

PREDICTING 3-YEAR PERSISTENT OR RECURRENT MAJOR DEPRESSIVE EPISODE WITH MACHINE LEARNING

BRUNO BRAGA MONTEZANO¹; AMANDA RODRIGUES FIALHO²; KAREN JANSEN³

¹Universidade Católica de Pelotas – brunodoyt@gmail.com

²Universidade Católica de Pelotas – amandarfialho@hotmail.com

³Universidade Católica de Pelotas – karen.jansen@ucpel.edu.br

1. INTRODUCTION

Depression is a common mental disorder throughout the world. It is estimated that more than 300 million people have this disorder. Among the worst outcomes, depression can lead to suicide, around 800,000 people die from suicide each year (World Health Organization, 2018). A large proportion of people diagnosed with major depressive disorder end up, throughout their lives, suffering from recurrent episodes of depression. Studies show a prevalence ranging between 7.3% and 31.5% of recurrent depressive episodes (Skodol, 2011; Bukh, 2016).

Studies show that severe depression can cause more harmful damages when compared to mild depression (Yuki Tomonaga, 2013; Bukh, 2016). The severity of the disease is related to high financial costs, both for the individual and for public bodies. Severe depressive episodes are associated with higher expenses with medication, hospitalization, loss of ability to work (Yuki Tomonaga, 2013). The severity of depression is related to a higher suicide risk (Hawton et al., 2013). Furthermore, subjects who had a severe condition in previous depressive episodes had a five and a half times greater risk of developing a recurrent episode compared to mild and moderate depression. Therefore, severe depression can be understood as a predictor for the recurrence of depressive episodes (Bukh, 2016).

Longitudinal designs are considered the most appropriate to identify the incidence and risk factors for disorders. In this sense, longitudinal studies have investigated risk factors that can predict the recurrence or persistence of depressive episodes, such as substance dependence, anxiety disorders and personality disorders (Skodol, 2011). Most studies assess the independent effects of these disorders for the recurrence or persistence of depressive episodes, using linear analyses; few studies with non-linear models were found in the literature. A study found better accuracy in a predictive model defined by machine learning than in the conventional model by logistic regression to predict persistence, chronicity, hospitalization, disability and suicide attempt among subjects with major depressive disorder (MDD).

A study with machine learning using elastic nets included anxiety disorders, demographic variables, life events and substance use as the model variables with the best ability to predict recurrence or persistence of depression (Librenza, 2020). The identification of predictors of recurrence and persistence of depressive episodes

in MDD can be an important ally to warn about risks and collaborate to clinical decisions.

Being aware of the predictive characteristics for the recurrence or persistence of depression can help in the construction of intervention models focused on more specific therapeutic processes for this population. In addition, the identification of predictive factors can help to identify subjects diagnosed with MDD, who, even after treatment or spontaneous remission of symptoms, have a higher risk of relapse. Therefore, the aim of the present study is to predict recurrent or persistent depressive episodes, in addition to predicting severe recurrent or persistent depressive episodes using a machine learning method.

2. METHODS

This is a prospective cohort study. The first phase took place between 2012 and 2015 and the second phase between 2017 and 2018. All participants signed an informed consent form before participating in the study.

Individuals aged between 18 and 60 years were evaluated in an outpatient service for Research and Extension in Mental Health at the Catholic University of Pelotas. Individuals were evaluated by trained psychologists to perform a clinical diagnostic interview - Mini International Neuropsychiatric Interview Plus (MINI-Plus). We initially evaluated 925 subjects, of which 585 were diagnosed with major depressive disorder in the first phase of the study. Of these 468 subjects were re-evaluated in the second phase of the study. Subjects diagnosed with bipolar disorder, psychotic disorders, suicide plan or attempt in the last month were excluded. Thus resulting in a total of 368 eligible subjects.

In the first wave of the study, a questionnaire prepared by the research team was conducted with questions related to sociodemographic data and questions related to mental disorders and family history of psychiatric disorders. The Brazilian Association of Population Studies (ABEP) instrument was used to assess issues related to socioeconomic status. We used the MINI-Plus to assess the psychiatric disorders.

The Alcohol Smoking and Substance Involvement Screening Test was used to assess substance use/abuse. The Millon III Multiaxial Clinical Inventory was used to assess personality traits. The Childhood Trauma Questionnaire was used to assess the history of abuse and/or neglect during childhood.

In the second wave, to assess the recurrent and/or persistent depressive episode, the Mini International Neuropsychiatric Interview Plus (MINI-Plus) was used. To assess the severity of depressive symptoms, the Beck Depression Inventory II (BDI II) was used, which aims to outline the presence and intensity of depressive symptoms.

The same sociodemographic, clinical, comorbid disorders and substance use variables were used as predictors in all predictive models and all predictor variables were collected at baseline.

Outcomes derived from the BDI II score in the follow-up.

We trained elastic net models using the glmnet package (Version 4.1.1) (Friedman et al., 2010) available through the caret package (Version 6.0.88) (Kuhn, 2021) in R (Version 4.1.0) (R Core Team, 2021). Three predictive models were developed, each considering a different approach, as follows: (1) prediction of recurrence/persistence of depression (no depression vs. any depression severity); (2) prediction of severe depression without moderate cases (no depression vs. severe depression); (3) prediction of severe depression without non-depressed subjects (mild and moderate depression vs. severe depression). The area under the curve (AUC) and balanced accuracy were used to measure model performance.

3. RESULTS AND DISCUSSION

368 subjects were included in the study. Of these, the majority were female (82.6%), the average age was 37 years, 77.7% were white, the average years of education was 11 years, 56.5% belonged to the low economic status, 58.7% lived with a partner, and 62.5% were currently studying or working. In regards to clinical symptoms, 51.1% presented suicide risk, 23.9% dependence or abuse of alcohol, marijuana 4.1%, other drugs 28.3%, hypnotics 5.2% and tobacco 22%. Most of the sample did not have anxiety disorders.

The present study consists of three models. The models results are summarized below:

- Predicting recurrent or persistent depression: The model showed an AUC of 0.78 [0.68; 0.88] with a balanced accuracy of 69%. The five predictors were: abuse/dependence of marijuana, suicide risk, cluster A and cluster B of personality disorders and specific phobia.
- Predicting severe depression versus non-depressed subjects: In order to avoid the problem of the great similarity of the intermediate groups (mild and moderate depression) with the outcome when modeling severe depression, we chose to drop the subjects with intermediate levels of depression in this second model. The predictive model had an AUC of 0.81 [0.70; 0.91] with a balanced accuracy of 71%. In the top five most predictive variables, there were: suicide risk, specific phobia, gender (female), abuse/dependence of marijuana and generalized anxiety disorder.
- Predicting severe depression among depressed subjects: The model performed with an AUC of 0.65 [0.50; 0.80] with a balanced accuracy of 62%. Top five predictive features consisted of: paternal psychiatric medication, agoraphobia, abuse/dependence of other drugs, obsessive-compulsive disorder and cluster A. This model, however, had a larger number of important variables in the prediction, possibly due to the great similarity of the outcome and non-outcome groups.

4. CONCLUSIONS

These are possibly innovative findings when developing models for predicting severe depression with machine learning, due to the advantage of understanding complex outcomes from these algorithms. In addition to preliminarily demonstrating the possibility of understanding the severity of a case of depression as a spectrum, even in predictive models. New studies are suggested to improve the creation of predictive models for the recurrence of mood episodes, especially depressive ones.

5. REFERENCES

- BUKH, J. D.; ANDERSEN, P. K.; KESSING, L. V.. Rates and predictors of remission, recurrence and conversion to bipolar disorder after the first lifetime episode of depression – a prospective 5-year follow-up study. **Psychological Medicine**, [S.L.], v. 46, n. 6, p. 1151-1161, 8 jan. 2016. Cambridge University Press (CUP). <http://dx.doi.org/10.1017/s0033291715002676>.
- HAWTON, Keith; COMABELLA, Carolina Casañas I; HAW, Camilla; SAUNDERS, Kate. Risk factors for suicide in individuals with depression: a systematic review. **Journal Of Affective Disorders**, [S.L.], v. 147, n. 1-3, p. 17-28, maio 2013. Elsevier BV. <http://dx.doi.org/10.1016/j.jad.2013.01.004>.
- LIBRENZA-GARCIA, Diego; PASSOS, Ives Cavalcante; FEITEN, Jacson Gabriel; LOTUFO, Paulo A.; GOULART, Alessandra C.; SANTOS, Itamar de Souza; VIANA, Maria Carmen; BENSEÑOR, Isabela M.; BRUNONI, Andre Russowsky. 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**, [S.L.], p. 1-9, 4 jun. 2020. Cambridge University Press (CUP). <http://dx.doi.org/10.1017/s0033291720001579>.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE. Determinantes Sociais e Riscos para a Saúde, Doenças Crônicas Não Transmissíveis e Saúde Mental. 2018. Disponível em: https://www.paho.org/bra/index.php?option=com_content&view=article&id=5671:folha-informativa-suicidio&Itemid=839. Acesso em: 28 jul. 2021.
- SKODOL, Andrew E.; GRILO, Carlos M.; KEYES, Katherine M.; GEIER, Timothy; GRANT, Bridget F.; HASIN, Deborah S.. Relationship of Personality Disorders to the Course of Major Depressive Disorder in a Nationally Representative Sample. **American Journal Of Psychiatry**, [S.L.], v. 168, n. 3, p. 257-264, mar. 2011. American Psychiatric Association Publishing. <http://dx.doi.org/10.1176/appi.ajp.2010.10050695>.
- TOMONAGA, Yuki; HAETTENSCHWILER, Josef; HATZINGER, Martin; HOLSBOER-TRACHSLER, Edith; RUFER, Michael; HEPP, Urs; SZUCS, Thomas D.. The Economic Burden of Depression in Switzerland. **Pharmacoeconomics**, [S.L.], v. 31, n. 3, p. 237-250, 16 fev. 2013. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s40273-013-0026-9>.