depressed and non-depressed participants

Considering both baseline and follow-up, 1085 participants presented with a history of depression, while 12 387 participants have not experienced any depressive episode. The elastic net model had an AUC of 0.79 (0.76–0.82) with a balanced accuracy of 73%. The model retained all variables except past or present history of smoking. In the five top features selected, there were four psychiatric comorbidities (SAD, OCD, GAD, and PD) and the self-reported health evaluation.

Prediction of incident depression

There was a total of 499 participants with a new depressive episode at follow-up. The model was trained to differentiate these incidents cases from the non-depressed patients at wave 2. The model had an AUC of 0.71 (0.66–0.77) and a balanced accuracy of 68%. Among the five top variables, there were two comorbidities (OCD and GAD), two clinical features (use of antidepressants and use of benzodiazepines), and sex. Past or present history of smoking and educational level were discarded by the model.

Distinguishing chronic depression and non-depressed patients

At wave 2, 160 patients that were depressed at wave 1 persisted in a depressive episode. The model was trained to differentiate chronically (persistent or recurrent) depressed participants from those without a depressive episode, and had an AUC of 0.90 (0.86–0.95), with a balanced accuracy of 82%. OCD and GAD were the most relevant features, with sex, self-report health, and negative life events following. The only variable discarded by the model was the self-reported race.

Sensitivity analyses

Since the use of antidepressants may be a confounder, we repeated the analysis for outcomes 1 and 3 without this variable to check if the performance could be inflated by its inclusion. The results can be seen in Fig. 3 and Table 2. The model to distinguish depressed from non-depressed maintained the same AUC 0.79 (0.78–0.81), while the model to distinguish non-depressed participants from those with a chronic depressive course had an absolute increase from 0.90 (0.86–0.95) to 0.91 (0.89–0.94), although with a significant overlap in the confidence intervals. The same two variables previously excluded (history of smoking for model 1 and ethnicity for model 3) were also excluded in the sensitivity analysis models. A sensitivity analysis for GAD, SAD, and OCD for the three comparisons can also be found in the online Supplementary material.

Models with 100 random splits

To assess the stability of the predictive models, we repeated the analysis with 100 random splits for the training/testing sets

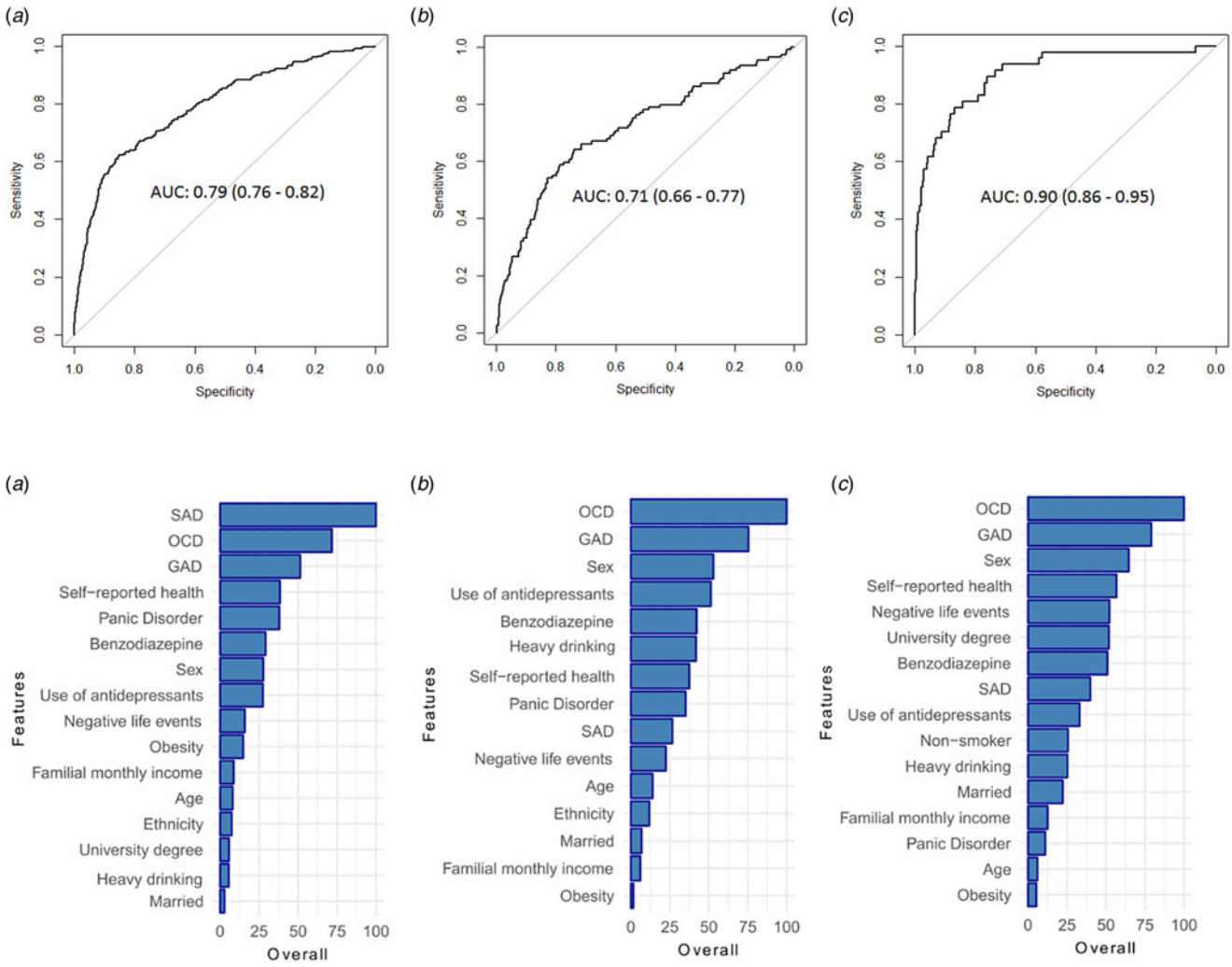

Fig. 2. ROC curve and AUC value for the predictive models of depression courses and variables selected by the elastic net model with relative relevance weights. Models differentiating (*a*) participants with depression from non-depressed participants; (*b*) participants with incident depression from participants who did not develop depression; (*c*) participants without depression from those with chronic depression.

(online Supplementary Tables S5–S7 and Fig. S4) obtaining similar results and confidence intervals for the AUC.

Discussion

The present study investigated three machine learning models for classifying cases of depression and predicting incidence, and chronicity of depression within the ELSA-Brasil cohort, using baseline variables from wave 1 (2008–2010) as predictors, and the courses defined at wave 2 (2012–2014). The present study is the first to assess depression prognosis in a large sample using machine learning techniques. Particularly, we designed predictive models to distinguish (a) participants with depression from those without depression; (b) participants with incident depression from those without depression; and (c) participants with chronic depression (persistent or recurrent) from those without depression. We obtained AUCs ranging from 0.71 to 0.90, and balanced accuracies ranging from 68% to 82%.

Our first predictive model can be used to screen subjects with or at-risk of developing depression in a populational sample with a small set of features, easily accessible to the clinician. The evaluation of only individuals with depression requires extensive screening and diagnostic procedures that usually incur in great costs and small samples, many times unsuited for a proper machine learning analysis. Our second model can detect patients who will develop depression in the follow-up, and then can be used to monitor these cases, enabling both preventive measures and early intervention. Finally, our third model can be used to detect which patients will have a chronic course of depression, marked by either a persistent or recurrent trajectory, which are more likely to have a course marked by functioning impairment and poor outcomes, and that are likely to need tertiary care. In the context of a developing country, these may help to better allocate resources and improve assessment.

The models presented low PPV, given the low prevalence of depression in the sample, and high NPV. Nevertheless, in the context of a developing country with scarce resources, the algorithm may serve as a screening tool that can aid to prioritize resources in cases at-risk. Two other studies also evaluated depression in population-based cohorts. Wang et al. assessed 28 059 individuals from waves 1 and 2 from the National Epidemiological Survey on Alcohol and Related conditions, obtaining a model with C

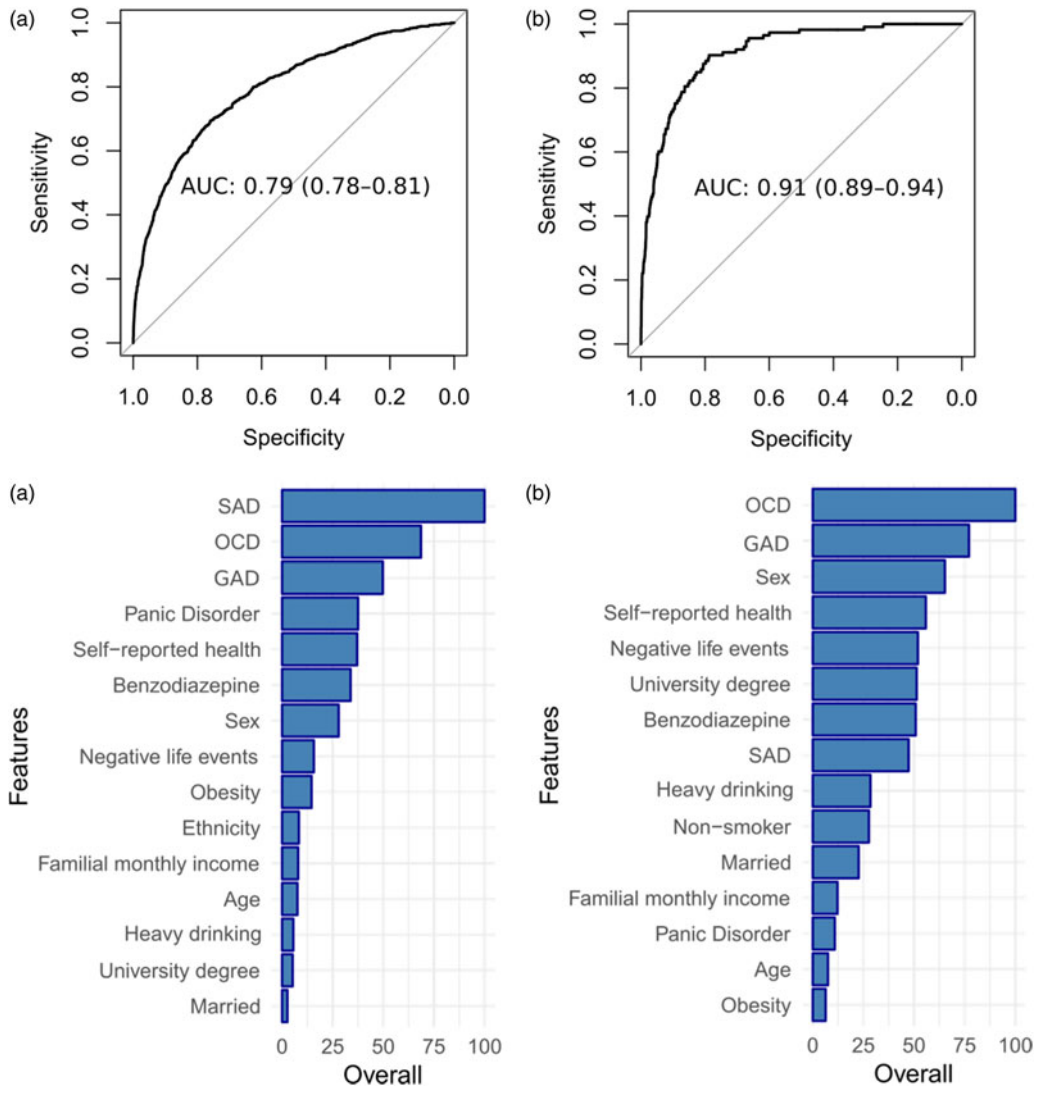

Fig. 3. ROC curves and selected variables with their relative relevance weights for sensitivity analysis for outcomes 1 and 3. (*a*) Depressed *v.* non-depressed patients. (*b*) Chronic depression *v.* non-depressed patients.

statistics of 0.7538 to discriminate those individuals that will develop a first-onset major depression (Wang et al., 2014). Rosellini et al. used the same dataset to predict the risk of several psychiatric disorders, obtaining an AUC of 0.73 to predict depression. We found no studies predicting depression in population-based cohorts in developing countries (Rosellini et al., 2020).

Predicting which individuals are at-risk to convert to depression can enable timely and personalized preventive strategies to take place, shifting our focus from only treating acute episodes to directly intervening in the course of the disorder. This may yield a substantial impact to ease the burden directly associated with depression, such as cognitive and functioning impairments (Lépine & Briley, 2011), high risk for suicidal behavior (Bostwick & Pankratz, 2000; Ösby, Brandt, Correia, Ekbom, & Sparén, 2001), and decreased quality of life (Brenes, 2007). In addition, it could impact also in mortality and disability rates, as well as in the economic and family burden associated with the disorder (Kessler, 2012). For example, depression is a risk factor for clinical diseases such as diabetes (Brown, Majumdar, Newman, & Johnson, 2005), coronary heart disease (Gan et al., 2014; O'Neil et al., 2016), and autoimmune diseases(Andersson et al., 2015), with patients being twice as likely to die prematurely when compared to subjects without depression (Lépine & Briley, 2011). Our findings show a potential application of machine learning in predicting incidence, persistence, and remission of depressive episodes at an individual level. In addition, the models developed in this study are easy to implement, since all variables can be accessed at any moment by a clinician, without incurring in additional costs. Interventions focused on the course of the disorder can be designed to target the relevant factors selected in the predictive models, since from all included variables, only age, sex, and ethnicity are not modifiable (Andersson et al., 2015).

Comorbidities were between the most relevant predictive features in all models. This is in accordance with a previous study by our group that showed large effect sizes for OCD and anxiety disorders to predict incident and persistent depression, using traditional statistical methods (Brunoni et al., 2020). The present study differs from our previous one as we employed more variables, tested more outcomes, and used a machine learning approach. Our findings are also in accordance with a recent

Table 2. Performance metrics for the sensitivity analysis for outcomes 1 and 3

Sensitivity analysis (SA)						
Model	Sensitivity	Specificity	PPV	NPV	Balanced accuracy	AUC
A-SA	0.67	0.78	0.21	0.96	0.72	0.79 (0.78–0.81)
C-SA	0.83	0.83	0.06	1.00	0.83	0.91 (0.89–0.94)
Comparison 1						
			No-depression			Depression
No antidepressants			12 143			921
Antidepressants			694			164
$\chi^2 = 161.4$, p value $<2.2 \times 10^{-16}$.						
Comparison 3						
			No persistent depression			Persistent depression
No antidepressants			12 143			126
Antidepressants			694			34

$\chi^2 = 72.06$, p value $<2.2 \times 10^{-16}$.

meta-analysis of 66 prospective studies that showed that anxiety disorders predict depressive disorder, with effect sizes of 2.58 (1.81–5.2) for GAD, 2.06 (1.71–3.97) for SAD, and 5.60 (4.21–6.01) for OCD (Jacobson & Newman, 2017). Some authors also consider the presence of high anxiety traits as a phenotype with an increased predisposition to stress-induced depression (Weger & Sandi, 2018). While SAD was the most important feature to differentiate depressed from non-depressed patients, it had an intermediate relevance for the other outcomes.

Regarding medication use, the use of benzodiazepines had an intermediate relevance for the three models. The use of antidepressants had an intermediate relevance for most models, except for the one predicting incident depression, in which it was the fourth more relevant feature. Of note, we decided to include use of antidepressants in all our models because patients may be using these for other reasons than being depressed, such as treatment of chronic pain, anxiety disorders, and obsessive-compulsive disorder. For example, there is evidence that subjects with GAD not treated with antidepressants have a higher risk to develop a depressive episode (Goodwin & Gorman, 2002). When performing a sensitivity analysis removing this variable for models 1 and 3, the first model had a mild decrease in performance, while the third remained at a similar value.

Among all sociodemographic features, the most relevant for all models was sex, being the third most relevant feature to predict incident and chronic depression, and the sixth more relevant to distinguish between depressed and non-depressed patients. The role of sex in depression is well-known, with women being twice as likely to develop depression (Kuehner, 2003), although there is no conclusive evidence of its role in remission, recurrence, or persistence (Salk, Hyde, & Abramson, 2017). The difference in incidence seems to be higher during adolescence, the period which was not included in our population. Having a university degree was the sixth more important variable to predict chronic depression, although it was discarded by the model of incident depression. Other sociodemographic variables had an intermediate to small relevance in the models. Age, for example, was among the five less relevant features in all models. This could be explained by the age range of our sample (35–74 years) and the fact the incidence is higher during adolescence and early adulthood (Saluja et al., 2004).

Depression rates observed in the present study were lower than expected in Brazil, in which higher depression rates have been reported (Brunoni et al., 2020). Our depression rates might be lower due to several reasons, such as differences in variables associated with depression (such as age and socioeconomic position) and occupational cohort characteristics, as civil servants enrolled have advantages such as job stability and healthcare access not necessarily available to the Brazilian population. Therefore, ecological replication of our findings is warranted.

Our study had some limitations. Since this is an occupational cohort, it is uncertain if the findings can be generalized to a community sample. Due to the nature of the sample, unemployment, which is known as a risk factor for both depression and a more pernicious course of depressive symptoms, could not be included as a predictor. Although the lack of a large set of features can be considered a limitation, since other variables could improve further the model performance, it can also be seen as an advantage, since a small set of features that are easy and fast to collect makes a more feasible tool that can be used in large populations with small costs. In addition, we had a large sample size available, which makes the machine learning process more robust. Another limitation lies in the fact that the machine learning model uses the features to predict the outcome class, but these features are not necessarily causal factors for it, and thus the directionality predictors–outcome is hard to establish, especially regarding psychiatric comorbidities. It is also important to notice that the selected variables and their relevance are bound to the model, population, and predictors selected, and using different algorithms or parameters may alter the selected features and their relevance. In this sense, although machine learning can provide some insights into pathophysiological mechanisms or risk factors, it is not an appropriate and conclusive analysis for this purpose. Finally, an important limitation is the short follow-up

period, which may have influenced the high rates of false positives found, and the fact that the CIS-R only evaluates the week before the assessment. Because of that, relevant data to the courses of the depressive disorder may be lost and even depressive episodes may not be accounted for, while for a more reliable determination of depressive trajectories, more frequent evaluations and longer periods of follow-up are required.

Conclusion

In the present study, we developed three predictive models of depressive course in an occupational cohort, using machine learning techniques. Using a small number of clinical and sociodemographic predictors, we showed that it is possible to distinguish non-depressed participants from those with depression, including incident and chronic cases, with high model performance. In addition, we also showed that clinical variables seem to be, at least for this sample, more relevant than sociodemographic variables. Knowing beforehand which individuals will have a depressive episode, and within these, which will have a more chronic and debilitating course, could help improve how we assess patients in clinical settings, shifting our focus from treating acute episodes to preventing them.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291720001579.

Financial support. The ELSA-Brasil study was supported by the Brazilian Ministry of Health and CNPq (grants 01060010.00RS, 01060212.00BA, 01060300.00ES, 01060278.00MG, 01060115.00SP, 01060071.00RJ).

Conflict of interest. The authors declare that they have no competing interests.

References

Andersson, N., Gustafsson, L., Okkels, N., Taha, F., Cole, S., Munk-Jørgensen, P., & Goodwin, R. (2015). Depression and the risk of autoimmune disease: A nationally representative, prospective longitudinal study. *Psychological Medicine*, *45*(16), 3559–3569. doi: 10.1017/S0033291715001488.

Andrews, G. (2008). Reducing the burden of depression. *The Canadian Journal of Psychiatry*, *53*, 420–427.

Aquino, E. M. L., Barreto, S. M., Bensenor, I. M., Carvalho, M. S., Chor, D., Duncan, B. B., … Szklo, M. (2012). Brazilian Longitudinal Study of Adult Health (ELSA-Brasil): Objectives and design. *American Journal of Epidemiology*, *175*, 315–324.

Beard, J. R., Tracy, M., Vlahov, D., & Galea, S. (2008). Trajectory and socioeconomic predictors of depression in a prospective study of residents of New York City. *Annals of Epidemiology*, *18*, 235–243.

Berk, R. A. (2016). *Statistical learning from a regression perspective*. Springer International Publishing / Springer Nature Switzerland AG.

Bostwick, J. M., & Pankratz, V. S. (2000). Reviews and overviews affective disorders and suicide risk: A reexamination. *American Journal of Psychiatry*, *157*, 1925–1932.

Brenes, G. A. (2007). Anxiety, depression, and quality of life in primary care patients. *Primary Care Companion to the Journal of Clinical Psychiatry*, *9*, 437–443.

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. *20th International Conference on Pattern Recognition*, pp. 3121–3124. IEEE.

Brown, L. C., Majumdar, S. R., Newman, S. C., & Johnson, J. A. (2005). History of depression increases risk of type 2 diabetes in younger adults. *Diabetes Care*, *28*, 1063–1067.

Brunoni, A. R., Nunes, M. A., Figueiredo, R., Barreto, S. M., da Fonseca, M. d. J., Lotufo, P. A., & Benseñor, I. M. (2013). Patterns of benzodiazepine and antidepressant use among middle-aged adults. The Brazilian longitudinal study of adult health (ELSA-Brasil). *Journal of Affective Disorders*, *151*, 71–77.

Brunoni, A. R., Santos, I. S., Passos, I. C., Goulart, A. C., Koyanagi, A., Carvalho, A. F., … Benseñor, I. M. (2020). Socio-demographic and psychiatric risk factors in incident and persistent depression: An analysis in the occupational cohort of ELSA-Brasil. *Journal of Affective Disorders*, *263*, 252–257.

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. doi: 10.1016/j.neunet.2018.07.011.

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, *15*, 233–234.

Chor, D., de Alves, M. G., Giatti, L., Cade, N. V., Nunes, M. A., del Molina, M. C. B., … de Oliveira, L. C. (2013). Questionnaire development in ELSA-Brasil: Challenges of a multidimensional instrument. *Revista de Saúde Pública*, *47*, 27–36.

Das-Munshi, J., Castro-Costa, E., Dewey, M. E., Nazroo, J., & Prince, M. (2014). Cross-cultural factorial validation of the Clinical Interview Schedule-Revised (CIS-R); findings from a nationally representative survey (EMPIRIC). *International Journal of Methods in Psychiatric Research*, *23*, 229–244.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*, 91–118.

Gan, Y., Gong, Y., Tong, X., Sun, H., Cong, Y., Dong, X., … Lu, Z. (2014). Depression and the risk of coronary heart disease: A meta-analysis of prospective cohort studies. *BMC Psychiatry*, *14*, 371. doi: 10.1186/s12888-014-0371-z.

Goodwin, R. D., & Gorman, J. M. (2002). Psychopharmacologic treatment of generalized anxiety disorder and the risk of major depression. *American Journal of Psychiatry*, *159*, 1935–1937.

Head, J., Stansfeld, S. A., Ebmeier, K. P., Geddes, J. R., Allan, C. L., Lewis, G., & Kivimäki, M. (2013). Use of self-administered instruments to assess psychiatric disorders in older people: Validity of the General Health Questionnaire, the Center for Epidemiologic Studies Depression Scale and the self-completion version of the revised Clinical Interview Sch. *Psychological Medicine*, *43*, 2649–2656.

Jacobson, N. C., & Newman, M. G. (2017). Anxiety and depression as bidirectional risk factors for one another: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *143*, 1155–1200.

Kessler, R. C. (2012). The costs of depression. *Psychiatric Clinics of North America*, *35*, 1–14.

Kuehner, C. (2003). Gender differences in unipolar depression: An update of epidemiological findings and possible explanations. *Acta Psychiatrica Scandinavica*, *108*, 163–174.

Kupfer, D. J., Frank, E., & Phillips, M. L. (2012). Major depressive disorder: New clinical, neurobiological, and treatment perspectives. *The Lancet*, *379*, 1045–1055.

Lépine, J. P., & Briley, M. (2011). The increasing burden of depression. *Neuropsychiatric Disease and Treatment*, *7*(Suppl 1), 3–7. doi: 10.2147/NDT.S19617.

Lewis, G., Pelosi, A. J., Araya, R., & Dunn, G. (1992). Measuring psychiatric disorder in the community: A standardized assessment for use by lay interviewers. *Psychological Medicine*, *22*, 465–486.

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231.

Musliner, K. L., Munk-Olsen, T., Eaton, W. W., & Zandi, P. P. (2016). Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes. *Journal of Affective Disorders*, *192*, 199–211.

Nunes, M. A., Pinheiro, A. P., Bessel, M., Brunoni, A. R., Kemp, A. H., Benseñor, I. M., … Schmidt, M. I. (2016). Common mental disorders and sociodemographic characteristics: Baseline findings of the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *Revista Brasileira de Psiquiatria*, *38*, 91–97.

O'Neil, A., Fisher, A. J., Kibbey, K. J., Jacka, F. N., Kotowicz, M. A., Williams, L. J., … Pasco, J. A. (2016). Depression is a risk factor for incident coronary heart disease in women: An 18-year longitudinal study. *Journal of Affective Disorders*, *196*, 117–124.

Ösby, U., Brandt, L., Correia, N., Ekbom, A., & Sparén, P. (2001). Excess mortality in bipolar and unipolar disorder in Sweden. *Archives of General Psychiatry*, *58*, 844.

Piccinelli, M., Tessari, E., Bortolomasi, M., Piasere, O., Semenzin, M., Garzotto, N., & Tansella, M. (1997). Efficacy of the alcohol use disorders identification test as a screening tool for hazardous alcohol intake and related disorders in primary care: A validity study. *BMJ*, *314*, 420–420.

Rai, D., Zitko, P., Jones, K., Lynch, J., & Araya, R. (2013). Country- and individual-level socioeconomic determinants of depression: Multilevel cross-national comparison. *British Journal of Psychiatry*, *202*, 195–203.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rosellini, A. J., Liu, S., Anderson, G. N., Sbi, S., Tung, E. S., & Knyazhanskaya, E. (2020). Developing algorithms to predict adult onset internalizing disorders: An ensemble learning approach. *Journal of Psychiatric Research*, *121*, 189–196.

Salk, R. H., Hyde, J. S., & Abramson, L. Y. (2017). Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms. *Psychological Bulletin*, *143*, 783–822.

Saluja, G., Iachan, R., Scheidt, P. C., Overpeck, M. D., Sun, W., & Giedd, J. N. (2004). Prevalence of and risk factors for depressive symptoms among young adolescents. *Archives of Pediatrics & Adolescent Medicine*, *158*, 760.

Skapinakis, P., Weich, S., Lewis, G., Singleton, N., & Araya, R. (2006). Socio-economic position and common mental disorders. *British Journal of Psychiatry*, *189*, 109–117.

Spijker, J., de Graaf, R., Bijl R, V., Beekman, A. T. F., Ormel, J., & Nolen, W. A. (2004). Determinants of persistence of major depressive episodes in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Journal of Affective Disorders*, *81*, 231–240.

van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). New York: Chapman and Hall/CRC. https://doi.org/10.1201/9780429492259COPY

Wang, J., Sareen, J., Patten, S., Bolton, J., Schmitz, N., & Birney, A. (2014). A prediction algorithm for first onset of major depression in the general population: Development and validation. *Journal of Epidemiology and Community Health*, *68*, 418–424.

Weger, M., & Sandi, C. (2018). High anxiety trait: A vulnerable phenotype for stress-induced depression. *Neuroscience & Biobehavioral Reviews*, *87*, 27–37.