

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM PSIQUIATRIA E  
CIÊNCIAS DO COMPORTAMENTO

Bruno Braga Montezano

Predição de transtorno bipolar e desfechos funcionais  
em adultos jovens: um acompanhamento de cinco  
anos

Porto Alegre

2023

Bruno Braga Montezano

**Predição de transtorno bipolar e desfechos funcionais em adultos jovens: um acompanhamento de cinco anos**

Dissertação submetida ao Programa de Pós-Graduação em Psiquiatria e Ciências do Comportamento da Faculdade de Medicina da UFRGS, como requisito parcial para obtenção do título de Mestre em Psiquiatria e Ciências do Comportamento, com ênfase em Avaliação e Indicadores de Desfechos em Saúde Mental.

Orientador: Dr. Ives Cavalcante Passos  
Coorientadora: Dr<sup>a</sup>. Karen Jansen

Porto Alegre

2023

### CIP - Catalogação na Publicação

Montezano, Bruno Braga  
Predição de transtorno bipolar e desfechos  
funcionais em adultos jovens: um acompanhamento de  
cinco anos / Bruno Braga Montezano. -- 2023.  
151 f.  
Orientador: Ives Cavalcante Passos.

Coorientadora: Karen Jansen.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Faculdade de Medicina, Programa de  
Pós-Graduação em Psiquiatria e Ciências do  
Comportamento, Porto Alegre, BR-RS, 2023.

1. Transtorno bipolar. 2. Aprendizado de máquina.  
3. Incidência. 4. Medicina de precisão. 5. XGBoost. I.  
Passos, Ives Cavalcante, orient. II. Jansen, Karen,  
coorient. III. Título.

Bruno Braga Montezano

## **Predição de transtorno bipolar e desfechos funcionais em adultos jovens: um acompanhamento de cinco anos**

Dissertação submetida ao Programa de Pós-Graduação em Psiquiatria e Ciências do Comportamento da Faculdade de Medicina da UFRGS, como requisito parcial para obtenção do título de Mestre em Psiquiatria e Ciências do Comportamento, com ênfase em Avaliação e Indicadores de Desfechos em Saúde Mental.

Trabalho aprovado. Porto Alegre, 08 de setembro de 2023:

---

**Ives Cavalcante Passos**

Orientador

---

**Karen Jansen**

Coorientadora

---

**Adriane Ribeiro Rosa**

Relatora

---

**Marcelo Pio de Almeida Fleck**

Convidado 1

---

**Pedro Lemos Ballester**

Convidado 2

Porto Alegre

2023

*Este trabalho é dedicado à minha família.*

# Agradecimentos

Aos meus pais Eduardo e Renata pelos ensinamentos, parceria e absoluto suporte. À minha namorada, Camila, por todo apoio afetivo, cuidado e parceria. À minha irmã, Manuela, por sempre estar presente com boas energias e alegrando meus dias. Ao meu orientador, professor Ives Cavalcante Passos, por todo suporte durante o período do curso de mestrado como um guia nesta jornada. Estes agradecimentos se estendem à minha coorientadora Karen Jansen, que tem me apoiado desde o início de minha carreira acadêmica. À Claudia, secretária do Programa de Pós-Graduação em Psiquiatria e Ciências do Comportamento (PPGPSIQ) da UFRGS por toda ajuda com sua agilidade invejável. Também agradeço a Mila, do apoio administrativo do PPGPSIQ. Também agradeço à banca composta pelos doutores Adriane Rosa, Marcelo Fleck, Pedro Ballester e Daniela Braga por aceitarem o convite para ler e avaliar a presente dissertação, muito obrigado. Obrigado por confiarem em meu trabalho. Agradeço aos colegas e amigos que sempre ajudam como rede de apoio. Vocês são essenciais. Às instituições que fizeram este trabalho possível: Universidade Federal do Rio Grande do Sul, Hospital de Clínicas de Porto Alegre, e Universidade Católica de Pelotas.

*“O que permanece em seu estado natural não poderia ser infeliz;  
caso contrário,  
poder-se-ia dizer que o homem deve queixar-se de não voar como as aves,  
de não andar com quatro patas como os quadrúpedes,  
de não ter a cabeça armada de chifres como os touros.”  
(Erasmus de Roterdã, 1511)*

# Resumo

O transtorno bipolar (TB) é uma doença psiquiátrica crônica associada à altas taxas de morbidade e mortalidade. Estudos anteriores demonstram redução significativa da expectativa de vida, além de risco aumentado para doença cardiovascular e morte por suicídio. Apesar de ser um transtorno com início precoce, existe um atraso de até 10 anos entre o início de sintomas e o diagnóstico adequado. Como consequência do crescimento da psiquiatria de precisão, pesquisas têm explorado o uso de técnicas de aprendizado de máquina para prever TB, com foco em diagnóstico diferencial. No entanto, grande parte destes estudos são baseados em amostras clínicas pequenas, com curtos períodos de acompanhamento. A presente dissertação visa construir um modelo de classificação binária capaz de prever casos incidentes de TB em um intervalo de cinco anos através de características sociodemográficas e clínicas em uma amostra de adultos jovens, a partir de um grande estudo de coorte populacional. Avaliamos 1.091 sujeitos sem TB com 18 a 24 anos de idade no *baseline* a partir de uma amostra comunitária de jovens adultos da cidade de Pelotas (RS). O diagnóstico de TB no *follow-up* foi construído com base na *Mini International Neuropsychiatric Interview 5.0*. Cento e noventa preditores demográficos, sociais, clínicos e ambientais foram incluídos no *pipeline* de pré-processamento e modelagem. Utilizamos o algoritmo *XGBoost*, estado-da-arte para dados tabulares, com validação cruzada *5-fold* repetida por cinco vezes junto à seleção de variáveis e métodos de sobreamostragem para criar um modelo que pudesse prever quais sujeitos desenvolveriam TB cinco anos mais tarde. Quarenta e nove indivíduos (4,49%) receberam um diagnóstico de TB cinco anos depois. O melhor modelo selecionado com base no processo de validação cruzada apresentou um desempenho aceitável com um valor de AUC de teste de 0,786 (IC 95%: 0,686, 0,887). O modelo final incluiu dez preditores, sendo eles: sentimento de fracasso, tristeza, episódio depressivo atual na primeira avaliação, problemas de estresse auto-relatado, autoconfiança, uso de cocaína ao longo da vida, status socioeconômico, frequência sexual, relacionamento com parceiro fixo e taquilalia. Um teste de permutação com 10.000 permutações demonstrou performance de AUC significativamente melhor do modelo construído comparado a classificadores aleatórios ( $p < 0,001$ ). Os resultados do estudo trazem insights relevantes no que tange à compreensão do TB como um fenômeno latente, em especial, considerando que a depressão maior é comumente a primeira manifestação da doença, em linha com sintomas depressivos sendo os principais preditores no modelo apresentado. Visando uma melhor caracterização do TB, sugerimos que estudos futuros concentrem-se em fazer um acompanhamento sistemático que leve em conta estas características durante outras etapas do desenvolvimento, bem como investir em estudos que utilizem populações de risco específicas. Além disso, a inclusão de dados digitais de saúde, informações biológicas e neuropsicológicas pode ajudar no aprimoramento de novos modelos preditivos.

**Palavras-chaves:** Transtorno bipolar. Aprendizado de máquina. *XGBoost*. Incidência. Medicina de precisão.



# Abstract

Bipolar disorder (BD) is a chronic psychiatric illness associated with high rates of morbidity and mortality. Previous studies demonstrate a significant reduction in life expectancy, as well as an increased risk for cardiovascular disease and death by suicide. Despite being an early-onset disorder, there is a delay of up to 10 years between symptom onset and adequate diagnosis. As a consequence of the growth of precision psychiatry, research has explored the use of machine learning techniques to predict BD, with a focus on differential diagnosis. However, a large portion of these studies are based on small clinical samples with short follow-up periods. The present thesis aims to build a binary classification model capable of predicting incident cases of BD within a 5-year interval through sociodemographic and clinical characteristics in a sample of young adults from a large population cohort study. We evaluated 1,091 individuals without BD aged 18–24 at baseline from a community sample of young adults in the city of Pelotas (RS). The diagnosis of BD in the follow-up was based on the Mini International Neuropsychiatric Interview 5.0. One hundred and ninety demographic, social, clinical, and environmental predictors were included in the preprocessing and modeling pipeline. We used the state-of-the-art XGBoost algorithm for tabular data, with 5-fold cross-validation repeated five times, along with variable selection and oversampling methods, to create a model that could predict which subjects would develop BD five years later. Forty-nine individuals (4.49%) received a BD diagnosis five years later. The best model based on the cross-validation procedure showed acceptable performance with a test AUC value of 0.786 (95% CI: 0.686, 0.887). The final model included ten predictors, namely, feeling like a failure, sadness, current depressive episode at baseline, self-reported stress problems, self-confidence, lifetime cocaine use, socioeconomic status, sexual frequency, relationship with a fixed partner, and tachylalia. A permutation test with 10,000 permutations demonstrated significantly better AUC performance of the built model compared to random classifiers ( $p < 0.001$ ). The study results provide relevant insights regarding the understanding of BD as a latent phenomenon, particularly considering that major depression is commonly the first manifestation of the disease, in line with depressive symptoms being the main predictors in the presented model. In order to better characterize BD, we suggest that future studies focus on systematic follow-up that takes these characteristics into account during other stages of development, as well as investing in studies that use specific at-risk populations. Additionally, the inclusion of digital health data, biological and neuropsychological information can help improve new predictive models.

**Keywords:** Bipolar disorder. Machine learning. *XGBoost*. Incidence. Precision Medicine.

# Lista de Figuras

Figura 1	– Flutuações de humor típicas de um quadro diagnóstico de transtorno de humor bipolar tipo II. . . . .	19
Figura 2	– Esquema ilustrativo dos mecanismos de transcrição do DNA e tradução do RNA mensageiro (mRNA) em proteína, processado chamado de expressão gênica. . . . .	22
Figura 3	– <i>Tradeoff</i> de viés-variância em modelos de aprendizado de máquina. Trata-se da propriedade de um modelo que a variância de um parâmetro estimado através de amostras pode ser reduzido ao aumentar o viés dos parâmetros estimados. . . . .	24
Figura 4	– Os quatro V's do <i>big data</i> . . . . .	25
Figura 5	– Fluxograma do tamanho amostral do estudo com adultos jovens de 18 a 24 anos da cidade de Pelotas/RS para exploração dos casos incidentes de transtorno bipolar. . . . .	33
Figura 6	– Diagrama do algoritmo de <i>gradient tree boosting</i> desde a entrada de dados até a predição. . . . .	37
Figura 7	– Frequência cumulativa do número de pacotes em R hospedados no <i>Comprehensive R Archive Network</i> de 2008 a 2022. . . . .	39
Figura 8	– Fluxo de trabalho com uso de repositórios <i>git</i> para versionamento de código. . . . .	40
Figura 9	– Distribuição das áreas abaixo da curva ROC (ROC AUC) a partir do teste de permutação com 10.000 rearranjos. A linha tracejada indica a AUC nos dados originais de teste (ROC AUC = 0,786). . . . .	43
Figura 10	– Histograma da área abaixo da curva ROC (ROC AUC) baseada em 1.000 divisões de conjuntos de treino e teste aleatórios. A média e IC 95% encontrados foram de 0,723 [0,719–0,726]. Esta análise demonstra a robustez da capacidade preditiva do modelo de <i>boosting</i> selecionado. A linha contínua indicata a média da ROC AUC e as linhas tracejadas indicam os IC 95% da média. . . . .	44

- Figura 11 – Valores SHAP para cada variável de acordo com valor da variável baseado no ajuste final do modelo na base de treino ( $n = 763$ ). Variáveis binárias devem ser interpretadas em vermelho como positivo e azul como negativo. O eixo  $y$  representa a variável, ordenada pela maior contribuição absoluta entre todas as observações, independentemente da direção da associação. Cada ponto representa um participante na base de treino. Em observações com valor SHAP menor que zero, a variável comportou-se como fator protetivo, de modo contrário, fator de risco. A cor de preenchimento dos pontos representa o valor da variável para um dado indivíduo (vermelho corresponde a maiores valores e azul corresponde a menores valores). Os valores são relativos sem interpretação a partir das estimativas individuais. . . . . 44
- Figura 12 – Gráfico de importância de variáveis baseado no modelo com remoção de variáveis relacionadas diretamente ao humor (ROC AUC de teste: 0,644). Histórico de transtorno mental materno e problemas de estresse auto-relatados foram as duas variáveis mais importantes do modelo. . . . . 46
- Figura 13 – Associações lineares do escore de severidade de depressão (escore total do BDI) no início do estudo com a) desempenho cognitivo (escore do MOCA) e b) prejuízo funcional (escore da FAST) cinco anos depois. As linhas de regressão ajustadas mostram uma correlação negativa da gravidade dos sintomas depressivos com o desempenho cognitivo ( $r = -0,18$ ,  $n = 959$ ), e uma correlação positiva com prejuízo funcional ( $r = 0,24$ ,  $n = 982$ ). Indivíduos com episódios atuais de humor no acompanhamento foram excluídos da análise ( $n = 107$ ), a fim de minimizar o efeito direto dos sintomas de humor nessas medidas, resultando em uma subamostra de 984 indivíduos. Abreviações: MOCA (*Montreal Cognitive Assessment*). FAST (*Functioning Assessment Short Test*). . . . . 46
- Figura 14 – Correlações do escore de severidade da depressão (escore total do Inventário de Depressão de Beck [BDI]) no início do estudo com prejuízo funcional (escore global da *Functioning Assessment Short Test* [FAST]) cinco anos depois, estratificado por a) sexo e b) risco de suicídio no início do estudo. a) Entre as mulheres, um coeficiente de correlação positivo ( $r$ ) de 0,283 sugere uma associação moderada. Por outro lado, entre os homens, a correlação ( $r = 0,174$ ) indica uma relação positiva mais fraca. b) Entre os indivíduos com risco de suicídio no início do estudo, observa-se forte correlação positiva ( $r = 0,47$ ). Em contraste, entre aqueles sem risco de suicídio, houve uma relação positiva moderada ( $r = 0,217$ ). Indivíduos com episódios atuais de humor no acompanhamento foram excluídos da análise ( $n = 984$ ), a fim de minimizar o efeito direto dos sintomas de humor no funcionamento. . . . . 47

- Figura 15 – Associação entre severidade dos sintomas depressivos na segunda avaliação e probabilidade predita pelo modelo preditivo de xgboost ( $n = 328$ ). Coeficiente de correlação de Pearson e  $p$ -valor associado estão descritos no corpo do gráfico. A curva foi ajustada a partir do método de *local polynomial regression* (LOESS). . . . . 48
- Figura 16 – Idade do primeiro episódio depressivo na vida (auto-relato) entre sujeitos com e sem risco de suicídio (baseado no módulo c da MINI). O gráfico mostra uma tendência que sujeitos com risco de suicídio apresentam início mais precoce de depressão ( $n = 31$ ). Oitenta e cinco sujeitos apresentaram um episódio depressivo no baseline e 54 apresentaram valores ausentes na idade de início da depressão, resultando em uma subamostra final de 31 indivíduos. O teste  $t$  de Student foi usado e sugere uma diferença significativa entre os grupos ( $\mu_{semrisco} = 21,76$ ,  $\mu_{comrisco} = 18,50$ ,  $\Delta = 3,26$ , IC 95% [0,32–6,21],  $t(20,20) = 2,31$ ,  $p = 0,032$ ,  $d = 1,03$ , IC 95% [0,09, 1,94]). Abreviações: MINI (*Mini International Neuropsychiatric Interview*). . . . . 49

# Lista de Tabelas

Tabela 1 – Construtos avaliados e quantidade de itens de cada instrumento incluído na modelagem de incidência de transtorno bipolar em jovens adultos. . . .	36
Tabela 2 – Hiperparâmetros finais do modelo de XGBoost para predição de incidência de transtorno bipolar após 5 anos. . . . .	42

# Lista de abreviaturas e siglas

AM	Aprendizado de máquina
ASSIST	<i>Alcohol, Smoking and Substance Involvement Screening Test</i>
AUC	<i>Area Under the Receiver Operating Characteristic Curve</i>
CI	<i>Confidence interval</i>
CRAN	<i>Comprehensive R Archive Network</i>
GBM	<i>Gradient boosting machine</i>
HCL-32	<i>Hypomania Checklist 32</i>
IA	Inteligência artificial
IC	Intervalo de confiança
MDQ	<i>Mood Disorders Questionnaire</i>
MINI	<i>Mini International Neuropsychiatric Interview</i>
AM	<i>Machine learning</i>
MRI	Ressonância magnética
OMS	Organização Mundial da Saúde
PLN	Processamento de linguagem natural
SCID	<i>Structured Clinical Interview for DSM-5 Disorders</i>
TB	Transtorno bipolar
TCC	Terapia cognitivo-comportamental
TDM	Transtorno depressivo maior
TEA	Transtorno do espectro autista
TUS	Transtorno por uso de substância
UCPel	Universidade Católica de Pelotas
UFRGS	Universidade Federal do Rio Grande do Sul
XGBoost	<i>Extreme Gradient Boosting</i>

# Lista de símbolos

$\alpha$	Letra grega alfa
$\beta$	Letra grega beta
$\mu$	Letra grega mu, convencionalmente usado para denotar média populacional ou valor esperado no campo da probabilidade e estatística

# Sumário

	<b>Apresentação</b> . . . . .	<b>17</b>
<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>19</b>
<b>1.1</b>	<b>Transtorno bipolar</b> . . . . .	<b>19</b>
<b>1.2</b>	<b>Aprendizado de máquina</b> . . . . .	<b>23</b>
<b>1.3</b>	<b>Psiquiatria de precisão</b> . . . . .	<b>24</b>
<b>1.4</b>	<b>Aprendizado de máquina e transtorno bipolar</b> . . . . .	<b>27</b>
<b>2</b>	<b>JUSTIFICATIVA</b> . . . . .	<b>30</b>
<b>3</b>	<b>OBJETIVOS</b> . . . . .	<b>31</b>
<b>3.1</b>	<b>Objetivo geral</b> . . . . .	<b>31</b>
<b>3.2</b>	<b>Objetivos específicos</b> . . . . .	<b>31</b>
<b>4</b>	<b>MÉTODOS</b> . . . . .	<b>32</b>
<b>4.1</b>	<b>Amostra e coleta de dados</b> . . . . .	<b>32</b>
<b>4.2</b>	<b>Variáveis</b> . . . . .	<b>33</b>
4.2.1	Desfecho . . . . .	33
4.2.2	Preditores . . . . .	34
4.2.2.1	Dados sociodemográficos e ambientais . . . . .	34
4.2.2.2	Dados sobre uso de substâncias . . . . .	35
4.2.2.3	Dados clínicos . . . . .	35
4.2.2.4	Dados sobre comportamento sexual . . . . .	36
4.2.2.5	Instrumentos psicométricos . . . . .	36
<b>4.3</b>	<b>Análise de dados</b> . . . . .	<b>37</b>
4.3.1	<i>Gradient tree boosting</i> . . . . .	37
4.3.2	<i>SHAP values</i> . . . . .	38
4.3.3	Implementação . . . . .	38
4.3.4	Versionamento . . . . .	40
<b>5</b>	<b>RESULTADOS</b> . . . . .	<b>41</b>
<b>6</b>	<b>ASPECTOS ÉTICOS</b> . . . . .	<b>50</b>
<b>6.1</b>	<b>Coleta e manejo de dados</b> . . . . .	<b>50</b>
<b>6.2</b>	<b>Predição de desfechos clínicos em saúde mental</b> . . . . .	<b>50</b>
<b>7</b>	<b>CONCLUSÃO</b> . . . . .	<b>52</b>



	<b>Bibliografia</b> . . . . .	<b>53</b>
	<b>APÊNDICES</b>	<b>67</b>
	<b>APÊNDICE A – OUTROS TRABALHOS</b> . . . . .	<b>68</b>
A.1	<b>Artigos publicados</b> . . . . .	<b>68</b>
A.2	<b>Artigos aceitos</b> . . . . .	<b>69</b>
A.3	<b>Artigos submetidos</b> . . . . .	<b>69</b>
A.4	<b>Capítulos de livro</b> . . . . .	<b>70</b>
A.5	<b>Preprints</b> . . . . .	<b>70</b>
	<b>APÊNDICE B – IMPLEMENTAÇÃO EM R</b> . . . . .	<b>71</b>
B.1	<b>Pré-processamento dos dados</b> . . . . .	<b>71</b>
B.2	<b>xgboost e validação cruzada</b> . . . . .	<b>74</b>
B.3	<b>Teste de permutação</b> . . . . .	<b>77</b>
	<b>APÊNDICE C – COMPROVANTE DE SUBMISSÃO DO ARTIGO</b> . . . . .	<b>79</b>
	<b>APÊNDICE D – ARTIGO 1</b> . . . . .	<b>80</b>
	<b>APÊNDICE E – MATERIAL SUPLEMENTAR DO ARTIGO 1</b> . . . . .	<b>113</b>
	<b>APÊNDICE F – FIGURAS DO CORPO DO ARTIGO</b> . . . . .	<b>147</b>

# Apresentação

O transtorno bipolar (TB) é uma doença psiquiátrica associada a uma alta morbidade e mortalidade [1]. A prevalência populacional ao longo da vida é de aproximadamente 2,4%, sendo 0,6% para TB tipo I, 0,4% para TB tipo II e 1,4% para indivíduos com quadros subsindrômicos [2]. Estudos anteriores descrevem que pacientes com TB apresentam uma redução significativa na expectativa de vida de cerca de 10 anos em relação à população geral [3]. Doença cardiovascular é o principal fator associado à mortalidade precoce no TB; ainda assim, mortes por suicídio são mais comumente observadas no TB do que em outras condições em saúde mental, com estes pacientes apresentando de vinte a trinta vezes maiores taxas de morte por suicídio em relação à população geral [1, 4, 3]. Além disso, pacientes com TB apresentam prejuízos funcionais e psicossociais significativos, também representando um importante custo financeiro [1]. Os Estados Unidos da América reportaram que os custos totais associados ao TB tipo I excederam US\$200 bilhões no ano de 2015 [5].

Mesmo que a maioria dos pacientes com TB apresentem sintomas clínicos antes dos 25 anos de idade, há um atraso significativo de 6–10 anos entre o início dos sintomas e o diagnóstico correto [6, 7]. Além disso, o diagnóstico tardio está associado a uma maior duração de doença não tratada, que está, em última análise, ligada a um prognóstico pior em termos de hospitalizações, funcionamento e recorrência de episódios de humor [8]. Altos índices de comorbidade psiquiátrica, diagnóstico diferencial complexo, início habitual com sintomas depressivos, e redução do comportamento de busca de ajuda são algumas das razões para o atraso no reconhecimento adequado do TB [9, 1, 6]. Entretanto, o diagnóstico de TB é eminentemente clínico, com evidências limitadas para apoiar o uso de neuroimagens ou biomarcadores de laboratório durante a investigação clínica [1].

Levando em consideração este contexto, a ascensão do campo da psiquiatria de precisão, com o uso de ferramentas de *big data* e aprendizado de máquina, representa uma promessa, que pode trazer uma revolução em termos de diagnóstico, seleção de tratamento e prognóstico na saúde mental [10, 11]. Até esta data, vários estudos exploraram o uso destas técnicas no TB, com base em fontes de dados distintas (incluindo neuroimagens, dados clínicos e sociodemográficos, biomarcadores periféricos, testes neuropsicológicos, dados genéticos, entre outros), sendo que a maioria destes modelos se concentra em tarefas de classificação que ajudam no diagnóstico diferencial entre o TB e outras condições psiquiátricas, como esquizofrenia, depressão grave e indivíduos saudáveis [12, 13]. No entanto, a maioria destes estudos apresentam desempenhos de classificação modestos, são baseados em amostras pequenas e clínicas, originárias de desenhos transversais para coleta de dados, ou apresentam períodos curtos de acompanhamento (*follow-up*) [12]. Todas essas limitações podem comprometer a generalização e a tradução dos resultados de tais investigações para ambientes clínicos e de saúde pública

[13], ainda impossibilitando uma implementação mais efetiva das descobertas até então.

Dessa forma, levando em conta estas lacunas, a presente dissertação consiste de um estudo longitudinal de acompanhamento de jovens adultos em um intervalo de cinco anos, onde investigou-se a incidência de TB a partir do uso de técnicas de *tree boosting* para classificação de dados tabulares. O artigo resultante da pesquisa foi submetido à revista *Bipolar Disorders* com o título “*Bipolar disorder prediction and functional outcomes in young adults: a 5-year follow-up study*”. O principal trabalho da dissertação foi desenvolvido entre os anos de 2022 e 2023 no Laboratório de Psiquiatria Molecular, localizado no Centro de Pesquisas Experimentais do Hospital de Clínicas de Porto Alegre (UFRGS), sob orientação do professor Ives Cavalcante Passos (UFRGS) e coorientação da professora Karen Jansen (UCPel).

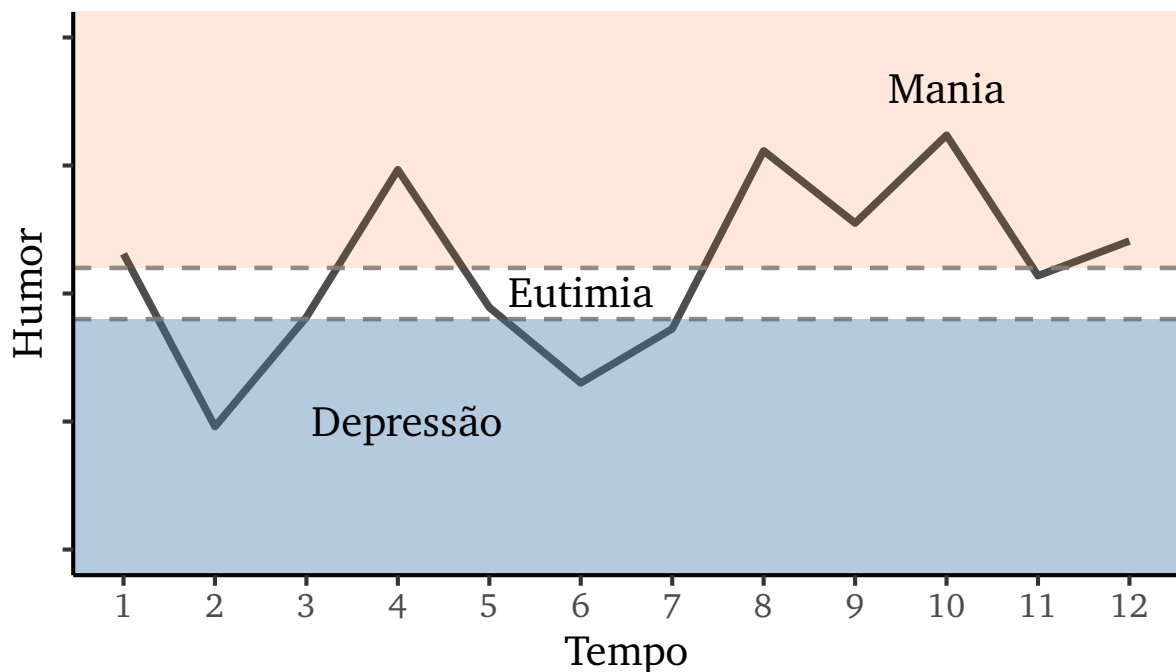
# 1 Introdução

## 1.1 Transtorno bipolar

O transtornos bipolares são um grupo complexo de transtornos crônicos e severos que incluem o TB tipo I, caracterizado pela presença de um episódio maníaco, e o TB tipo II, caracterizado pela presença de um episódio hipomaníaco e um episódio depressivo maior [1].

Os episódios de humor dentro de um quadro de TB podem flutuar de modo não-linear em diferentes intensidades (Figura 1). Os episódios maníacos e hipomaníacos são caracterizados por períodos de aumento exacerbado de energia e atividade ou nos níveis de irritabilidade, podendo estar acompanhado de delírios de grandiosidade, diminuição da necessidade do sono, fuga de ideias, taquilalia, distração aumentada, agitação psicomotora, maior impulsividade e exposição a comportamentos de risco. Já os episódios depressivos, presentes apenas nos quadros de TB tipo II, são identificados a partir de tristeza profunda, anedonia, alterações nos padrões de sono, mudanças no apetite, entre outros sintomas característicos [14].

Figura 1 – Flutuações de humor típicas de um quadro diagnóstico de transtorno de humor bipolar tipo II.



Fonte: Adaptado de Muzina, Colangelo, Manning e Calabrese [15].

Apesar da Iniciativa Mundial de Pesquisa em Saúde Mental<sup>1</sup> veicular estimativas de prevalência ao longo da vida e durante 12 meses de 2,4% e 1,5% respectivamente, taxas de prevalências podem variar de acordo com o país por questões metodológicas ou diferenças culturais. Ademais, a prevalência de TB tipo I é similar entre homens e mulheres, enquanto o TB tipo II ocorre mais frequentemente entre mulheres [17].

Os instrumentos de rastreio de TB mais estudados são o *Mood Disorders Questionnaire* (MDQ) e o *Hypomania Checklist 32* (HCL-32). Ambos possuem sensibilidade (MDQ: 80%; HCL-32: 82%) e especificidade (MDQ: 70%; HCL-32: 57%) razoáveis, considerando o contexto de aplicação. Estas escalas devem ser utilizadas com intuito de mapear sujeitos potencialmente em risco para o desenvolvimento da doença, para então, serem encaminhados à uma avaliação clínica mais aprofundada quanto ao TB [1].

O TB, além de trazer consequências funcionais e cognitivas e perda de qualidade de vida [18, 19], tem como característica a permanência destes prejuízos também em sujeitos em períodos de eutímia [20]. Sabe-se que algumas características como número de episódios de humor, duração da doença e número de hospitalizações podem indicar maiores ou menores graus de déficits funcionais e cognitivos [19]. Agentes farmacológicos como lurasidona, mifepristona e eritropoietina, terapia cognitivo-comportamental (TCC) em combinação a psicoeducação, terapia interpessoal e de ritmo social e terapia de nutrição, exercício e bem-estar são possíveis meios de tratamento para atenuação dos sintomas previamente citados [21].

Além das consequências descritas, o TB possui a maior taxa de suicídio entre todos os transtornos psiquiátricos, com taxas de 20 a 30 vezes maiores do que a população geral [22]. Pesquisas apontam que até 20% dos indivíduos com TB (em sua maioria, não-tratados) morrem por suicídio, com taxas de 20–60% para ao menos uma tentativa de suicídio ao longo da vida [23]. Diversos estudos demonstram que o lítio apresenta efeito significativo na redução de tentativas de suicídio e mortes por suicídio quando comparado a antidepressivos e outros estabilizadores de humor em pacientes com TB [22, 24]. A eficácia também se estende à ideação suicida [25]. Esse efeito mostra-se mais eficaz no uso prolongado de lítio em relação ao uso no curto prazo, além de ter sua eficácia demonstrada na população geral com uso de lítio na água em estudos ecológicos [26].

Sabe-se que transtornos por uso de substância (TUS) são altamente comórbidos com TB, geralmente requerendo abordagens multidisciplinares de tratamento para atenuar os prejuízos no desfecho clínico. Existem algumas evidências limitadas que medicamentos como lítio, valproato e citicolina podem ser efetivos na redução do uso de substâncias, mas vários estudos no campo acompanham tamanhos amostrais pequenos [27]. Apesar do conhecimento em torno dos TUS e TB, uma metanálise baseada em 11 estudos de desenho transversal verificou

<sup>1</sup> A Iniciativa Mundial de Pesquisa em Saúde Mental é um projeto colaborativo da Organização Mundial da Saúde, da Universidade de Harvard, da Universidade de Michigan e de pesquisadores do mundo todo para coordenar a análise e a implementação de pesquisas epidemiológicas sobre distúrbios mentais e comportamentais e abuso de substâncias em todas as regiões da OMS [16].

uma associação fraca entre transtorno por uso de cannabis e tentativas de suicídio no TB [28].

Transtornos de personalidade do Cluster B (antissocial, borderline, histriônico e narcisista) possuem uma alta incidência entre pacientes com TB, prejudicando o prognóstico clínico. Além disso, por vezes o diagnóstico diferencial entre transtorno de personalidade antissocial e TB é dificultado por conta dos possíveis comportamentos agressivos e impulsivos em ambas doenças, juntamente à outros sintomas sobrepostos [29].

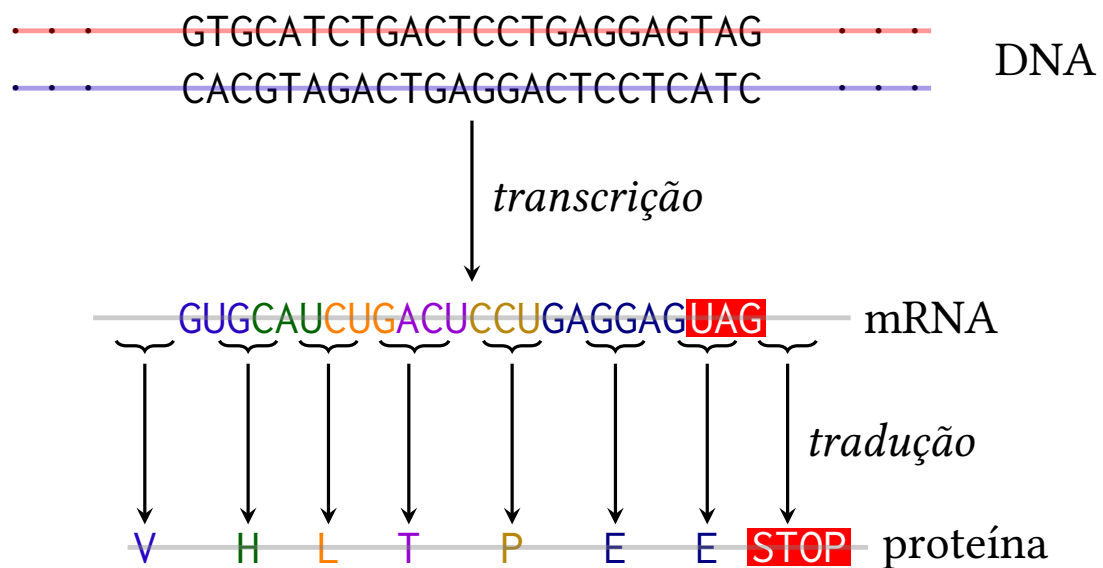
O transtorno do espectro autista (TEA) mostra-se como um grupo heterogêneo de transtornos neurodesenvolvimentais que afeta em torno de 2,6% da população. Sabe-se que a prevalência de TB em sujeitos no TEA é de aproximadamente 5 a 8%. Em geral, a sintomatologia do transtorno bipolar em sujeitos com TEA são inespecíficas e acabam diferindo do observado na população geral, dificultando o diagnóstico adequado. Logo, instrumentos psicométricos específicos são ainda um campo de desenvolvimento no diagnóstico de TB em indivíduos com TEA [30].

Um estudo de revisão de literatura, Lucidi et al. [31] buscaram mapear possíveis associações de características da microbiota intestinal com o TB. Alguns marcadores como por exemplo menor diversidade microbiana podem indicar vulnerabilidade para o desenvolvimento de TB em sujeitos geneticamente predispostos. Pacientes com TB também apresentam menores quantidades de microorganismos anti-inflamatórios do que pacientes controle. Em geral, estudos observacionais verificam maiores níveis de desregulação na microbiota em sujeitos com TB.

No início de 2022, uma revisão sistemática e metanálise de ensaios clínicos randomizados foi publicada a fim de avaliar o tratamento de lítio no TB em adultos. O estudo demonstrou que ainda, mesmo após anos de estudo, o lítio segue sendo o padrão ouro para o tratamento de TB. O artigo identificou eficácia do lítio no tratamento de mania aguda mesmo com sintomas psicóticos concomitantes, além de eficácia condicionada a outros agentes para depressão bipolar aguda, e eficaz como monoterapia na fase de manutenção para prevenir mania [25]. Prevenção de novos episódios a partir do uso de psicofármacos [32]. Também no ano de 2022, uma revisão sistemática analisou a eficácia de diferentes tratamentos para o TB de acordo com a trajetória de doença [33].

Levando em conta que o TB é uma doença psiquiátrica complexa e de causa multifatorial, sabemos que existe uma base genética significativa [1]. Nos últimos anos, a genômica tem se mostrado uma ferramenta útil para compreender estes mecanismos [34]. Estudos genéticos têm identificado variantes genéticas associadas ao TB e sugerido diferentes vias biológicas envolvidas na patogênese do transtorno. Além disso, a genômica também tem proporcionado novas oportunidades para o desenvolvimento de biomarcadores diagnósticos, preditivos e terapêuticos. Diversos estudos de genômica do TB foram realizados em todo o mundo, incluindo estudos de associação do genoma completo (GWAS) [35], assim como análises de expressão gênica [36] (Figura 2). No entanto, a compreensão completa da genômica do TB ainda está em

Figura 2 – Esquema ilustrativo dos mecanismos de transcrição do DNA e tradução do RNA mensageiro (mRNA) em proteína, processado chamado de expressão gênica.



evolução, e novas pesquisas são necessárias para identificar os fatores genéticos subjacentes e melhorar a compreensão da biologia da doença.

Com o intuito de avançar a compreensão da etiologia biológica do transtorno bipolar, um estudo de associação genômica realizado por Mullins et al. [37] com 41.917 casos de TB e 371.549 controles foi capaz de identificar 64 loci genômicos associados. Encontrou-se enriquecimento significativo de sinais em genes que codificam alvos de antipsicóticos, bloqueadores de canais de cálcio, antiepilépticos e anestésicos. Além disso, a integração de dados de loci de traços quantitativos de expressão implicou 15 genes fortemente ligados ao TB por meio da expressão gênica, codificando alvos potenciais para terapias, como HTR6, MCHR1, DCLK3 e FURIN. As análises dos subtipos do TB indicaram alta correlação genética entre o TB tipo I e II e identificaram loci associados adicionais. Estes dados trazem novas informações sobre a etiologia biológica do transtorno e demonstram genes de interesse para futuros estudos.

Exposição à poluição do ar tem sido associada a uma série de problemas de saúde, incluindo transtornos de humor e risco aumentado de suicídio. De acordo com um estudo de 2019 publicado por Braithwaite et al. [38], a exposição crônica à poluição do ar pode alterar o equilíbrio de neurotransmissores no cérebro e levar a mudanças na função cerebral que estão associadas a transtornos de humor, como depressão e ansiedade. Além disso, a poluição do ar também pode aumentar o estresse oxidativo, o que pode agravar a inflamação crônica e aumentar o risco de doenças psiquiátricas. O estudo de Braithwaite et al. [38] também encontrou uma associação entre a exposição à poluição do ar e o risco aumentado de suicídio, sugerindo que a saúde mental pode ser afetada negativamente pela poluição do ar. É importante destacar que esses achados são apenas uma parte da literatura em desenvolvimento sobre o assunto e

que mais estudos precisam ser feitos para entender completamente a natureza da associação entre poluição do ar e transtornos de humor e suicídio. Para além disso, outros fatores ambientais podem ser preditores para o desenvolvimento de TB. Estes incluem estresse psicossocial [39], exposição à luz artificial [40] e alterações climáticas [41, 42]. A interação complexa entre fatores genéticos e ambientais também pode influenciar o desenvolvimento do TB. A questão multifatorial enfatiza a necessidade de considerar uma ampla gama de fatores ao avaliar o risco de desenvolvimento de transtornos de humor, incluindo o TB. Além disso, é importante lembrar que a idade de início, a gravidade e a apresentação dos sintomas do TB podem variar amplamente entre indivíduos e que uma abordagem personalizada para o tratamento pode ser necessária para levar em conta essas diferenças.

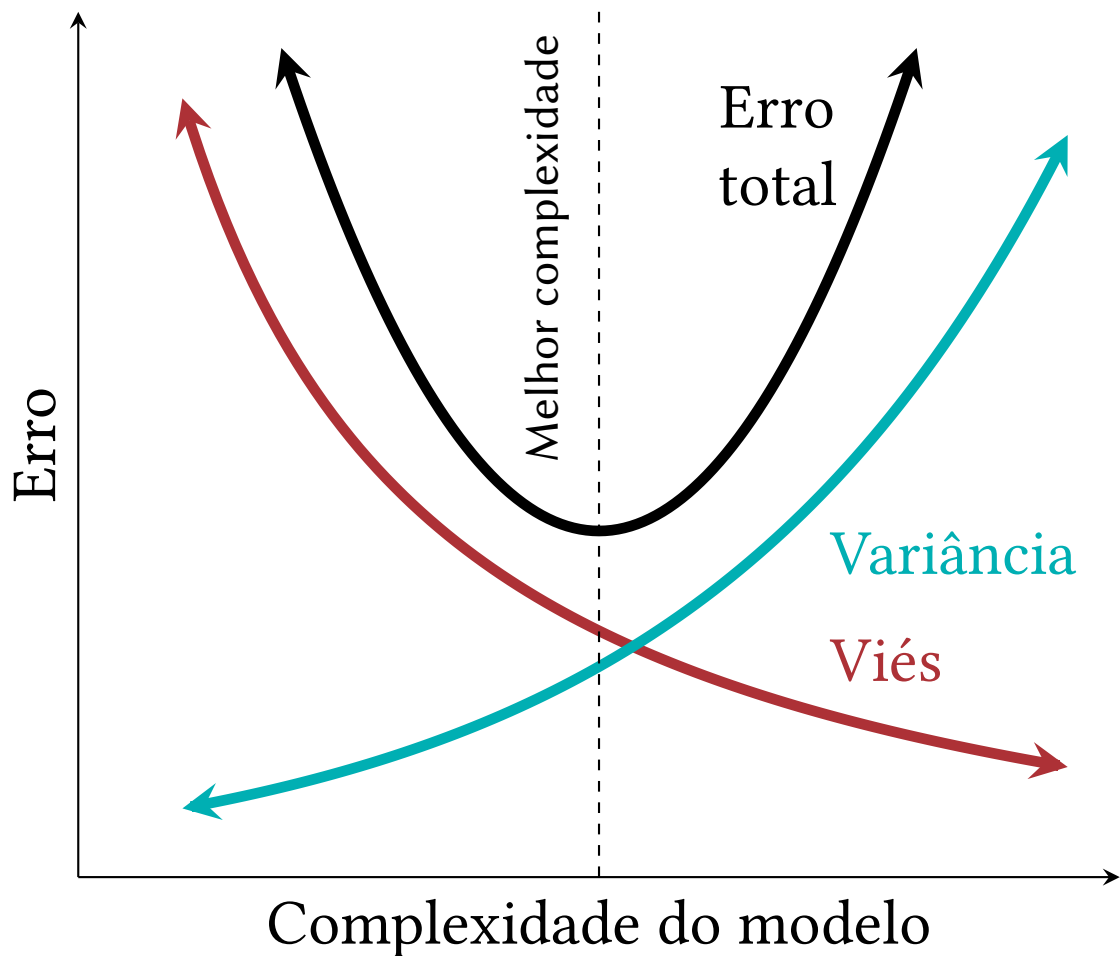
Levando em conta que a suicidalidade é um fenômeno presente na vida de pacientes com TB, vê-se a necessidade de estudar as possíveis estratégias de enfrentamento relacionadas à ideação e tentativa de suicídio. Um estudo de Poyraz, Özdemir, Şen, Sağlam, Enginkaya e Tomruk [43] verificou uma prevalência de aproximadamente 34% de tentativa de suicídio em pacientes com TB, e frequência de ideação suicida ao longo da vida de 60%. Quando comparados aos controles, os pacientes com TB apresentaram maior uso de desengajamento comportamental e estratégias religiosas. Os pacientes com tentativas de suicídio prévia apresentaram pior prognóstico, através de um agravamento do curso da doença, além de um início mais precoce do transtorno, com mais episódios de humor e mais disfuncionais, reforçando a compreensão do TB em sua forma latente [43].

## 1.2 Aprendizado de máquina

Aprendizado de máquina (AM), ou *machine learning* (AM) é um campo da ciência da computação dedicado a compreender e construir métodos que são capazes de aprender padrões a partir de dados de modo a melhorar o desempenho em um conjunto de tarefas específicas [44]. O termo foi utilizado pela primeira vez em 1959 pelo cientista da computação da IBM, Arthur Samuel, que trabalhava com programas capazes de jogar damas [45]. O AM trabalha com diferentes paradigmas: o aprendizado supervisionado, aprendizado não-supervisionado e aprendizado por reforço (*reinforcement learning*) [46]. Em geral, os modelos de aprendizado estatístico trabalham com duas grandezas relevantes: o viés e a variância [46]. Variância está relacionada com a mudança no modelo caso ele fosse construído com dados de treino diferentes. Espera-se portanto, que um determinado modelo, seja teórico ou baseado nos dados (*data-driven*) possa ter uma estrutura semelhante independentemente dos dados usados para ser treinado. Já o viés está relacionado com o grau de erro que é introduzido pela simplificação que o modelo imputa em dado problema real. Em vista disso, o modelo ideal apresenta baixa variância, juntamente com um baixo viés. O balanço entre viés e variância em um modelo de AM pode ser observado na Figura 3.



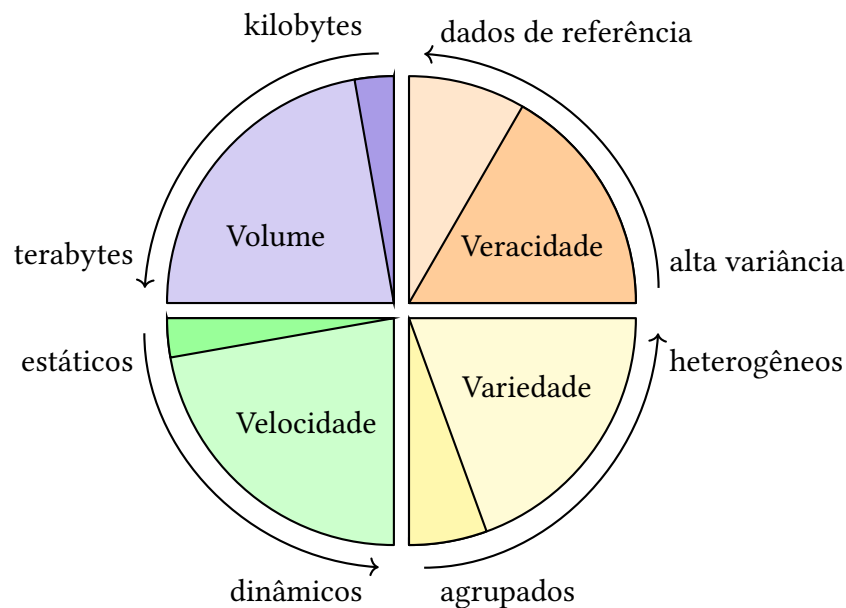
Figura 3 – *Tradeoff* de viés-variância em modelos de aprendizado de máquina. Trata-se da propriedade de um modelo que a variância de um parâmetro estimado através de amostras pode ser reduzido ao aumentar o viés dos parâmetros estimados.



O AM permite que as máquinas aprendam por si mesmas a partir de exemplos e ajustem suas ações ao longo do tempo, o que as torna capazes de realizar tarefas complexas de forma autônoma. Em muitos casos, o AM supera a performance humana em tarefas que exigem habilidades como reconhecimento de padrões, previsão e classificação [47, 48]. A aplicação do aprendizado de máquina em áreas como a saúde tem potencial para revolucionar a forma como esse campo é gerenciado e fornecer soluções mais eficientes e precisas. Neste trabalho, explora-se como o AM pode ser aplicado ao campo da psiquiatria de precisão para o rastreamento de transtornos mentais, especificamente o TB.

### 1.3 Psiquiatria de precisão

A psiquiatria digital é um conceito recente que traz consigo novos olhares para os diagnósticos e intervenções em saúde mental. Este campo traz consigo diversas influências da medicina de precisão como um todo, além de grandes áreas como o aprendizado de má-

Figura 4 – Os quatro V's do *big data*.

Fonte: Adaptado de Rajan [50].

quina e o *big data*. O *big data* consiste, em geral, de dados com várias dimensões, esparsos, com presença de ruído que necessitam de métodos matemáticos e estatísticos suficientemente complexos para extrair padrões e descobertas a partir destes dados. Além da característica de grande volume, o *big data* também carrega noções de velocidade, no que diz respeito à rapidez e métodos de coleta dos dados (coleta passiva ou ativa de dados digitais, *surveys* online, entrevistas presenciais, etc). Além disso, também nota-se questões relacionadas à variedade dos dados, dentro do campo da psiquiatria, pode-se observar dados de neuroimagem, exames clínicos variados, dados tabulares (planilhas e questionários), prontuários (físicos e eletrônicos), dados de voz, entre outros. Deve-se levar em conta também os aspectos de incerteza e veracidade dos dados, incluindo boas práticas de gestão e manipulação de dados, atenção aos diferentes tipos de dados sendo utilizados e suas formas de representação, etc. Estes atributos podem ser visualizados de forma esquemática na Figura 4. A psiquiatria de precisão dentro do contexto clínico almeja integrar dados biológicos e ambientais para personalizar planos de tratamento. Pode auxiliar no monitoramento de drogas terapêuticas, resposta ao tratamento, além da possibilidade de implementação de serviços como calculadoras de risco em sistemas de cuidado em saúde [49].

Levando em conta o papel do AM na psiquiatria de precisão, pode-se pontuar algumas oportunidades e desafios. Considerando que apenas metade dos pacientes apresentam sucesso com as escolhas de tratamento farmacológico atuais [51], questionamentos sobre a necessidade de intervenções mais específicas a nível individual devem ser feitos. Dessa forma, o AM é capaz de contribuir no campo das predições clínicas em pacientes individualmente, por exemplo, para sugerir doses personalizadas de um determinado psicofármaco. Também, o AM pode

ser utilizado para prever desfechos em saúde mental (como diagnósticos psiquiátricos, prejuízo funcional, severidade de quadros clínicos [52], etc) com estimativas *out-of-sample*<sup>2</sup>, além de ser adequado para dados observacionais por conta das técnicas necessitarem menos pressupostos, e capacidade de manejar vários desfechos de uma vez só, como por exemplo: uma rede neural capaz de produzir três saídas: a severidade de determinado transtorno psiquiátrico, a dosagem de um medicamento e abordagem psicoterápica mais adequada. Técnicas de aprendizado estatístico são capazes, ainda, de captar doenças psiquiátricas de modo mais dinâmico e abrangente que a forma estática adotada por guias diagnósticos como DSM-5 e CID-10, produzindo ferramentas que auxiliam sistemas dimensionais de compreensão da doença mental. No entanto, algumas barreiras também devem ser consideradas. Dentro de *pipelines* de AM, é comum que várias etapas e escolhas interdependentes estejam envolvidas. Esta flexibilidade pode aumentar o risco de que os achados obtidos a partir deste conjunto de passos complexos não possa ser reproduzido outras vezes. Dificuldades como disponibilidade e manejo de dados, coleta de dados longitudinais com presença de confundidores relevantes e heterogeneidade dos dados também são importantes focos de atenção para pesquisadores em AM na psiquiatria [53].

Em 2019, uma revisão sobre AM, big data e suas implicações para modelagem de doenças e descobertas terapêuticas em psiquiatria apresentou resultados que indicam a IA como opção viável para construir preditores úteis para desfechos em saúde mental e na personalização de novas terapêuticas. [54]. Técnicas de modelagem preditiva também vêm sendo utilizadas para prever problemas de saúde mental em adolescentes, como demonstra um estudo de 2020 com uma grande amostra sueca de 7.638 gêmeos, onde foi construído um modelo para previsão de problemas de saúde mental com uma ROC AUC<sup>3</sup> de 0.739 [55]. Outras iniciativas voltadas à saúde mental na adolescência como a IMAGEN, estudo longitudinal de neuroimagem e genética em adolescentes com foco em comportamentos de reforço, também discutem o papel da identificação de marcadores biológicos no auxílio da psiquiatria de precisão, focando no agrupamento de sintomas baseado nos mecanismos psicobiológicos compartilhados, juntamente à preditores de curso da doença e resposta ao tratamento em grupos clínicos [56].

No campo da farmacogenômica, estuda-se os efeitos do corpo na droga, assim como o efeito da droga no corpo. Apesar de sua importância, a área apresentou pouco desenvolvimento em sua aplicabilidade clínica [57]. Com o advento do enfoque do aprendizado de máquina e suas técnicas, agregado ao aumento da quantidade de dados disponíveis, novos estudos vêm sendo desenvolvidos tanto para predição de tratamento, como para predição de diagnóstico e prognóstico e detecção de biomarcadores (tanto com dados de neuroimagem quanto dados

<sup>2</sup> Dados *out-of-sample* são aqueles que não pertencem à amostra de treino do modelo (dados externos), ou seja, não foram utilizados para sua construção. Em outras palavras, são os dados não disponíveis que deseja-se quantificar as estimativas.

<sup>3</sup> Métrica que demonstra o desempenho de um modelo de classificação considerando todos os limiares de classificação.

demográficos e clínicos) [58]. Na atualidade, processamento de linguagem natural (PLN)<sup>4</sup> tem ganhado espaço na saúde mental. Uma revisão sistemática de 2021 que avaliou 58 artigos de PLN na saúde mental apontou que a maior parte dos estudos confirmam hipóteses clínicas ao invés de desenvolver novos dados empíricos. Em vista disso, Le Glaz, Haralambous, Kim-Dufor, Lenca, Billot, Ryan, Marsh, DeVyllder, Walter, Berrouiguet e Lemey [59] reconhecem o papel do PLN em dados não-explorados, mas salienta questões éticas que devem ser discutidas ao decorrer dos avanços no uso da técnica [59].

## 1.4 Aprendizado de máquina e transtorno bipolar

Na literatura, alguns estudos têm dado atenção à necessidade de fazer uso de ferramentas tecnológicas para o *screening* de um transtorno grave como o TB. O estudo de Claude, Houenou, Duchesnay e Favre [60] revisou a literatura científica para avaliar a aplicação de técnicas de AM em dados de ressonância magnética (MRI) em pacientes com TB. Eles identificaram 47 estudos que utilizaram algoritmos de aprendizado de máquina baseados em dados de MRI, e descobriram que a precisão relatada para a classificação do TB variou amplamente, de 57% a 100% em comparação com controles saudáveis. No entanto, apesar das limitações metodológicas em estudos anteriores e da necessidade de replicação em grandes amostras multicêntricas, esta revisão sugere que o uso de técnicas de AM em dados de MRI pode levar a futuros diagnósticos assistidos por computador do TB e pode ajudar a prever a resposta ao tratamento. No entanto, mais pesquisas são necessárias para validar esses resultados e estabelecer a eficácia clínica dessas técnicas.

Com base no posicionamento do *International Society for Bipolar Disorders Big Data Task Force* [13], a utilização de *big data analytics* pode ser uma estratégia promissora para o TB. Os resultados da revisão sugerem que as aplicações de AM em *big data* podem fornecer calculadoras de risco para ajudar na tomada de decisões de tratamento e prever o prognóstico clínico individual, incluindo construtos relacionados ao suicídio, como ideação ou tentativa de suicídio. Essa abordagem pode avançar o diagnóstico ao permitir a descoberta de fenótipos *data-driven* relevantes, além de prever a transição para o transtorno em indivíduos de alto risco. No entanto, há vários desafios metodológicos que precisam ser abordados para traduzir esses resultados para ambientes clínicos, incluindo a heterogeneidade dos dados, a falta de validação externa e replicabilidade de alguns estudos, o custo e a falta de financiamento apropriado.

Em um estudo recente de Sawalha et al. [61], foram aplicadas técnicas de AM em uma bateria de testes cognitivos para identificar pacientes com TB em seu primeiro episódio. Os resultados mostraram que as deficiências cognitivas podem aparecer em estágios precoces do TB e persistir em estágios posteriores da doença, sugerindo que os mesmos déficits podem existir tanto para casos de primeiro episódio quanto para pacientes crônicos de TB. Esses dé-

<sup>4</sup> PLN é a área da inteligência artificial e linguística que visa estudar a compreensão automática de línguas humanas naturais por *softwares* através de métodos estatísticos.

ficits cognitivos podem servir como marcadores precoces de TB, e a aplicação de AM em testes cognitivos pode fornecer uma ferramenta útil para a detecção precoce do transtorno [62].

Recentemente, tem havido um avanço na medicina de precisão e um estudo Siqueira Rotenberg, Borges-Júnior, Lafer, Salvini e Silva Dias [62] visou usar abordagens de AM como um possível preditor de recaídas depressivas em pacientes com TB. Quatro algoritmos de AM foram aplicados a um conjunto de dados de pacientes do STEP-BD (*Systematic Treatment Enhancement Program for Bipolar Disorder*) em uma coorte de 800 pacientes. Os algoritmos de AM apresentaram desempenho razoável nas previsões, variando de 61% a 80% no F-score. Os três sintomas de humor mais importantes observados na visita de recaída (no modelo de *random forest*) foram: anedonia, humor depressivo e alterações na energia. Portanto, os resultados sugerem que a aplicação de modelos de medicina de precisão a tomada de decisões médicas no tratamento do TB e prevenção de futuras recaídas.

É válido notar que os sintomas identificados como preditores de recaídas depressivas no estudo acima, são, de fato, comumente usados como critérios diagnósticos para depressão maior. Essa observação levanta uma questão importante sobre a utilidade clínica dessas variáveis na previsão de recaídas em pacientes com TB. Enquanto os resultados deste estudo demonstraram que esses sintomas têm relevância na previsão de recaídas, é essencial reconhecer que seu uso como preditores pode não oferecer uma contribuição substancialmente nova para o campo da medicina de precisão. Embora essas variáveis possam ser informativas, sua aplicabilidade clínica pode ser limitada, uma vez que não fornecem novos insights ou abordagens terapêuticas significativamente distintas daquelas já estabelecidas no tratamento do TB. Esta consideração é crucial ao avaliar a implementação prática de tais modelos. Portanto, levando em conta a importância de mapear preditores discriminativos, é necessário ponderar a utilidade clínica real desses preditores e explorar maneiras pelas quais a medicina de precisão pode oferecer abordagens mais inovadoras e personalizadas para o tratamento e prevenção do TB, que vão além dos critérios diagnósticos convencionais.

Uma revisão de escopo publicada em 2021 [63] identificou 33 estudos que usaram diferentes modelos de AM, incluindo modelos de classificação e regressão. Dados de MRI foram os mais usados para classificar pacientes com TB, com uma acurácia máxima de 98% [63]. Outra função importante que os modelos de IA podem ter no campo do TB é a de diferenciar transtornos do TB ou diferenciar casos de depressão unipolar e bipolar. Um estudo de Bayes, Spoelma, Hadzi-Pavlovic e Parker [64] discutiu o uso de AM na diferenciação de TB e transtorno de personalidade borderline baseado em construtos cognitivos e comportamentais, estratégias de regulação emocional e práticas parentais durante a infância. O estudo encontrou em torno de 73,5% de acurácia na classificação geral de TB e transtorno de personalidade borderline. Na diferenciação entre depressão unipolar e bipolar, estudos demonstraram boa capacidade de classificação através do uso de dados de MRI [65], biomarcadores sanguíneos [66] e assinaturas de biomarcadores [67].

É importante notar que a acurácia de 98% mencionada no estudo do parágrafo anterior [63] para a classificação de pacientes com TB com base em dados de MRI é notável, mas deve ser interpretada com cautela. Essa alta acurácia pode sugerir uma solução definitiva para a diferenciação de pacientes com TB, o que não reflete necessariamente a realidade clínica. A literatura em medicina de precisão e inteligência artificial muitas vezes apresenta resultados promissores em conjuntos de dados específicos, mas a aplicabilidade clínica desses resultados pode variar consideravelmente em cenários do mundo real. Para mais, estudos de classificação em transtornos de humor com MRI apresentam diversas limitações metodológicas [68], onde pode-se citar também a falta de processos de validação e estimação de incerteza (erro), uso inadequado de métricas de desempenho dos modelos a depender de critérios como desbalanço de classe, por exemplo.

Existem diversos estudos que utilizam approaches de machine learning para predição de TB com base em marcadores biológicos, clínicos e neuropsicológicos [69], além de dados de MRI [70]. A revisão sistemática de Colombo, Calesella, Mazza, Melloni, Morelli, Scotti, Benedetti, Bollettini e Vai [69] — que incluiu 81 estudos na revisão e 65 na metanálise — mostrou que apesar de existirem modelos capazes de alcançar altas acurácias na diferenciação de TB e outras doenças psiquiátricas, ainda há um caminho para percorrer no avanço das práticas metodológicas adotadas. Por exemplo, um estudo mostrou que a combinação de dados de MRI estruturais e funcionais mostrou boa acurácia na predição do TB [70]. Outro estudo utilizou biomarcadores cognitivos e imunológicos para prever o diagnóstico de TB e esquizofrenia, obtendo resultados promissores [71]. Além disso, a neuroimagem tem sido amplamente utilizada na diferenciação de primeiro episódio de psicose [72]. Dados de neuroimagem podem ser combinados com outros dados clínicos para desenvolver modelos de predição do risco de conversão para TB ou esquizofrenia em indivíduos com alto risco de desenvolver a doença [73]. Outro campo de estudo em AM no TB é o PLN. Estudos recentes têm explorado a relação entre linguagem natural e transtornos psiquiátricos, incluindo TB. Por exemplo, foi mostrado que a análise de textos escritos por pacientes com TB pode fornecer informações valiosas para a avaliação da doença e do estado de humor dos pacientes [74]. Por fim, calculadoras de risco no transtorno bipolar são uma ferramenta útil para ajudar na tomada de decisão clínica. Essas calculadoras combinam dados clínicos e demográficos para prever o risco de recorrência de episódios de humor em pacientes com TB. Utilizando modelos de AM, essas calculadoras podem ser aprimoradas e ajustadas para aumentar sua precisão na previsão do risco de recaída e orientar o tratamento adequado para pacientes com TB [75].

## 2 Justificativa

Apesar da literatura apresentar diversos estudos almejando prever o fenômeno do desenvolvimento de TB em várias populações, a implementação de técnicas que referem-se ao estado da arte para classificação de desfechos em saúde a partir de dados tabulares é escassa. Novas abordagens podem auxiliar no progresso da psiquiatria de precisão no que tange o mapeamento e diagnóstico de TB através da concretização de ferramentas mais rápidas, e mais importante, escaláveis, expandindo o acesso e eficiência em saúde pública para identificação de novos sujeitos em risco ou em início de TB.

Para mais, o estudo da caracterização de quadros diagnósticos de TB a partir de técnicas de aprendizado estatístico robustas em conjunto com uso de métodos de *interpretable AI*, podem possibilitar a descoberta de novos preditores sociais, demográficos ou clínicos na direção da construção de marcadores de risco para o TB em populações de jovens adultos. Pode-se portanto, além da consumação de novos aparatos de *screening* diagnóstico, trazer novos *insights* clínicos para profissionais de saúde mental.

Entende-se que na literatura, alguns estudos já exploram o potencial de diferentes métodos computacionais para predição de TB, porém em sua grande maioria são estudos que utilizam dados com alto custo de coleta ou pouco acessíveis em um contexto clínico cotidiano (como dados de neuroimagem e amostras de sangue). O presente trabalho traz consigo a meta de trazer uma nova compreensão quanto à possibilidade de dar caminho à uma ferramenta de *screening* com dados de coleta facilitada construída a partir de algoritmos que possam funcionar dentro uma escala comunitária.

## 3 Objetivos

### 3.1 Objetivo geral

Desenvolver um modelo preditivo para prever incidência de transtorno bipolar após 5 anos em populações de adultos jovens através de dados sociodemográficos, clínicos e ambientais.

### 3.2 Objetivos específicos

- a) Produzir um *gradient boosting machine* capaz de realizar previsões com valor aceitável de AUC como métrica de avaliação de desempenho, utilizando técnicas de validação cruzada e amostragem para estimar parâmetros mais robustos de eficácia do modelo;
- b) Explorar meios de prevenção do início de transtorno bipolar por meio do reconhecimento de marcadores sociais, demográficos e clínicos que antecedem o desenvolvimento da doença;
- c) Identificar preditores mais discriminativos quanto ao desfecho de transtorno bipolar através de técnicas de interpretação de resultados em modelos de *machine learning*;
- d) Mapear efeitos dos preditores mais importantes para previsão no prognóstico dos pacientes com transtornos de humor no que tange desfechos secundários de funcionalidade e cognição.



## 4 Métodos

### 4.1 Amostra e coleta de dados

O estudo trata-se de uma coorte prospectiva que fez uso de uma amostra comunitária de jovens adultos entre 18 e 24 anos no *baseline*, residentes da área urbana da cidade de Pelotas, Rio Grande do Sul, Brasil. Os dados foram coletados pelo grupo de pesquisa do Programa de Pós-Graduação em Saúde e Comportamento (PPGSC) da UCPel. A primeira coleta do estudo aconteceu entre os anos de 2007 e 2009, e a amostra foi selecionada por meio de amostragem por conglomerados, levando em conta 89 setores censitários aleatoriamente selecionados de um total de 448 setores [76].

Os seguintes critérios de inclusão foram considerados no *baseline*:

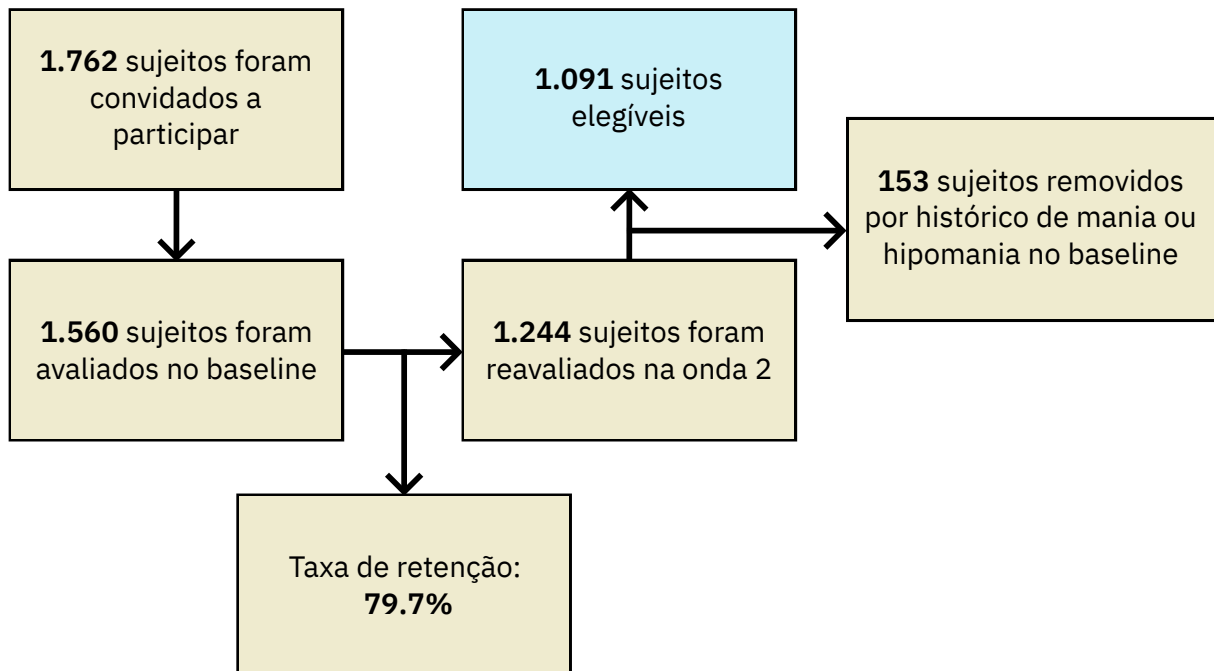
- a) Idade entre 18 e 24 anos;
- b) Residente da zona urbana de Pelotas.

Déficit cognitivo severo (avaliado a partir do julgamento clínico do avaliador) que pudesse causar dificuldades na compreensão dos instrumentos de avaliação do estudo foi considerado o único critério de exclusão.

Todos os sujeitos elegíveis ( $n = 1.762$ ) foram convidados para participar, destes, 1.560 aceitaram e consentiram com sua participação. Entrevistadores treinados conduziram uma entrevista presencial na casa dos participantes, para que a confidencialidade dos dados fosse garantida. Na primeira etapa de coleta, os dados foram coletados por meio de questionários em papel impressos com os instrumentos de pesquisa e os critérios diagnósticos para os transtornos mentais avaliados. Posteriormente, estes questionários passaram pela aplicação do método de dupla digitação, em que dois digitadores transferem os questionários para uma planilha digital, e então estes dados são pareados e comparados a fim de garantir maior confiabilidade do processo de transferência. Os conflitos de digitação foram resolvidos pelo consenso da equipe responsável. Ao final, os dados foram providos através de uma planilha em formato csv.

O *follow-up* (acompanhamento) ocorreu de 2012 a 2014, ou seja, em média, cinco anos após a primeira avaliação. Os participantes do *baseline* ( $n = 1.560$ ) foram convidados para uma segunda avaliação. Todos os entrevistadores realizavam reuniões semanais para discussão das avaliações, com foco em diagnósticos de TB não conclusivos. 1.244 indivíduos foram localizados e consentiram em ser reavaliados (taxa de retenção: 79,7%), e 14 (0,9%) sujeitos foram identificados como perdas por óbito. Dado que o objetivo do estudo é explorar os casos incidentes de TB, os sujeitos com história de mania ou hipomania na linha de base (*baseline*)

Figura 5 – Fluxograma do tamanho amostral do estudo com adultos jovens de 18 a 24 anos da cidade de Pelotas/RS para exploração dos casos incidentes de transtorno bipolar.



do estudo foram excluídos, resultando em um tamanho amostral final de 1.091 indivíduos. O fluxograma para obtenção do tamanho amostral final pode ser observado na Figura 5.

Na segunda fase do estudo, os dados foram coletados através de *tablets* com sistema operacional *Android* por meio do aplicativo livre e de código aberto *Open Data Kit<sup>5</sup> Collect* (versão 1.23.3). Os formulários foram posteriormente processados e transferidos para planilhas eletrônicas (arquivos separados por vírgula — *csv*) a partir de arquivos em formato *xml*. As tabelas foram copiadas para o *IBM SPSS Statistics for Windows* na versão 22, onde descrições e rótulos para cada variável do banco de dados final foram adicionados. Ao final, a base de dados foi disponibilizada para uso interno do grupo de pesquisa do PPGSC em formato *sav*.

## 4.2 Variáveis

### 4.2.1 Desfecho

O desfecho caracteriza-se pelo diagnóstico de TB dentro dos cinco anos seguintes à avaliação do *baseline*. O diagnóstico foi realizado a partir dos critérios do DSM-IV [77] e do CID-10 [78] por meio da entrevista diagnóstica estruturada *Mini International Neuropsychiatric Interview* (versão 5.0.0) [79]. Trata-se de um instrumento psicométrico voltado para avaliação de diferentes transtornos mentais em estudos epidemiológicos ou ensaios clínicos multicêntricos. Este instrumento foi traduzido e validado para o português brasileiro por Amorim [80]. Todos

<sup>5</sup> <https://opendatakit.org/>

os aplicadores do instrumento foram treinados previamente por psicólogos com experiência na aplicação. Para o desfecho do presente trabalho, utilizou-se os módulos A e D, referentes aos episódios depressivos e aos episódios maníacos ou hipomaníacos, respectivamente. Na segunda onda de avaliação, em casos de dúvida diagnóstica por parte dos avaliadores treinados, um psiquiatra com experiência em transtornos de humor reavaliou o participante.

Uma limitação da medida de desfecho utilizada refere-se aos registros de história de episódio depressivo maior passado no *baseline*. Na primeira etapa do estudo, dados sobre a história de depressão eram coletados apenas se o sujeito entrevistado fechava critérios diagnósticos para episódio depressivo atual. Logo, não foi possível checar quanto à história de episódios depressivos em sujeitos potencialmente depressivos, porém eutímicos. Dessa forma, considerando que o estudo almejou verificar a taxa de incidência de TB em cinco anos, optou-se por excluir todos os participantes que apresentaram episódio maníaco ou hipomaníaco, seja atual ou passado, no *baseline*.

## 4.2.2 Preditores

O presente projeto considerou diversas variáveis como preditores candidatos para o TB em adultos jovens. Características sociodemográficas, clínicas e ambientais construíram o conjunto de *features*<sup>6</sup> utilizadas para as predições. Previamente às etapas de filtragem e seleção de variáveis, 190 variáveis estavam presentes no conjunto de dados.

### 4.2.2.1 Dados sociodemográficos e ambientais

As variáveis sociodemográficas e ambientais incluídas no *pipeline* de AM estão descritas a seguir. A variável de sexo foi incluída pois apesar de não se observar diferenças entre a prevalência de TB entre os sexos, o TB tipo II parece ocorrer mais frequentemente em mulheres [81]. A cor da pele, apesar de existirem poucos estudos que investigam uma relação direta entre cor da pele e TB, alguns pesquisadores apontam para disparidades entre raças no que diz respeito ao tratamento e pesquisa em pacientes com TB [82]. Por esta questão, e a fim de observar possíveis interações da variável de cor da pele com demais preditores, incluiu-se no estudo.

Apesar de estudos não verificarem diferença de idade no início do TB tipo I e II [83], estudos na literatura indicam que a distribuição da idade de início do TB se comporte de modo trimodal, ou seja, alguns sujeitos apresentam um início mais precoce (em torno dos 17 anos), intermediário (em média, 26 anos) ou tardio ( $\mu = 46,3$  anos) [84]. Deste modo, esta variável esteve presente no fluxo de modelagem.

O status socioeconômico foi abarcado tendo em vista que estudos mostram que adolescentes com TB apresentam maior risco de problemas legais, risco aumentado para transtorno

<sup>6</sup> Termo utilizado no campo do AM para denominar as variáveis independentes de um modelo de aprendizado estatístico.

de estresse pós-traumático e transtorno de déficit de atenção e hiperatividade comórbidos [85]. Ademais, maior status socioeconômico pode predizer um melhor prognóstico em adolescentes com TB [86]. E o status socioeconômico pode ser um indicador inicial de transtornos de humor, levando em conta que sujeitos com TDM e TB apresentam piores desfechos socioeconômicos durante todo período de vida profissional [87]. Quanto à ocupação, sabe-se que um diagnóstico errôneo pode acarretar em tratamentos inadequados ao caso de TB, o que pode levar à uma cascata de desfechos negativos [88].

#### 4.2.2.2 Dados sobre uso de substâncias

O uso e abuso de substâncias são comuns em indivíduos com TB, e podem ter um impacto significativo na gravidade e na frequência dos episódios de humor [89, 90]. Alguns estudos sugerem que o abuso de substâncias pode estar associado a uma maior incidência de episódios maníacos, enquanto outros indicam que o abuso de substâncias pode aumentar o risco de depressão em indivíduos com TB. O uso de AM pode ser uma ferramenta útil para identificar padrões e prever episódios de humor em indivíduos com TB, levando em consideração o histórico de uso de substâncias e outros fatores relevantes. Combinar essas informações pode permitir que os médicos personalizem o tratamento para cada paciente e ajudem a prevenir recaídas.

No presente projeto, variáveis relacionadas ao uso de substâncias foram empregadas nas rotinas de modelagem. As seguintes substâncias foram avaliadas para possível abuso ou dependência: maconha, cocaína, crack, anfetaminas, inalantes, sedativos, alucinógenos, opióides, substâncias ilícitas e outras substâncias relatadas pelo entrevistado. O instrumento utilizado para tal foi o *Alcohol, Smoking and Substance Involvement Screening Test* (ASSIST). Variáveis referentes ao uso das substâncias ao longo da vida também foram coletadas (descritas acima). Outros dados coletados foram: idade que o participante experimentou drogas pela primeira vez, uso de drogas injetáveis sem prescrição ao longo da vida e uso de medicação para problemas de estresse nos últimos 30 dias.

#### 4.2.2.3 Dados clínicos

Variáveis clínicas e diagnósticos psiquiátricos no *baseline* também foram levados em conta na modelagem de incidência de TB. Os diagnósticos foram realizados baseando-se nos critérios do DSM-IV por meio da *Mini International Neuropsychiatric Interview 5.0* (MINI) por psicólogos treinados. Em casos de dúvidas diagnósticas nos possíveis casos de TB, os sujeitos eram reavaliados por psiquiatra especializado em transtornos de humor.

Os diagnósticos de transtornos mentais incluídos foram: transtorno de personalidade antissocial, transtorno de pânico atual (com e sem agorafobia), transtorno de pânico ao longo da vida, ataques pobres ao longo da vida, agorafobia sem pânico, transtorno de ansiedade social, episódio depressivo atual, episódio depressivo melancólico atual, episódio depressivo

Tabela 1 – Construtos avaliados e quantidade de itens de cada instrumento incluído na modelagem de incidência de transtorno bipolar em jovens adultos.

<b>Instrumento</b>	<b>Construto primário</b>	<b>Número de itens</b>
<i>Beck Depression Inventory</i>	Sintomas depressivos	21
<i>Hypomania Checklist</i>	Sintomas maníacos e hipomaníacos	32
<i>Social Readjustment Rating Scale</i>	Eventos estressores	26
<i>Beck Scale for Suicide Ideation</i>	Ideação suicida	21

passado, distímia atual, transtorno obsessivo-compulsivo atual, transtorno de estresse pós-traumático atual. Entre os transtornos alimentares, foram incluídos: anorexia nervosa atual, bulimia nervosa atual, anorexia nervosa purgativa. Risco de suicídio atual e também foram considerados possíveis preditores. Variáveis relacionadas ao tratamento foram usadas: consulta psiquiátrica ou psicológica ao longo da vida, tratamento psicoterapêutico atual com psiquiatra ou psicólogo, interrupção de tratamento antes da alta.

#### 4.2.2.4 Dados sobre comportamento sexual

A literatura traz consigo algumas associações de TB com comportamentos sexuais. Um estudo realizado na Colômbia [91] demonstrou que pacientes com esquizofrenia e TB, em geral, relatam não possuir uma vida sexual ativa, e quando a têm, apresentam poucas estratégias comportamentais de prevenção em saúde sexual.

As variáveis de comportamento sexual avaliadas foram: idade da primeira relação sexual, teve relação sexual na última semana, uso de preservativo na última relação sexual, uso de álcool antes da última relação sexual, número de parceiros sexuais no último ano, foi forçado a ter relações sexuais ao longo da vida, e se já teve relação sexual ao longo da vida (dicotômica).

#### 4.2.2.5 Instrumentos psicométricos

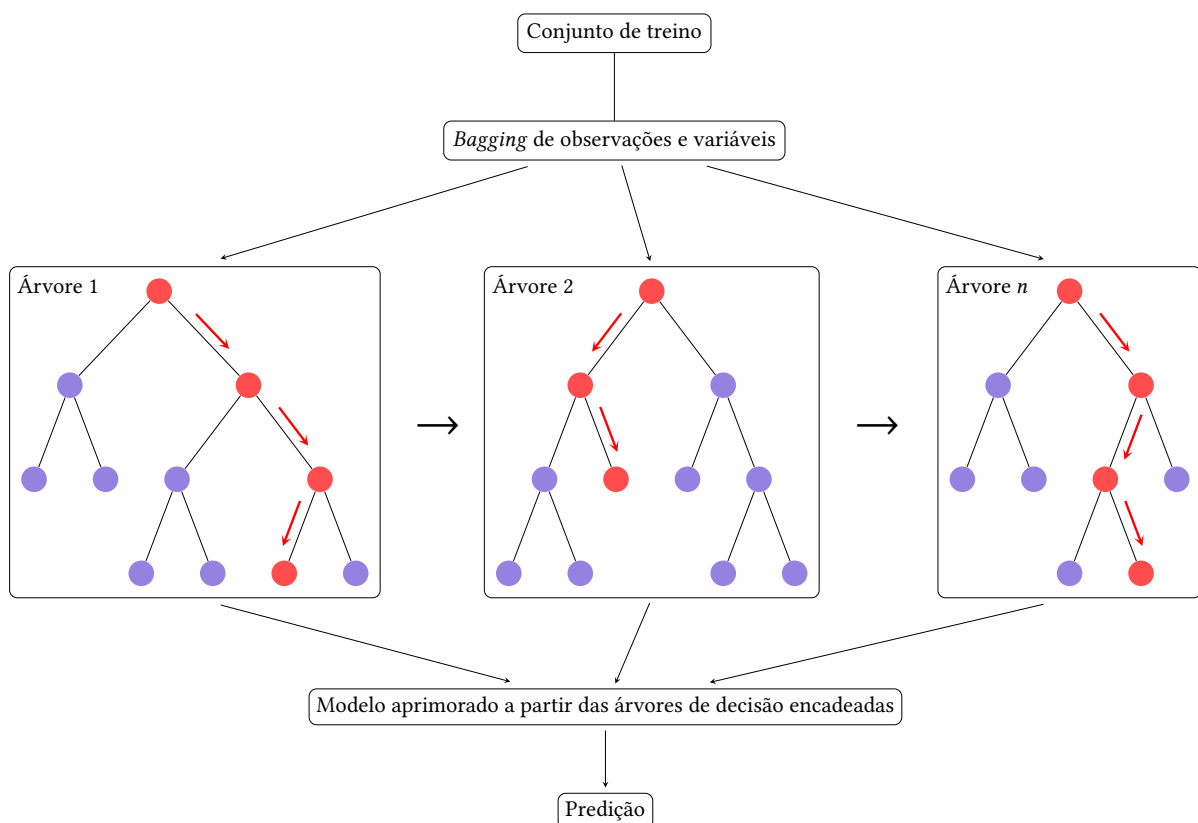
Com intuito de incluir características de humor, efeito de eventos estressores na vida e ideação suicida. Os seguintes instrumentos foram utilizados: *Beck Depression Inventory* (BDI), *Hypomania Checklist* (HCL-32), *Social Readjustment Rating Scale* (SRRS) e *Beck Scale for Suicide Ideation* (BSS). A Tabela 1 resume o construto avaliado e a quantidade de itens de cada um dos instrumentos. No *pipeline* de modelagem, apenas os itens de forma individual foram inseridos no conjunto de dados. Optou-se pela retirada dos escores totais dos instrumentos pela possível perda de parcimônia, e para manutenção da facilidade na coleta das informações em uma implementação de calculadora de risco com dados acessíveis.

## 4.3 Análise de dados

### 4.3.1 Gradient tree boosting

*Gradient tree boosting* (*gradient boosted trees* ou GBMs) é uma técnica de AM que se utiliza da combinação de várias árvores de decisão<sup>7</sup> para melhorar a acurácia das predições. A técnica é bastante utilizada em áreas que se beneficiam do reconhecimento de padrões, como reconhecimento de fala, detecção de anomalias, previsão de séries temporais, etc. O GBM usa um algoritmo de *boosting*<sup>8</sup> que combina a predição de várias árvores de decisão simples, cada uma construída sequencialmente, com intuito de corrigir os erros do modelo anterior. O processo de construção das árvores e correção de erros (diminuição dos resíduos) continua até que um determinado critério seja atingido [92]. O funcionamento de um modelo de *gradient tree boosting* pode ser visualizado na Figura 6.

Figura 6 – Diagrama do algoritmo de *gradient tree boosting* desde a entrada de dados até a predição.



Fonte: Adaptado de Riebesell [93].

<sup>7</sup> Árvore de decisão no AM é um algoritmo de aprendizado supervisionado que constrói um modelo preditivo na forma de uma árvore, na qual cada nó corresponde a uma decisão sobre uma variável e cada folha representa uma predição.

<sup>8</sup> *Boosting* é uma técnica de machine learning que combina múltiplos modelos de aprendizado fracos para criar um modelo forte, em que cada modelo é treinado para corrigir as falhas do modelo anterior.

### 4.3.2 SHAP *values*

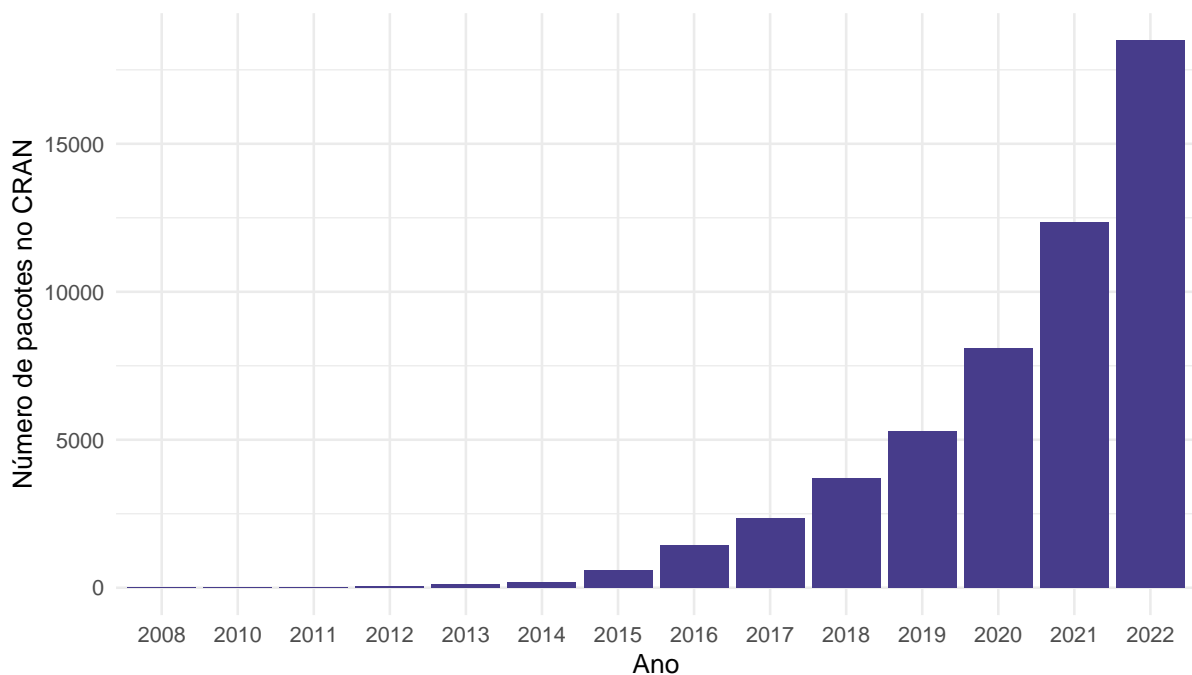
Os SHAP *values* são uma técnica de interpretação avançada frequentemente empregada em modelos de machine learning para fornecer insights sobre como cada característica ou variável contribui para as previsões do modelo. Esses valores são baseados em conceitos da teoria dos jogos e visam atribuir um valor específico a cada característica, quantificando sua influência nas predições do modelo [94].

O valor SHAP pode ser positivo ou negativo. Maiores valores indicam influência no aumento das probabilidades preditas pelo modelo e valores negativos indicam uma diminuição na probabilidade predita pelo modelo em uma determinada observação. No presente estudo, SHAP *values* foram usados como uma abordagem para responder a perguntas como: “Quais fatores têm maior impacto na probabilidade de início do TB?” e “Como as diferentes características afetam as previsões do modelo de maneira individual e coletiva?”. Maiores detalhes sobre implementação do método podem ser checadas na seção 4.3.3.

### 4.3.3 Implementação

A análise de dados da presente dissertação foi realizada a partir da escrita de *scripts* implementados na linguagem de programação R [95] na versão 4.3.1. A linguagem R foi escolhida para implementação dos métodos por se tratar de uma linguagem funcional de alto nível, além de ser gratuita, livre e *open source*, ao contrário de outras linguagens que impõe restrições comerciais, como licenças não-livres ou pagas. O R também possui a vantagem de ser extremamente extensível e aglutinar uma grande comunidade de usuários, possibilitando o uso de pacotes de funções para diversos métodos tradicionais e *bleeding edge* de análise estatística e computacional. O *Comprehensive R Archive Network* (CRAN), uma rede de servidores ao redor do mundo que hospeda versões de códigos, documentação e pacotes de R tem hospedado cada vez mais pacotes com o passar dos anos, atingindo taxas de crescimento altas e constantes, chegando ao patamar de 18.489 pacotes até agosto de 2022 (Figura 7).

Figura 7 – Frequência cumulativa do número de pacotes em R hospedados no *Comprehensive R Archive Network* de 2008 a 2022.



Fonte: <https://cran.r-project.org/>

Os *scripts* foram escritos através do editor de texto Neovim [96] na versão 0.7.2 e do ambiente de desenvolvimento integrado RStudio [97] na versão 2022.07.1 para sistemas Linux.

Alguns pacotes contidos no metapacote *tidyverse* [98] foram utilizados para facilitar o processo de manejo e estruturação dos dados. O pacote *dplyr* [99] foi usado para a manipulação de dados no que diz respeito à criação de subconjuntos, elaboração de novas variáveis, filtragem de dados, combinação de dados e aplicação de técnicas para resumo de dados. Para a tarefa de ler arquivos em formatos de *softwares* estrangeiros ao R, como arquivos *sav* do *SPSS* [100], optou-se pelo pacote *haven*, capaz de carregar e interpretar estruturas de dados de cada formato. De modo a facilitar a aplicação de conceitos de linguagem funcional, utilizou-se o pacote *purrr* [101]. O pacote *tidyr* [102] foi usado para tarefas de padronizar tabelas e dados estruturados, de modo que as tabelas sejam estruturados de tal maneira que: cada coluna é uma variável, cada linha é uma observação e cada célula identifica um único valor. Os gráficos científicos do projeto foram elaborados a partir do pacote *ggplot2* [103], focado em criar gráficos seguindo os princípios da “Gramática dos Gráficos” [104].

Para a modelagem estatística, o *framework* *tidymodels* [105] foi usado para implementação do *pipeline* de AM proposto. Este *framework* é caracterizado pelos pacotes: *rsample* [106] como infraestrutura para amostragem e divisão de dados; *parsnip* [107] que provê uma interface unificada com uma gama de diferentes algoritmos para implementação de modelos com opções de *backends* diversos; *recipes* [108] para facilitar o processo de pré-processamento



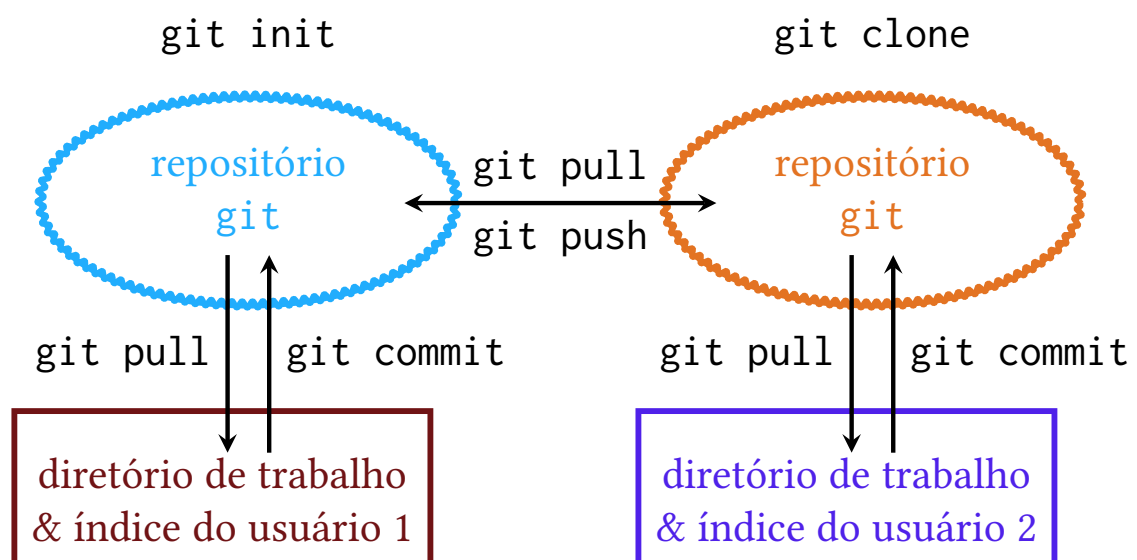
de dados por meio de técnicas de *feature engineering*; *workflows* [109] para agrupar o pré-processamento, modelagem e pós-processamento em um fluxo único; *tune* [110] para auxiliar na otimização de hiperparâmetros e etapas de pré-processamento; *yardstick* [111] para medir a efetividade de modelos a partir de métricas de desempenho; *broom* [112] para converter informações de objetos de saída de modelos em formatos mais simplificados; e o pacote *dials* [113] para criar e manejar parâmetros de tunagem e grades de parâmetros. O pacote *finetune* [114] teve o papel de possibilitar o uso das técnicas de *racing* com ANOVA na tunagem de hiperparâmetros do *XGBoost*.

Com propósito de implementar o algoritmo de *XGBoost* através de programação em R, utilizou-se o pacote *xgboost* (versão 1.7.5) [115], disponível no CRAN. Para calcular os valores SHAPley, o pacote *SHAPforxgboost* (versão 0.1.3) [116] foi usado. Uma parcela das implementações em código para as análises do estudo estão descritas no Apêndice B.

#### 4.3.4 Versionamento

Para o processo de versionamento de código, utilizou-se o sistema de controle de versão livre e de código aberto *git*<sup>9</sup> para registrar mudanças nos arquivos do fluxo de análise de dados do projeto. O projeto *git* foi desenvolvido por Linus Torvalds no ano de 2005 para utilização no desenvolvimento do kernel Linux, em conjunto com outros desenvolvedores do kernel. O projeto é distribuído sob a licença GPL-2.0-only. O fluxo com *git* permite colaboração entre pesquisadores de forma remota, assim como histórico de mudanças e criação de novas versões de código (Figura 8).

Figura 8 – Fluxo de trabalho com uso de repositórios *git* para versionamento de código.



<sup>9</sup> <https://git-scm.com/>

## 5 Resultados

O manuscrito do presente trabalho foi hospedado no dia 3 de abril de 2023 no repositório de preprints MedRxiv<sup>10</sup>. O artigo em versão final pode ser visualizado no Apêndice D (corpo do manuscrito) e E (material suplementar). O estudo foi recentemente apresentado em um simpósio na Conferência Anual da Sociedade Internacional de Transtorno Bipolar (junho de 2023)<sup>11</sup>, visando divulgar e discutir os achados com a comunidade científica. O artigo está submetido para a revista *Bipolar Disorders* da editora *Wiley* como um artigo original e está em revisão por pares no presente momento (Apêndice C). A revista possui um fator de impacto<sup>12</sup> de 5.4. A revista foi fundada em 1999 com intuito de divulgar pesquisas com relevância sobre mecanismos básicos, aspectos clínicos ou tratamento do transtorno bipolar. O presente estudo enquadrou-se no escopo da revista levando em conta o estudo do TB, tendo em vista a investigação de fatores ambientais, sociais, psicológicos e epidemiológicos que pudessem associar-se ao TB. Ainda, este estudo traz avanços em métodos de pesquisa e técnicas de interpretação de resultados para maior transferência dos achados ao conhecimento clínico dos profissionais de saúde, objetivo que também é almejado pela *Bipolar Disorders*.

O estudo visou avaliar uma amostra de 1.244 adultos jovens que foram reavaliados após um intervalo de cinco anos para a incidência de TB. Levando em conta que os sujeitos com história de mania ou hipomania no *baseline* foram excluídos, a amostra final consistiu de 1.091 sujeitos, após a remoção de 153 indivíduos (12,3%). Dos 1.091, 49 (4,49%) adultos jovens receberam diagnóstico de TB cinco anos após a primeira avaliação.

O processo de validação do modelo seguiu um fluxo tradicional de construção de modelo preditivo. O primeiro passo foi a divisão dos dados entre bases de treino e teste através de *stratified sampling*<sup>13</sup>. No presente trabalho, obtivemos uma base de treino com 763 sujeitos e uma base de teste com 328 sujeitos. Os hiperparâmetros do melhor modelo na validação cruzada podem ser observados na Tabela 2.

<sup>10</sup> <https://doi.org/10.1101/2023.03.31.22282507>

<sup>11</sup> <https://isbd2023.org/>

<sup>12</sup> Método bibliométrico que avalia a importância de um periódico científico. Trata-se do número médio de citações de artigos científicos publicados em determinado periódico.

<sup>13</sup> O uso da estratificação foi feito para manter a proporção entre as classes (TB e não-TB) em ambos os conjuntos de dados.

Tabela 2 – Hiperparâmetros finais do modelo de XGBoost para predição de incidência de transtorno bipolar após 5 anos.

Hiperparâmetro	Definição	Valor final
<i>mtry</i>	Número ou proporção de preditores que serão amostrados aleatoriamente em cada divisão ao criar as árvores de decisão.	9
<i>trees</i>	Um número inteiro para o número de árvores contidas no modelo.	1429
<i>min_n</i>	Um número inteiro para o número mínimo de sujeitos em um nó que é necessário para que o nó seja dividido ainda mais.	10
<i>tree_depth</i>	Um número inteiro para a profundidade máxima da árvore (ou seja, número de divisões).	14
<i>learn_rate</i>	Um número para a taxa na qual o algoritmo de boosting se adapta de iteração para iteração.	0.0017
<i>loss_reduction</i>	Um número para a redução na função de perda necessária para dividir ainda mais.	0.6120
<i>sample_size</i>	Número ou proporção de sujeitos dos dados de treino que está exposto ao ajuste dos modelos.	0.9075

Ao final do processo de validação cruzada, o modelo foi retreinado em todo o conjunto de treino com os hiperparâmetros finais e avaliado no conjunto de teste. O modelo de XGBoost demonstrou um desempenho satisfatório com uma ROC AUC de teste de 0,786 (IC 95%: 0,686, 0,887). Em geral, o modelo foi capaz de prever corretamente as instâncias negativas (sujeitos que não desenvolveram TB), apesar de ter apresentado certa dificuldade na predição das instâncias positivas, ou seja, acertar os sujeitos que de fato desenvolveram TB na segunda avaliação.

No contexto deste estudo, é relevante destacar que lidamos com um desequilíbrio significativo de classes, onde a classe positiva consiste naqueles sujeitos que desenvolveram TB na segunda avaliação, enquanto a classe negativa compreende aqueles que não desenvolveram a condição. Esse desequilíbrio é uma característica comum em problemas de predição médica, onde os eventos de interesse são relativamente raros em comparação com os não eventos. Uma métrica de avaliação crucial em cenários de desequilíbrio de classe é o valor preditivo positivo (VPP). O VPP representa a proporção de predições positivas corretas em relação ao total de predições positivas feitas pelo modelo. No nosso caso, o VPP mediria a capacidade do modelo de XGBoost em identificar corretamente os sujeitos que realmente desenvolveram TB no *follow-up*, entre todos aqueles que foram previstos como positivos. É importante reconhecer que, devido ao desequilíbrio de classe, o modelo pode ter um alto desempenho em relação à métrica de acurácia geral, mas ainda assim ter dificuldade em prever adequadamente a classe positiva. Isso pode resultar em um VPP menor do que o desejado, o que significa que um número significativo de predições positivas pode ser falsos positivos. Portanto, é fundamental que, ao avaliar a utilidade clínica deste modelo, se considere não apenas a ROC AUC, mas também o VPP, pois ele fornece informações específicas sobre a capacidade do modelo em identificar corretamente os indivíduos em risco de desenvolver TB. Melhorar o VPP pode ser de grande importância em cenários de desequilíbrio de classe, pois isso reduziria os diagnósticos incorretos e permitiria uma intervenção mais direcionada e eficaz em pacientes em risco.

Para avaliar a estabilidade do modelo, duas abordagens foram adotadas: teste de permutação [117] e diferentes divisões (*splits*) de bases de treino e teste. O teste de permutação serviu para comparar a distribuição do desempenho da ROC AUC em arranjos aleatórios do desfecho na base de treino com os dados de teste originais em 10.000 permutações. No teste de permutação, o modelo performou significativamente melhor do que modelos de classificação

aleatórios ( $p < 0,001$ ). A distribuição do teste de permutação pode ser observada na Figura 9. Para as diferentes divisões nos conjuntos de treino e teste, os modelos ajustados seguiram os mesmos hiperparâmetros do modelo final do estudo (Tabela 2). A média da AUC foi de 0,723 (IC 95%: 0,719; 0,726) e a análise foi capaz de demonstrar certa estabilidade no desempenho preditivo do modelo e combinação de hiperparâmetros selecionados (Figura 10).

O modelo final consiste de dez variáveis que foram selecionadas na validação cruzada a partir do algoritmo Boruta [118], que faz uso de modelos de *random forest* para mapear subconjuntos de variáveis mais preditivas de um determinado desfecho. As dez variáveis mais importantes, ou seja, associadas ao desfecho foram: sentir-se um fracasso, sintoma de tristeza, episódio depressivo no *baseline*, sintomas de estresse auto-relatados, autoconfiança, uso de cocaína ao longo da vida, status socioeconômico, frequência sexual (variável dicotômica de frequência semanal), ter um parceiro e taquialia (Figura 11).

Figura 9 – Distribuição das áreas abaixo da curva ROC (ROC AUC) a partir do teste de permutação com 10.000 rearranjos. A linha tracejada indica a AUC nos dados originais de teste (ROC AUC = 0,786).

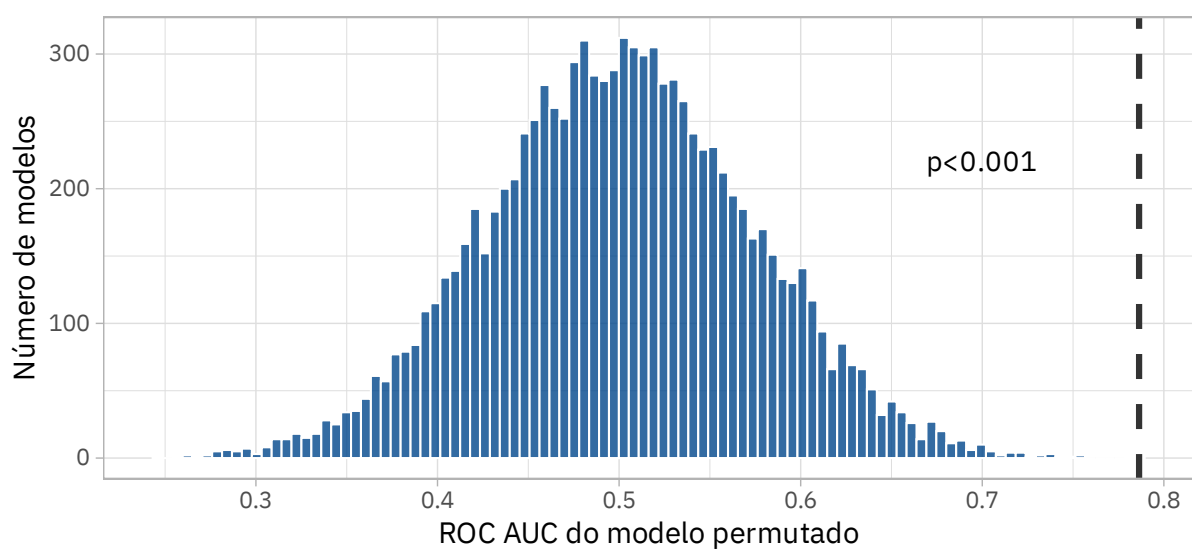


Figura 10 – Histograma da área abaixo da curva ROC (ROC AUC) baseada em 1.000 divisões de conjuntos de treino e teste aleatórios. A média e IC 95% encontrados foram de 0,723 [0,719–0,726]. Esta análise demonstra a robustez da capacidade preditiva do modelo de *boosting* selecionado. A linha contínua indicata a média da ROC AUC e as linhas tracejadas indicam os IC 95% da média.

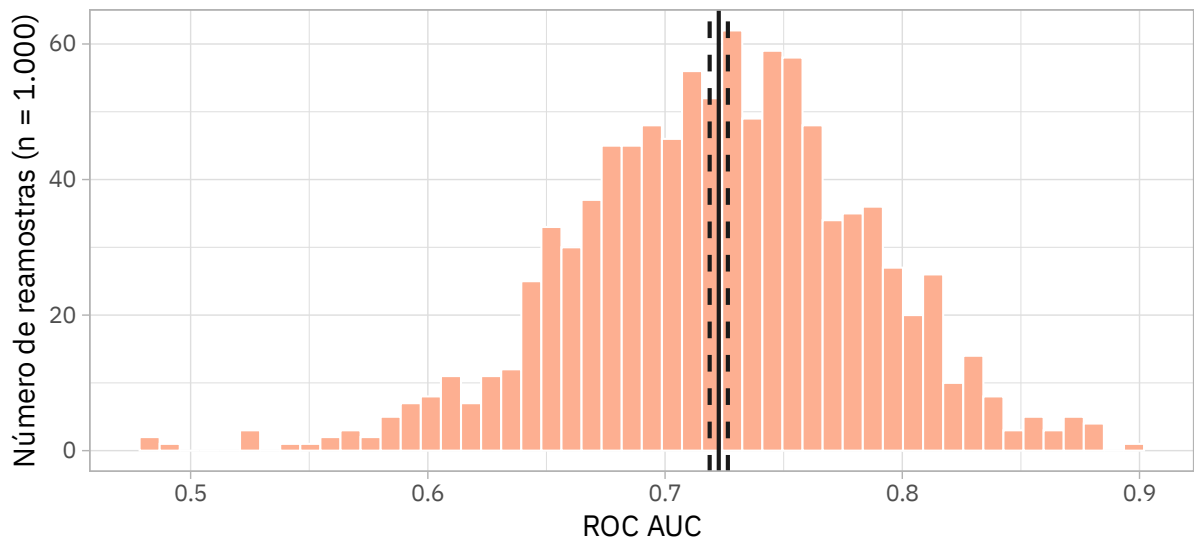
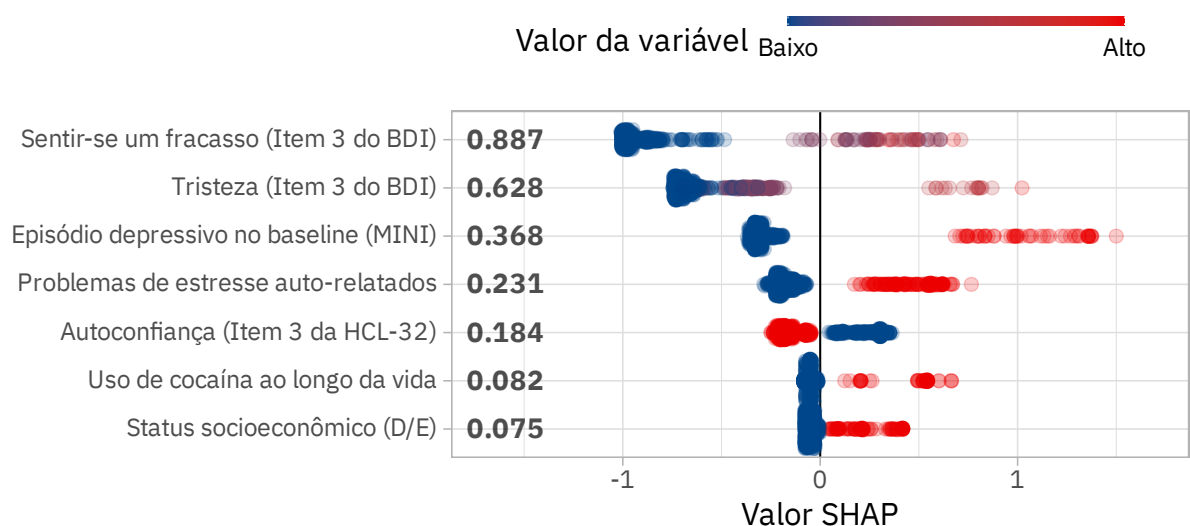


Figura 11 – Valores SHAP para cada variável de acordo com valor da variável baseado no ajuste final do modelo na base de treino ( $n = 763$ ). Variáveis binárias devem ser interpretadas em vermelho como positivo e azul como negativo. O eixo  $y$  representa a variável, ordenada pela maior contribuição absoluta entre todas as observações, independentemente da direção da associação. Cada ponto representa um participante na base de treino. Em observações com valor SHAP menor que zero, a variável comportou-se como fator protetivo, de modo contrário, fator de risco. A cor de preenchimento dos pontos representa o valor da variável para um dado indivíduo (vermelho corresponde a maiores valores e azul corresponde a menores valores). Os valores são relativos sem interpretação a partir das estimativas individuais.



Dentro deste contexto, os resultados parecem reforçar uma compreensão do TB como um fenômeno latente onde a manifestação mais relevante é a depressão maior precoce. Estes achados acabam por trazer novas perguntas sobre a trajetória e progressão de pacientes de humor a partir de outras óticas, como investigações e associações sobre o construto da depressão com a funcionalidade e a cognição. Na mesma linha, buscou-se explorar os resultados encontrados a partir do construto da depressão como fator associado ao desenvolvimento de TB, e mais, compreender qual o impacto ou associação de tais características depressivas na trajetória dos transtornos de humor.

Em relação ao modelo construído sem a inclusão de variáveis relacionadas ao humor, todo o fluxo de construção do modelo foi executado novamente, com nova estimação de hiperparâmetros por meio da validação cruzada. Ao final, um novo modelo foi ajustado em todo conjunto de treino e testado na base de testes com os hiperparâmetros provenientes da validação. Para este modelo, a importância de variáveis estimada foi baseada no método de ganho (*gain*) ao invés de SHAP, pois o objetivo com este segundo modelo foi focar na magnitude da importância de cada *feature*, e não necessariamente na direção da associação ou predição individual em cada observação. O *gain* é o ganho na acurácia introduzido por uma variável nos ramos (*branches*) da árvore de decisão em que ela se encontra. A ideia é que antes de adicionar uma nova variável  $x$  ao ramo, haviam alguns sujeitos classificados erroneamente; depois de adicionar a divisão nesta *feature*, há dois novos ramos (*branches*), e cada um destes ramos é mais preciso do que o formato prévio da árvore (ou seja, separa melhor os sujeitos positivos e negativos). O modelo apresentou desempenho preditivo inferior ao modelo principal, principalmente na ROC AUC (0,786 vs. 0,644). As características mais importantes para prever o início de TB que não foram diretamente associadas com medidas de humor podem ser vistas na Figura 12. Problemas de saúde mental materna foram as características mais relevantes, seguidas por problemas de estresse auto-relatados, ocupação no último ano, tentativa de suicídio em relações sociais círculo e abuso ou dependência de álcool. Associações lineares entre os escores de gravidade da depressão no início do estudo e o desempenho cognitivo, bem como o comprometimento funcional anos depois, foram examinados. A análise excluiu indivíduos com episódios atuais de humor no *follow-up* ( $n = 107$ ) para minimizar a influência direta dos sintomas de humor nas medidas. A amostra final foi composta por 984 indivíduos. A análise de regressão revelou uma correlação negativa entre a gravidade dos sintomas depressivos e o desempenho cognitivo ( $\beta = -0,11$ , IC 95% [-0,14, -0,07],  $t(957) = -5,51$ ,  $p < 0,001$ ,  $r = -0,18$ ). Isso indica que escores basais mais altos de depressão foram associados a desempenho cognitivo inferior cinco anos depois (Figura 13a). Além disso, foi observada uma correlação positiva entre a gravidade dos sintomas depressivos e o comprometimento funcional ( $\beta = 0,23$ , IC 95% [0,17, 0,30],  $t(980) = 7,55$ ,  $p < 0,001$ ,  $r = 0,24$ ). Isso sugere que pontuações basais mais altas de depressão foram associadas a um aumento do comprometimento funcional cinco anos depois (Figura 13b).

Figura 12 – Gráfico de importância de variáveis baseado no modelo com remoção de variáveis relacionadas diretamente ao humor (ROC AUC de teste: 0,644). Histórico de transtorno mental materno e problemas de estresse auto-relatados foram as duas variáveis mais importantes do modelo.

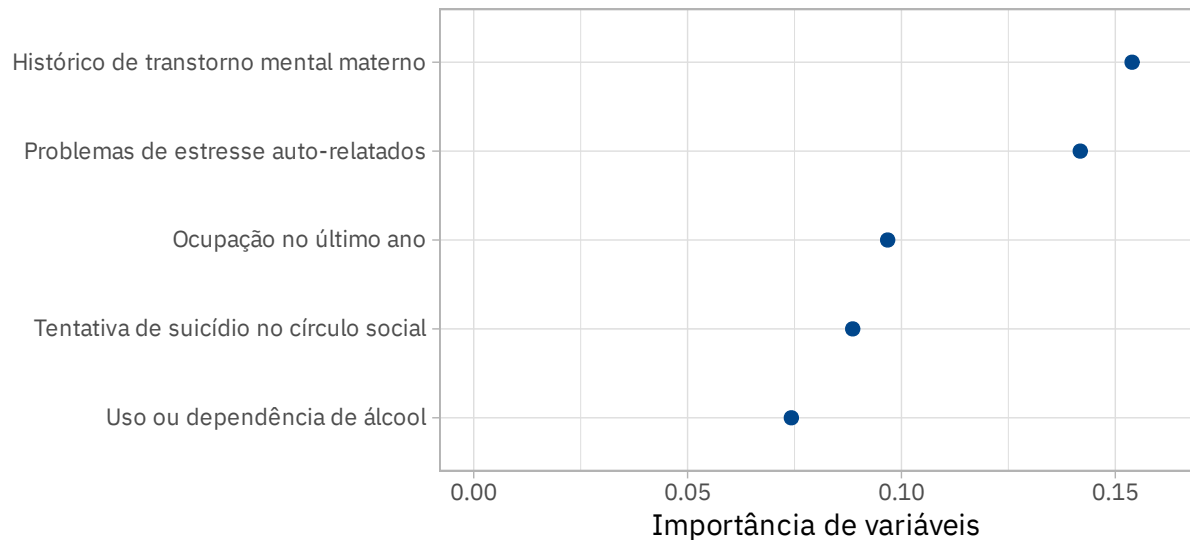
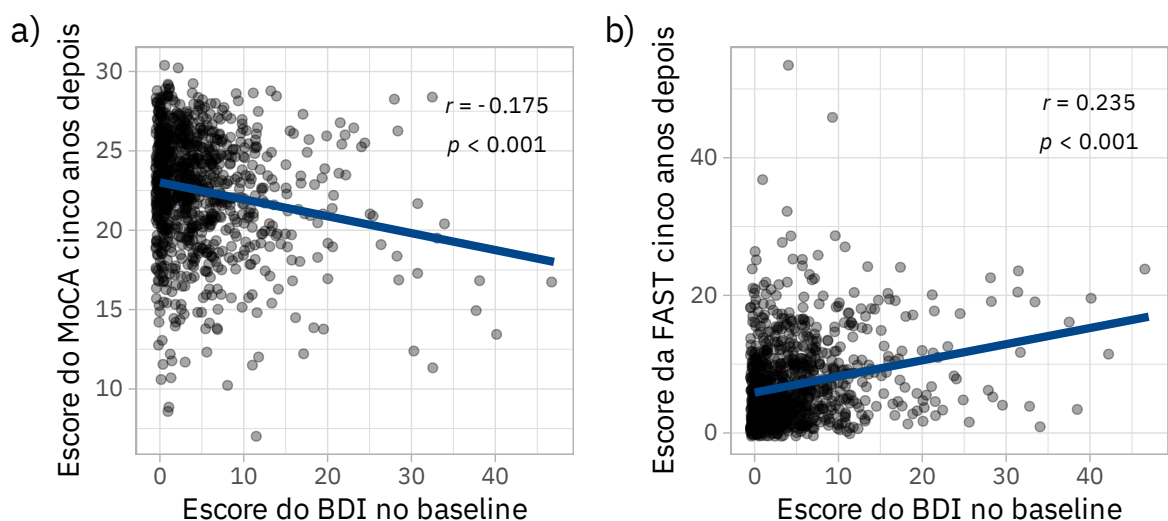
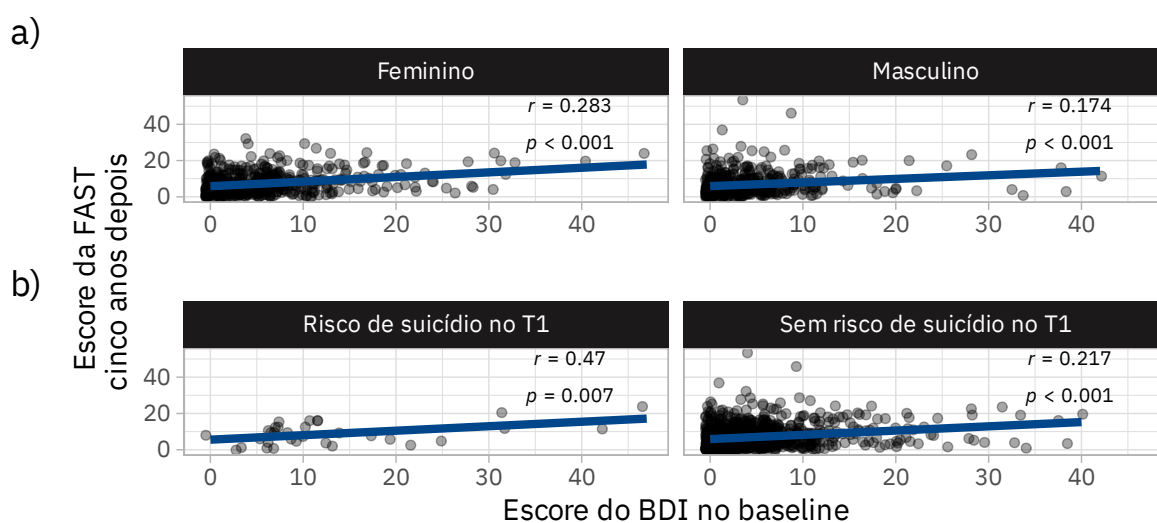


Figura 13 – Associações lineares do escore de severidade de depressão (escore total do BDI) no início do estudo com a) desempenho cognitivo (escore do MoCA) e b) prejuízo funcional (escore da FAST) cinco anos depois. As linhas de regressão ajustadas mostram uma correlação negativa da gravidade dos sintomas depressivos com o desempenho cognitivo ( $r = -0,18$ ,  $n = 959$ ), e uma correlação positiva com prejuízo funcional ( $r = 0,24$ ,  $n = 982$ ). Indivíduos com episódios atuais de humor no acompanhamento foram excluídos da análise ( $n = 107$ ), a fim de minimizar o efeito direto dos sintomas de humor nessas medidas, resultando em uma subamostra de 984 indivíduos. Abreviações: MoCA (*Montreal Cognitive Assessment*). FAST (*Functioning Assessment Short Test*).



Na análise de subgrupo, entre as mulheres, uma correlação positiva moderada ( $r = 0,283$ ,  $p < 0,001$ ) sugere uma associação entre a severidade da depressão basal e o comprometimento funcional. Por outro lado, entre os homens, foi observada uma relação positiva mais fraca ( $r = 0,174$ ,  $p < 0,001$ ) [Figura 14a]. Quando estratificados pelo risco de suicídio no início do estudo, os indivíduos com risco de suicídio exibiram uma forte correlação positiva ( $r = 0,47$ ,  $p = 0,007$ ) entre a gravidade da depressão no *baseline* e o prejuízo funcional (Figura 14b). Em contraste, aqueles sem risco de suicídio mostraram uma relação positiva moderada ( $r = 0,217$ ,  $p < 0,001$ ).

Figura 14 – Correlações do escore de severidade da depressão (escore total do Inventário de Depressão de Beck [BDI]) no início do estudo com prejuízo funcional (escore global da *Functioning Assessment Short Test* [FAST]) cinco anos depois, estratificado por a) sexo e b) risco de suicídio no início do estudo. a) Entre as mulheres, um coeficiente de correlação positivo ( $r$ ) de 0,283 sugere uma associação moderada. Por outro lado, entre os homens, a correlação ( $r = 0,174$ ) indica uma relação positiva mais fraca. b) Entre os indivíduos com risco de suicídio no início do estudo, observa-se forte correlação positiva ( $r = 0,47$ ). Em contraste, entre aqueles sem risco de suicídio, houve uma relação positiva moderada ( $r = 0,217$ ). Indivíduos com episódios atuais de humor no acompanhamento foram excluídos da análise ( $n = 984$ ), a fim de minimizar o efeito direto dos sintomas de humor no funcionamento.

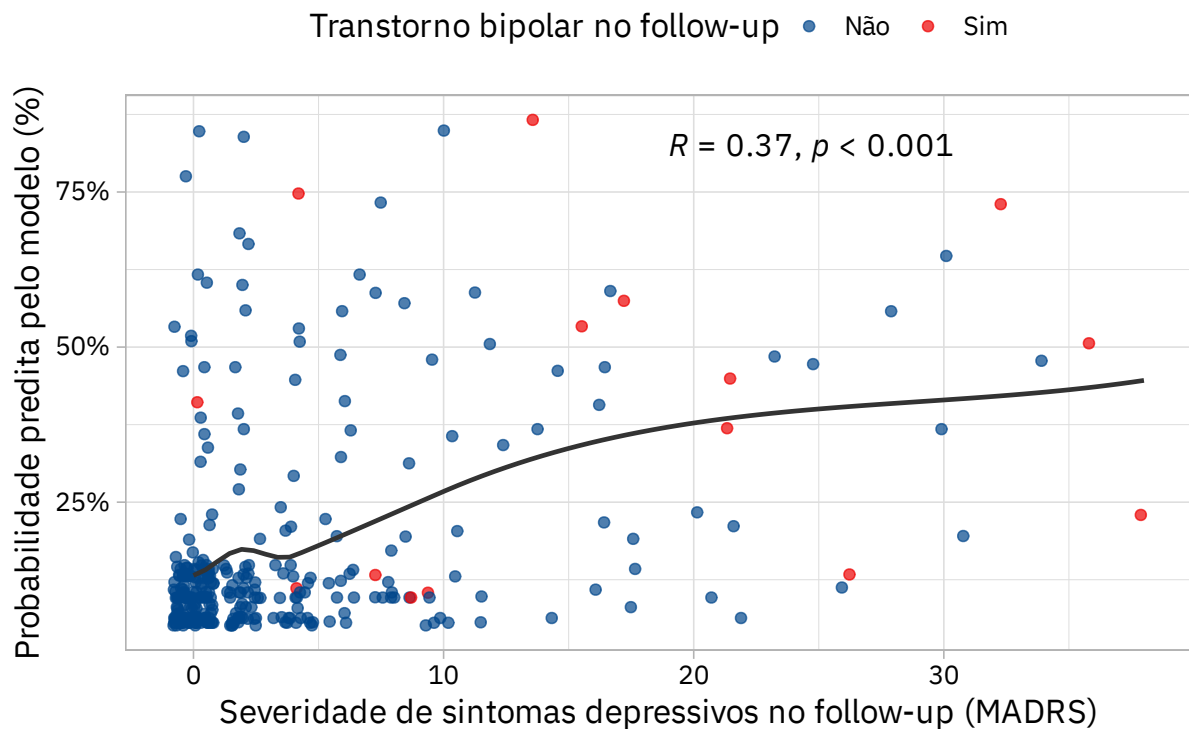


Ademais, ao avaliar as probabilidades previstas pelo modelo principal nas observações do conjunto de teste ( $n = 328$ ), observou-se uma correlação (produto-momento de Pearson) significativa entre a probabilidade prevista pelo modelo e a severidade de sintomas depressivos no *follow-up* segundo a Montgomery-Åsberg Depression Rating Scale (MADRS), com uma associação moderada ( $r = 0,37$ , IC 95% [0,28, 0,46],  $t(325) = 7,29$ ,  $p < 0,001$ ). A associação pode ser visualizada na Figura 15. Este resultado pode explicar parcialmente a dificuldade do modelo na previsão dos casos positivos, dada que a severidade de depressão foi um padrão bastante presente nos dados onde o modelo foi treinado. Outro ponto importante de se observar (Figura 15)



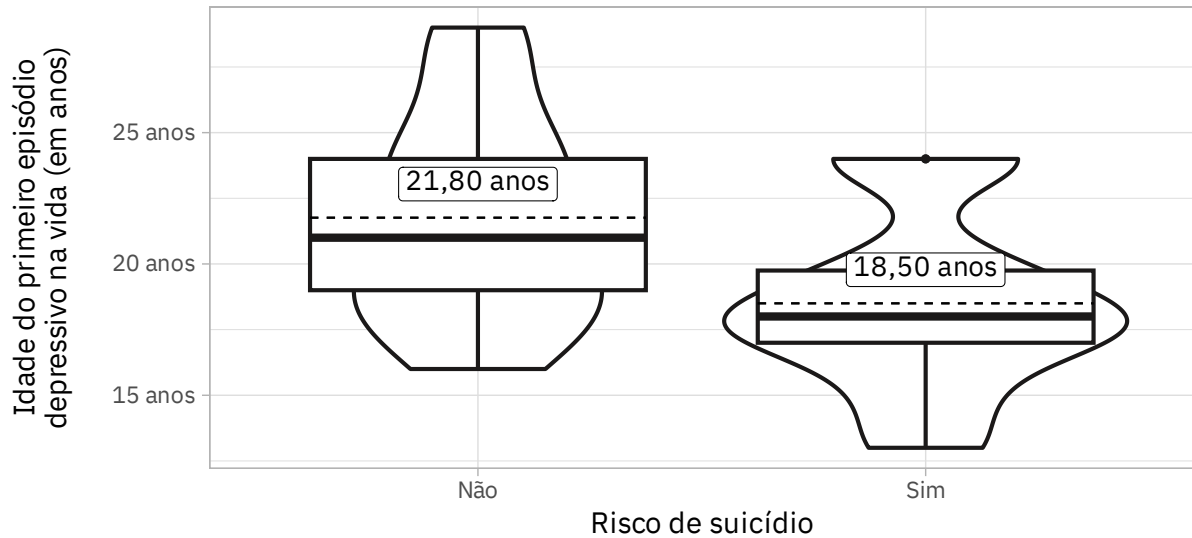
é que entre os sujeitos da base de teste que apresentaram TB após cinco anos é que não existe um padrão linear entre a depressão e a predição, o que denota uma grande heterogeneidade na gravidade de sintomatologia depressiva nos sujeitos em risco para TB.

Figura 15 – Associação entre severidade dos sintomas depressivos na segunda avaliação e probabilidade predita pelo modelo preditivo de XGBoost ( $n = 328$ ). Coeficiente de correlação de Pearson e  $p$ -valor associado estão descritos no corpo do gráfico. A curva foi ajustada a partir do método de *local polynomial regression* (LOESS).



Estudos anteriores na literatura indicam uma associação entre início precoce de transtornos mentais e piores desfechos ao longo da vida [119, 120], em especial para o TDM, onde observa-se inclusive maiores níveis de suicidalidade nesta população [121]. Portanto, objetivou-se explorar tais associações no conjunto de dados do presente projeto. Investigamos a relação entre a idade auto-relatada no primeiro episódio depressivo e o risco de suicídio entre os sujeitos do estudo. A Figura 16 ilustra a comparação entre os indivíduos com e sem risco de suicídio ( $n = 31$ ). Nossos achados revelam uma tendência, indicando que os indivíduos com risco de suicídio tendem a experimentar um início mais precoce da depressão em comparação com aqueles sem (média no grupo sem risco de suicídio = 21,76 anos, média no grupo com risco de suicídio = 18,50 anos,  $\Delta = 3,26$ , IC 95% [0,32, 6,21],  $t(20, 20) = 2,31$ ,  $p = 0,032$ ;  $d$  de Cohen = 1,03, IC 95% [0,09, 1,94]).

Figura 16 – Idade do primeiro episódio depressivo na vida (auto-relato) entre sujeitos com e sem risco de suicídio (baseado no módulo c da MINI). O gráfico mostra uma tendência que sujeitos com risco de suicídio apresentam início mais precoce de depressão ( $n = 31$ ). Oitenta e cinco sujeitos apresentaram um episódio depressivo no baseline e 54 apresentaram valores ausentes na idade de início da depressão, resultando em uma subamostra final de 31 indivíduos. O teste  $t$  de Student foi usado e sugere uma diferença significativa entre os grupos ( $\mu_{semrisco} = 21,76$ ,  $\mu_{comrisco} = 18,50$ ,  $\Delta = 3,26$ , IC 95% [0,32–6,21],  $t(20, 20) = 2,31$ ,  $p = 0,032$ ,  $d = 1,03$ , IC 95% [0,09, 1,94]). Abreviações: MINI (*Mini International Neuropsychiatric Interview*).



## 6 Aspectos éticos

### 6.1 Coleta e manejo de dados

O projeto que possibilitou a coleta dos dados utilizados no presente trabalho foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Católica de Pelotas pelo número de protocolo 2008/118. Para mais, os participantes da pesquisa assinaram um Termo de Consentimento Livre e Esclarecido.

Os adultos jovens que aceitaram participar da coleta responderam à uma entrevista clínica para avaliar a presença de alguma psicopatologia. Os sujeitos que fecharam diagnóstico para algum transtorno psiquiátrico durante a entrevista foram encaminhados para o serviço de saúde pública local. Em relação ao encaminhamento, o estabelecimento de saúde específico era decidido de acordo com a gravidade do quadro do sujeito, juntamente com as possibilidades de deslocamento e disponibilidade do indivíduo.

Os dados foram manejados a partir das melhores práticas de importação e limpeza de dados, onde foram construídos fluxos (*pipelines*) reproduzíveis que podem ser replicados a partir de códigos abertos escritos na linguagem de programação R.

### 6.2 Predição de desfechos clínicos em saúde mental

Há uma grande discussão ética em torno do avanço de tecnologias preditivas de diagnóstico em saúde. Estudos na literatura já apontam limitações e riscos éticos da medicina personalizada ao falar sobre privacidade, confidencialidade, disponibilidade de suporte ao paciente [122]. O consentimento e a forma de gestão de dados também pode ser uma lacuna a ser trabalhada na área, levando em conta o aumento na quantidade de biobancos e *databases* com grandes volumes de dados [123], em especial na psiquiatria onde trabalha-se com dados extremamente sensíveis [124]. Outro aspecto importante é a inclusão de pessoas com deficiência no campo da medicina e da psiquiatria personalizada [125], e isto não é exclusividade desta população, mas de vários subgrupos sociais pouco representados que acabam não sendo estudados de forma adequada a partir de suas particularidades. Nesta linha, podemos destacar o estigma e o viés algorítmico como possíveis desafios a serem lidados pela comunidade científica. O viés algorítmico muitas vezes pode ser explicado pela falta de dados sobre determinada população, já o estigma pode ser replicado em pesquisa a partir do ponto de vista individual, público ou social, manifestações estas que acabam refletindo na construção de conhecimento [126]. E por mais que existam iniciativas com uso de técnicas de AM em diversas frentes da psiquiatria, alguns objetos de estudo acabam se apresentando mais multifacetados do que ou-

tros. Um exemplo é o suicídio, que apesar da literatura demonstrar contribuições relevantes para a área no que tange predição de ideação ou tentativas de suicídio, ainda existe muita discussão sobre a melhor maneira de implementação de tais ferramentas na prática clínica, especialmente ao explorar questões éticas importantes [127].

## 7 Conclusão

Em resumo, desenvolveu-se um modelo com técnicas de AM capaz de prever o diagnóstico de TB em uma população específica de adultos jovens em um período de aproximadamente cinco anos. O modelo incorpora variáveis clínicas, socioambientais, de uso de substâncias, relacionadas à vida sexual e demográficas coletadas por meio de amostra probabilística. Nossos achados também destacaram associações significativas entre a gravidade dos sintomas depressivos e os desfechos funcionais e cognitivos. No entanto, para uma assinatura mais abrangente do diagnóstico de TB, estudos futuros devem se concentrar na realização de acompanhamentos sistemáticos para rastrear indivíduos em vários estágios de desenvolvimento. Além disso, adotar abordagens multimodais pode aumentar ainda mais a capacidade preditiva e a utilidade clínica desses modelos [128]. Além disso, os esforços de pesquisa devem visar populações específicas em risco, como indivíduos com depressão ou filhos de pais com TB. Além disso, a incorporação de dados digitais de saúde, marcadores biológicos, informações neuropsicológicas e o uso de técnicas de neuroimagem são promissores para o desenvolvimento de modelos futuros com maior aplicabilidade.

É importante ressaltar que a “maior aplicabilidade” de modelos como este reside em sua capacidade de complementar a prática clínica e auxiliar as práticas tradicionais da psiquiatria e psicologia. Embora nossos resultados forneçam uma base exploratória para a predição de diagnósticos, a verdadeira utilidade clínica emerge quando esses modelos são integrados às decisões de tratamento e prevenção. A aplicação clínica eficaz desses modelos requer uma abordagem multidisciplinar, onde psiquiatras, psicólogos e outros profissionais da saúde trabalham em conjunto com cientistas de dados e cientistas da computação, por exemplo. Além disso, a coleta contínua de dados clínicos e o monitoramento do progresso do paciente podem ajudar a calibrar e aprimorar os modelos ao longo do tempo para possibilitar abordagens inovadoras como a “clínica digital” [129].

Considerando os aspectos abordados no decorrer dessa dissertação, entende-se que os fenômenos do diagnóstico e da compreensão da trajetória do TB são complexos e demandam estudos robustos para a captação de melhores insights. Nesse sentido, o TB ainda é palco de muita discussão no que tange sua caracterização e identificação.

# Bibliografia

1. McIntyre RS, Berk M, Brietzke E, Goldstein BI, López-Jaramillo C, Kessing LV, Malhi GS, Nierenberg AA, Rosenblat JD, Majeed A, Vieta E, Vinberg M, Young AH e Mansur RB. Bipolar disorders. *Lancet* (London, England). 2020 Dec; 396:1841–56. DOI: [10.1016/S0140-6736\(20\)31544-0](https://doi.org/10.1016/S0140-6736(20)31544-0). eprint: [33278937](https://eprints.elsevier.com/33278937)
2. Merikangas KR, Jin R, He JP, Kessler RC, Lee S, Sampson NA, Viana MC, Andrade LH, Hu C, Karam EG, Ladea M, Medina-Mora ME, Ono Y, Posada-Villa J, Sagar R, Wells JE e Zarkov Z. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of General Psychiatry*. 2011 Mar; 68:241–51. DOI: [10.1001/archgenpsychiatry.2011.12](https://doi.org/10.1001/archgenpsychiatry.2011.12). eprint: [21383262](https://eprints.elsevier.com/21383262)
3. Kessing LV, Vradi E e Andersen PK. Life expectancy in bipolar disorder. *Bipolar Disorders*. 2015 Aug; 17:543–8. DOI: [10.1111/bdi.12296](https://doi.org/10.1111/bdi.12296). eprint: [25846854](https://eprints.elsevier.com/25846854)
4. Plans L, Barrot C, Nieto E, Rios J, Schulze TG, Papiol S, Mitjans M, Vieta E e Benabarre A. Association between completed suicide and bipolar disorder: A systematic review of the literature. *Journal of Affective Disorders*. 2019 Jan; 242:111–22. DOI: [10.1016/j.jad.2018.08.054](https://doi.org/10.1016/j.jad.2018.08.054). eprint: [30173059](https://eprints.elsevier.com/30173059)
5. Cloutier M, Greene M, Guerin A, Touya M e Wu E. The economic burden of bipolar I disorder in the United States in 2015. *Journal of Affective Disorders*. 2018 Jan; 226:45–51. DOI: [10.1016/j.jad.2017.09.011](https://doi.org/10.1016/j.jad.2017.09.011). eprint: [28961441](https://eprints.elsevier.com/28961441)
6. Yatham LN, Kennedy SH, Parikh SV, Schaffer A, Bond DJ, Frey BN, Sharma V, Goldstein BI, Rej S, Beaulieu S, Alda M, MacQueen G, Milev RV, Ravindran A, O'Donovan C, McIntosh D, Lam RW, Vazquez G, Kapczinski F, McIntyre RS, Kozicky J, Kanba S, Lafer B, Suppes T, Calabrese JR, Vieta E, Malhi G, Post RM e Berk M. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disorders*. 2018 Mar; 20:97–170. DOI: [10.1111/bdi.12609](https://doi.org/10.1111/bdi.12609). eprint: [29536616](https://eprints.elsevier.com/29536616)
7. Scott J e Leboyer M. Consequences of delayed diagnosis of bipolar disorders. *L'Encephale*. 2011 Dec; 37:3. DOI: [10.1016/S0013-7006\(11\)70048-3](https://doi.org/10.1016/S0013-7006(11)70048-3). eprint: [22212870](https://eprints.elsevier.com/22212870)
8. Altamura AC, Buoli M, Caldiroli A, Caron L, Melter CC, Dobra C, Cigliobianco M e Quarantini FZ. Misdiagnosis, duration of untreated illness (DUI) and outcome in bipolar patients with psychotic symptoms: A naturalistic study. *Journal of Affective Disorders*. 2015 Aug; 182:70–5. DOI: [10.1016/j.jad.2015.04.024](https://doi.org/10.1016/j.jad.2015.04.024). eprint: [25978716](https://eprints.elsevier.com/25978716)
9. McIntyre RS e Calabrese JR. Bipolar depression: the clinical characteristics and unmet needs of a complex disorder. *Current Medical Research and Opinion*. 2019 Nov; 35:1993–2005. DOI: [10.1080/03007995.2019.1636017](https://doi.org/10.1080/03007995.2019.1636017). eprint: [31311335](https://eprints.elsevier.com/31311335)

10. Fernandes BS, Williams LM, Steiner J, Leboyer M, Carvalho AF e Berk M. The new field of 'precision psychiatry'. *BMC Medicine*. 2017 Apr; 15:80. DOI: [10.1186/s12916-017-0849-x](https://doi.org/10.1186/s12916-017-0849-x). eprint: [28403846](https://eprints.bmc.com/28403846)
11. Passos IC, Mwangi B e Kapczinski F. Big data analytics and machine learning: 2015 and beyond. *Lancet Psychiatry*. 2016 Jan; 3:13–5. DOI: [10.1016/S2215-0366\(15\)00549-0](https://doi.org/10.1016/S2215-0366(15)00549-0). eprint: [26772057](https://eprints.lancet.com/26772057)
12. Librenza-Garcia D, Kotzian BJ, Yang J, Mwangi B, Cao B, Lima LNP, Bermudez MB, Boeira MV, Kapczinski F e Passos IC. The impact of machine learning techniques in the study of bipolar disorder: A systematic review. *Neuroscience and Biobehavioral Reviews*. 2017 Sep; 80:538–54. DOI: [10.1016/j.neubiorev.2017.07.004](https://doi.org/10.1016/j.neubiorev.2017.07.004). eprint: [28728937](https://eprints.neurosci.com/28728937)
13. Passos IC, Ballester PL, Barros RC, Librenza-Garcia D, Mwangi B, Birmaher B, Brietzke E, Hajek T, Jaramillo CL, Mansur RB, Alda M, Haarman BCM, Isometsa E, Lam RW, McIntyre RS, Minuzzi L, Kessing LV, Yatham LN, Duffy A e Kapczinski F. Machine learning and big data analytics in bipolar disorder: A position paper from the International Society for Bipolar Disorders Big Data Task Force. *Bipolar Disorders*. 2019 Nov; 21:582–94. DOI: [10.1111/bdi.12828](https://doi.org/10.1111/bdi.12828). eprint: [31465619](https://eprints.bipolar.com/31465619)
14. American Psychiatric Association, Cordioli AV, Silva CTB da, Passos IC, Kieling C e Barcellos MT. *DSM-5 - Manual Diagnóstico e Estatístico de Transtornos Mentais*. Brazil: Artmed, 2014 May. Disponível em: <https://psychiatry.org/psychiatrists/practice/dsm>
15. Muzina DJ, Colangelo E, Manning JS e Calabrese JR. Differentiating bipolar disorder from depression in primary care. *Cleveland Clinic Journal of Medicine*. 2007 Feb; 74:89. DOI: [10.3949/ccjm.74.2.89](https://doi.org/10.3949/ccjm.74.2.89). eprint: [17333635](https://eprints.clevelandclinic.com/17333635)
16. World Mental Health Survey Initiative | Survey Research Center. [Online; acesso em 16 de agosto de 2022]. 2022 Aug. Disponível em: <https://www.src.isr.umich.edu/projects/world-mental-health-survey-initiative>
17. Carvalho AF, Firth J e Vieta E. Bipolar Disorder. *New England Journal of Medicine*. 2020 Jul; 383:58–66. DOI: [10.1056/NEJMra1906193](https://doi.org/10.1056/NEJMra1906193)
18. Solé B, Jiménez E, Torrent C, Reinares M, Del Mar Bonnin C, Torres I, Varo C, Grande I, Valls E, Salagre E, Sanchez-Moreno J, Martinez-Aran A, Carvalho AF e Vieta E. Cognitive Impairment in Bipolar Disorder: Treatment and Prevention Strategies. *International Journal of Neuropsychopharmacology*. 2017 Aug; 20:670–80. DOI: [10.1093/ijnp/pyx032](https://doi.org/10.1093/ijnp/pyx032). eprint: [28498954](https://eprints.oxfordjournals.com/28498954)
19. Cardoso T, Bauer IE, Meyer TD, Kapczinski F e Soares JC. Neuroprogression and Cognitive Functioning in Bipolar Disorder: A Systematic Review. *Current Psychiatry Reports*. 2015 Sep; 17:75. DOI: [10.1007/s11920-015-0605-x](https://doi.org/10.1007/s11920-015-0605-x). eprint: [26257147](https://eprints.cupress.com/26257147)

20. Léda-Rêgo G, Bezerra-Filho S e Miranda-Scippa Â. Functioning in euthymic patients with bipolar disorder: A systematic review and meta-analysis using the Functioning Assessment Short Test. *Bipolar Disorders*. 2020 Sep; 22:569–81. DOI: [10.1111/bdi.12904](https://doi.org/10.1111/bdi.12904). eprint: [32243046](https://eprints.org/32243046)
21. Del Mar Bonnín C, Reinares M, Martínez-Arán A, Jiménez E, Sánchez-Moreno J, Solé B, Montejo L e Vieta E. Improving Functioning, Quality of Life, and Well-being in Patients With Bipolar Disorder. *International Journal of Neuropsychopharmacology*. 2019 Aug; 22:467–77. DOI: [10.1093/ijnp/pyz018](https://doi.org/10.1093/ijnp/pyz018). eprint: [31093646](https://eprints.org/31093646)
22. Miller JN e Black DW. Bipolar Disorder and Suicide: a Review. *Current Psychiatry Reports*. 2020 Jan; 22:6. DOI: [10.1007/s11920-020-1130-0](https://doi.org/10.1007/s11920-020-1130-0). eprint: [31955273](https://eprints.org/31955273)
23. Dome P, Rihmer Z e Gonda X. Suicide Risk in Bipolar Disorder: A Brief Review. *Medicina (Kaunas, Lithuania)*. 2019 Jul; 55:403. DOI: [10.3390/medicina55080403](https://doi.org/10.3390/medicina55080403). eprint: [31344941](https://eprints.org/31344941)
24. Benard V, Vaiva G, Masson M e Geoffroy PA. Lithium and suicide prevention in bipolar disorder. *L'Encephale*. 2016 Jun; 42:234–41. DOI: [10.1016/j.encep.2016.02.006](https://doi.org/10.1016/j.encep.2016.02.006). eprint: [27000268](https://eprints.org/27000268)
25. Fountoulakis KN, Tohen M e Zarate Jr. CA. Lithium treatment of Bipolar disorder in adults: A systematic review of randomized trials and meta-analyses. *European Neuropsychopharmacology : the Journal of the European College of Neuropsychopharmacology*. 2022 Jan; 54:100–15. DOI: [10.1016/j.euroneuro.2021.10.003](https://doi.org/10.1016/j.euroneuro.2021.10.003). eprint: [34980362](https://eprints.org/34980362)
26. Del Matto L, Muscas M, Murru A, Verdolini N, Anmella G, Fico G, Corponi F, Carvalho AF, Samalin L, Carpiniello B, Fagiolini A, Vieta E e Pacchiarotti I. Lithium and suicide prevention in mood disorders and in the general population: A systematic review. *Neuroscience and Biobehavioral Reviews*. 2020 Sep; 116:142–53. DOI: [10.1016/j.neubiorev.2020.06.017](https://doi.org/10.1016/j.neubiorev.2020.06.017). eprint: [32561344](https://eprints.org/32561344)
27. Preuss UW, Schaefer M, Born C e Grunze H. Bipolar Disorder and Comorbid Use of Illicit Substances. *Medicina (Kaunas, Lithuania)*. 2021 Nov; 57:1256. DOI: [10.3390/medicina57111256](https://doi.org/10.3390/medicina57111256). eprint: [34833474](https://eprints.org/34833474)
28. Bartoli F, Crocamo C e Carrà G. Cannabis use disorder and suicide attempts in bipolar disorder: A meta-analysis. *Neuroscience and Biobehavioral Reviews*. 2019 Aug; 103:14–20. DOI: [10.1016/j.neubiorev.2019.05.017](https://doi.org/10.1016/j.neubiorev.2019.05.017). eprint: [31121199](https://eprints.org/31121199)
29. Carbone EA, Filippis R de, Caroleo M, Calabrò G, Staltari FA, Destefano L, Gaetano R, Steardo Jr. L e De Fazio P. Antisocial Personality Disorder in Bipolar Disorder: A Systematic Review. *Medicina (Kaunas, Lithuania)*. 2021 Feb; 57:183. DOI: [10.3390/medicina57020183](https://doi.org/10.3390/medicina57020183). eprint: [33672619](https://eprints.org/33672619)



30. Dunalska A, Rzeszutek M, Dębowska Z e Bryńska A. Comorbidity of bipolar disorder and autism spectrum disorder - review paper. *Psychiatria Polska*. 2021 Dec; 55:1421–31. DOI: [10.12740/PP/OnlineFirst/122350](https://doi.org/10.12740/PP/OnlineFirst/122350). eprint: [35472236](https://eprints.pwr.edu.pl/eprint/35472236)
31. Lucidi L, Pettorruso M, Vellante F, Di Carlo F, Ceci F, Santovito MC, Di Muzio I, Fornaro M, Ventriglio A, Tomasetti C, Valchera A, Gentile A, Kim YK, Martinotti G, Fraticelli S, Di Giannantonio M e De Berardis D. Gut Microbiota and Bipolar Disorder: An Overview on a Novel Biomarker for Diagnosis and Treatment. *International Journal of Molecular Sciences*. 2021 Apr; 22:3723. DOI: [10.3390/ijms22073723](https://doi.org/10.3390/ijms22073723). eprint: [33918462](https://eprints.pwr.edu.pl/eprint/33918462)
32. Nestsiarovich A, Gaudiot CES, Baldessarini RJ, Vieta E, Zhu Y e Tohen M. Preventing new episodes of bipolar disorder in adults: Systematic review and meta-analysis of randomized controlled trials. *European Neuropsychopharmacology : the Journal of the European College of Neuropsychopharmacology*. 2022 Jan; 54:75–89. DOI: [10.1016/j.euroneuro.2021.08.264](https://doi.org/10.1016/j.euroneuro.2021.08.264). eprint: [34489127](https://eprints.pwr.edu.pl/eprint/34489127)
33. Marchionatti LE, Antonelli-Salgado T, Erthal IN, Lopes SLS, Lucas LS, Rebouças DB e Passos IC. Bipolar disorder treatment according to illness trajectory: a systematic review of clinical trials. *Psychiatry Research*. 2022 Jun; 312:114572. DOI: [10.1016/j.psychres.2022.114572](https://doi.org/10.1016/j.psychres.2022.114572). eprint: [35490572](https://eprints.pwr.edu.pl/eprint/35490572)
34. Zhang C, Xiao X, Li T e Li M. Translational genomics and beyond in bipolar disorder. *Molecular Psychiatry*. 2021 Jan; 26:186–202. DOI: [10.1038/s41380-020-0782-9](https://doi.org/10.1038/s41380-020-0782-9). eprint: [32424235](https://eprints.pwr.edu.pl/eprint/32424235)
35. Prata DP, Costa-Neves B, Cosme G e Vassos E. Unravelling the genetic basis of schizophrenia and bipolar disorder with GWAS: A systematic review. *Journal of Psychiatric Research*. 2019 Jul; 114:178–207. DOI: [10.1016/j.jpsychires.2019.04.007](https://doi.org/10.1016/j.jpsychires.2019.04.007). eprint: [31096178](https://eprints.pwr.edu.pl/eprint/31096178)
36. Munkholm K, Vinberg M, Berk M e Kessing LV. State-related alterations of gene expression in bipolar disorder: a systematic review. *Bipolar Disorders*. 2012 Nov; 14:684–96. DOI: [10.1111/bdi.12005](https://doi.org/10.1111/bdi.12005). eprint: [23043691](https://eprints.pwr.edu.pl/eprint/23043691)
37. Mullins N et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics*. 2021 Jun; 53:817–29. DOI: [10.1038/s41588-021-00857-4](https://doi.org/10.1038/s41588-021-00857-4). eprint: [34002096](https://eprints.pwr.edu.pl/eprint/34002096)
38. Braithwaite I, Zhang S, Kirkbride JB, Osborn DPJ e Hayes JF. Air Pollution (Particulate Matter) Exposure and Associations with Depression, Anxiety, Bipolar, Psychosis and Suicide Risk: A Systematic Review and Meta-Analysis. *Environmental Health Perspectives*. 2019 Dec; 127:126002. DOI: [10.1289/EHP4595](https://doi.org/10.1289/EHP4595). eprint: [31850801](https://eprints.pwr.edu.pl/eprint/31850801)

39. Post RM e Leverich GS. The role of psychosocial stress in the onset and progression of bipolar disorder and its comorbidities: the need for earlier and alternative modes of therapeutic intervention. *Development and Psychopathology*. 2006; Fall;18:1181–211. DOI: [10.1017/S0954579406060573](https://doi.org/10.1017/S0954579406060573). eprint: [17064434](https://eprints.org/17064434)
40. Bauer M, Glenn T, Alda M, Andreassen OA, Angelopoulos E, Ardaur R, Baethge C, Bauer R, Baune BT, Bellivier F, Belmaker RH, Berk M, Bjella TD, Bossini L, Bersudsky Y, Cheung EYW, Conell J, Del Zompo M, Dodd S, Etain B, Fagiolini A, Frye MA, Fountoulakis KN, Garneau-Fournier J, Gonzalez-Pinto A, Gottlieb JF, Harima H, Hassel S, Henry C, Iacovides A, Isometsä ET, Kapczinski F, Kliwicky S, König B, Krogh R, Kunz M, Lafer B, Larsen ER, Lewitzka U, Lopez-Jaramillo C, MacQueen G, Manchia M, Marsh W, Martinez-Cengotitabengoa M, Melle I, Monteith S, Morken G, Munoz R, Nery FG, O'Donovan C, Osher Y, Pfennig A, Quiroz D, Ramesar R, Rasgon N, Reif A, Ritter P, Rybakowski JK, Sagduyu K, Miranda-Scippa Â, Severus E, Simhandl C, Stein DJ, Strejilevich S, Sulaiman AH, Suominen K, Tagata H, Tatebayashi Y, Torrent C, Vieta E, Viswanath B, Wanchoo MJ, Zetin M e Whybrow PC. Influence of light exposure during early life on the age of onset of bipolar disorder. *Journal of Psychiatric Research*. 2015 May; 64:1–8. DOI: [10.1016/j.jpsychires.2015.03.013](https://doi.org/10.1016/j.jpsychires.2015.03.013). eprint: [25862378](https://eprints.org/25862378)
41. Bauer M et al. Association between polarity of first episode and solar insolation in bipolar I disorder. *Journal of Psychosomatic Research*. 2022 Sep; 160:110982. DOI: [10.1016/j.jpsychores.2022.110982](https://doi.org/10.1016/j.jpsychores.2022.110982). eprint: [35932492](https://eprints.org/35932492)
42. Parker GB, Hadzi-Pavlovic D e Graham RK. Examining for any impact of climate change on the association between seasonality and hospitalization for mania. *Journal of Affective Disorders*. 2017 Jan; 208:431–5. DOI: [10.1016/j.jad.2016.11.006](https://doi.org/10.1016/j.jad.2016.11.006). eprint: [27846412](https://eprints.org/27846412)
43. Poyraz CA, Özdemir A, Şen CÇ, Sağlam NGU, Enginkaya S e Tomruk N. The Impact of Coping Strategies on Suicide Attempts and Suicidal Ideation in Bipolar Disorder. *Journal of Nervous and Mental Disease*. 2021 Aug; 209:564–70. DOI: [10.1097/NMD.0000000000001347](https://doi.org/10.1097/NMD.0000000000001347). eprint: [33867505](https://eprints.org/33867505)
44. Mitchell T. *Machine Learning*. en. McGraw-Hill Series in Computer Science. New York, NY: McGraw-Hill Professional, 1997 Mar
45. Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*. 1959 Jul; 3:210–29. DOI: [10.1147/rd.33.0210](https://doi.org/10.1147/rd.33.0210)
46. James G, Witten D, Hastie T, Tibshirani R e Taylor J. *An Introduction to Statistical Learning*. Cham, Switzerland: Springer International Publishing, 2023. Disponível em: <https://link.springer.com/book/10.1007/978-3-031-38747-0>

47. Farrell CJ. Identifying mislabelled samples: Machine learning models exceed human performance. *Ann. Clin. Biochem.* 2021 Nov; 58:650–2. DOI: [10.1177/00045632211032991](https://doi.org/10.1177/00045632211032991). eprint: [34210147](https://doi.org/10.1177/00045632211032991)
48. Mazurowski MA, Buda M, Saha A e Bashir MR. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *J. Magn. Reson. Imaging* 2019 Apr; 49:939–54. DOI: [10.1002/jmri.26534](https://doi.org/10.1002/jmri.26534). eprint: [30575178](https://doi.org/10.1002/jmri.26534)
49. Zanardi R, Prestifilippo D, Fabbri C, Colombo C, Maron E e Serretti A. Precision psychiatry in clinical practice. *International Journal of Psychiatry in Clinical Practice.* 2021 Mar; 25:19–27. DOI: [10.1080/13651501.2020.1809680](https://doi.org/10.1080/13651501.2020.1809680). eprint: [32852246](https://doi.org/10.1080/13651501.2020.1809680)
50. Rajan K. Materials Informatics: The Materials “Gene” and Big Data. *Annual Review of Materials Research* 2015; 45:153–69
51. Wong EHF, Yocca F, Smith MA e Lee CM. Challenges and opportunities for drug discovery in psychiatric disorders: the drug hunters’ perspective. *International Journal of Neuropsychopharmacology.* 2010 Oct; 13:1269–84. DOI: [10.1017/S1461145710000866](https://doi.org/10.1017/S1461145710000866). eprint: [20716397](https://doi.org/10.1017/S1461145710000866)
52. Fialho AR, **Montezano BB**, Ballester PL, Azevedo Cardoso T de, Mondin TC, Moreira FP, Souza LDdM, Silva RA da e Jansen K. Predicting 3-year persistent or recurrent major depressive episode using machine learning techniques. *Psychiatry Research Communications.* 2022 Sep; 2:100055. DOI: [10.1016/j.psychom.2022.100055](https://doi.org/10.1016/j.psychom.2022.100055)
53. Bzdok D e Meyer-Lindenberg A. Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry Cognitive Neuroscience and Neuroimaging.* 2018 Mar; 3:223–30. DOI: [10.1016/j.bpsc.2017.11.007](https://doi.org/10.1016/j.bpsc.2017.11.007). eprint: [29486863](https://doi.org/10.1016/j.bpsc.2017.11.007)
54. Tai AMY, Albuquerque A, Carmona NE, Subramaniepillai M, Cha DS, Sheko M, Lee Y, Mansur R e McIntyre RS. Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. *Artificial Intelligence in Medicine.* 2019 Aug; 99:101704. DOI: [10.1016/j.artmed.2019.101704](https://doi.org/10.1016/j.artmed.2019.101704). eprint: [31606109](https://doi.org/10.1016/j.artmed.2019.101704)
55. Tate AE, McCabe RC, Larsson H, Lundström S, Lichtenstein P e Kuja-Halkola R. Predicting mental health problems in adolescence using machine learning techniques. *PloS One.* 2020 Apr; 15:e0230389. DOI: [10.1371/journal.pone.0230389](https://doi.org/10.1371/journal.pone.0230389). eprint: [32251439](https://doi.org/10.1371/journal.pone.0230389)
56. Quinlan EB, Banaschewski T, Barker GJ, Bokde ALW, Bromberg U, Büchel C, Desrivières S, Flor H, Frouin V, Garavan H, Heinz A, Brühl R, Martinot JL, Martinot MLP, Nees F, Orfanos DP, Paus T, Poustka L, Hohmann S, Smolka MN, Fröhner JH, Walter H, Whelan R, Schumann G e Consortium I. Identifying biological markers for improved precision medicine in psychiatry. *Molecular Psychiatry.* 2020 Feb; 25:243–53. DOI: [10.1038/s41380-019-0555-5](https://doi.org/10.1038/s41380-019-0555-5). eprint: [31676814](https://doi.org/10.1038/s41380-019-0555-5)

57. Pardiñas AF, Owen MJ e Walters JTR. Pharmacogenomics: A road ahead for precision medicine in psychiatry. *Neuron*. 2021 Dec; 109:3914–29. DOI: [10.1016/j.neuron.2021.09.011](https://doi.org/10.1016/j.neuron.2021.09.011). eprint: [34619094](https://eprints.oxfordjournals.org/doi/abs/10.1016/j.neuron.2021.09.011)
58. Lin E, Lin CH e Lane HY. Precision Psychiatry Applications with Pharmacogenomics: Artificial Intelligence and Machine Learning Approaches. *International Journal of Molecular Sciences*. 2020 Feb; 21:969. DOI: [10.3390/ijms21030969](https://doi.org/10.3390/ijms21030969). eprint: [32024055](https://eprints.oxfordjournals.org/doi/abs/10.3390/ijms21030969)
59. Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, Marsh J, DeVyllder J, Walter M, Berrouguet S e Lemey C. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. *Journal of Medical Internet Research*. 2021 May; 23:e15708. DOI: [10.2196/15708](https://doi.org/10.2196/15708). eprint: [33944788](https://eprints.oxfordjournals.org/doi/abs/10.2196/15708)
60. Claude LA, Houenou J, Duchesnay E e Favre P. Will machine learning applied to neuroimaging in bipolar disorder help the clinician? A critical review and methodological suggestions. *Bipolar Disorders*. 2020 Jun; 22:334–55. DOI: [10.1111/bdi.12895](https://doi.org/10.1111/bdi.12895). eprint: [32108409](https://eprints.oxfordjournals.org/doi/abs/10.1111/bdi.12895)
61. Sawalha J, Cao L, Chen J, Selvitella A, Liu Y, Yang C, Li X, Zhang X, Sun J, Zhang Y, Zhao L, Cui L, Zhang Y, Sui J, Greiner R, Li XM, Greenshaw A, Li T e Cao B. Individualized identification of first-episode bipolar disorder using machine learning and cognitive tests. *Journal of Affective Disorders*. 2021 Mar; 282:662–8. DOI: [10.1016/j.jad.2020.12.046](https://doi.org/10.1016/j.jad.2020.12.046). eprint: [33445089](https://eprints.oxfordjournals.org/doi/abs/10.1016/j.jad.2020.12.046)
62. Siqueira Rotenberg L de, Borges-Júnior RG, Lafer B, Salvini R e Silva Dias R da. Exploring machine learning to predict depressive relapses of bipolar disorder patients. *Journal of Affective Disorders*. 2021 Dec; 295:681–7. DOI: [10.1016/j.jad.2021.08.127](https://doi.org/10.1016/j.jad.2021.08.127). eprint: [34509784](https://eprints.oxfordjournals.org/doi/abs/10.1016/j.jad.2021.08.127)
63. Jan Z, Ai-Ansari N, Mousa O, Abd-Alrazaq A, Ahmed A, Alam T e Househ M. The Role of Machine Learning in Diagnosing Bipolar Disorder: Scoping Review. *Journal of Medical Internet Research*. 2021 Nov; 23:e29749. DOI: [10.2196/29749](https://doi.org/10.2196/29749). eprint: [34806996](https://eprints.oxfordjournals.org/doi/abs/10.2196/29749)
64. Bayes A, Spoelma MJ, Hadzi-Pavlovic D e Parker G. Differentiation of bipolar disorder versus borderline personality disorder: A machine learning approach. *Journal of Affective Disorders*. 2021 Jun; 288:68–73. DOI: [10.1016/j.jad.2021.03.082](https://doi.org/10.1016/j.jad.2021.03.082). eprint: [33845326](https://eprints.oxfordjournals.org/doi/abs/10.1016/j.jad.2021.03.082)
65. Han KM, De Berardis D, Fornaro M e Kim YK. Differentiating between bipolar and unipolar depression in functional and structural MRI studies. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 2019 Apr; 91:20–7. DOI: [10.1016/j.pnpbp.2018.03.022](https://doi.org/10.1016/j.pnpbp.2018.03.022). eprint: [29601896](https://eprints.oxfordjournals.org/doi/abs/10.1016/j.pnpbp.2018.03.022)

66. Tomasik J, Han SYS, Barton-Owen G, Mirea DM, Martin-Key NA, Rustogi N, Lago SG, Olmert T, Cooper JD, Ozcan S, Eljasz P, Thomas G, Tuytten R, Metcalfe T, Schei TS, Farrag LP, Friend LV, Bell E, Cowell D e Bahn S. A machine learning algorithm to differentiate bipolar disorder from major depressive disorder using an online mental health questionnaire and blood biomarker data. *Translational Psychiatry*. 2021 Jan; 11:41. DOI: [10.1038/s41398-020-01181-x](https://doi.org/10.1038/s41398-020-01181-x). eprint: [33436544](https://eprints.psych.ac.uk/33436544)
67. Wollenhaupt-Aguiar B, Librenza-Garcia D, Bristot G, Przybylski L, Stertz L, Burque RK, Ceresér KM, Spanemberg L, Caldieraro MA, Frey BN, Fleck MP, Kauer-Sant'Anna M, Passos IC e Kapczinski F. Differential biomarker signatures in unipolar and bipolar depression: A machine learning approach. *Australian and New Zealand Journal of Psychiatry*. 2020 Apr; 54:393–401. DOI: [10.1177/0004867419888027](https://doi.org/10.1177/0004867419888027). eprint: [31789053](https://eprints.psych.ac.uk/31789053)
68. Kim YK e Na KS. Application of machine learning classification for structural brain MRI in mood disorders: Critical review from a clinical perspective. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* 2018 Jan; 80:B. DOI: [10.1016/j.pnpbp.2017.06.024](https://doi.org/10.1016/j.pnpbp.2017.06.024). eprint: [28648568](https://eprints.psych.ac.uk/28648568)
69. Colombo F, Calesella F, Mazza MG, Melloni EMT, Morelli MJ, Scotti GM, Benedetti F, Bollettini I e Vai B. Machine learning approaches for prediction of bipolar disorder based on biological, clinical and neuropsychological markers: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*. 2022 Apr; 135:104552. DOI: [10.1016/j.neubiorev.2022.104552](https://doi.org/10.1016/j.neubiorev.2022.104552). eprint: [35120970](https://eprints.psych.ac.uk/35120970)
70. Li H, Cui L, Cao L, Zhang Y, Liu Y, Deng W e Zhou W. Identification of bipolar disorder using a combination of multimodality magnetic resonance imaging and machine learning techniques. *BMC Psychiatry*. 2020 Oct; 20:488. DOI: [10.1186/s12888-020-02886-5](https://doi.org/10.1186/s12888-020-02886-5). eprint: [33023515](https://eprints.psych.ac.uk/33023515)
71. Fernandes BS, Karmakar C, Tamouza R, Tran T, Yearwood J, Hamdani N, Laouamri H, Richard JR, Yolken R, Berk M, Venkatesh S e Leboyer M. Precision psychiatry with immunological and cognitive biomarkers: a multi-domain prediction for the diagnosis of bipolar disorder or schizophrenia using machine learning. *Translational Psychiatry*. 2020 May; 10:162. DOI: [10.1038/s41398-020-0836-4](https://doi.org/10.1038/s41398-020-0836-4). eprint: [32448868](https://eprints.psych.ac.uk/32448868)
72. Li Z, Li W, Wei Y, Gui G, Zhang R, Liu H, Chen Y e Jiang Y. Deep learning based automatic diagnosis of first-episode psychosis, bipolar disorder and healthy controls. *Computerized Medical Imaging and Graphics : the Official Journal of the Computerized Medical Imaging Society*. 2021 Apr; 89:101882. DOI: [10.1016/j.compmedimag.2021.101882](https://doi.org/10.1016/j.compmedimag.2021.101882). eprint: [33684730](https://eprints.psych.ac.uk/33684730)
73. Achalia R, Sinha A, Jacob A, Achalia G, Kaginalkar V, Venkatasubramanian G e Rao NP. A proof of concept machine learning analysis using multimodal neuroimaging and neurocognitive measures as predictive biomarker in bipolar disorder. *Asian Journal of Psychiatry*. 2020 Apr; 50:101984. DOI: [10.1016/j.ajp.2020.101984](https://doi.org/10.1016/j.ajp.2020.101984). eprint: [32143176](https://eprints.psych.ac.uk/32143176)

74. Harvey D, Lobban F, Rayson P, Warner A e Jones S. Natural Language Processing Methods and Bipolar Disorder: Scoping Review. *JMIR Mental Health*. 2022 Apr; 9:e35928. DOI: [10.2196/35928](https://doi.org/10.2196/35928). eprint: [35451984](https://eprints.jmir.org/eprint/35451984)
75. Ribeiro JS, Pereira D, Salagre E, Coroa M, Oliveira PS, Santos V, Madeira N, Grande I e Vieta E. Risk Calculators in Bipolar Disorder: A Systematic Review. *Brain Sciences*. 2020 Aug; 10:525. DOI: [10.3390/brainsci10080525](https://doi.org/10.3390/brainsci10080525). eprint: [32781733](https://eprints.jmir.org/eprint/32781733)
76. Instituto Brasileiro de Geografia e Estatística. Censo Brasileiro de 2010. 2010. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html>
77. American Psychiatric Association. DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. 4<sup>a</sup> ed. Arlington, TX: American Psychiatric Press, 1994 May
78. World Health Organization. The ICD-10 classification of mental and behavioural disorders. Genève, Switzerland: World Health Organization, 1993
79. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R e Dunbar GC. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*. 1998; 59:20. eprint: [9881538](https://eprints.jmir.org/eprint/9881538). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/9881538>
80. Amorim P. Mini International Neuropsychiatric Interview (MINI): validação de entrevista breve para diagnóstico de transtornos mentais. *Brazilian Journal of Psychiatry*. 2000 Sep; 22:106–15. DOI: [10.1590/S1516-44462000000300003](https://doi.org/10.1590/S1516-44462000000300003)
81. Müller JK e Leweke FM. Bipolar disorder: clinical overview. *Medizinische Monatsschrift Fur Pharmazeuten*. 2016 Sep; 39:363. eprint: [29956510](https://eprints.jmir.org/eprint/29956510). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/29956510>
82. Akinhanmi MO, Biernacka JM, Strakowski SM, McElroy SL, Berry JEB, Merikangas KR, Assari S, McInnis MG, Schulze TG, LeBoyer M, Tamminga C, Patten C e Frye MA. Racial disparities in bipolar disorder treatment and research: a call to action. *Bipolar Disorders*. 2018 Sep; 20:506–14. DOI: [10.1111/bdi.12638](https://doi.org/10.1111/bdi.12638). eprint: [29527766](https://eprints.jmir.org/eprint/29527766)
83. Dell’Osso B, Grancini B, Vismara M, De Cagna F, Maggi M, Molle M, Cremaschi L, Miller S, Ketter TA e Altamura AC. Age at onset in patients with bipolar I and II disorder: a comparison of large sample studies. *Journal of Affective Disorders*. 2016 Sep; 201:57–63. DOI: [10.1016/j.jad.2016.04.009](https://doi.org/10.1016/j.jad.2016.04.009). eprint: [27177297](https://eprints.jmir.org/eprint/27177297)
84. Bolton S, Warner J, Harriss E, Geddes J e Saunders KEA. Bipolar disorder: Trimodal age-at-onset distribution. *Bipolar Disorders*. 2021 Jun; 23:341–56. DOI: [10.1111/bdi.13016](https://doi.org/10.1111/bdi.13016). eprint: [33030292](https://eprints.jmir.org/eprint/33030292)

85. Lu W, Dimick MK, Fiksenbaum LM, Timmins V, Mitchell RHB, Zou Y e Goldstein BI. Clinical correlates of socioeconomic status in adolescent bipolar disorder. *Comprehensive Psychiatry*. 2020 Aug; 101:152186. DOI: [10.1016/j.comppsy.2020.152186](https://doi.org/10.1016/j.comppsy.2020.152186). eprint: [32504872](https://eprints.org/32504872)
86. Diler RS, Merranko JA, Hafeman D, Goldstein TR, Goldstein BI, Hower H, Gill MK, Axelson DA, Ryan N, Strober M, Keller MB, Yen S, Hunt JI, Weinstock LM, Iyengar S e Birmaher BB. Higher socioeconomic status and less parental psychopathology improve prognosis in youths with bipolar disorder. *Journal of Affective Disorders*. 2022 Apr; 302:185–93. DOI: [10.1016/j.jad.2022.01.058](https://doi.org/10.1016/j.jad.2022.01.058). eprint: [35033593](https://eprints.org/35033593)
87. Hakulinen C, Musliner KL e Agerbo E. Bipolar disorder and depression in early adulthood and long-term employment, income, and educational attainment: A nationwide cohort study of 2,390,127 individuals. *Depression and Anxiety*. 2019 Nov; 36:1080–8. DOI: [10.1002/da.22956](https://doi.org/10.1002/da.22956). eprint: [31508865](https://eprints.org/31508865)
88. Nasrallah HA. Consequences of misdiagnosis: inaccurate treatment and poor patient outcomes in bipolar disorder. *Journal of Clinical Psychiatry*. 2015 Oct; 76:e1328. DOI: [10.4088/JCP.14016tx2c](https://doi.org/10.4088/JCP.14016tx2c). eprint: [26528666](https://eprints.org/26528666)
89. Messer T, Lammers G, Müller-Siecheneder F, Schmidt RF e Latifi S. Substance abuse in patients with bipolar disorder: A systematic review and meta-analysis. *Psychiatry Res*. 2017 Jul; 253:338–50. DOI: [10.1016/j.psychres.2017.02.067](https://doi.org/10.1016/j.psychres.2017.02.067). eprint: [28419959](https://eprints.org/28419959)
90. Lalli M, Brouillette K, Kapczinski F e Azevedo Cardoso T de. Substance use as a risk factor for bipolar disorder: A systematic review. *J. Psychiatr. Res*. 2021 Dec; 144:285–95. DOI: [10.1016/j.jpsychires.2021.10.012](https://doi.org/10.1016/j.jpsychires.2021.10.012). eprint: [34710665](https://eprints.org/34710665)
91. Correa AMP, Carrillo RAA, Vega DCS, Cano SG, Arango LGA, Builes LFT, García ÁMA, Villa EU, Aguirre-Acevedo DC e López-Jaramillo C. Sexual and Reproductive Health in Patients with Schizophrenia and Bipolar Disorder. *Rev. Colomb. Psiquiatr. (Engl. Ed)*. 2020 Jan; 49:15–22. DOI: [10.1016/j.rcp.2018.04.007](https://doi.org/10.1016/j.rcp.2018.04.007). eprint: [32081203](https://eprints.org/32081203)
92. Friedman JH. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* 2001; 29:1189–232. DOI: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451). Disponível em: <https://doi.org/10.1214/aos/1013203451>
93. Riebesell J. TikZ Diagrams on Physics and Machine Learning. [Online; acesso em 25 de agosto de 2022]. 2022 Aug. Disponível em: <https://tikz.netlify.app>
94. Lundberg S e Lee SI. A Unified Approach to Interpreting Model Predictions. arXiv 2017 May. DOI: [10.48550/arXiv.1705.07874](https://doi.org/10.48550/arXiv.1705.07874). eprint: [1705.07874](https://eprints.org/1705.07874)
95. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2022. Disponível em: <https://www.R-project.org/>
96. Arruda T de. Neovim (version 0.7.2). 2022. Disponível em: <https://neovim.io/>

97. RStudio Team. RStudio: Integrated Development Environment for R. RStudio, PBC. Boston, MA, 2022. Disponível em: <http://www.rstudio.com/>
98. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K e Yutani H. Welcome to the tidyverse. *Journal of Open Source Software* 2019; 4:1686. DOI: [10.21105/joss.01686](https://doi.org/10.21105/joss.01686)
99. Wickham H, François R, Henry L e Müller K. dplyr: A Grammar of Data Manipulation. R package version 1.0.9. 2022. Disponível em: <https://CRAN.R-project.org/package=dplyr>
100. IBM Corp. IBM SPSS Statistics for Windows. Versão 25.0. Armonk, NY: IBM Corp, 2017. Disponível em: <https://hadoop.apache.org>
101. Henry L e Wickham H. purrr: Functional Programming Tools. R package version 0.3.4. 2020. Disponível em: <https://CRAN.R-project.org/package=purrr>
102. Wickham H e Girlich M. tidyr: Tidy Messy Data. R package version 1.2.0. 2022. Disponível em: <https://CRAN.R-project.org/package=tidyr>
103. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. Disponível em: <https://ggplot2.tidyverse.org>
104. Wilkinson L, Wills D, Rope D, Norton A e Dubbs R. The Grammar of Graphics (Statistics and Computing). New York, NY, USA: Springer, 2005 Jul. Disponível em: <https://www.amazon.com/Grammar-Graphics-Statistics-Computing/dp/0387245448>
105. Kuhn M e Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. 2020. Disponível em: <https://www.tidymodels.org>
106. Silge J, Chow F, Kuhn M e Wickham H. rsample: General Resampling Infrastructure. R package version 1.1.0. 2022. Disponível em: <https://CRAN.R-project.org/package=rsample>
107. Kuhn M e Vaughan D. parsnip: A Common API to Modeling and Analysis Functions. R package version 1.0.0. 2022. Disponível em: <https://CRAN.R-project.org/package=parsnip>
108. Kuhn M e Wickham H. recipes: Preprocessing and Feature Engineering Steps for Modeling. R package version 1.0.1. 2022. Disponível em: <https://CRAN.R-project.org/package=recipes>
109. Vaughan D. workflows: Modeling Workflows. R package version 1.0.0. 2022. Disponível em: <https://CRAN.R-project.org/package=workflows>



110. Kuhn M. tune: Tidy Tuning Tools. R package version 1.0.0. 2022. Disponível em: <https://CRAN.R-project.org/package=tune>
111. Kuhn M e Vaughan D. yardstick: Tidy Characterizations of Model Performance. R package version 1.0.0. 2022. Disponível em: <https://CRAN.R-project.org/package=yardstick>
112. Robinson D, Hayes A e Couch S. broom: Convert Statistical Objects into Tidy Tibbles. R package version 1.0.0. 2022. Disponível em: <https://CRAN.R-project.org/package=broom>
113. Kuhn M e Frick H. dials: Tools for Creating Tuning Parameter Values. R package version 1.0.0. 2022. Disponível em: <https://CRAN.R-project.org/package=dials>
114. Kuhn M. finetune: Additional Functions for Model Tuning. R package version 0.2.0. 2022. Disponível em: <https://CRAN.R-project.org/package=finetune>
115. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T, Li M, Xie J, Lin M, Geng Y, Li Y e Yuan J. xgboost: Extreme Gradient Boosting. R package version 1.6.0.1. 2022. Disponível em: <https://github.com/dmlc/xgboost>
116. Liu Y e Just A. SHAPforxgboost: SHAP Plots for 'XGBoost'. R package version 0.1.1. 2021. Disponível em: <https://CRAN.R-project.org/package=SHAPforxgboost>
117. Fisher R. The design of experiments, oliver and boyd ltd. London, UK 1935
118. Kursa MB, Jankowski A e Rudnicki WR. Boruta – A System for Feature Selection. *Fundamenta Informaticae* 2010; 101:271–85. DOI: [10.3233/fi-2010-288](https://doi.org/10.3233/fi-2010-288). Disponível em: <https://doi.org/10.3233/fi-2010-288>
119. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S e Ustün TB. Age of onset of mental disorders: a review of recent literature. *Curr. Opin. Psychiatry* 2007 Jul; 20:359–64. DOI: [10.1097/YCO.0b013e32816ebc8c](https://doi.org/10.1097/YCO.0b013e32816ebc8c). eprint: [17551351](https://eprints.psychiatry.org/17551351)
120. Girolamo G de, Dagani J, Purcell R, Cocchi A e McGorry PD. Age of onset of mental disorders and use of mental health services: needs, opportunities and obstacles. *Epidemiol. Psychiatr. Sci.* 2012 Mar; 21:47–57. DOI: [10.1017/s2045796011000746](https://doi.org/10.1017/s2045796011000746). eprint: [22670412](https://eprints.psychiatry.org/22670412)
121. Korczak DJ e Goldstein BI. Childhood onset major depressive disorder: course of illness and psychiatric comorbidity in a community sample. *J. Pediatr.* 2009 Jul; 155:118–23. DOI: [10.1016/j.jpeds.2009.01.061](https://doi.org/10.1016/j.jpeds.2009.01.061). eprint: [19394039](https://eprints.psychiatry.org/19394039)
122. Salari P e Larijani B. Ethical Issues Surrounding Personalized Medicine: A Literature Review. *Acta Medica Iranica.* 2017 Mar; 55:209–17. eprint: [28282721](https://eprints.psychiatry.org/28282721). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/28282721>

123. Conboy C. Consent and Privacy in the Era of Precision Medicine and Biobanking Genomic Data. *American Journal of Law & Medicine*. 2020 May; 46:167–87. DOI: [10.1177/0098858820933493](https://doi.org/10.1177/0098858820933493). eprint: [32659188](https://eprints.org/32659188)
124. Ahmed E e Hens K. Microbiome in Precision Psychiatry: An Overview of the Ethical Challenges Regarding Microbiome Big Data and Microbiome-Based Interventions. *AJOB Neuroscience*. 2021 Aug; 1-17. DOI: [10.1080/21507740.2021.1958096](https://doi.org/10.1080/21507740.2021.1958096). eprint: [34379050](https://eprints.org/34379050)
125. Sabatello M. Precision medicine, health disparities, and ethics: the case for disability inclusion. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*. 2018 Apr; 20:397–9. DOI: [10.1038/gim.2017.120](https://doi.org/10.1038/gim.2017.120). eprint: [28771252](https://eprints.org/28771252)
126. Walsh CG, Chaudhry B, Dua P, Goodman KW, Kaplan B, Kavuluru R, Solomonides A e Subbian V. Stigma, biomarkers, and algorithmic bias: recommendations for precision behavioral health with artificial intelligence. *JAMIA Open*. 2020 Jan; 3:9–15. DOI: [10.1093/jamiaopen/ooz054](https://doi.org/10.1093/jamiaopen/ooz054). eprint: [32607482](https://eprints.org/32607482)
127. Linthicum KP, Schafer KM e Ribeiro JD. Machine learning in suicide science: Applications and ethics. *Behavioral Sciences & the Law*. 2019 May; 37:214–22. DOI: [10.1002/bsl.2392](https://doi.org/10.1002/bsl.2392). eprint: [30609102](https://eprints.org/30609102)
128. Kline A, Wang H, Li Y, Dennis S, Hutch M, Xu Z, Wang F, Cheng F e Luo Y. Multimodal machine learning in precision health: A scoping review. *npj Digital Med*. 2022 Nov; 5:1–14. DOI: [10.1038/s41746-022-00712-8](https://doi.org/10.1038/s41746-022-00712-8)
129. Zimerman A, **Montezano BB**, Vecchia GFD, Kapczinski F e Passos IC. The Dawn of Digital Psychiatry. *Digital Mental Health: A Practitioner's Guide*. Ed. por Passos IC, Rabelo-da-Ponte FD e Kapczinski F. Cham: Springer International Publishing, 2023 :1–11. DOI: [10.1007/978-3-031-10698-9\\_1](https://doi.org/10.1007/978-3-031-10698-9_1). Disponível em: [https://doi.org/10.1007/978-3-031-10698-9\\_1](https://doi.org/10.1007/978-3-031-10698-9_1)
130. Kapczinski F, **Montezano BB**, Nardi AE, Lledo PM e Katsimpardi L. Autophagy-based antidepressants? *Brazilian Journal of Psychiatry* 2023; 45:203–4. DOI: [10.47626/1516-4446-2023-3127](https://doi.org/10.47626/1516-4446-2023-3127)
131. Kapczinski F, **Montezano BB** e Azevedo Cardoso T de. Latent bipolar disorder. *Lancet* 2023 Jun; 401:2109. DOI: [10.1016/S0140-6736\(23\)00402-6](https://doi.org/10.1016/S0140-6736(23)00402-6)
132. **Montezano BB**, Cardoso TdA, Souza LDdM, Moreira FP, Mondin TC, Silva RA da e Jansen K. Associations between general sleep quality and measures of functioning and cognition in subjects recently diagnosed with bipolar disorder. *Psychiatry Research Communications* 2023 Jun; 3:100121. DOI: [10.1016/j.psycom.2023.100121](https://doi.org/10.1016/j.psycom.2023.100121)

133. Rakovski C, Ballester PL, **Montezano BB**, Dias de Mattos Souza L, Jansen K, Azevedo da Silva R, Mondin TC, Pedrotti Moreira F, De Boni RB, Frey BN, Kapczinski F e Azevedo Cardoso T de. Predictors of illicit substance abuse/dependence during young adulthood: A machine learning approach. *Journal of Psychiatric Research*. 2023 Jan; 157:168–73. DOI: [10.1016/j.jpsychires.2022.11.030](https://doi.org/10.1016/j.jpsychires.2022.11.030)
134. Bücken J, Amoretti S, Vieta E, Czepielewski L, Scheibe D, **Montezano BB**, Goularte J e Rosa A. Validity and Reliability of the Brazilian Version of the Cognitive Reserve Assessment Scale in Health (CRASH). *Brazilian Journal of Psychiatry* 2023; 0:1–17. DOI: [10.47626/1516-4446-2023-3198](https://doi.org/10.47626/1516-4446-2023-3198)
135. Carvalho Q de, Ziani PR, **Montezano BB**, Goularte JF e Rosa AR. Validation of the Brazilian version of the Processes of Change Questionnaire in Weight Management in adults with overweight and obesity in Brazil. medRxiv 2023. DOI: [10.1101/2023.03.20.23287450](https://doi.org/10.1101/2023.03.20.23287450). eprint: <https://www.medrxiv.org/content/early/2023/03/24/2023.03.20.23287450.full.pdf>. Disponível em: <https://www.medrxiv.org/content/early/2023/03/24/2023.03.20.23287450>

# Apêndices

# APÊNDICE A – Outros trabalhos

Além do trabalho que compõe a dissertação, outros artigos foram produzidos durante o período que compreendeu o mestrado.

## A.1 Artigos publicados

1. Kapczinski F, **Montezano BB**, Nardi AE, Lledo PM e Katsimpardi L. Autophagy-based antidepressants? *Brazilian Journal of Psychiatry* 2023; 45:203–4. DOI: [10.47626/1516-4446-2023-3127](https://doi.org/10.47626/1516-4446-2023-3127).
2. Kapczinski F, **Montezano BB** e Azevedo Cardoso T de. Latent bipolar disorder. *Lancet* 2023 Jun; 401:2109. DOI: [10.1016/S0140-6736\(23\)00402-6](https://doi.org/10.1016/S0140-6736(23)00402-6).
3. **Montezano BB**, Cardoso TdA, Souza LDdM, Moreira FP, Mondin TC, Silva RA da e Jansen K. Associations between general sleep quality and measures of functioning and cognition in subjects recently diagnosed with bipolar disorder. *Psychiatry Research Communications* 2023 Jun; 3:100121. DOI: [10.1016/j.psycom.2023.100121](https://doi.org/10.1016/j.psycom.2023.100121).
4. Fialho AR, **Montezano BB**, Ballester PL, Azevedo Cardoso T de, Mondin TC, Moreira FP, Souza LDdM, Silva RA da e Jansen K. Predicting 3-year persistent or recurrent major depressive episode using machine learning techniques. *Psychiatry Research Communications*. 2022 Sep; 2:100055. DOI: [10.1016/j.psycom.2022.100055](https://doi.org/10.1016/j.psycom.2022.100055).
5. Bartz de Ávila G, **Montezano BB**, Dias de Mattos Souza L, Azevedo Cardoso T de, Azevedo da Silva R, Mondin TC, Moreira FP e Jansen K. Personality disorders as predictors for the conversion from major depressive disorder to bipolar disorder: A prospective cohort study. *Psychiatry Research Communications*. 2022 Sep; 2:100068. DOI: [10.1016/j.psycom.2022.100068](https://doi.org/10.1016/j.psycom.2022.100068).
6. Rakovski C, Ballester PL, **Montezano BB**, Dias de Mattos Souza L, Jansen K, Azevedo da Silva R, Mondin TC, Pedrotti Moreira F, De Boni RB, Frey BN, Kapczinski F e Azevedo Cardoso T de. Predictors of illicit substance abuse/dependence during young adulthood: A machine learning approach. *Journal of Psychiatric Research*. 2023 Jan; 157:168–73. DOI: [10.1016/j.jpsychires.2022.11.030](https://doi.org/10.1016/j.jpsychires.2022.11.030).

## A.2 Artigos aceitos

1. Bücken J, Amoretti S, Vieta E, Czepielewski L, Scheibe D, **Montezano BB**, Goularte J e Rosa A. Validity and Reliability of the Brazilian Version of the Cognitive Reserve Assessment Scale in Health (CRASH). *Brazilian Journal of Psychiatry* 2023; 0:1–17. DOI: [10.47626/1516-4446-2023-3198](https://doi.org/10.47626/1516-4446-2023-3198).
2. de Aguiar KR, **Montezano BB**, Feiten JG, Watts D, Zimmerman A, Mondin TC, da Silva RA, Souza LDdM, Cardoso TA, Jansen K, Kapczinski F, Passos IC. Predicting functional impairment in patients with mood disorder: A 5-year follow-up. Aceito como *Research Article* na revista *Psychiatry Research*.

## A.3 Artigos submetidos

1. Gnielka V, **Montezano BB**, Baldez DP, Shintani AO, Rabelo-da-Ponte FD, Menezes AMB, Wehrmeister FC, Gonçalves H, Kunz M, Kauer-Sant'Anna M, Watts D, Kapczinski F, Passos IC. Perinatal factors as predictors of bipolar disorder onset in young adulthood: a 22-year birth cohort. Submetido como *Original Article* para a revista *Brazilian Journal of Psychiatry*.
2. Ribas LH, **Montezano BB**, Nieves M, Kampmann LB, Jansen K. The relationship between high parental stress scores and emotional and behavioral problems in school-age children: a systematic review and meta-analysis. Submetido como *Regular Article* para a revista *Pediatrics*.
3. Rodrigues Jorge AC, **Montezano BB**, Aguiar KA, Noronha LT, Baldez DP, Watts D, Menezes AMB, Wehrmeister FC, Gonçalves H, Kunz M, Kapczinski F, Passos IC. Cannabis lifetime use on bipolar onset in young adults: a two-decade birth cohort. Manuscrito em processo de finalização.
4. Ache ALS, **Montezano BB**, Spanemberg L, Fleck MP. Video-based treatment for mental health symptoms at the COVID-19 pandemic. Manuscrito em processo de finalização.
5. Anzolin AP, **Montezano BB**, Feiten JG, Ceresér KMM, Goularte JF, Pinto JV, Abreu PB, Passos IC, Sant'Anna MK. Naturalist study of ketamin in the treatment of depression: evaluation of suicide risk and serum measurements of SIRT3, SUPAR, PCRus, interleukin-6, complete blood count, leptin, lipid profile, and blood glucose. Manuscrito em processo de construção.
6. Moreira V, **Montezano BB**, Mombach J, Bazzo G, Dimer D, Kuhn DM, Barros BM, Rösler A, Mendoza MR. Detection, assessment and mitigation of bias and unfairness in machine

learning models for healthcare: a systematic review. Manuscrito em processo de construção.

## A.4 Capítulos de livro

1. Zimerman A, **Montezano BB**, Vecchia GFD, Kapczinski F e Passos IC. The Dawn of Digital Psychiatry. *Digital Mental Health: A Practitioner's Guide*. Ed. por Passos IC, Rabelo-da-Ponte FD e Kapczinski F. Cham: Springer International Publishing, 2023 :1–11. DOI: [10.1007/978-3-031-10698-9\\_1](https://doi.org/10.1007/978-3-031-10698-9_1). Disponível em: [https://doi.org/10.1007/978-3-031-10698-9\\_1](https://doi.org/10.1007/978-3-031-10698-9_1).
2. Shintani AO, **Montezano BB**, Passos IC. Suicídio. *Psicofármacos: consulta rápida*. Ed. por Cordioli AV, Gallois CB e Passos IC. Artmed, 2023. Capítulo publicado e disponível em formato impresso.
3. **Montezano BB**, Lorenzoni IC, Vecchia, GD, Passos IC. O uso da inteligência artificial para identificar doenças psiquiátricas. *Psiquiatria digital: diagnóstico e manejo*. Ed. por Passos IC e Gallois CB. Artmed, 2023. Capítulo aprovado.
4. Roza TH, **Montezano BB**, Antonelli-Salgado T, Passos IC. Predição de risco de suicídio utilizando técnicas de inteligência artificial. *Psiquiatria digital: diagnóstico e manejo*. Ed. por Passos IC e Gallois CB. Artmed, 2023. Capítulo finalizado.
5. Cardoso TA, **Montezano BB**, Meller M. Substance use disorder in the course of bipolar disorder. *Bipolar Disorder: An Evidence-Based Clinical Guide*. Ed. por Passos IC, Berk M e Kapczinski F. Springer Nature, 2023. Capítulo finalizado.

## A.5 Preprints

1. **Montezano BB**, A. Cardoso T de, Mattos Souza LD de, Moreira FP, Mondin TC, Silva RA da e Jansen K. Associations between general sleep quality and measures of functioning and cognition in subjects recently diagnosed with bipolar disorder. medRxiv. 2022 Jul :2022.07.24.22277972. DOI: [10.1101/2022.07.24.22277972](https://doi.org/10.1101/2022.07.24.22277972). eprint: [2022.07.24.22277972](https://www.medrxiv.org/content/early/2022/07/24/2022.07.24.22277972).
2. Carvalho Q de, Ziani PR, **Montezano BB**, Goularte JF e Rosa AR. Validation of the Brazilian version of the Processes of Change Questionnaire in Weight Management in adults with overweight and obesity in Brazil. medRxiv 2023. DOI: [10.1101/2023.03.20.23287450](https://doi.org/10.1101/2023.03.20.23287450). eprint: <https://www.medrxiv.org/content/early/2023/03/24/2023.03.20.23287450.full.pdf>. Disponível em: <https://www.medrxiv.org/content/early/2023/03/24/2023.03.20.23287450>.

# APÊNDICE B – Implementação em R

## B.1 Pré-processamento dos dados

```

1 # Ler dados brutos a partir do arquivo em sav
2 raw <- haven::read_sav(file = "data/coorte-t1-t2-24-08-17.sav",
3   encoding = "utf8",
4   user_na = FALSE)
5
6 # Criar objeto com dados processados
7 processed_data <- raw |>
8   # Limpar nome de variáveis (snake case, remover caracteres especiais, etc)
9   janitor::clean_names() |>
10  # Calcular escore total da HCL-32
11  dplyr::rowwise() |>
12  dplyr::mutate(hclscore_t1 = sum(dplyr::c_across(
13    dplyr::matches("^hcl3"),
14    na.rm = TRUE)) |>
15  dplyr::ungroup() |>
16  # Transformar rotulos do SPSS em fatores do R
17  haven::as_factor() |>
18  # Remover sujeitos com episodio atual ou passado de (hipo)mania no baseline
19  dplyr::filter(perdas == "completou_avaliacao" &
20    hipo_t1 == "nao" & mania_t1 == "nao") |>
21  # Colocar nivel positivo como referencia do fator
22  dplyr::mutate(outcome = relevel(bipolar_conferido, ref = "Sim")) |>
23  # Selecionar desfecho do follow-up, ID unico, idade e resto das features
24  # do baseline
25  dplyr::select(outcome,
26    rec,
27    idade = idade_t1_2,
28    dplyr::matches("t1$"),
29    -dnasc_t1) |>
30  # Remover sufixo das variáveis para melhor manejo
31  dplyr::rename_with(~(x) stringr::str_remove(x, "_t1")) |>
32  # Selecionar variáveis relevantes baseado em expertise teorica
33  dplyr::select(
34    rec, outcome, sexo, idade, cpele_dic, abep3, trabatu, tabaco2, alcool2,
35    maconha2, cocaina2, crack2, anfet2, inalante2, sedativos2, alucinog2,
36    opioides2, outrasub2, ilicidas2, distat, rsat, agress, afogof, apoio,
37    apisepa, prfalec, religiaodic, hospner, probner, medic, smae, spai,
38    sirmao, mmae, mpai, savo, tent1, tent2, tent3, tent4, tent5, tent6,
39    relacao, abep5, cinto, capecet, acidente, dirigibb, estano, trabdin,
40    doenca, tocat, teptat, pancfo, pansfo, cons, trat, interr, bsi1:bsi21,

```



```
41   aldtenta, algmata, evesta:evestz, hcl31:hcl332, bdi1:bdi21,
42   qlusoua:qlusouj, injecao, idadrog, idadeere, ultimare, camisin, sexobb,
43   parceiro, pesrel, vezgravi, forcsex, anorat, buliat, apurg, tpsanc,
44   acesso, edmat, edmmel, edmps, hipo, e5panvi, e6panat, e7sint, agospan,
45   fobsoa
46 )|>
47 # Realizar "feature engineering" e correcoes pequenas na codificacao
48 dplyr::mutate(
49   # Sexo masculino como referencia
50   sexo = relevel(sexo, ref = c("masculino")),
51   # Dicotomizar variavel de cor da pele
52   cpele_dic = as.factor(dplyr::if_else(
53     cpele_dic == "branco",
54     "branco",
55     "nao_branco"
56   )),
57   # Dicotomizar variavel sobre acidente na vida
58   acidente = as.factor(
59     dplyr::if_else(acidente == "nao_sofreu_nenhum_acidente",
60                   "nao", "sim")
61   ),
62   # Dicotomizar variavel de ingestao de alcool ao dirigir
63   dirigibb = as.factor(dplyr::if_else(
64     dirigibb %in% c("nao_andou_em_nenhum_veiculo", "nenhum", "1_vez"),
65     "nao",
66     "sim"
67   )),
68   # Dicotomizar variavel sobre uso de cinto de seguranca
69   cinto = as.factor(dplyr::if_else(
70     cinto %in% c("nao_andou_de_moto", "quase_sempre", "sempre"),
71     "sim",
72     "nao"
73   )),
74   # Dicotomizar variavel sobre uso de capacete ao andar de moto
75   capecet = as.factor(dplyr::if_else(
76     capecet %in% c("nao_andou_de_moto", "quase_sempre", "sempre"),
77     "sim",
78     "nao"
79   )),
80   # Dicotomizar variavel sobre envolvimento em briga fisica
81   agress = as.factor(dplyr::if_else(
82     agress %in% c("sim_mas_nao_me_machuquei", "sim_e_sai_machucado"),
83     "sim",
84     "nao"
85   )),
86   # Recodificar variavel de religiao
87   religiaodic = as.factor(dplyr::if_else(religiaodic == 1, "sim", "nao")),
```

```
88 # Recodificar variavel sobre sexo forçado
89 forcsex = as.factor(dplyr::if_else(
90   forcsex == "sim",
91   "sim",
92   dplyr::if_else(forcsex == "nao", "nao", NA_character_)
93 )),
94 # Criar variavel dicotomica sobre gravidez
95 vezgravi = as.factor(dplyr::if_else(
96   vezgravi == "nenhuma",
97   "nao",
98   dplyr::if_else(vezgravi %in% c("1_vez", "2_vezes_ou_mais"),
99     "sim", NA_character_)
100 )),
101 # Criar variavel sobre frequencia de relacao sexual
102 ultimare = as.factor(dplyr::if_else(
103   ultimare == "ha_menos_de_uma_semana",
104   "sim",
105   "nao"
106 )),
107 # Recodificar variavel sobre ja ter usado substancia por injecao
108 injecao = as.factor(dplyr::if_else(injecao == "nao_nunca",
109   "nao",
110   "sim")),
111 # Recodificar variavel se ja interrompeu tratamento antes do termino
112 interr = as.factor(dplyr::if_else(interr == "sim",
113   "sim",
114   "nao",
115   missing = "nao"))
116 ) |>
117 # Padronizando valores dos itens da HCL-32
118 dplyr::mutate(dplyr::across(dplyr::starts_with("hcl3"),
119   \ (x) as.factor(
120     dplyr::case_when(x %in% c("0", "nao") ~ "nao",
121       x == "sim" ~ "sim",
122       TRUE ~ "nao")
123   ))) |>
124 # Recodificando variaveis dos eventos vitais
125 dplyr::mutate(dplyr::across(
126   c(dplyr::starts_with("evest")),
127   \ (x) as.factor(dplyr::if_else(x == "sim",
128     "sim",
129     "nao",
130     missing = "nao"))
131 )) |>
132 # Recodificando variaveis de possivel abuso/dependencia de substancias
133 dplyr::mutate(dplyr::across(
134   c(tabaco2, alcool2, maconha2, cocaina2, crack2, anfet2, inalante2,
```

```
135     sedativos2, alucinog2, opioides2, outrasub2, ilicitas2),
136     \ (x) as.factor(dplyr::if_else(x == "sem_uso/_uso_ocasional",
137     "nao", "sim"))
138   ))
139
140 # Exportar os dados processados para um arquivo no formato rds
141 readr::write_rds(x = processed_data,
142     file = "data/processed_data_bd.rds",
143     compress = "none")
```

## B.2 XGBoost e validação cruzada

```
1 # Ler dados processados previamente
2 da <- readr::read_rds(file = "data/processed_data_bd.rds")
3
4 # Atribuir um valor para o seed
5 set.seed(1)
6
7 # Realizar divisao dos dados
8 # 70% para a base de treino e 30% para base de testes
9 initial_split <- rsample::initial_split(da, prop = 0.7, strata = "outcome")
10 train <- rsample::training(initial_split)
11 test <- rsample::testing(initial_split)
12
13 # Atribuir seed mais uma vez
14 set.seed(1)
15
16 # Criar receita para o pipeline com os seguintes passos:
17 # 1) Especificar a variavel ID
18 # 2) Remover features com mais de 10% de valores ausentes
19 # 3) Imputar features categoricas com a moda
20 # 4) Imputar features numericas com a mediana
21 # 5) Remover features que sao altamente esparsas ou desbalanceadas
22 # 6) Utilzar o algoritmo ROSE para oversampling das observacoes
23 # 7) Criar variaveis dummy para as features nominais/categoricas
24 xgb_rec <- train |>
25   recipes::recipe(outcome ~ .) |>
26   recipes::update_role(rec, new_role = "id") |>
27   recipes::step_filter_missing(recipes::all_predictors(), threshold = 0.1) |>
28   recipes::step_nzv(recipes::all_predictors()) |>
29   recipes::step_impute_mode(recipes::all_nominal_predictors()) |>
30   recipes::step_impute_median(recipes::all_numeric_predictors()) |>
31   themis::step_rose(recipes::all_outcomes(),
32     minority_prop = 0.5,
33     seed = 1) |>
34   recipes::step_dummy(recipes::all_nominal_predictors(),
```

```
35         one_hot = FALSE) |>
36   recipeselectors::step_select_boruta(
37     recipes::all_predictors(),
38     outcome = "outcome",
39   )
40
41 # Utilizar 5-fold CV repetido por 5 vezes usando amostragem estratificada
42 # baseada no desfecho de TB
43 set.seed(1)
44 folds <- rsample::vfold_cv(data = train,
45                             v = 5,
46                             repeats = 5,
47                             strata = "outcome")
48
49 # Especificar o algoritmo de XGBoost e os hiperparametros que serao tunados
50 # no processo de validacao cruzada
51 xgb_spec <- parsnip::boost_tree(
52   mtry = tune::tune(),
53   trees = tune::tune(),
54   min_n = tune::tune(),
55   tree_depth = tune::tune(),
56   loss_reduction = tune::tune(),
57   learn_rate = tune::tune(),
58   sample_size = tune::tune()
59 ) |>
60   parsnip::set_mode("classification") |>
61   parsnip::set_engine("xgboost", tree_method = "gpu_hist")
62
63
64 # Criar o workflow a partir da receita criada e da especificacao do modelo
65 xgb_wf <- workflows::workflow() |>
66   workflows::add_model(xgb_spec) |>
67   workflows::add_recipe(xgb_rec)
68
69
70 # Atribuir novamente a seed
71 set.seed(1)
72
73 # As reamostragens serao criadas a partir dos folds previamente criados.
74 # Detalhes tecnicos sobre grid search por meio de racing com modelos de
75 # ANOVA sao descritos em Kuhn (2014)
76 xgb_rs <- xgb_wf |>
77   finetune::tune_race_anova(
78     resamples = folds,
79     # Grid com 15 combinacoes de hiperparametros
80     grid = 15,
81     control = finetune::control_race(
```

```
82     save_pred = TRUE, # Salvar predicoes
83     verbose = TRUE, # Output mais informativo
84     verbose_elim = TRUE,
85     allow_par = TRUE, # Permitir paralelizacao
86     randomize = TRUE, # Aleatorizar ordem de execucao dos folds
87     alpha = 0.05,
88     burn_in = 3 # Quantas reamostragens para iniciar a filtragem
89   ),
90   metrics = yardstick::metric_set(yardstick::roc_auc,
91                                   yardstick::bal_accuracy,
92                                   yardstick::ppv)
93 )
94
95 # Plotar a corrida das diferentes combinacoes
96 xgb_rs |>
97   finetune::plot_race()
98
99 # Checar melhor combinacao de hiperparametros a partir da AUC
100 xgb_rs |>
101   tune::show_best(metric = "roc_auc",
102                   n = 10) |>
103   tibble::view()
104
105 # Selecionar melhor combinacao de hiperparametros pela AUC
106 xgb_best_param <- xgb_rs |>
107   tune::select_best(metric = "roc_auc")
108
109 # Inserir hiperparametros escolhidos no workflow
110 xgb_final_wf <- xgb_wf |>
111   tune::finalize_workflow(xgb_best_param)
112
113 # Ajustar modelo final na base de treino e estimar medidas de desempenho
114 # na base de testes
115 xgb_last_fit <- xgb_final_wf |>
116   tune::last_fit(initial_split,
117                 metrics = yardstick::metric_set(
118                   yardstick::roc_auc, # AUC
119                   yardstick::bal_accuracy, # Acuracia balanceada
120                   yardstick::ppv, # Valor preditivo positivo
121                   yardstick::npv, # Valor preditivo negativo
122                   yardstick::sensitivity, # Sensibilidade
123                   yardstick::specificity, # Especificidade
124                   yardstick::f_meas, # F1 score
125                   yardstick::pr_auc # AUC da precisao-recall
126                 ))
127
128 # Checar as medidas calculadas
```

```
129 xgb_last_fit |>
130   tune::collect_metrics()
131
132 # Extrair objeto do modelo final
133 xgb_fit_engine <- xgb_last_fit |>
134   hardhat::extract_fit_engine()
```

### B.3 Teste de permutação

```
1 # Carregar sessão do último ajuste do modelo
2 load("../cache/final_assessment.RData")
3
4 # Carregar modelo final
5 modelo <- xgboost::xgb.load("../cache/final_model")
6
7 # Calcular valor final da AUC na base de teste
8 original_auc <- xgb_test |>
9   dplyr::select(outcome) |>
10  dplyr::bind_cols(xgb_preds |>
11    tibble::as_tibble_col(column_name = ".pred_Sim")) |>
12  yardstick::roc_auc(outcome, .pred_Sim) |>
13  dplyr::pull(.estimate)
14
15 # Extrair matriz de dados de teste processada
16 xgb_test_processed <- xgb_rec |>
17   recipes::prep() |>
18   recipes::bake(new_data = xgb_test) |>
19   dplyr::select(-outcome, -rec) |>
20   as.matrix()
21
22 # Estimar desempenho no teste de permutação
23 # com dez mil repetições
24 set.seed(1)
25 permuted_performance <- replicate(10000, {
26
27   test_permuted <- xgb_rec |>
28     recipes::prep() |>
29     recipes::bake(new_data = xgb_test) |>
30     dplyr::mutate(outcome = sample(outcome))
31
32   test_matrix_permuted <- test_permuted |>
33     dplyr::select(-outcome, -rec) |>
34     as.matrix()
35
36   permuted_preds <- predict(modelo, test_matrix_permuted)
37
```

```
38 binded_permut <- test_permuted |>
39   dplyr::select(outcome) |>
40   dplyr::bind_cols(permuted_preds |>
41     tibble::as_tibble_col(column_name = ".pred_Sim"))
42
43 yardstick::roc_auc(truth = outcome, estimate = .pred_Sim,
44   data = binded_permut) |>
45   dplyr::pull(.estimate)
46 })
47
48 # Checar estrutura do objeto
49 str(permuted_performance)
50
51 # Calcular o p-valor do teste de permutacao
52 p_value <- mean(permuted_performance >= original_auc)
53
54 # Plotar os resultados
55 ggplot2::ggplot(data.frame(permuted_performance),
56   ggplot2::aes(x = permuted_performance)) +
57   ggplot2::geom_histogram(fill = "blue", alpha = 0.5) +
58   ggplot2::geom_vline(xintercept = original_auc, color = "red",
59     linetype = "dashed",
60     linewidth = 2) +
61   ggplot2::xlab("Permuted Area Under the ROC Curve") +
62   ggplot2::ylab("Count") +
63   ggplot2::scale_x_continuous(breaks = seq(0, 1, 0.1)) +
64   ggplot2::theme_minimal(24)
```

# APÊNDICE C – Comprovante de submissão do artigo



Bruno Braga Montezano <brunobmontezano@gmail.com>

---

**Submission Confirmation: Bipolar Disorders - Manuscript BDI-2023-0165 [email ref: SE-6-a]**

1 message

---

Harsritha S <onbehalf@manuscriptcentral.com>

Thu, Jul 27, 2023 at 2:11 AM

Reply-To: bdiedoffice@wiley.com

To: brunobmontezano@gmail.com

Cc: brunobmontezano@gmail.com, gnielkav@gmail.com, ashintani@gmail.com, kyarardeaguiar@gmail.com, roza.h.thiago@gmail.com, taianeacardoso@hotmail.com, luciano.souza@ucpel.edu.br, fernanda.moreira@ucpel.edu.br, ricardo.silva@ucpel.edu.br, thaisemondin@gmail.com, flavio.kapczinski@gmail.com, karen.jansen@ucpel.edu.br, ivescp1@gmail.com

27-Jul-2023

Re: BDI-2023-0165 "Bipolar disorder prediction and functional outcomes in young adults: a 5-year follow-up study"

Dear Dr. Montezano,

Thank you for submitting your manuscript to Bipolar Disorders. The manuscript is now being given editorial consideration.

Co-authors: Please contact Bipolar Disorders' Editorial Office as soon as possible if you disagree with being listed as a co-author.

Please note, our journal is currently transitioning to Wiley's Research Exchange submission portal.

If you submitted this manuscript through our Research Exchange site, you can view the status of your manuscript by logging into the submission site at [wiley.atyponrex.com/journal/BDI](http://wiley.atyponrex.com/journal/BDI).

If you submitted this manuscript through ScholarOne, you can view the status of your manuscript by checking your Author Center after logging in to <https://mc.manuscriptcentral.com/bdi>.

This journal offers a number of license options, information about this is available here: <https://authorservices.wiley.com/author-resources/Journal-Authors/licensing/index.html>. All co-authors are required to confirm that they have the necessary rights to grant in the submission, including in light of each co-author's funder policies. For example, if you or one of your co-authors received funding from a member of Coalition S, you may need to check which licenses you are able to sign.

Thank you for submitting your manuscript to Bipolar Disorders.

Kind regards

Ms. Harsritha S

Editorial Office, Bipolar Disorders

[bdiedoffice@wiley.com](mailto:bdiedoffice@wiley.com)

Professor Gin S. Malhi, MD

Editor-in-Chief, Bipolar Disorders

<http://wileyonlinelibrary.com/journal/bdi>



# APÊNDICE D – Artigo 1

**Title:** Bipolar disorder prediction and functional outcomes in young adults: a 5-year follow-up study

**Running title:** Bipolar prediction and functioning

**Authors:** Bruno Braga Montezano<sup>1,2,\*</sup>, Vanessa Gnielka<sup>1</sup>, Augusto Ossamu Shintani<sup>1,2</sup>, Kyara Rodrigues de Aguiar<sup>1,2</sup>, Thiago Henrique Roza<sup>2</sup>, Taiane de Azevedo Cardoso<sup>3</sup>, Luciano Dias de Mattos Souza<sup>4</sup>, Fernanda Pedrotti Moreira<sup>4</sup>, Ricardo Azevedo da Silva<sup>4</sup>, Thaíse Campos Mondin<sup>4</sup>, Flávio Kapczinski<sup>1,2,3,5</sup>, Karen Jansen<sup>4</sup>, Ives Cavalcante Passos<sup>1,2</sup>

**Affiliations:** <sup>1</sup>Laboratory of Molecular Psychiatry, Centro de Pesquisa Experimental (CPE) and Centro de Pesquisa Clínica (CPC), Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, RS, Brazil; <sup>2</sup>Graduate Program in Psychiatry and Behavioral Sciences, Department of Psychiatry, School of Medicine, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil; <sup>3</sup>Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada; <sup>4</sup>Graduate Program in Health and Behavior, Universidade Católica de Pelotas, Pelotas, RS, Brazil; <sup>5</sup>Instituto Nacional de Ciência e Tecnologia Translacional em Medicina (INCT-TM), Porto Alegre, RS, Brazil

## **Acknowledgements**

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This work received financial supports from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [grant number 481998/2012-8-APQ]; FAPERGS (Fundação de Amparo à Pesquisa do

\*Corresponding author

Estado do Rio Grande do Sul) [grant number 19/2551-0000689-6]; and Instituto Nacional de Ciência e Tecnologia Translacional em Medicina (INCT-TM) [grant numbers 2014/50891-1/FAPESP; 465458/2014-9/CNPQ], to conduct the study design and data collection. Dr. Passos is a CNPq research fellow. This work received financial supports from Fundo de Incentivo à Pesquisa e Eventos (FIPE) do HCPA, Brazil. Dr. Kapczinski reports personal fees from Janssen — Johnson & Johnson, as well as grants from the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (# 20/2551-0000250-0), the Canada Foundation for Innovation, the Social Sciences and Humanities Research Council of Canada, the Canadian Institutes of Health Research, and the Natural Sciences and Engineering Research Council of Canada, outside the submitted work.

#### **Abstract**

**Objectives:** To develop a predictive model for bipolar disorder (BD) onset in young adults based on 5-year follow-up data and to assess the long-term functional implications of depressive symptom severity.

**Methods:** A total of 1,091 young adults without a current or past episode of mania or hypomania at baseline were assessed through follow-up interviews. The XGBoost algorithm with preprocessing steps was trained on cross-validation. Additionally, linear regression analyses were conducted to investigate the associations between depression severity and functional impairment, as well as cognitive performance after five years.

**Results:** The final model demonstrated favorable predictive metrics for BD onset (ROC AUC: 78.6%; sensitivity: 37.5%; specificity: 92%), with depressive symptom-related variables being the most influential predictors. Moreover, severe baseline depressive symptoms were associated with lower cognitive performance ( $\beta=-0.11$ , 95% CI [-0.14, -0.07],  $p<0.001$ ,  $r=-0.18$ ) and increased functional impairment ( $\beta=0.23$ , 95% CI [0.17, 0.30],  $p<0.001$ ,  $r=0.24$ ) after five years. Subgroup analyses revealed gender and suicide risk-related variations in these associations.

**Conclusions:** The final classifier demonstrated acceptable classification performance for BD onset in young adults, also reinforcing a latent understanding of BD trajectory. Additionally, the severity of depressive symptoms was found to have long-term implications for cognitive performance and functional outcomes. These findings highlight the importance of early detection and intervention for individuals at risk of BD. Further research should focus on replication studies and exploring multidimensional markers for enhanced predictive models in BD.

**Keywords:** Mood Disorder. Supervised Machine Learning. Cohort Studies. Risk Calculator. Precision Psychiatry.

## 1. Introduction

Bipolar disorder (BD) is a chronic psychiatric condition associated with high morbidity and mortality<sup>1</sup>. The global lifetime prevalence of BD is approximately 2.4%, being 0.6% for bipolar I disorder, 0.4% for bipolar II disorder and 1.4% for individuals with subthreshold presentations<sup>2</sup>. Previous investigations describe that these patients present a significant reduction of life expectancy of about 10 years relative to the general population<sup>3</sup>. Cardiovascular disease is the main factor associated with premature mortality in BD; nonetheless, deaths by suicide are more commonly reported in BD than in other mental health conditions, with these patients presenting a twenty to thirty times higher chance of dying by suicide<sup>1,3,4</sup>. In addition, patients with BD present significant functional and psychosocial impairment, also representing an important economic cost<sup>1</sup>. For instance, evidence from the United States described that the total costs associated with bipolar I disorder exceeded \$200 billion in the year of 2015<sup>5</sup>.

Even though the majority of the patients with BD present clinical symptoms before the age of 25, there is a significant delay of 6-10 years between the onset of the symptoms and the correct diagnosis<sup>6,7</sup>. Furthermore, delayed diagnosis is associated with longer duration of untreated illness, which is ultimately linked to a poorer prognosis in terms of hospitalizations, functioning and recurrence of episodes<sup>8</sup>. High rates of psychiatric comorbidity, difficulty in the differential diagnosis, usual onset with depressive symptoms, and reduced help-seeking behavior are some of the reasons for the delay in the proper recognition of BD<sup>1,7,9</sup>.

Nevertheless, the diagnosis of BD is eminently clinical, with limited evidence to support the use of neuroimaging or laboratory biomarkers during clinical investigation<sup>1</sup>.

Taking into account this context, the rise of the concept of precision psychiatry, with the use of big data and machine learning tools represents a promise, which may ultimately bring a revolution in terms of diagnosis, treatment selection and prognosis in the field of mental health<sup>10,11</sup>. To this date several studies have explored the use of these techniques in BD, based on distinct data sources (including neuroimaging, clinical and sociodemographic data, peripheral biomarkers, neuropsychological tests, genetics, among others), with the majority of these models being focused on classification tasks that help in the differential diagnosis between BD and other psychiatric conditions such as schizophrenia, major depression and healthy individuals<sup>12,13</sup>. Nevertheless, most of these studies present modest classification performances, are based on small and clinical samples, originary from cross-sectional procedures of data collection, or present short periods of follow-up<sup>12</sup>. All these limitations may compromise the generalizability and the translation of the results of such investigations to clinical and public health settings<sup>13</sup>.

Thus, considering these gaps, the present study aims to create a binary classification model capable of predicting incident cases of BD in a 5-year interval through sociodemographic and clinical features in a sample of young adults, from a large and population-based cohort study. As secondary aims, the study explores

associations between depression severity and functional and cognitive outcomes,  
as well as the relationship between suicidality and age at depression onset.

## **2. Methods**

### **2.1. Participants**

This was a prospective cohort study that collected sociodemographic and clinical information from a population-based sample of young adults aged between 18 and 24 years, living in the urban area of the city of Pelotas, located in southern Brazil. The first phase took place between 2007 and 2009, and the sample was selected through cluster sampling, considering eighty-nine randomly selected census-based sectors from 448 total sectors<sup>14</sup>.

The following inclusion criteria were considered at baseline: (1) age between 18 and 24 years old; (2) live in the urban area. Severe cognitive disability (assessed through clinical judgment) that could cause difficulties in understanding study instruments was considered the only exclusion criteria. All eligible subjects ( $n = 1,762$ ) were invited to participate, of which 1,560 accepted and consented to participate. Trained interviewers conducted a face-to-face interview at the participants' homes, so that data confidentiality was ensured. Data were collected through printed paper questionnaires with research instruments and diagnostic criteria for mental disorders.

The follow-up occurred from 2012 to 2014, that is, an average interval of five years after the first assessment. The participants from baseline ( $n = 1,560$ ) were invited for a second data collection. All interviewers met weekly to discuss the assessments, focusing on those who were uncertain about the BD diagnosis.

In these situations, a psychiatrist was recruited to carry out the reassessment. 1,244 individuals were located and consented to be reevaluated (79.7% of retention), and 14 (0.9%) were lost due to death. Since the present study aims to predict BD incidence, subjects who met diagnostic criteria for a lifetime manic or hypomanic episode were excluded ( $n = 153$ ), resulting in a total sample of 1,091 subjects (Figure S1). Unlike the baseline, data were collected through tablets using Open Data Kit (ODK), an open-source mobile data collection platform<sup>15</sup>. The forms were filled out offline, and the data were later backed up to computers through secure data transfer protocols. This study was approved by the Research Ethics Committee of Universidade Católica de Pelotas under protocol number 2008/118. The subjects who presented any psychiatric diagnosis in the clinical interview were referred for specialized treatment in the local health system. All participants signed a printed informed consent form and could withdraw from the study at any time.

## **2.2. Main outcome**

The BD diagnosis was built with modules A and D from Mini International Neuropsychiatric Interview 5.0 (MINI), in order to assess current or past depressive episodes and current or past manic or hypomanic episodes, respectively. The BD diagnoses were reassessed in those cases where the diagnosis was questionable. For MINI description and psychometric properties, see Supporting Information.



### **2.3. Predictors**

One hundred and ninety features were included from the original dataset before preprocessing steps. These variables include demographic, social, clinical, and environmental characteristics. The following features were included in the modeling pipeline:

- a) Sociodemographic and environmental variables: Sex, skin color, age, socioeconomic status (3 levels and 5 levels), current occupation, currently studying, worked for money, has a partner, has a religion, access to psychotherapy, knows someone who attempted suicide or committed suicide, involvement in physical fights, family gun ownership, social support, has divorced parents, has any deceased parents, has someone close by already deceased, individual and family stress problems, lives with parents, family suicide attempts, seat belt wearing, helmet use when riding a motorcycle, suffered an accident that led to an emergency room, drove or took a ride with a drunk driver.
- b) Substance use variables: Indicative of substance abuse or dependence (tobacco, alcohol, cannabis, cocaine, crack, amphetamines, inhalants, sedatives, hallucinogens, opioids, illicit drugs, any other substances) assessed by Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) and lifetime use features (same substances cited above), age that first used drugs, injected drugs use, use of medication for stress

problems in the last 30 days.

- c) Clinical variables: mental disorder diagnoses (anxiety, mood and personality disorders), eating disorders, current suicide risk, serious organic disease, lifetime psychiatrist or psychologist visit, lifetime psychotherapeutic treatment, interrupted treatment.
- d) Sex-related variables: age of first sexual intercourse, sexual intercourse in the last week (sex frequency), condom use, alcohol use before sexual intercourse, number of sexual partners, number of pregnancies, lifetime sexual abuse, lifetime sexual intercourse (binary).
- e) Psychometric instrument items: Beck Depression Inventory (BDI) [21 items], Hypomania Checklist (HCL-32) [32 items], Social Readjustment Rating Scale (SRRS) [26 items], Beck Scale for Suicide Ideation (BSS) [21 items].

#### **2.4. Machine learning analysis**

Aiming to predict new cases of BD in young adults, using features previously described, we created an ML pipeline to generate a predictive model using supervised learning. We used a vastly used machine learning algorithm for tabular data called tree gradient boosting. Tree gradient boosting is part of what is called ensemble algorithms — joining many models to make predictions together — in statistical learning methods (see Supporting Information).

All analyses were done using R and the XGBoost library (see Supporting Information for details).

### **2.5. Preprocessing**

Before any preprocessing routine was performed, the data was divided into two subsets. A training set, consisting of 70% of the sample, and a test set with the remaining samples (30% of the total samples). The entire test set was isolated until the end of all tuning and validation procedures to build the model, to then be used to simulate the model performance on new data.

### **2.6. Feature selection**

The Boruta system was implemented for feature selection in the present pipeline. It consists of a random forest based algorithm that iteratively removes features that are statistically less important than random synthetic features (artificial noise). For each iteration, removed variables are prevented from being considered for the next iteration<sup>16</sup>. Boruta is considered a wrapper method as it takes into account a subset of variables with different combinations in each iteration.

### **2.7. Class imbalance**

For this paper, an algorithm named ROSE (Random Over-Sampling Examples) was used. ROSE is a smoothed-bootstrap-based technique that creates new artificial observations in data in order to minimize or eliminate class imbalance<sup>17</sup>. For more details on software used, see Supporting Information.

### **2.8. Cross-validation**

The CV process was used to tune *XGBoost* hyperparameters described earlier. We used the *k*-fold CV technique with 5 folds repeated five times. In order to optimize the hyperparameter combinations, we used a racing method proposed by Kuhn<sup>18</sup>.

In [Figure 1](#), the CV procedure can be visualized inside the orange area. For each fold, the Boruta and ROSE algorithms were applied just in training folds, leaving the testing fold untouched in order to properly estimate model error. A maximum of 25 runs (5-fold CV repeated up to five times) was considered for hyperparameter tuning. At the end, the remaining models of the ANOVA racing process were evaluated, and the model with the highest validation AUC was chosen to be tested in a testing set from the initial data split.

### **2.9. Model interpretability**

The performance metrics used for model evaluation are available in the Supporting Information. In order to make model predictions more interpretable, we used SHAPley values. SHAPley values show how to fairly distribute the total output among all features. Beyond that, SHAP (SHapley Additive exPlanations) allow for explanation on individual predictions<sup>19</sup>. In addition to the use of feature importance visualization, partial dependence plots (PDP) were also employed. They are able to show the marginal effect a feature has on the predicted outcome of a model. For further details, see Supporting Information.

### **2.10. Resampling methods**

In order to assess uncertainty in the model, a permutation test and random re samples of training and testing splits were used. To assess the statistical significance of the model performance, a permutation test was conducted (see Supporting Information). To assess model reliability, 1,000 random resamples of training and testing splits were used. For each iteration, the model was retrained on the resampled training set with the same hyperparameters from the original model and tested on the resampled testing set. The mean and 95% confidence intervals were calculated in order to evaluate whether the resampled estimates were similar to the main model performance.

### **2.11. Secondary outcomes**

In order to explore secondary cognitive and functional outcomes related to depression severity (BDI score), we used Functioning Assessment Short Test (FAST) and Montreal Cognitive Assessment (MoCA). Since the results on feature selection and feature importance for the main classification model highlighted three depression-related features (i.e. past failure, sadness and depressive episode at baseline), we investigated associations of depressive symptoms severity (BDI total score) with functional impairment (FAST global score) and cognitive performance (MoCA score). For that, we used Pearson's product-moment correlation coefficient, simple linear regression and fitted regression lines to visualize potential association trends. BDI had 2 missing

values, and MoCA had 23 subjects with missing values. No values were imputed in the linear models. We used a subsample of subjects without current mood episodes at baseline ( $n = 984$ ). Additional analysis on suicidality (suicide risk at baseline assessed by MINI C module) and age at onset of first depressive episode in life (self-report at baseline) was done based on two-sample Student's  $t$ -test. The subsample of subjects with depressive episode at baseline consisted of 84 subjects, of which the ones with missing values at depression age at onset variable ( $n = 54$ ) were removed, resulting in a sample of 31 subjects.

#### **2.12. Predictive model without mood-related features**

In the same direction of further investigating BD prediction regarding other features, we built another predictive model using the same procedure of the first ML model with the removal of all mood-related features from the pipeline to avoid capturing direct associations of diagnostic criteria and the outcome. The following features were excluded: depressive episode at baseline, melancholic depressive episode at baseline, past depressive episode (measured at baseline), all BDI items (depressive symptoms) and all HCL-32 items (hypomania-related symptoms). The filtering process removed 56 features in total.

### 3. Results

The present study aimed to create a model to predict BD onset on young adults based on 5-year follow-up data. We assessed 1,091 subjects at follow-up interviews who had no current or past episode of mania or hypomania at the first assessment. Of these, 4.49% ( $n = 49$ ) young adults received a diagnosis of BD five years later. Descriptive tables of demographic features at baseline are presented in [Table 1](#). Additional summaries on remaining features included in the modeling pipeline with absolute and relative frequencies, and missing value counts are available in [Tables S1](#) (clinical and environmental features with at least  $\geq 1$  missing), [S2](#) (BDI items) and [S3](#) (BSS items). [Table S4](#) shows the selected hyperparameter set from the CV.

XGBoost showed an acceptable performance predicting BD five years before the diagnosis with a test set ROC AUC of 0.786 [95% CI: 0.686, 0.887] and precision-recall curve of 0.208 ([Figure S2](#)). The other performance metrics using a cut-off of 0.5 for class decision boundary<sup>1</sup> can be seen in [Table S5](#), in addition to the performance of the model without mood-related features.

The six most relevant baseline features in BD prediction were past failure (BDI item 4), sadness (BDI item 1), current depressive episode (assessed by MINI interview), self-reported stress problems, self-confidence (HCL-32 item 3) and lifetime cocaine use. Feature importance can be seen in more detail in [Figure 2](#).

---

<sup>1</sup> Class decision boundary separates the data points into classes, where the algorithm switches from one class to another. In the present paper, a threshold of 0.5 was used. If a prediction had a probability  $\geq 0.5$ , it was classified as a positive instance, otherwise, as a negative one.

Given the importance of interpreting the model trajectory to a given prediction, to visualize the influence of each feature on the prediction of a specific sample, a force plot was built. The SHAPley values for each training observation are shown in [Figure S3](#). Partial dependence plots for the continuous remaining features after feature selection are available in [Figure S4](#).

In addition to the main pipeline, 1,000 different random training and testing splits were sampled in order to fit the final model. The estimates can be visualized in [Figure S5](#). After calculating mean and confidence intervals, we observed a resampled AUC of 0.723 (95% CI [0.719, 0.726]). Along with the random splits, we also performed a permutation test as proposed by Fisher (1935)<sup>20</sup> to compare the distribution of ROC AUC performance of random rearrangements of the outcome with the original test data using 10,000 permutations. We observed statistical difference between the original and permuted models ( $p < 0.001$ ). The distribution of the permuted AUCs is available in [Figure S6](#). This result shows that our model predictions for BD incidence after the five years are significantly more accurate than random classifiers.

Regarding the second model without the inclusion of mood-related variables, it presented lower predictive performance compared to the main model ([Table S5](#)), especially on ROC AUC (0.786 vs. 0.644). The most important features for predicting BD onset that were not directly associated with mood measures can be seen in [Figure 3](#). Maternal mental health issues was the most relevant feature, followed by self-reported stress problems, occupation in last year,



suicide attempt in social circle and alcohol abuse or dependence.

Linear associations between depression severity scores at baseline and cognitive performance as well as functional impairment years later were examined. The analysis excluded subjects with current mood episodes at follow-up ( $n = 107$ ) to minimize the direct influence of mood symptoms on the measures. The final sample consisted of 984 individuals. The regression analysis revealed a negative correlation between depressive symptom severity and cognitive performance ( $\beta = -0.11$ , 95% CI [-0.14, -0.07],  $t(957) = -5.51$ ,  $p < 0.001$ ,  $r = -0.18$ ). This indicates that higher baseline depression scores were associated with lower cognitive performance five years later (Figure 4a). Furthermore, a positive correlation was observed between depressive symptom severity and functional impairment ( $\beta = 0.23$ , 95% CI [0.17, 0.30],  $t(980) = 7.55$ ,  $p < 0.001$ ,  $r = 0.24$ ). This suggests that higher baseline depression scores were associated with increased functional impairment five years later (Figure 4b). In subgroup analysis, among females, a moderate positive correlation ( $r = 0.283$ ,  $p < 0.001$ ) suggests an association between baseline depression severity and functional impairment. Conversely, among males, a weaker positive relationship ( $r = 0.174$ ,  $p < 0.001$ ) was observed (Figure S7a). When stratified by suicide risk at baseline, individuals with suicide risk exhibited a strong positive correlation ( $r = 0.47$ ,  $p = 0.007$ ) between baseline depression severity and functional impairment (Figure S7b). In contrast, those without suicide risk showed a moderate positive relationship ( $r = 0.217$ ,  $p < 0.001$ ). We investigated the relationship between self-reported age at

first depressive episode and suicide risk among study subjects. [Figure 5](#) illustrates the comparison between individuals with suicide risk and those without ( $n = 31$ ). Our findings reveal a trend, indicating that subjects with suicide risk tend to experience earlier onset of depression compared to those without (mean in group without suicide risk = 21.76, mean in group with suicide risk = 18.50,  $\Delta = 3.26$ , 95% CI [0.32, 6.21],  $t(20.20) = 2.31$ ,  $p = 0.032$ ; Cohen's  $d = 1.03$ , 95% CI [0.09, 1.94]).

#### 4. Discussion

Our aim was to develop a predictive model with acceptable classification performance for forecasting BD within a 5-year interval. The final model demonstrated favorable metrics (ROC AUC: 78.6%; sensitivity: 37.5%; specificity: 92%), indicating a good predictive capacity. To the best of our knowledge, this is the second Brazilian study to investigate the prediction of BD incidence. A previous study investigated the development of a prediction model, with the use of elastic net algorithms, to identify participants who would develop BD over the follow-up, in a large community birth cohort, from the city of Pelotas in Brazil<sup>21</sup>. According to the results of this investigation, the model with the best performance (ROC AUC: 82%) predicted BD at the age of 22 years, using clinical and sociodemographic data from the age of 18 years<sup>21</sup>. Based on a scoping review<sup>22</sup> that aimed to explore the role of ML in BD diagnosis, no included study reported the use of boosted trees algorithms in order to predict BD ( $n = 33$ ). The mean ROC AUC value of the models was 0.81. However, it is important to note that the review did not evaluate the implementation of resampling techniques such as CV or bootstrapping in these studies. Incorporating these techniques could potentially mitigate any inflated performances observed<sup>22</sup>. A recent systematic review by Colombo and colleagues<sup>23</sup> provided meta-analysis on ML approaches for BD prediction using different types of markers. The pooled accuracy for BD prediction based on 65 studies ( $n = 11,336$ ;  $n_{BD} = 3,903$ ) was 0.77 [95% CI: 0.75, 0.80], similar to our study accuracy of 0.89 and ROC AUC of 0.79 [95% CI: 0.69, 0.89].

A recent systematic review on clinical prediction models in psychiatry pointed to several aspects that predictive models could improve, such as overfitting prevention, generalizability and clinical utility<sup>24</sup>. In this study, we employed statistical learning techniques to predict BD using a larger sample size compared to many previous studies, despite having a relatively low value of events per variable (EPV) at approximately 5.8, which may have increased the model instability. In an attempt to minimize lack of external validation — since it is a known issue in predictive modeling in psychiatry — due to heterogeneity of measures between different samples, the present study also estimated uncertainty with resampling methodology. Both the permutation test and resamples of data splitting showed acceptable model reliability while maintaining an average ROC AUC above 0.7 on resamples and better classification when compared to random classifiers.

The current study aligns with previous findings indicating that depressive symptoms serve as key predictors for BD conversion<sup>25,26</sup>. Interestingly, the prediction model developed in this study identified three primary markers that are closely associated with depression: past failure, sadness, and early depressive episodes. These findings suggest that these factors may serve as prodromal symptoms of the disorder<sup>27,28</sup> or even indicate a genetic predisposition to emotional distress<sup>29</sup>. Moreover, these results reinforce the perspective of BD as a progressive trajectory, with the first mood episode serving as a crucial milestone for the onset of this complex disorder<sup>30–32</sup>. In the majority of cases, the initial mood

episode in patients with BD is depressive, often occurring years before a manic episode<sup>30,33,34</sup>. Consequently, the model developed in this study also holds utility in facilitating the differential diagnosis between unipolar and bipolar depression. By considering additional predictors, it becomes possible to assess the likelihood of a given current depressive episode belonging to a unipolar or bipolar clinical condition. Although, the literature in that matter has controversial evidence. In a meta-analysis that included 31 studies, depression severity was not a risk factor for conversion from unipolar to BD<sup>35</sup>, however this result could be due to heterogeneity in the studies.

Since our main classifier found depression-related features as the most associated ones with BD onset, we investigated the effect of depression severity on cognitive performance and global functioning. We found that more severe baseline depressive symptoms were associated with poorer cognitive performance and also worse global functional impairment five years later. Females and subjects with suicide risk at baseline presented a stronger association on depression severity and functional impairment. Besides an increased risk for BD, severe depression is also associated with worse cognitive performance<sup>36</sup>, especially in untreated illness. The same pattern is described in the literature for lifetime major depressive disorder related to worse cognitive performance, and subjects with severe psychosocial impairment may perform even worse on cognitive tasks<sup>37</sup>. Although, Szymkowicz et al. (2021) reports weak associations between depression severity and cognitive performance in young to middle-aged men

without clinical depression<sup>38</sup>. Other findings suggest that cognitive impairment and global functioning impairment are associated with the severity of depressive symptoms in bipolar depression<sup>39</sup>.

We explored the relationship between suicidality and age at depression onset and found that subjects that presented an earlier onset for the first depressive episode in life had a higher prevalence of suicide risk. This finding is particularly concerning for individuals with BD since they are already at high risk for suicide due to the nature of the disease<sup>4</sup>. Several studies in the literature have indicated that experiencing depression at a young age may contribute to a more complex and challenging course of illness, including increased suicidality<sup>40</sup>. The emergence of BD after an initial episode of depression can significantly impact the overall prognosis and treatment outcomes<sup>41</sup> with increased risk of recurrent mood episodes and more severe symptomatology. Furthermore, suicidality might be magnified in these specific cases of early depression onset with conversion to BD, particularly in settings with severe depressive symptomatology<sup>42</sup>. The confluence of factors, such as the inherently high suicide risk associated with BD and the early onset of depression, may lead to poorer outcomes for those affected. It is crucial for healthcare professionals to be vigilant and identify these high-risk cases promptly in order to act on preventive treatment plans.

Our study also evaluated lifetime cocaine use as a significant predictor, in addition to alcohol abuse or dependence at baseline in the model without depression-related features. Numerous studies have examined the relationship

between substance use and the development of mood disorders. A cross-sectional study in 2013 found that individuals with primary substance use disorder (SUD) had a higher likelihood of subsequently developing mood disorders, with an average onset of 11 years after SUD initiation<sup>43</sup>. The study identified a particularly elevated risk of developing BD among individuals with drug dependence. A systematic review published in 2021 supported the role of substance use as a predictor for BD and (hypo)manic symptoms, and alcohol use was reported as a risk factor for BD in approximately 43% of studies<sup>44</sup>. Furthermore, specific research on cocaine use and BD has been conducted. A prospective study examined lifetime cocaine use as a potential predictor for conversion from major depressive disorder to BD<sup>45</sup>. The analysis revealed a 3.41-fold increased risk of conversion to BD in individuals who reported lifetime cocaine use at baseline. Similarly, a systematic review reported a five-fold increased risk of BD development in individuals with lifetime cocaine use<sup>46</sup>. Besides the increased risk for BD onset, SUD comorbidity can also lead to poorer outcomes such as increased number of suicide attempts, accelerated BD progression and worse cognitive symptoms<sup>47</sup>, requiring specific treatment plans<sup>48</sup>. Therefore, our findings align with existing literature, contributing to the growing body of evidence on this topic.

There are several noteworthy strengths in this study. Firstly, our sample consisted of young adults aged 18 to 24 years. As per the literature, symptoms of BD typically manifest before the age of 25, making our sample relevant for early

disease identification. This critical period of development in our sample population increases the likelihood of capturing early onset symptoms. Additionally, whenever doubts arose regarding the diagnosis based on the standardized diagnostic interview (MINI), we sought the expertise of an external psychiatrist to provide confirmation. This approach ensured a thorough and reliable characterization of the diagnosis, establishing a gold standard. Furthermore, the average follow-up period of five years exceeded the durations observed in other studies within this field<sup>49</sup>. The use of a state-of-the-art algorithm that was not commonly used in previous studies with comparable aims is also a positive factor.

The study benefits from a large sample collected through a probabilistic sampling method from a city in southern Brazil, encompassing approximately 343,651 inhabitants. These factors contribute to the robustness of our model. However, predicting the outcome presents challenges due to two main reasons: 1) the rarity of the outcome and 2) the possibility of participants without BD at follow-up developing BD in the future. Addressing these issues statistically is a common challenge in studies within this field. Another important consideration when interpreting the presented results is the generalizability and applicability of the model. Precision psychiatry studies are emerging, and although we have achieved satisfactory results in this work (test ROC AUC 0.78), we acknowledge that the presented data primarily serve scientific purposes and provide a foundation for future improvements. Our study demonstrates the potential for developing a calculator that could be implemented in basic health systems in the



near future. The information presented here is particularly relevant for patients exhibiting characteristics identified as potentially important for BD diagnosis, such as early onset depressive episode, severe depressive symptoms, family history of mental health issues, and substance use. Such a tool holds promise for robust screening, enabling early intervention, ensuring better prognosis, and preventing the progression to more severe clinical conditions.

In summary, we have developed a binary model using a state-of-the-art algorithm capable of predicting the diagnosis of BD in a specific population of young adults over a period of approximately five years. The model incorporates clinical, socio environmental, substance use, sex-related variables, and demographic data collected through a probabilistic sample. Our findings have also highlighted significant associations between the severity of depressive symptoms and functional and cognitive outcomes. However, for a more comprehensive signature of BD diagnosis, future studies should focus on conducting systematic follow-ups to track subjects across various developmental stages. Further, embracing multimodal approaches may further enhance the predictive capacity and clinical utility of these models<sup>50</sup>. Also, research efforts should target specific at-risk populations, such as individuals with depression or children of parents with BD. Additionally, incorporating digital health data, biological markers, neuropsychological information, and using neuroimaging techniques hold promise for the development of future models with greater applicability.

## References

1. McIntyre RS, Berk M, Brietzke E, et al. Bipolar disorders. *Lancet Lond Engl.* 2020;396(10265):1841-1856. doi:10.1016/S0140-6736(20)31544-0
2. Merikangas KR, Jin R, He JP, et al. Prevalence and Correlates of Bipolar Spectrum Disorder in the World Mental Health Survey Initiative. *Arch Gen Psychiatry.* 2011;68(3):241. doi:10.1001/archgenpsychiatry.2011.12
3. Kessing LV, Vradi E, Andersen PK. Life expectancy in bipolar disorder. *Bipolar Disord.* 2015;17(5):543-548. doi:10.1111/bdi.12296
4. Plans L, Barrot C, Nieto E, et al. Association between completed suicide and bipolar disorder: A systematic review of the literature. *J Affect Disord.* 2019;242:111-122. doi:10.1016/j.jad.2018.08.054
5. Cloutier M, Greene M, Guerin A, Touya M, Wu E. The economic burden of bipolar I disorder in the United States in 2015. *J Affect Disord.* 2018;226:45-51. doi:10.1016/j.jad.2017.09.011
6. Scott J, Leboyer M. Consequences of delayed diagnosis of bipolar disorders. *L'Encephale.* 2011;37(Suppl):3. doi:10.1016/S0013-7006(11)70048-3
7. Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord.* 2018;20(2):97-170. doi:10.1111/bdi.12609
8. Altamura AC, Buoli M, Caldiroli A, et al. Misdiagnosis, duration of untreated illness (DUI) and outcome in bipolar patients with psychotic symptoms: A naturalistic study. *J Affect Disord.* 2015;182:70-75. doi:10.1016/j.jad.2015.04.024
9. McIntyre RS, Calabrese JR. Bipolar depression: the clinical characteristics and unmet needs of a complex disorder. *Curr Med Res Opin.* 2019;35(11):1993-2005. doi:10.1080/03007995.2019.1636017
10. Fernandes BS, Williams LM, Steiner J, Leboyer M, Carvalho AF, Berk M. The new field of "precision psychiatry." *BMC Med.* 2017;15(1):80. doi:10.1186/s12916-017-0849-x
11. Passos IC, Mwangi B, Kapczynski F. Big data analytics and machine learning: 2015 and beyond. *Lancet Psychiatry.* 2016;3(1):13-15. doi:10.1016/S2215-0366(15)00549-0
12. Librenza-Garcia D, Kotzian BJ, Yang J, et al. The impact of machine learning techniques in the study of bipolar disorder: A systematic review. *Neurosci Biobehav Rev.* 2017;80:538-554. doi:10.1016/j.neubiorev.2017.07.004
13. Passos IC, Ballester PL, Barros RC, et al. Machine learning and big data analytics in bipolar disorder: A position paper from the International Society for Bipolar Disorders Big Data Task Force. *Bipolar Disord.* 2019;21(7):582-594. doi:10.1111/bdi.12828
14. Brazilian Institute of Geography and Statistics. Brazilian 2010 Census. Published online 2010. <https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html>
15. Hartung C, Lerer A, Anokwa Y, Tseng C, Brunette W, Borriello G. Open Data Kit: Tools to Build Information Services for Developing Regions. In: *Proceedings of the*

- 4th ACM/IEEE International Conference on Information and Communication Technologies and Development - ICTD '10. ACM Press; 2010.  
doi:10.1145/2369220.2369236
16. Kurska MB, Jankowski A, Rudnicki WR, Boruta – A System for Feature Selection. *Fundam Informaticae*. 2010;101(4):271-285. doi:10.3233/fi-2010-288
  17. Menardi G, Torelli N. Training and assessing classification rules with imbalanced data. *Data Min Knowl Discov*. 2014;28(1):92-122. doi:10.1007/s10618-012-0295-5
  18. Kuhn M. Futility Analysis in the Cross-Validation of Machine Learning Models. *ArXiv E-Prints*. Published online May 2014. doi:10.48550/arXiv.1405.6974
  19. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. *ArXiv E-Prints*. Published online May 2017. doi:10.48550/arXiv.1705.07874
  20. Fisher RA. *The Design of Experiments*. Oliver and Boyd; 1935.
  21. Rabelo-da-Ponte FD, Feiten JG, Mwangi B, et al. Early identification of bipolar disorder among young adults - a 22-year community birth cohort. *Acta Psychiatr Scand*. 2020;142(6):476-485. doi:10.1111/acps.13233
  22. Jan Z, Al-Ansari N, Mousa O, et al. The Role of Machine Learning in Diagnosing Bipolar Disorder: Scoping Review. *J Med Internet Res*. 2021;23(11):e29749. doi:10.2196/29749
  23. Colombo F, Calesella F, Mazza MG, et al. Machine learning approaches for prediction of bipolar disorder based on biological, clinical and neuropsychological markers: A systematic review and meta-analysis. *Neurosci Biobehav Rev*. 2022;135:104552. doi:10.1016/j.neubiorev.2022.104552
  24. Meehan AJ, Lewis SJ, Fazel S, et al. Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. *Mol Psychiatry*. 2022;27:2700-2708. doi:10.1038/s41380-022-01528-4
  25. Hafeman DM, Merranko J, Goldstein TR, et al. Assessment of a Person-Level Risk Calculator to Predict New-Onset Bipolar Spectrum Disorder in Youth at Familial Risk. *JAMA Psychiatry*. 2017;74(8):841-847. doi:10.1001/jamapsychiatry.2017.1763
  26. Perich T, Lau P, Hadzi-Pavlovic D, et al. What clinical features precede the onset of bipolar disorder? *J Psychiatr Res*. 2015;62:71-77. doi:10.1016/j.jpsychires.2015.01.017
  27. Faedda GL, Baldessarini RJ, Marangoni C, et al. An International Society of Bipolar Disorders task force report: Precursors and prodromes of bipolar disorder. *Bipolar Disord*. 2019;21(8):720-740. doi:10.1111/bdi.12831
  28. Van Meter AR, Burke C, Youngstrom EA, Faedda GL, Correll CU. The Bipolar Prodrome: Meta-Analysis of Symptom Prevalence Prior to Initial or Recurrent Mood Episodes. *J Am Acad Child Adolesc Psychiatry*. 2016;55(7):543-555. doi:10.1016/j.jaac.2016.04.017
  29. Smeland OB, Wang Y, Frei O, et al. Genetic Overlap Between Schizophrenia and Volumes of Hippocampus, Putamen, and Intracranial Volume Indicates Shared Molecular Genetic Mechanisms. *Schizophr Bull*. 2018;44(4):854-864. doi:10.1093/schbul/sbx148
  30. Duffy A, Horrocks J, Doucette S, Keown-Stoneman C, McCloskey S, Grof P. The developmental trajectory of bipolar disorder. *Br J Psychiatry J Ment Sci*. 2014;204(2):122-128. doi:10.1192/bjp.bp.113.126706

31. Kapczinski F, Montezano BB, de Azevedo Cardoso T. Latent bipolar disorder. *Lancet*. 2023;401(10394):2109. doi:10.1016/S0140-6736(23)00402-6
32. O'Donovan C, Alda M. Depression Preceding Diagnosis of Bipolar Disorder. *Front Psychiatry*. 2020;11:500. doi:10.3389/fpsy.2020.00500
33. Mesman E, Nolen WA, Reichart CG, Wals M, Hillegers MHJ. The Dutch bipolar offspring study: 12-year follow-up. *Am J Psychiatry*. 2013;170(5):542-549. doi:10.1176/appi.ajp.2012.12030401
34. Mesman E, Nolen WA, Keijsers L, Hillegers MHJ. Baseline dimensional psychopathology and future mood disorder onset: findings from the Dutch Bipolar Offspring Study. *Acta Psychiatr Scand*. 2017;136(2):201-209. doi:10.1111/acps.12739
35. Kessing LV, Willer I, Andersen PK, Bukh JD. Rate and predictors of conversion from unipolar to bipolar disorder: A systematic review and meta-analysis. *Bipolar Disord*. 2017;19(5):324-335. doi:10.1111/bdi.12513
36. Galimberti C, Bosi MF, Volontè M, Giordano F, Dell'Osso B, Viganò CA. Duration of untreated illness and depression severity are associated with cognitive impairment in mood disorders. *Int J Psychiatry Clin Pr*. 2020;24(3):227-235. doi:10.1080/13651501.2020.1757116
37. de Nooij L, Harris MA, Adams MJ, et al. Cognitive functioning and lifetime major depressive disorder in UK Biobank. *Eur Psychiatry*. 2020;63(1):e28. doi:10.1192/j.eurpsy.2020.24
38. Szymkowicz SM, Dotson VM, Vanderploeg RD. Weak associations between depressive symptom severity, depressive symptom clusters, and cognitive performance in young to middle-aged men without clinical depression. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*. 2021;28(6):921-935. doi:10.1080/13825585.2020.1840505
39. Kapczinski NS, Narvaez JC, Magalhães PV, et al. Cognition and functioning in bipolar depression. *Rev Bras Psiquiatr*. 2016;38(3):201-206. doi:10.1590/1516-4446-2014-1558
40. de Girolamo G, Dagani J, Purcell R, Cocchi A, McGorry PD. Age of onset of mental disorders and use of mental health services: needs, opportunities and obstacles. *Epidemiol Psychiatr Sci*. 2012;21(1):47-57. doi:10.1017/s2045796011000746
41. Kalman JL, Loohuis LMO, Vreeker A, et al. Characterisation of age and polarity at onset in bipolar disorder. *Br J Psychiatry*. 2021;219(6):659-669. doi:10.1192/bjp.2021.102
42. Lage RR, de Assis da Silva R, Tancini MB, Nardi AE, Mograbi DC, Cheniaux E. Suicidal Ideation in Bipolar Disorder: The Relation with Clinical and Sociodemographic Variables. *Psychiatr Q*. 2022;93(2):453-461. doi:10.1007/s11126-021-09965-0
43. Kenneson A, Funderburk JS, Maisto SA. Substance use disorders increase the odds of subsequent mood disorders. *Drug Alcohol Depend*. 2013;133(2):338-343. doi:10.1016/j.drugalcdep.2013.06.011
44. Lalli M, Brouillette K, Kapczinski F, de Azevedo Cardoso T. Substance use as a risk factor for bipolar disorder: A systematic review. *J Psychiatr Res*. 2021;144:285-295. doi:10.1016/j.jpsychires.2021.10.012

45. Cardoso TA, Jansen K, Mondin TC, et al. Lifetime cocaine use is a potential predictor for conversion from major depressive disorder to bipolar disorder: A prospective study. *Psychiatry Clin Neurosci*. 2020;74(8):418-423. doi:10.1111/pcn.13012
46. Marangoni C, Hernandez M, Faedda GL. The role of environmental exposures as risk factors for bipolar disorder: A systematic review of longitudinal studies. *J Affect Disord*. 2016;193:165-174. doi:10.1016/j.jad.2015.12.055
47. Preuss UW, Schaefer M, Born C, Grunze H. Bipolar Disorder and Comorbid Use of Illicit Substances. *Med Kaunas*. 2021;57(11):1256. doi:10.3390/medicina57111256
48. Grunze H, Schaefer M, Scherk H, Born C, Preuss UW. Comorbid Bipolar and Alcohol Use Disorder-A Therapeutic Challenge. *Front Psychiatry*. 2021;12:660432. doi:10.3389/fpsy.2021.660432
49. Ribeiro JS, Pereira D, Salagre E, et al. Risk Calculators in Bipolar Disorder: A Systematic Review. *Brain Sci*. 2020;10(8):525. doi:10.3390/brainsci10080525
50. Kline A, Wang H, Li Y, et al. Multimodal machine learning in precision health: A scoping review. *Npj Digit Med*. 2022;5(1):171. doi:10.1038/s41746-022-00712-8

## Tables

**Table 1.** Sociodemographic features measured at baseline grouped by bipolar disorder diagnosis in the follow-up ( $n = 1091$ ).

Features	With BD ( $n = 49$ )	Without BD ( $n = 1,042$ )	Overall ( $n = 1,091$ )	<i>p</i> -value
<b>Sex</b>				0.302
Male	16 (32.7%)	457 (43.9%)	473 (43.4%)	
Female	33 (67.3%)	585 (56.1%)	618 (56.5%)	
<b>Age</b>				0.353
Mean (SD)	20.9 (2.28)	20.5 (2.10)	20.5 (2.11)	
Median [Min, Max]	21.0 [17.0, 25.0]	20.0 [16.0, 26.0]	20.0 [16.0, 26.0]	
Missing	0 (0%)	8 (0.8%)	8 (0.7%)	
<b>Socioeconomic status</b>				0.635
Upper	14 (28.6%)	401 (38.5%)	415 (38.0%)	
Middle	26 (53.1%)	509 (48.8%)	535 (49.0%)	
Lower	9 (18.4%)	132 (12.7%)	141 (12.9%)	
<b>Skin color</b>				0.800
White	34 (69.4%)	765 (73.4%)	799 (73.2%)	
Non-white	15 (30.6%)	273 (26.2%)	288 (26.4%)	
Missing	0 (0%)	4 (0.4%)	4 (0.4%)	
<b>Currently working</b>				0.965
No	8 (16.3%)	245 (23.5%)	253 (23.2%)	
Yes	14 (28.6%)	380 (36.5%)	394 (36.1%)	
Missing	27 (55.1%)	417 (40.0%)	444 (40.7%)	
<b>Has a partner</b>				0.953
No	13 (26.5%)	247 (23.7%)	260 (23.8%)	
Yes	33 (67.3%)	696 (66.8%)	729 (66.8%)	

---

Features	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)	p-value
Missing	3 (6.1%)	99 (9.5%)	102 (9.3%)	

---

The  $p$ -values were calculated from  $t$ -tests for numeric features and  $\chi^2$  tests for categorical features. Abbreviations: BD (Bipolar disorder). Standard deviation (SD).

**Figure legends**

**Figure 1.** Machine learning pipeline flowchart. The figure shows data splitting, preprocessing routine, feature selection, cross-validation, model fitting, model assessment and feature importance steps using XGBoost algorithm.

**Figure 2.** SHAPley values for each feature included in the final model. Each y-axis tick represents a feature, sorted by the highest absolute contribution across all observations, regardless of the direction of the association. Each dot represents a participant in the training set ( $n = 763$ ). Observations with SHAPley values lower than zero behaved as protective factors, otherwise they were risk factors. The fill color represents the value of the variable for a given individual (red corresponds to higher values, and blue corresponds to lower values). Categorical features are coded as 0/1. The bold text next to each feature name represents the average absolute contribution (irrespective of direction).

**Figure 3.** Feature importance plot based on the model with removal of mood-related features (test ROC AUC: 0.644). Maternal mental health issues and self-reported stress problems were the two most important variables.

**Figure 4.** Linear associations of depression severity score (BDI total score) at baseline with a) cognitive performance (MoCA score) and b) functional impairment (FAST score) five years later. The fitted regression lines show a negative correlation of depressive symptoms severity with cognitive performance ( $r = -0.18$ ,  $n = 959$ ), and a positive correlation with functional impairment ( $r = 0.24$ ,  $n = 982$ ). Subjects with current mood episodes at follow-up were excluded for the analysis



( $n = 107$ ), in order to minimize the direct effect of mood symptoms on these measures, resulting in a subsample of 984 individuals. Abbreviations: MoCA (Montreal Cognitive Assessment). FAST (Functioning Assessment Short Test).

**Figure 5.** Self-reported age at first depressive episode among subjects with and without suicide risk (based on MINI C module on suicide risk). The plot shows a trend that subjects with suicide risk present earlier depression onset ( $n = 31$ ). Eighty-five subjects presented a depressive episode at baseline and 54 presented missing values at depression age at onset variable, resulting in a final subsample of 31 individuals. The Student's  $t$ -test was used and suggests a significant difference between the groups (mean in group No = 21.76, mean in group Yes = 18.50,  $\Delta = 3.26$ , 95% CI [0.32, 6.21],  $t(20.20) = 2.31$ ,  $p = 0.032$ ; Cohen's  $d = 1.03$ , 95% CI [0.09, 1.94]). Abbreviations: MINI (Mini International Neuropsychiatric Interview).

# APÊNDICE E – Material suplementar do Artigo 1

## Supporting Information

### ***Bipolar disorder prediction and functional outcomes in young adults: a 5-year follow-up study***

#### **Mini International Neuropsychiatric Interview (MINI) description and psychometric properties**

MINI is a short-term diagnostic interview designed for clinical assessment of mental disorders according to Diagnostic and Statistical Manual of Mental Disorders — Fourth Edition (DSM-IV) and ICD-10 criteria<sup>1,2</sup>.

Despite evaluating several disorders, MINI psychometric properties for the diagnosis of lifetime manic episode (sensitivity: 81.0%; specificity: 94.0%; positive predictive value: 76.0%; negative predictive value: 95.0%) and major depressive episode (sensitivity: 96.0%; specificity: 88.0%; positive predictive value: 87.0%; negative predictive value: 97.0%) are reliable when compared to DSM Structured Clinical Interview<sup>3</sup>.

#### **Tuned hyperparameters**

Hyperparameters are parameters whose values are used to control the algorithm training process. The following tree boosting hyperparameters were tuned<sup>4,5</sup>:

- a) *mtry*: Number of predictors that is randomly sampled at each split.
- b) *trees*: Number of trees contained in the ensemble.
- c) *min\_n*: Minimum number of observations in a node required for the node to be split further.
- d) *tree depth*: Maximum depth (number of splits) of each tree.
- e) *loss\_reduction*: Reduction in the loss function required to split further.
- f) *learn\_rate*: Step size at each iteration while moving toward a loss function optimization.
- g) *sample\_size*: Proportion of the data set used for modeling within an

iteration.

### **Tidymodels framework**

The implementation of the data modeling routines was carried out using the *tidymodels* framework<sup>6</sup>. The *tidymodels* is a R meta package made up of multiple packages that assist in different stages of a machine learning pipeline. In order to split the data and create cross-validation (CV) resamples, *rsample* package was used<sup>7</sup>, *parsnip* was used to access *XGBoost* functions in a unified manner<sup>5</sup>, *recipes* for preprocessing functions<sup>8</sup>, *workflows* to bundle the preprocessing, modeling and post-processing routines<sup>9</sup>, *yardstick* to easily calculate performance measures<sup>10</sup>. In the CV, fifteen hyperparameter combinations were used as candidate parameter sets. The values for each hyperparameter were randomly chosen based on an algorithm that attempts to maximize the determinant of the spatial correlation matrix between coordinates<sup>11</sup>.

### **Preprocessing steps**

Some preprocessing techniques were adopted to clean and tidy data prior to modeling. The following preprocessing steps were applied:

- 1) Remove all features with more than 10% of missing values.
- 2) Impute categorical features with mode.
- 3) Impute numeric features with median.
- 4) Remove near-zero variance features (few unique values relative to the number of observations and also a ratio of the frequency of the second most common value is large [ratio of 10]).
- 5) Create dummy variables with  $C - 1$  categories from categorical features.

### **Feature selection**

The feature selection process aims to automatically filter variables from the data matrix considering their relevance to the predictive modeling problem. Therefore, we can build more accurate and parsimonious models while, at the same time, saving computational resources through the use of less data in the next model fitting steps.

### **Class imbalance in bipolar disorder prediction**

Class imbalance is a common problem in classification modeling. It happens when we face a set of examples that presents a given level way more frequently than others. Since most ML classifiers assume data equally distributed, they tend to be more biased towards the majority class, causing bad performance on minority class classification.

This concept is especially important in the context of predicting mental disorders, as subjects who will present the disease will be exposed to greater health risk. Therefore, it is necessary that the classifiers of such outcomes can adequately predict this portion of the population. BD still has the aggravating factor of having a complex prognosis regarding the neuroprogression, which can be worsened by the length of disease<sup>12</sup>.

The *themis* and *ROSE* R packages were adopted to implement the ROSE method<sup>13,14</sup>.

### **Hyperparameter tuning racing method proposed by Kuhn<sup>15</sup>**

It consists in calculating the area under the receiver operating characteristic (ROC) curve for each parameter set across validation folds. After evaluating the parameter combinations for three resamples, a repeated measure ANOVA model is fitted. The combinations that are statistically different (based on  $\alpha$  level for one-sided confidence interval of 5%) from the best setting are excluded from further validation procedures. The ANOVA racing method was implemented via the *finetune* R package<sup>16</sup>.

### **Performance measures**

Aiming to evaluate the performance of the algorithm, some evaluation metrics were used. Firstly, the area under the receiver operating characteristic (ROC) curve (AUC) was used to diagnose the classifier's ability to predict correctly across multiple discrimination thresholds<sup>17</sup>.

Sensitivity and specificity were also used to assess model ability to correctly detect subjects with BD who have the disorder, and correctly detect

subjects that did not present BD who actually does not have BD, respectively<sup>18</sup>. Positive and negative predictive values (PPV and NPV) show the proportion of positive and negative predictions that are truly positive or negative, correspondingly.

Accuracy measures the proportion of correct predictions among the total number of observations evaluated<sup>19</sup>. In order to take into account the class imbalance previously described, we also used balanced accuracy as it inputs both sensitivity and specificity into its formula:

$$\text{Balanced accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \quad (1)$$

The F1 score is defined as the harmonic mean of PPV and sensitivity (Equation 2). This metric is able to demonstrate another model accuracy measure, being more robust in class imbalance scenarios.

$$F_1 = 2 \cdot \frac{\text{Positive predictive value} \cdot \text{Sensitivity}}{\text{Positive predictive value} + \text{Sensitivity}} \quad (2)$$

Additionally, we also calculated the following performance measures to assess the predictive model: precision-recall curve (PR AUC), positive predictive value (PPV) and negative predictive value (NPV).

#### **Permutation test**

This test provides an empirical estimate of the null distribution by randomly permuting the outcome labels while keeping the predictor variables intact. By comparing the observed performance to the permuted performance, we determined whether the model's accuracy exceeded what would be expected by chance. First, the original dataset was preprocessed. Subsequently, the dataset was divided into a training set and a separate test set for model evaluation. To evaluate the statistical significance of the observed ROC AUC, 10,000 permutations were performed, generating a distribution of permuted ROC AUC values. The  $p$ -value was calculated as the proportion of permuted ROC AUC

values greater than or equal to the observed ROC AUC.

### **SHAP and partial dependence plots (PDP) R packages**

PDP are able to show the marginal effect a feature has on the predicted outcome of a machine learning model<sup>20</sup>. In the present study, the SHAPley values were obtained using the R package *SHAPforxgboost*<sup>21</sup>. This package provides functions to create SHAP-related visualizations from a *XGBoost* model object. Additional visualizations for SHAP were generated with the *shapviz* package<sup>22</sup>. The PDP were built with the *pdp*<sup>23</sup> and *SHAPforxgboost*<sup>21</sup> R packages, along with the *ggplot2*<sup>24</sup> and *patchwork*<sup>25</sup> packages for plot composition.

### **R and XGBoost versions**

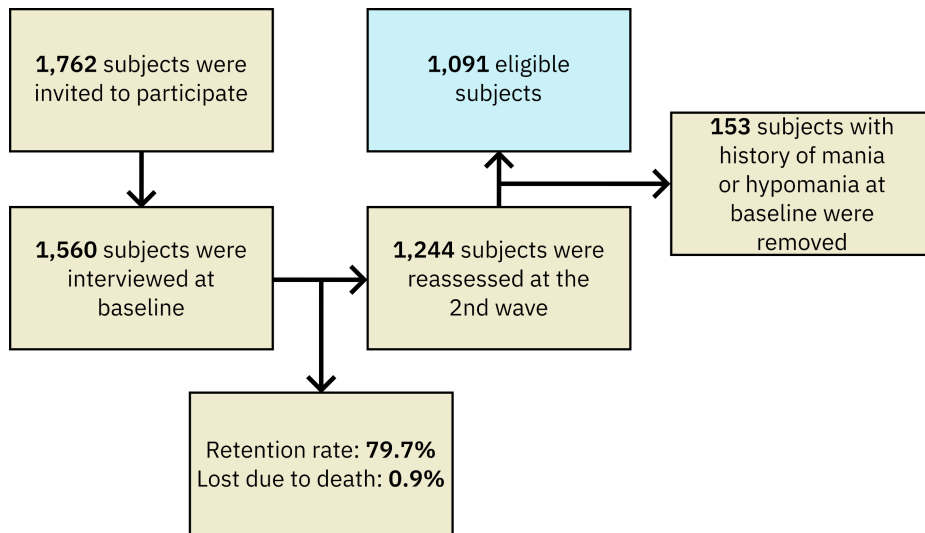
The analyses were implemented through the XGBoost library<sup>4</sup> in the R programming language on version 4.2.1<sup>26</sup>.

### **Tree gradient boosting**

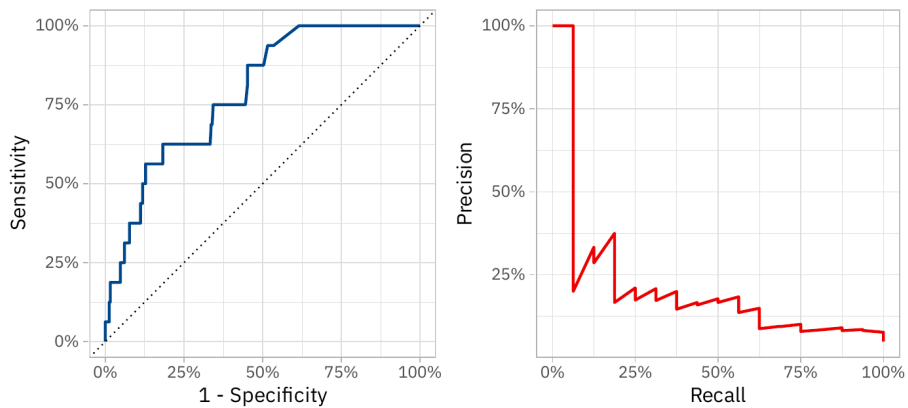
Boosting improves this concept by building a sequence of originally weak models into progressively more powerful models. Additionally, in gradient boosting techniques, the gradient of a loss function is used to choose the best approach to improve a weak learner<sup>27</sup>. In the context of gradient-boosted trees, weak learners are decision trees.

When it comes to tabular data, gradient boosting decision trees (GBDT) are seen as the state-of-the-art, reinforced by several competitions in the ML scenario. In addition, a study found that GBDT perform better than deep learning models across multiple tabular datasets, and also requires less hyperparameter tuning<sup>28</sup>.

**Figures and Tables**

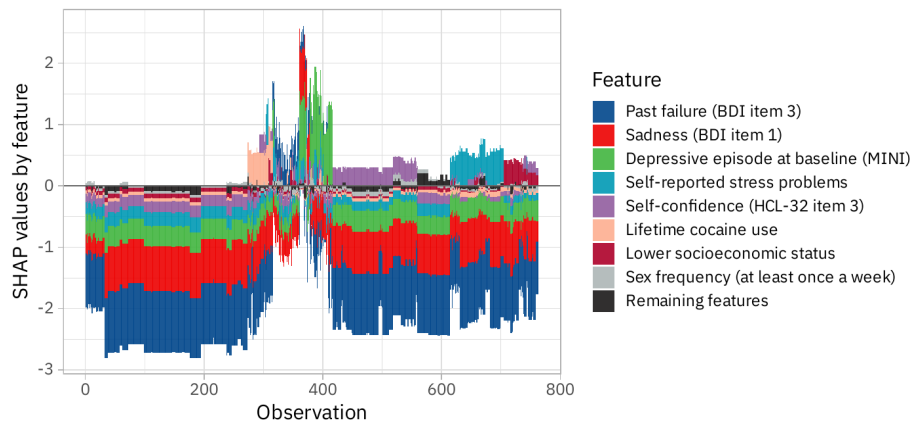


**Figure S1.** Sample size flow chart for identification of subjects that met criteria for final sample ( $n = 1,091$ ).

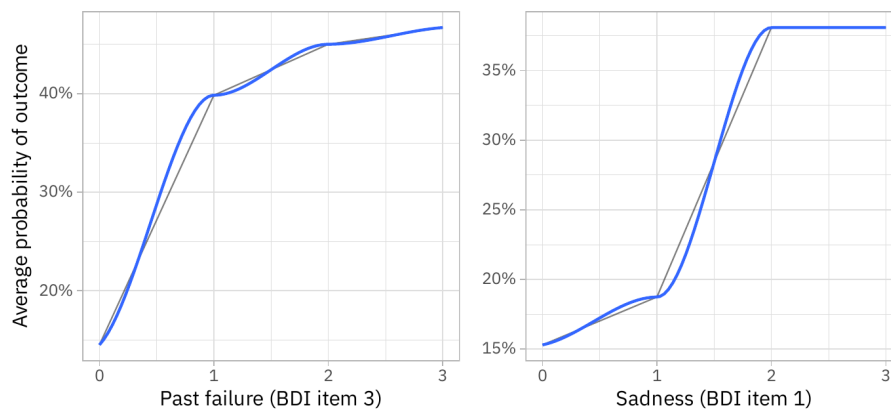


**Figure S2.** Receiver operating characteristic (ROC) curve and precision recall (PR) curve of the final model fitted on the training set with best parameter combination from cross-validation step,

assessed on test set, with an area under the ROC curve of 0.786 and area under the PR curve of 0.208.



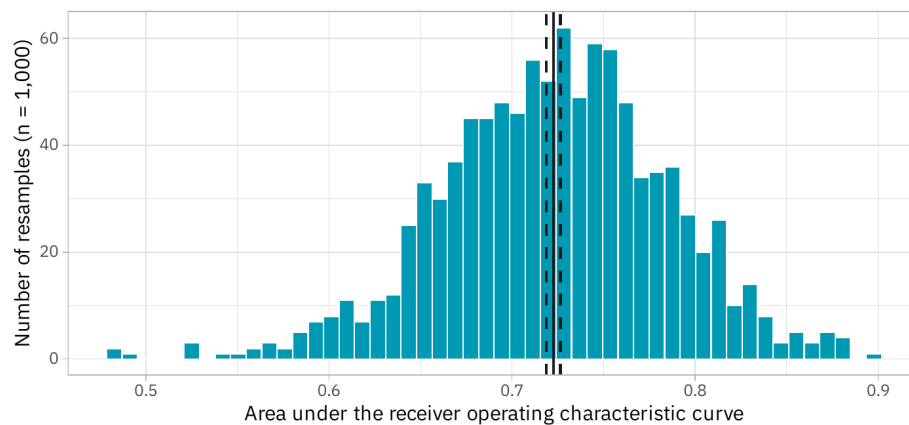
**Figure S3.** SHAPley force plot. The y-axis demonstrates the influence of each feature on current prediction based on SHAPley values. The x-axis represents all samples used to train the final model. The whole training set (n = 763) is presented. Current romantic relationship (having a partner), tachylalia (HCL-32 item 18) and middle socioeconomic status features were combined into the “Remaining features” category.



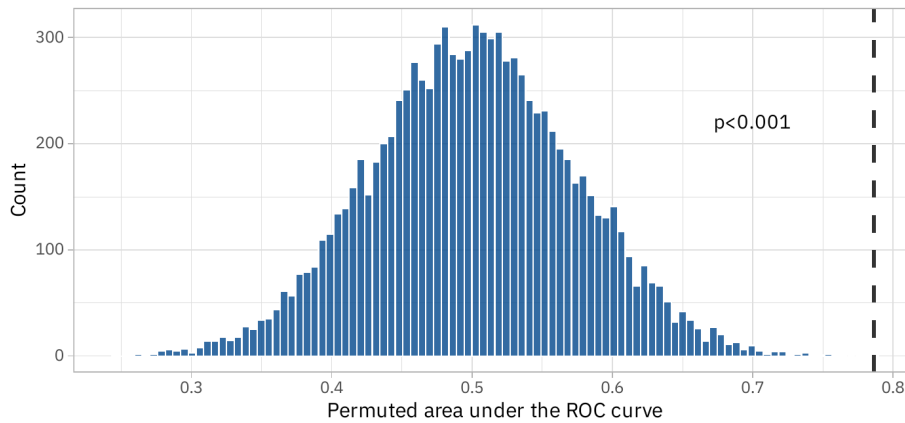
**Figure S4.** Partial dependence plots for continuous depression-related features. It shows the



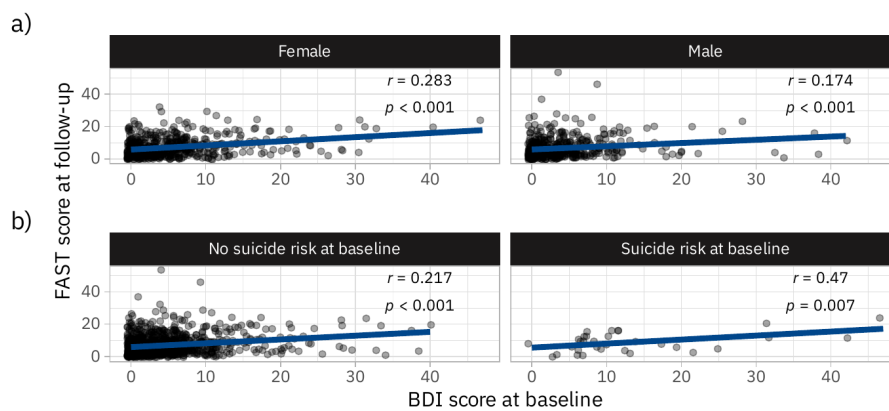
average trend of each feature. Other variables are held constant. The plots show an upward trend, which indicates that the higher the values of the variables of past failure and sadness, the higher the predicted probabilities for developing bipolar disorder after five years. The blue line indicates a regression line using the LOESS (locally weighted polynomial regression) method.



**Figure S5.** Histogram of the area under the receiver operating characteristic curve (ROC AUC) based on 1,000 random training and testing data splits. The ROC AUC mean and 95% CI found were 0.723 [0.719, 0.726]. This analysis is able to demonstrate the predictive performance robustness of the selected boosting model. The solid line indicates the ROC AUC mean and the dashed lines indicate 95% mean CIs.



**Figure S6.** Distribution of the areas under the receiver operating characteristic curves (ROC AUC) from the permutation test with 10,000 rearrangements. The dashed line indicates the AUC on the original test set (ROC AUC = 0.786).



**Figure S7.** Correlations of depression severity score (Beck Depression Inventory [BDI] total score) at baseline with functional impairment (Functioning Assessment Short Test [FAST] global functioning score) five years later stratified by a) gender and b) suicide risk at baseline. a) Among females, a positive correlation coefficient ( $r$ ) of 0.283 suggests a moderate association. Conversely, among males, the correlation ( $r = 0.174$ ) indicates a weaker positive relationship. b) Among individuals exhibiting suicide risk at baseline, a strong positive correlation ( $r = 0.47$ ) is

observed. In contrast, among those without suicide risk, there was a moderate positive relationship ( $r = 0.217$ ). Subjects with current mood episodes at follow-up were excluded for the analysis ( $n = 984$ ), in order to minimize the direct effect of mood symptoms on functioning.

**Table S1.** Additional clinical and environmental features collected at baseline stratified by bipolar disorder outcome at 5-year follow-up ( $n = 1,091$ ). All features presented at least one missing value.

Feature	With BD ( $n = 49$ )	Without BD ( $n = 1,042$ )	Overall ( $n = 1,091$ )
<b>Deceased parents</b>			
No	8 (16.3%)	271 (26.0%)	279 (25.6%)
Yes	41 (83.7%)	770 (73.9%)	811 (74.3%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Has a religion</b>			
No	14 (28.6%)	383 (36.8%)	397 (36.4%)
Yes	35 (71.4%)	658 (63.1%)	693 (63.5%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Psychiatric hospitalization (lifetime)</b>			
No	48 (98.0%)	1025 (98.4%)	1073 (98.4%)
Yes	1 (2.0%)	15 (1.4%)	16 (1.5%)
Missing	0 (0%)	2 (0.2%)	2 (0.2%)
<b>Stress problems (self-report)</b>			
No	30 (61.2%)	877 (84.2%)	907 (83.1%)
Yes	19 (38.8%)	163 (15.6%)	182 (16.7%)
Missing	0 (0%)	2 (0.2%)	2 (0.2%)
<b>Psychiatric medication use (last 30 days)</b>			
No	43 (87.8%)	996 (95.6%)	1039 (95.2%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
Yes	6 (12.2%)	38 (3.6%)	44 (4.0%)
Missing	0 (0%)	8 (0.8%)	8 (0.7%)
<b>Maternal problems</b>			
<b>stress</b>			
No	30 (61.2%)	751 (72.1%)	781 (71.6%)
Yes	19 (38.8%)	290 (27.8%)	309 (28.3%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Paternal problems</b>			
<b>stress</b>			
No	42 (85.7%)	966 (92.7%)	1008 (92.4%)
Yes	7 (14.3%)	75 (7.2%)	82 (7.5%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Sibling problems</b>			
<b>stress</b>			
No	42 (85.7%)	940 (90.2%)	982 (90.0%)
Yes	7 (14.3%)	100 (9.6%)	107 (9.8%)
Missing	0 (0%)	2 (0.2%)	2 (0.2%)
<b>Grandparents problems</b>			
<b>stress</b>			
No	43 (87.8%)	931 (89.3%)	974 (89.3%)
Yes	6 (12.2%)	110 (10.6%)	116 (10.6%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Partner attempt</b>			
<b>suicide</b>			

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
No	20 (40.8%)	310 (29.8%)	330 (30.2%)
Yes	0 (0%)	8 (0.8%)	8 (0.7%)
Missing	29 (59.2%)	724 (69.5%)	753 (69.0%)
<b>Offspring attempt</b>			
<b>suicide</b>			
No	19 (38.8%)	287 (27.5%)	306 (28.0%)
Yes	0 (0%)	0 (0%)	0 (0%)
Missing	30 (61.2%)	755 (72.5%)	785 (72.0%)
<b>Paternal attempt</b>			
<b>suicide</b>			
No	23 (46.9%)	338 (32.4%)	361 (33.1%)
Yes	0 (0%)	5 (0.5%)	5 (0.5%)
Missing	26 (53.1%)	699 (67.1%)	725 (66.5%)
<b>Maternal attempt</b>			
<b>suicide</b>			
No	21 (42.9%)	334 (32.1%)	355 (32.5%)
Yes	2 (4.1%)	16 (1.5%)	18 (1.6%)
Missing	26 (53.1%)	692 (66.4%)	718 (65.8%)
<b>Sibling attempt</b>			
<b>suicide</b>			
No	19 (38.8%)	315 (30.2%)	334 (30.6%)
Yes	3 (6.1%)	29 (2.8%)	32 (2.9%)
Missing	27 (55.1%)	698 (67.0%)	725 (66.5%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
<b>Friend attempt suicide</b>			
No	13 (26.5%)	155 (14.9%)	168 (15.4%)
Yes	10 (20.4%)	201 (19.3%)	211 (19.3%)
Missing	26 (53.1%)	686 (65.8%)	712 (65.3%)
<b>Lifetime intercourse sexual</b>			
No	3 (6.1%)	93 (8.9%)	96 (8.8%)
Yes	45 (91.8%)	942 (90.4%)	987 (90.5%)
Missing	1 (2.0%)	7 (0.7%)	8 (0.7%)
<b>Has a significant disease (self-report)</b>			
No	39 (79.6%)	883 (84.7%)	922 (84.5%)
Yes	10 (20.4%)	157 (15.1%)	167 (15.3%)
Missing	0 (0%)	2 (0.2%)	2 (0.2%)
<b>Has ever visited a psychiatrist or psychologist</b>			
No	35 (71.4%)	790 (75.8%)	825 (75.6%)
Yes	14 (28.6%)	249 (23.9%)	263 (24.1%)
Missing	0 (0%)	3 (0.3%)	3 (0.3%)
<b>Has ever been treated by a psychiatrist or psychologist</b>			
No	34 (69.4%)	817 (78.4%)	851 (78.0%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
Yes	8 (16.3%)	125 (12.0%)	133 (12.2%)
Missing	7 (14.3%)	100 (9.6%)	107 (9.8%)
<b>Knows someone who tried suicide (suicide attempt in social circle)</b>			
No	23 (46.9%)	669 (64.2%)	692 (63.4%)
Yes	26 (53.1%)	372 (35.7%)	398 (36.5%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Age that first used drugs</b>			
Mean (SD)	14.9 (2.59)	14.6 (2.34)	14.6 (2.35)
Median [Min, Max]	14.0 [10.0, 22.0]	15.0 [1.00, 22.0]	15.0 [1.00, 22.0]
Missing	8 (16.3%)	131 (12.6%)	139 (12.7%)
<b>Age of first sexual intercourse</b>			
Mean (SD)	15.9 (1.75)	15.8 (1.93)	15.8 (1.92)
Median [Min, Max]	16.0 [12.0, 19.0]	16.0 [10.0, 24.0]	16.0 [10.0, 24.0]
Missing	4 (8.2%)	108 (10.4%)	112 (10.3%)
<b>Sexual intercourse in the last week (sex frequency)</b>			
No	21 (42.9%)	318 (30.5%)	339 (31.1%)
Yes	24 (49.0%)	622 (59.7%)	646 (59.2%)
Missing	4 (8.2%)	102 (9.8%)	106 (9.7%)



Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
<b>Condom use during last sexual intercourse</b>			
No	19 (38.8%)	386 (37.0%)	405 (37.1%)
Yes	27 (55.1%)	558 (53.6%)	585 (53.6%)
Missing	3 (6.1%)	98 (9.4%)	101 (9.3%)
<b>Alcohol use before the last sexual intercourse</b>			
No	39 (79.6%)	782 (75.0%)	821 (75.3%)
Yes	7 (14.3%)	162 (15.5%)	169 (15.5%)
Missing	3 (6.1%)	98 (9.4%)	101 (9.3%)
<b>Had a partner at baseline</b>			
No	13 (26.5%)	247 (23.7%)	260 (23.8%)
Yes	33 (67.3%)	696 (66.8%)	729 (66.8%)
Missing	3 (6.1%)	99 (9.5%)	102 (9.3%)
<b>Number of people you had sex with in the last year</b>			
Mean (SD)	2.55 (3.59)	2.21 (2.93)	2.22 (2.96)
Median [Min, Max]	1.00 [1.00, 20.0]	1.00 [1.00, 28.0]	1.00 [1.00, 28.0]
Missing	5 (10.2%)	133 (12.8%)	138 (12.6%)
<b>Have you ever been or made someone pregnant?</b>			
No	29 (59.2%)	667 (64.0%)	696 (63.8%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
Yes	15 (30.6%)	266 (25.5%)	281 (25.8%)
Missing	5 (10.2%)	109 (10.5%)	114 (10.4%)
<b>Experienced forced sex</b>			
No	17 (34.7%)	649 (62.3%)	666 (61.0%)
Yes	2 (4.1%)	18 (1.7%)	20 (1.8%)
Missing	30 (61.2%)	375 (36.0%)	405 (37.1%)
<b>Anorexia nervosa at baseline</b>			
No	49 (100%)	1040 (99.8%)	1089 (99.8%)
Yes	0 (0%)	1 (0.1%)	1 (0.1%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Antisocial personality disorder at baseline</b>			
No	48 (98.0%)	1032 (99.0%)	1080 (99.0%)
Yes	1 (2.0%)	8 (0.8%)	9 (0.8%)
Missing	0 (0%)	2 (0.2%)	2 (0.2%)
<b>Do you have access to psychotherapy (public or private healthcare?)</b>			
Public	2 (4.1%)	34 (3.3%)	36 (3.3%)
Health insurance	1 (2.0%)	29 (2.8%)	30 (2.7%)
Private	3 (6.1%)	42 (4.0%)	45 (4.1%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
Other sources	1 (2.0%)	15 (1.4%)	16 (1.5%)
Missing	42 (85.7%)	922 (88.5%)	964 (88.4%)
<b>Melancholic depressive episode at baseline</b>			
No	35 (71.4%)	1005 (96.4%)	1040 (95.3%)
Yes	14 (28.6%)	36 (3.5%)	50 (4.6%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Past depressive episode (prior to baseline)</b>			
No	37 (75.5%)	1001 (96.1%)	1038 (95.1%)
Yes	12 (24.5%)	40 (3.8%)	52 (4.8%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)

Abbreviations: BD (Bipolar disorder).

**Table S2.** Beck Depression Inventory (BDI-II) items measured at baseline. Stratified by outcome at the 5-year follow-up.

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
<b>Item 1: Sadness</b>			
0	25 (51.0%)	839 (80.5%)	864 (79.2%)
1	16 (32.7%)	182 (17.5%)	198 (18.1%)
2	7 (14.3%)	16 (1.5%)	23 (2.1%)
3	1 (2.0%)	4 (0.4%)	5 (0.5%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 2: Pessimism</b>			
0	30 (61.2%)	870 (83.5%)	900 (82.5%)
1	17 (34.7%)	130 (12.5%)	147 (13.5%)
2	1 (2.0%)	22 (2.1%)	23 (2.1%)
3	1 (2.0%)	19 (1.8%)	20 (1.8%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 3: Past failure</b>			
0	34 (69.4%)	938 (90.0%)	972 (89.1%)
1	8 (16.3%)	65 (6.2%)	73 (6.7%)
2	4 (8.2%)	33 (3.2%)	37 (3.4%)
3	3 (6.1%)	5 (0.5%)	8 (0.7%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 4: Loss of pleasure</b>			

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
0	28 (57.1%)	846 (81.2%)	874 (80.1%)
1	15 (30.6%)	164 (15.7%)	179 (16.4%)
2	2 (4.1%)	13 (1.2%)	15 (1.4%)
3	4 (8.2%)	18 (1.7%)	22 (2.0%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 5: Guilty feelings</b>			
0	35 (71.4%)	911 (87.4%)	946 (86.7%)
1	7 (14.3%)	90 (8.6%)	97 (8.9%)
2	4 (8.2%)	27 (2.6%)	31 (2.8%)
3	3 (6.1%)	13 (1.2%)	16 (1.5%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 6: Punishment feelings</b>			
0	38 (77.6%)	902 (86.6%)	940 (86.2%)
1	6 (12.2%)	95 (9.1%)	101 (9.3%)
2	0 (0%)	14 (1.3%)	14 (1.3%)
3	5 (10.2%)	30 (2.9%)	35 (3.2%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 7: Self-dislike</b>			
0	32 (65.3%)	889 (85.3%)	921 (84.4%)
1	14 (28.6%)	134 (12.9%)	148 (13.6%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
2	3 (6.1%)	15 (1.4%)	18 (1.6%)
3	0 (0%)	3 (0.3%)	3 (0.3%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 8: Self-criticalness</b>			
0	26 (53.1%)	690 (66.2%)	716 (65.6%)
1	15 (30.6%)	277 (26.6%)	292 (26.8%)
2	7 (14.3%)	61 (5.9%)	68 (6.2%)
3	1 (2.0%)	13 (1.2%)	14 (1.3%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 9: Suicidal thoughts or wishes</b>			
0	44 (89.8%)	992 (95.2%)	1036 (95.0%)
1	4 (8.2%)	43 (4.1%)	47 (4.3%)
2	1 (2.0%)	3 (0.3%)	4 (0.4%)
3	0 (0%)	3 (0.3%)	3 (0.3%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 10: Crying</b>			
0	26 (53.1%)	892 (85.6%)	918 (84.1%)
1	15 (30.6%)	104 (10.0%)	119 (10.9%)
2	2 (4.1%)	9 (0.9%)	11 (1.0%)
3	6 (12.2%)	36 (3.5%)	42 (3.8%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 11: Agitation</b>			
0	24 (49.0%)	649 (62.3%)	673 (61.7%)
1	14 (28.6%)	233 (22.4%)	247 (22.6%)
2	5 (10.2%)	19 (1.8%)	24 (2.2%)
3	6 (12.2%)	140 (13.4%)	146 (13.4%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 12: Loss of interest</b>			
0	34 (69.4%)	877 (84.2%)	911 (83.5%)
1	14 (28.6%)	136 (13.1%)	150 (13.7%)
2	1 (2.0%)	23 (2.2%)	24 (2.2%)
3	0 (0%)	5 (0.5%)	5 (0.5%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 13: Indecisiveness</b>			
0	23 (46.9%)	775 (74.4%)	798 (73.1%)
1	13 (26.5%)	170 (16.3%)	183 (16.8%)
2	12 (24.5%)	86 (8.3%)	98 (9.0%)
3	1 (2.0%)	10 (1.0%)	11 (1.0%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 14: Worthlessness</b>			

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
0	32 (65.3%)	886 (85.0%)	918 (84.1%)
1	7 (14.3%)	87 (8.3%)	94 (8.6%)
2	7 (14.3%)	36 (3.5%)	43 (3.9%)
3	3 (6.1%)	32 (3.1%)	35 (3.2%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 15: Loss of energy</b>			
0	34 (69.4%)	876 (84.1%)	910 (83.4%)
1	10 (20.4%)	133 (12.8%)	143 (13.1%)
2	5 (10.2%)	24 (2.3%)	29 (2.7%)
3	0 (0%)	8 (0.8%)	8 (0.7%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 16: Changes in sleeping pattern</b>			
0	25 (51.0%)	724 (69.5%)	749 (68.7%)
1	16 (32.7%)	279 (26.8%)	295 (27.0%)
2	7 (14.3%)	32 (3.1%)	39 (3.6%)
3	1 (2.0%)	6 (0.6%)	7 (0.6%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 17: Irritability</b>			
0	20 (40.8%)	752 (72.2%)	772 (70.8%)
1	25 (51.0%)	258 (24.8%)	283 (25.9%)



Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
2	2 (4.1%)	28 (2.7%)	30 (2.7%)
3	2 (4.1%)	3 (0.3%)	5 (0.5%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 18: Changes in appetite</b>			
0	31 (63.3%)	874 (83.9%)	905 (83.0%)
1	12 (24.5%)	137 (13.1%)	149 (13.7%)
2	4 (8.2%)	22 (2.1%)	26 (2.4%)
3	2 (4.1%)	8 (0.8%)	10 (0.9%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 19: Concentration difficulty</b>			
0	33 (67.3%)	884 (84.8%)	917 (84.1%)
1	10 (20.4%)	113 (10.8%)	123 (11.3%)
2	3 (6.1%)	29 (2.8%)	32 (2.9%)
3	3 (6.1%)	15 (1.4%)	18 (1.6%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 20: Tiredness or fatigue</b>			
0	29 (59.2%)	829 (79.6%)	858 (78.6%)
1	19 (38.8%)	180 (17.3%)	199 (18.2%)
2	1 (2.0%)	24 (2.3%)	25 (2.3%)
3	0 (0%)	8 (0.8%)	8 (0.7%)

---

<b>Feature</b>	<b>With BD (n = 49)</b>	<b>Without BD (n = 1,042)</b>	<b>Overall (n = 1,091)</b>
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 21: Loss of interest in sex</b>			
0	40 (81.6%)	915 (87.8%)	955 (87.5%)
1	7 (14.3%)	94 (9.0%)	101 (9.3%)
2	2 (4.1%)	21 (2.0%)	23 (2.1%)
3	0 (0%)	11 (1.1%)	11 (1.0%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)

---

Abbreviations: BD (Bipolar disorder).

**Table S3.** Beck for Suicide Ideation Scale items measured at baseline. Stratified by outcome at the 5-year follow-up.

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
<b>Item 1: Wish to live</b>			
0	45 (91.8%)	1011 (97.0%)	1056 (96.8%)
1	3 (6.1%)	27 (2.6%)	30 (2.7%)
2	1 (2.0%)	0 (0%)	1 (0.1%)
Missing	0 (0%)	4 (0.4%)	4 (0.4%)
<b>Item 2: Wish to die</b>			
0	44 (89.8%)	1016 (97.5%)	1060 (97.2%)
1	3 (6.1%)	18 (1.7%)	21 (1.9%)
2	2 (4.1%)	6 (0.6%)	8 (0.7%)
Missing	0 (0%)	2 (0.2%)	2 (0.2%)
<b>Item 3: Reasons for living</b>			
0	45 (91.8%)	1006 (96.5%)	1051 (96.3%)
1	3 (6.1%)	25 (2.4%)	28 (2.6%)
2	0 (0%)	8 (0.8%)	8 (0.7%)
Missing	1 (2.0%)	3 (0.3%)	4 (0.4%)
<b>Item 4: Active attempt</b>			
0	46 (93.9%)	1027 (98.6%)	1073 (98.4%)
1	2 (4.1%)	13 (1.2%)	15 (1.4%)
2	1 (2.0%)	1 (0.1%)	2 (0.2%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 5: Passive attempt</b>			
0	46 (93.9%)	1024 (98.3%)	1070 (98.1%)
1	2 (4.1%)	12 (1.2%)	14 (1.3%)
2	1 (2.0%)	5 (0.5%)	6 (0.6%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 6: Duration of thoughts</b>			
0	47 (95.9%)	1037 (99.5%)	1084 (99.4%)
1	1 (2.0%)	4 (0.4%)	5 (0.5%)
2	1 (2.0%)	0 (0%)	1 (0.1%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 7: Frequency of ideation</b>			
0	49 (100%)	1038 (99.6%)	1087 (99.6%)
1	0 (0%)	3 (0.3%)	3 (0.3%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 8: Attitude toward ideation</b>			
0	46 (93.9%)	1027 (98.6%)	1073 (98.4%)
1	3 (6.1%)	11 (1.1%)	14 (1.3%)
2	0 (0%)	3 (0.3%)	3 (0.3%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 9: Control over action</b>			
0	49 (100%)	1037 (99.5%)	1086 (99.5%)
1	0 (0%)	3 (0.3%)	3 (0.3%)
2	0 (0%)	1 (0.1%)	1 (0.1%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 10: Deterrents to attempt</b>			
0	48 (98.0%)	1033 (99.1%)	1081 (99.1%)
1	1 (2.0%)	3 (0.3%)	4 (0.4%)
2	0 (0%)	5 (0.5%)	5 (0.5%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 11: Reasons for attempt</b>			
0	47 (95.9%)	1022 (98.1%)	1069 (98.0%)
1	1 (2.0%)	6 (0.6%)	7 (0.6%)
2	1 (2.0%)	13 (1.2%)	14 (1.3%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 12: Specificity of planning</b>			
0	48 (98.0%)	1038 (99.6%)	1086 (99.5%)
1	1 (2.0%)	3 (0.3%)	4 (0.4%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 13: Availability/opportunity</b>			
0	49 (100%)	1038 (99.6%)	1087 (99.6%)
1	0 (0%)	2 (0.2%)	2 (0.2%)
2	0 (0%)	1 (0.1%)	1 (0.1%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 14: Capability</b>			
0	46 (93.9%)	1034 (99.2%)	1080 (99.0%)
1	2 (4.1%)	6 (0.6%)	8 (0.7%)
2	1 (2.0%)	1 (0.1%)	2 (0.2%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 15: Expectancy</b>			
0	47 (95.9%)	1034 (99.2%)	1081 (99.1%)
1	2 (4.1%)	6 (0.6%)	8 (0.7%)
2	0 (0%)	1 (0.1%)	1 (0.1%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 16: Actual preparation</b>			
0	49 (100%)	1040 (99.8%)	1089 (99.8%)
1	0 (0%)	1 (0.1%)	1 (0.1%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)

Feature	With BD (n = 49)	Without BD (n = 1,042)	Overall (n = 1,091)
<b>Item 17: Suicidal note</b>			
0	49 (100%)	1041 (99.9%)	1090 (99.9%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 18: Final acts</b>			
0	49 (100%)	1040 (99.8%)	1089 (99.8%)
1	0 (0%)	1 (0.1%)	1 (0.1%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 19: Deception</b>			
0	48 (98.0%)	1028 (98.7%)	1076 (98.6%)
1	1 (2.0%)	7 (0.7%)	8 (0.7%)
2	0 (0%)	6 (0.6%)	6 (0.6%)
Missing	0 (0%)	1 (0.1%)	1 (0.1%)
<b>Item 20: Suicide attempt</b>			
0	44 (89.8%)	1017 (97.6%)	1061 (97.3%)
1	3 (6.1%)	20 (1.9%)	23 (2.1%)
2	2 (4.1%)	2 (0.2%)	4 (0.4%)
Missing	0 (0%)	3 (0.3%)	3 (0.3%)
<b>Item 21: Wish to die at last suicide attempt</b>			
0	0 (0%)	14 (1.3%)	14 (1.3%)
1	1 (2.0%)	6 (0.6%)	7 (0.6%)

---

<b>Feature</b>	<b>With BD (n = 49)</b>	<b>Without BD (n = 1,042)</b>	<b>Overall (n = 1,091)</b>
2	4 (8.2%)	9 (0.9%)	13 (1.2%)
Missing	44 (89.8%)	1013 (97.2%)	1057 (96.9%)

---

Abbreviations: BD (Bipolar disorder).



**Table S4.** Hyperparameter set chosen for the final model based on the highest area under the receiver operating characteristic curve in cross-validation routine.

Hyperparameter	Value on final model
<i>mtry</i>	9
<i>trees</i>	1429
<i>min_n</i>	10
<i>tree_depth</i>	14
<i>learn_rate</i>	0.0017
<i>loss_reduction</i>	0.6120
<i>sample_size</i>	0.9075

**Table S5.** Performance metrics from both *XGBoost* models applied on the testing set using a 0.5 threshold for positive classification. Main model ROC AUC: 0.786. Model without mood-related features ROC AUC: 0.644.

Performance metrics	Main model	Model without mood-related features
Sensitivity (recall)	0.375	0.312
Specificity	0.920	0.811
PPV	0.194	0.078
NPV	0.966	0.958
Balanced accuracy	0.647	0.562
Accuracy	0.893	0.786
F1-score	0.255	0.125

Abbreviations: PPV (Positive predictive value or precision). NPV (Negative predictive value).

**Supporting Information references**

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. American Psychiatric Press; 1994.
2. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders*. World Health Organization; 1993.
3. Amorim P. Mini International Neuropsychiatric Interview (MINI): validação de entrevista breve para diagnóstico de transtornos mentais. *Braz J Psychiatry*. 2000;22:106-115. doi:10.1590/S1516-44462000000300003
4. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *ArXiv E-Prints*. Published online March 2016. doi:10.1145/2939672.2939785
5. Kuhn M, Vaughan D. *Parsnip: A Common API to Modeling and Analysis Functions*.; 2022. <https://CRAN.R-project.org/package=parsnip>
6. Kuhn M, Wickham H. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*.; 2020. <https://www.tidymodels.org>
7. Silge J, Chow F, Kuhn M, Wickham H. *Rsample: General Resampling Infrastructure*.; 2022. <https://CRAN.R-project.org/package=rsample>
8. Kuhn M, Wickham H. *Recipes: Preprocessing and Feature Engineering Steps for Modeling*.; 2022. <https://CRAN.R-project.org/package=recipes>
9. Vaughan D. *Workflows: Modeling Workflows*.; 2022. <https://CRAN.R-project.org/package=workflows>
10. Kuhn M, Vaughan D. *Yardstick: Tidy Characterizations of Model Performance*.; 2022. <https://CRAN.R-project.org/package=yardstick>
11. Santner TJ, Williams BJ, Notz WI. *The Design and Analysis of Computer Experiments*. Springer; 2003. <https://link.springer.com/book/10.1007/978-1-4757-3799-8>
12. Librenza-Garcia D, Suh JS, Watts DP, et al. Structural and Functional Brain Correlates of Neuroprogression in Bipolar Disorder. *Curr Top Behav Neurosci*. 2021;48(197-213.); doi:10.1007/7854\_2020\_177
13. Hvitfeldt E. *Themis: Extra Recipes Steps for Dealing with Unbalanced Data*.; 2022. <https://CRAN.R-project.org/package=themis>
14. Lunardon N, Menardi G, Torelli N. ROSE: a Package for Binary Imbalanced Learning. *R J*. 2014;6(1):79-89. doi:10.32614/RJ-2014-008
15. Kuhn M. Futility Analysis in the Cross-Validation of Machine Learning Models. *ArXiv E-Prints*. Published online May 2014. doi:10.48550/arXiv.1405.6974
16. Kuhn M. *Finetune: Additional Functions for Model Tuning*.; 2022. <https://CRAN.R-project.org/package=finetune>
17. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27(8):861-874. doi:10.1016/j.patrec.2005.10.010
18. Yerushalmy J. Statistical Problems in Assessing Methods of Medical Diagnosis, with Special Reference to X-Ray Techniques. *Public Health Rep 1896-1970*. 1947;62(40):1432. doi:10.2307/4586294
19. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978;8(4):283-298. doi:10.1016/s0001-2998(78)80014-2
20. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat*. 2001;29(5):1189-1232. doi:10.1214/aos/1013203451
21. Liu Y, Just A. *SHAPforxgboost: SHAP Plots for "XGBoost."*; 2021.

- <https://CRAN.R-project.org/package=SHAPforxgboost>
22. Mayer M. *Shapviz: SHAP Visualizations.*; 2023.  
<https://CRAN.R-project.org/package=shapviz>
  23. Greenwell BM. pdp: An R Package for Constructing Partial Dependence Plots. *R J.* 2017;9(1):421-436.
  24. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York; 2016. <https://ggplot2.tidyverse.org>
  25. Pedersen TL. *Patchwork: The Composer of Plots.*; 2020.  
<https://CRAN.R-project.org/package=patchwork>
  26. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing; 2022. <https://www.R-project.org/>
  27. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning.* 2nd ed. Springer; 2021.
  28. Shwartz-Ziv R, Armon A. Tabular Data: Deep Learning is Not All You Need. Published online 2021. doi:10.48550/ARXIV.2106.03253

# APÊNDICE F – Figuras do corpo do artigo

Figure 1 – Machine learning pipeline flowchart. The figure shows data splitting, preprocessing routine, feature selection, cross-validation, model fitting, model assessment and feature importance steps using xgboost algorithm.

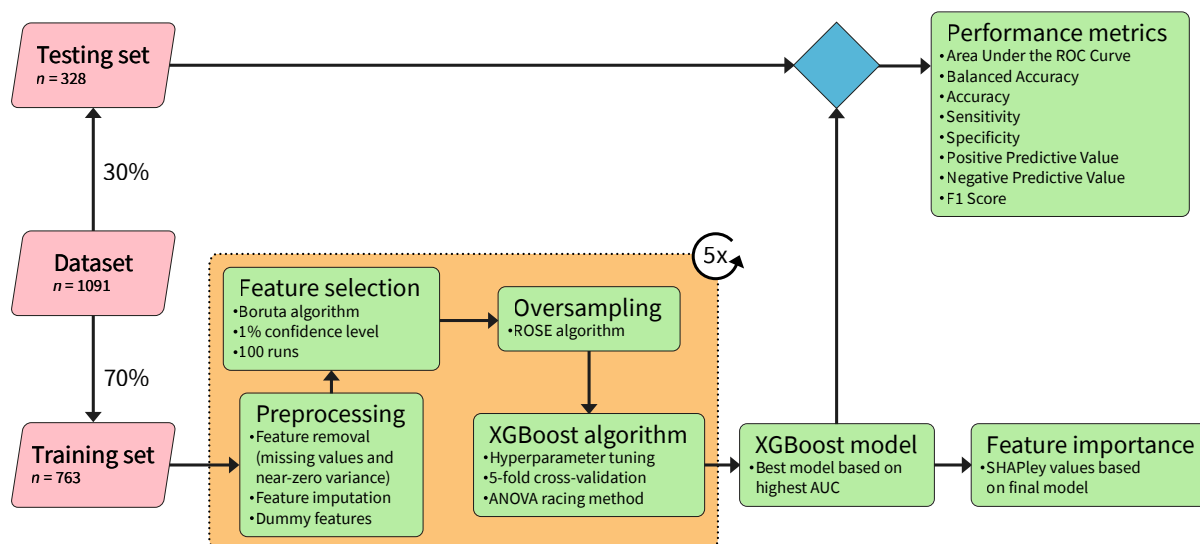


Figure 2 – SHAPley values for each feature included in the final model. Each *y*-axis tick represents a feature, sorted by the highest absolute contribution across all observations, regardless of the direction of the association. Each dot represents a participant in the training set (*n* = 763). Observations with SHAPley values lower than zero behaved as protective factors, otherwise they were risk factors. The fill color represents the value of the variable for a given individual (red corresponds to higher values, and blue corresponds to lower values). Categorical features are coded as 0/1. The bold text next to each feature name represents the average absolute contribution (irrespective of direction).

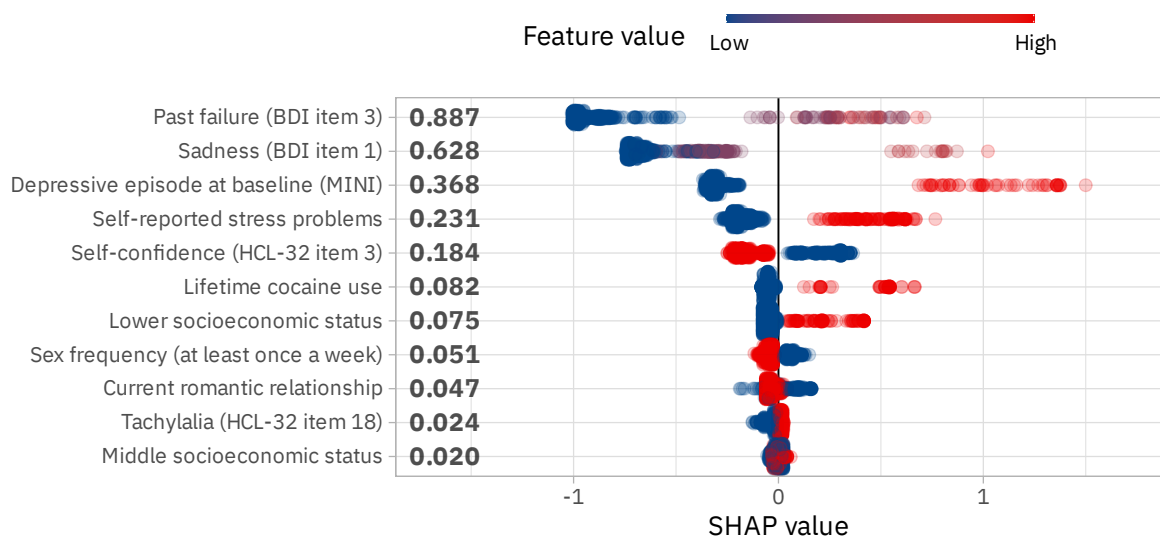


Figure 3 – Feature importance plot based on the model with removal of mood-related features (test ROC AUC: 0.644). Maternal mental health issues and self-reported stress problems were the two most important variables.

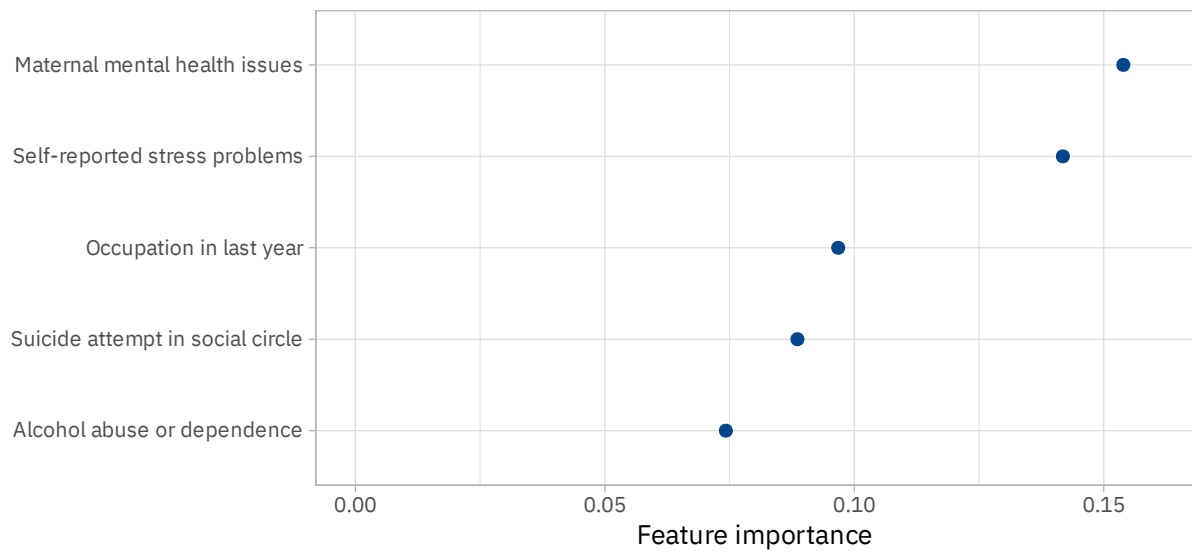


Figure 4 – Linear associations of depression severity score (BDI total score) at baseline with a) cognitive performance (MoCA score) and b) functional impairment (FAST score) five years later. The fitted regression lines show a negative correlation of depressive symptoms severity with cognitive performance ( $r = -0.18$ ,  $n = 959$ ), and a positive correlation with functional impairment ( $r = 0.24$ ,  $n = 982$ ). Subjects with current mood episodes at follow-up were excluded for the analysis ( $n = 107$ ), in order to minimize the direct effect of mood symptoms on these measures, resulting in a subsample of 984 individuals. Abbreviations: MoCA (Montreal Cognitive Assessment). FAST (Functioning Assessment Short Test).

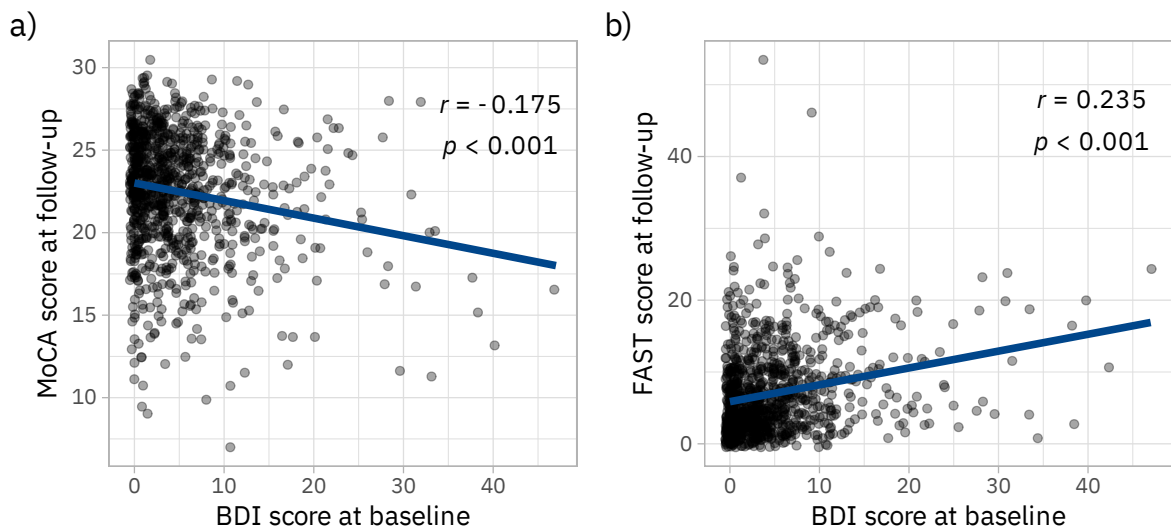


Figure 5 – Self-reported age at first depressive episode among subjects with and without suicide risk (based on MINI c module on suicide risk). The plot shows a trend that subjects with suicide risk present earlier depression onset ( $n = 31$ ). Eighty-five subjects presented a depressive episode at baseline and 54 presented missing values at depression age at onset variable, resulting in a final subsample of 31 individuals. The Student's  $t$ -test was used and suggests a significant difference between the groups (mean in group No = 21.76, mean in group Yes = 18.50,  $\Delta = 3.26$ , 95% CI [0.32, 6.21],  $t(20.20) = 2.31$ ,  $p = 0.032$ ; Cohen's  $d = 1.03$ , 95% CI [0.09, 1.94]). Abbreviations: MINI (Mini International Neuropsychiatric Interview).

