

An Interpretable Classification Model for Identifying Individuals with Attention Deficit Hyperactivity Disorder

N. Ventura¹, P. Loures¹, C. Nicola¹, D. M. Oliveira¹, M. Romano^{1,2}, D. M. Miranda^{1,2}, A. C. Silva¹, G. L. Pappa¹, W. Meira Jr¹

¹ Universidade Federal de Minas Gerais, Brazil

(nataventura, ploures, camilanicola, derickmath, ana.coutosilva, glpappa, meira)@dcc.ufmg.br

² Centro de Tecnologia em Medicina Molecular (CTMM) da UFMG, Minas Gerais, Brazil.

(romanosilva, debora.m.miranda)@gmail.com

Abstract. Attention Deficit Hyperactivity Disorder (ADHD) is a psychiatric condition that affects around 5% of children around the world. The primary attention procedure is traditionally based on analysis of ratings collected in questionnaires called psychometrics. This work aims to investigate interpretable classification models capable of not only accurately identifying individuals with ADHD, but also explain it, by providing the evidences that lead to the outcome. We compare the performance of Explainable Boosting Machine (EBM) with 3 other classical decision tree-based models and observed similar results, with the distinction of EBM being a more interpretable model. We also assess explanations quantitative and qualitatively, demonstrating how they may actually help psychiatrists in their practice.

CCS Concepts: • **Applied computing** → **Health care information systems**; • **Computing methodologies** → **Artificial intelligence**.

Keywords: machine learning, interpretability, healthcare

1. INTRODUÇÃO

O Transtorno de Deficit de Atenção com Hiperatividade é um transtorno do neurodesenvolvimento que afeta cerca de 5% das crianças e 2,5% dos adultos [Association et al. 2014]. Os principais sintomas são: desatenção, hiperatividade e impulsividade, e serão diagnósticos de transtorno quando o conjunto dos sintomas interferir na funcionalidade em mais de um ambiente por mais de 6 meses. Os portadores de TDAH cursam com piores desempenhos acadêmicos e profissionais, além de possuírem mais conflitos pessoais [Association et al. 2014]. O tratamento ajuda no controle dos sintomas e tem um impacto positivo na vida dos pacientes [Arnold et al. 2020]. Desta forma, é individualmente importante promover o acesso ao diagnóstico para toda a população afetada. Na realidade brasileira, estima-se que menos de 20% dos brasileiros com TDAH recebem tratamento adequado [Mattos et al. 2012].

No final de 2021, foi aprovada a lei N^o 14.254 que estabelece que o poder público deve desenvolver programas de identificação e acompanhamento integral para educandos com TDAH. Dessa forma, houve um súbito aumento da demanda de saúde mental no sistema público. A triagem para identificar indivíduos com TDAH pode utilizar classificadores que auxiliem profissionais generalistas a tomar a decisão de encaminhamento ao especialista ou iniciar tratamento com maior assertividade. Este trabalho se insere nesse contexto, ou seja, a automação do processo de triagem de indivíduos com TDAH.

Os autores agradecem pela ajuda no financiamento desse trabalho a FAPEMIG, CNPq e CAPES. Esse trabalho também foi parcialmente financiado pelo MASWeb, EUBra-BIGSEA, INCT-Cyber, ATMOSPHERE, CIIA-Saúde, AWS e pelo programa Google Latin America Research Awards.

Copyright©2020 Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

A tarefa de triagem é basicamente a classificação desses indivíduos que potencialmente tenham TDAH, o que não apenas deve ser acurado, entre outras propriedades típicas de um classificador, como também interpretável, o que é uma tendência clara na área [Lipton 2017; 2018]. A interpretabilidade é o grau em que um modelo pode ser entendido em termos humanos [Gilpin et al. 2018]. A explicabilidade, facilita o entendimento dos profissionais da saúde, que podem averiguar o diagnóstico dado pelo modelo e evitar que um diagnóstico seja dado por uma razão espúria. Um exemplo de diagnósticos errados por razões espúrias foi relatado no âmbito da identificação de melanomas, onde se constatou que o modelo tinha maiores chances de classificar uma pessoa como portadora de melanoma caso as fotos tivessem marcações de caneta cirúrgica [Winkler et al. 2019]. Dessa forma, modelos interpretáveis são importantes tanto para seu uso efetivo por profissionais, quanto para evitar erros.

Apesar de crucial, a interpretabilidade ainda é um grande desafio técnico incluindo um compromisso entre as dimensões de qualidade do modelo. Os modelos de Aprendizado de Máquina podem ser divididos entre opacos e transparentes. Nos modelos opacos não sabemos a razão pela qual o modelo chegou em uma determinada decisão, o que dificulta a interpretabilidade [Molnar 2020]. Enquanto os modelos transparentes são facilmente explicáveis, mas é comum que apresentem uma acurácia menor quando comparados a modelos opacos [Molnar 2020]. Deste modo, é importante investigar o balanço entre acurácia e interpretabilidade, o que motivou este trabalho a investigar a aplicação de um classificador transparente e criação de metáfora visual para a interpretação no âmbito da triagem de pacientes com TDAH, de modo a tornar mais acessível a identificação de pacientes com o transtorno.

O objetivo deste trabalho é desenvolver um classificador acurado capaz de direcionar crianças e adolescentes com TDAH para um tratamento especializado ou assertivo. Para isso utilizamos uma base de dados do laboratório de Medicina Molecular da Universidade Federal de Minas Gerais (UFMG), com dados de questionários de crianças com suspeita de TDAH.

2. TRIAGEM DE INDIVÍDUOS COM TDAH: DADOS E CARACTERIZAÇÃO

Como mencionado, o ponto de partida do nosso trabalho é uma base de dados proveniente do Projeto de Pesquisa “Avaliação multidimensional de indivíduos com transtorno de déficit de atenção e hiperatividade” aprovado pelo Comitê de Ética da Universidade Federal de Minas Gerais (CAE: 02899412.9.0000.5149). Essa base consiste em dados tabulares relativos a 345 crianças e adolescentes testados por psiquiatras a partir do método de diagnóstico *Kids Schedule for Affective Disorders and Schizophrenia for School Aged Children Present and Lifetime Version (K-SADS-PL)* [Brasil and Bordin 2010]. Um método considerado como padrão ouro para diagnóstico de TDAH em que psiquiatras avaliam a presença de sintomas de TDAH determinados pelo Diagnostic and Statistical Manual of Mental Disorders (DSM-V). Esses dados foram coletadas entre os anos de 2011 e 2020 e conta com dados provenientes de 23 diferentes tipo de instrumentos de coleta sobre o paciente, seu contexto familiar e socio-econômico. A Tabela I apresenta uma breve caracterização da base de dados.

Pacientes	Atributos	Prevalência TDAH	Dados Ausentes	Faixa Etária	Período Coleta
345	916	78,2%	40,9%	3 a 25 anos	2011 a 2020

Table I. Caracterização da base de dados.

A base, contudo, apresenta uma série de peculiaridades, discutidas a seguir. Como é apresentado na tabela, devido ao extenso tempo de coleta, a base conta com uma quantidade significativa de dados ausentes distribuídos de maneira irregular entre os pacientes, o que torna mais difícil a observação de padrões entre os pacientes. A base também utiliza diferentes tipagens de dados de maneira indiscriminada, ou seja, dados numéricos, ordinais, categóricos e de texto livre que são utilizados de forma não padronizada. Outras duas características relevantes são a utilização de diferentes escalas para avaliar seus atributos ordinais, uma vez que a base conta com diferentes psicometrias que avaliam o paciente

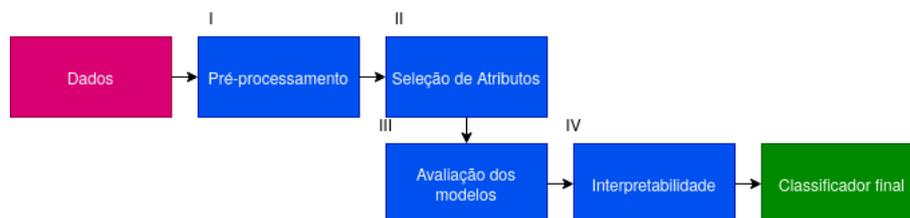


Fig. 1. Fluxograma dos passos realizados para construção do modelo.

a partir de diferentes fontes e perspectivas, assim como a presença de dados sensíveis que poderiam enviesar o classificador e relacionados às somas ou resultados de certos testes psicométricos. Como será discutido ao longo do trabalho, a interpretabilidade é peça fundamental para entender o impacto dessas e outras características da base.

3. METODOLOGIA

Conforme mostrado na Figura 1, a metodologia adotada neste trabalho pode ser dividida em quatro etapas. Nas 2 primeiras etapas, preparamos e selecionamos os atributos relevantes para a avaliação. A terceira etapa é a de classificação, onde particionamos a base de dados em dois conjuntos, um conjunto de treino e um conjunto de teste. Os modelos primeiramente realizam o aprendizado a partir do conjunto de treino, o qual possui respostas de pacientes e seus respectivos laudos dados por um psiquiatra. A partir dos resultados da classificação decidimos qual modelo a ser utilizado. Finalmente, na última etapa é construída a interpretação do modelo bem como sua avaliação. As quatro etapas são discutidas em mais detalhe a seguir.

3.1 Pré-processamento

Como apresentado na descrição dos dados, a base utilizada apresenta as seguintes características que demandam tratamento: (a) dados ausentes, (b) tipagens variada de dados, (c) diferentes escalas de psicometria, e (d) dados sensíveis. O impacto de cada característica foi minimizado como descrito a seguir:

- a Homogeneizamos as diferentes tipagens de dados por meio da padronização dos valores nulos nas tabelas. Utilizando a forma binária, a presença ou ausência de cada categoria, elimina a confusão entre dados categóricos e ordinais. Finalmente, os dados categoricos foram tratados a partir da utilização de One Hot Encoding e foram desconsiderados os dados de texto livre.
- b Os dados ausentes binários foram preenchidos com zero. O valor 0 foi utilizado por dois motivos: para evitar a introdução de valores negativos em operações multiplicativas e por ser um valor nulo em operações aditivas. Ou seja, buscamos um valor que minimize a distorção dos modelos. Esses valores também serão afetados pelos ajustes de escala psicométricas.
- c As diferentes escalas utilizadas por cada teste psicométrico foram “binarizadas”, de modo que a presença de qualquer característica foi classificada como verdadeiro (1) e a ausência como falso (0). Em seguida, foi somado 1 a esses valores (verdadeiro = 2 ausente = 1) para realizar o ajuste mencionado no tratamento de (a). A binarização das escalas buscou facilitar a interpretação e a tradução dos textos de diferentes psicometrias que atribuem diferentes valores textuais para denotar o grau que cada sintoma é percebido. Esse método de codificação foi comparado à utilização da escala original e não impactou a validade do instrumento, corroborando com resultados de estudos anteriores [Simms et al. 2019; Capik and Gozum 2015].
- d Foram excluídos os atributos considerados sensíveis ou discriminatórios (p.ex., o gênero do informante), assim como atributos que representam um cálculo feito com base em mais de uma

resposta de um teste psicométrico. Uma vez que o objetivo do trabalho depende da utilização e da seleção de questões individuais de diferentes psicometrias, não faz sentido utilizar cálculos baseados em questões não utilizadas.

Finalmente, foram eliminados os resultados do K-SADS-PL, uma vez que esse teste é voltado para o diagnóstico do TDAH a partir de testes aplicados por especialistas, os quais poderiam enviesar o resultado do trabalho.

3.2 Seleção de atributos

Após o pré-processamento dos dados, foi necessário selecionar os atributos a serem utilizados para a identificação do quadro clínico de TDAH no paciente. Para isso foi utilizada a correlação de Spearman [Spearman 1904] para avaliar quais atributos são mais correlacionados à presença de TDAH. Finalmente, foram selecionados os atributos que apresentem correlação maior que 1 desvio padrão da média, resultando em 216 atributos, a partir de 916 atributos iniciais.

3.3 Classificação

Essa seção discute a seleção dos modelos que serão utilizados para avaliar o modelo de classificação, em particular o EBM (3.3.1) e as métricas de desempenho que serão utilizadas para comparar esses modelos (3.3.2).

3.3.1 Modelos Selecionados. A tarefa de classificação consiste em prever, a partir da resposta aos formulários, se um indivíduo apresenta TDAH. Para esse fim, algoritmos de classificação baseados em árvore de decisão têm se mostrado superiores aos demais por se tratarem de dados tabulares, em especial para questionários de avaliação de TDAH [Gorishniy et al. 2021]. Assim, como o nosso foco é na interpretabilidade dos classificadores, utilizamos algoritmos baseados em árvore de decisão.

Dentre os modelos baseados em árvore de decisão, os modelos *Random Forest* (RF) [Cutler et al. 2012] e *Extreme Gradient Boosting* (XGBoost) [Chen et al. 2015] foram comparados ao modelo auto-interpretável chamado *Explainable Boosting Machine* (EBM) [Nori et al. 2019]. Também avaliamos o desempenho do modelo *Support Vector Machine* (SVM) [Cortes and Vapnik 1995] como um baseline clássico de classificação. Para todos os modelos foi realizado um *Random Search* [Andradóttir 2006] para otimização dos hiperparâmetros com k-fold cross-validation, conforme descrito em [Berrar 2018]. A métrica otimizada foi F1 e para cada modelo foram realizadas 500 iterações, 10 separações e 3 repetições.

3.3.2 Métricas de desempenho. Após a otimização dos hiperparâmetros, testamos os modelos. A base de dados foi particionada em um conjunto de treino e um conjunto de teste. O conjunto de treino possui 221 pontos, enquanto o conjunto de teste possui 114 pontos. Após a etapa de treino, foi realizado a etapa de teste. Nessa etapa, as métricas avaliadas foram: Precisão, revocação, especificidade, acurácia e F1-score.

Para cada modelo foi realizado um bootstrap, com 100 iterações, para obter o intervalo de confiança de 95%. Utilizamos esse intervalo de confiança para comparar se o desempenho dos modelos de classificação avaliados são significativamente distintos. Caso exista uma interseção entre os intervalos de confiança dos modelos, então poderemos escolher aquele que possui melhor interpretabilidade. Senão, terá que ser feita uma melhor análise do *trade-off* entre acurácia e interpretabilidade para a escolha do modelo.

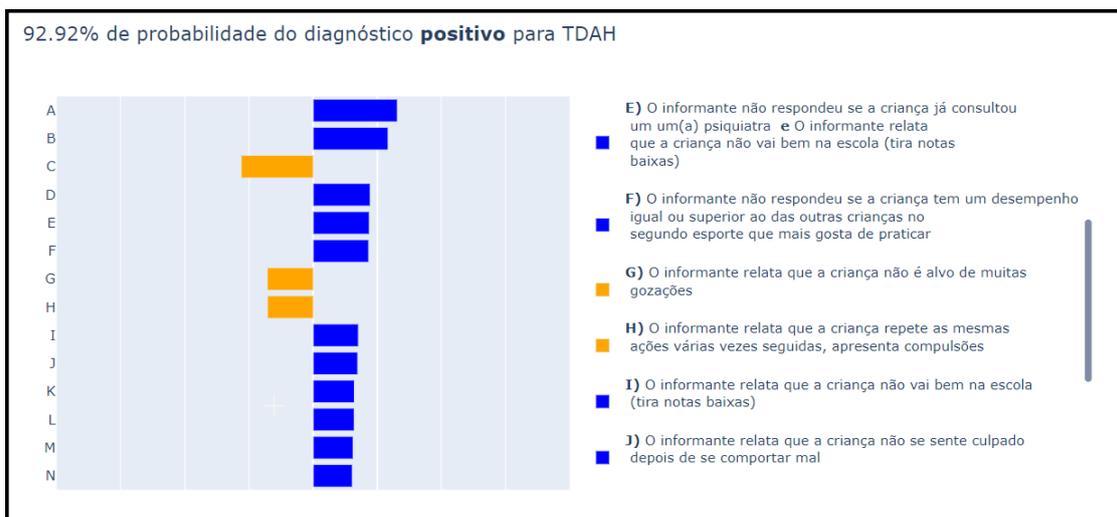


Fig. 2. Apresentação do modelo de interpretação. (Corte da imagem original para formatação do artigo.)

3.4 Interpretação

Para verificar se a interpretação do modelo atende os requisitos necessários, usaremos as métricas de estabilidade e fidelidade, além da análise por parte de especialistas. Em termos de desempenho, as diferenças entre os modelos não foram estatisticamente significativas. Modelos opacos são complexos e necessitam de um método adicional para gerar as explicações, que podem ser inconsistentes entre si [Ribeiro et al. 2016], afetando, portanto, a sua estabilidade. O EBM, por ser um modelo transparente, é mais estável e fidedigno e, por essa razão, foi o modelo escolhido.

Finalmente, foi proposta e implementada uma nova visualização para o modelo de interpretação, a qual é mais acessível ao público não especializado em computação. Essa visualização conta com um gráfico interativo que destaca o texto relacionado a cada um dos atributos psicométricos utilizados.

3.5 Métricas de interpretabilidade

Para a avaliação da interpretabilidade dos modelos, destacamos a métrica de estabilidade, que mensura o quanto as explicações diferem para pequenas alterações no conjunto de dados [Robnik-Šikonja and Bohanec 2018]. Para mensurar essa métrica, foram realizados 1500 testes para cada uma das 114 instâncias de teste da base de dados. Em cada uma dessas instâncias um dos atributos é aleatoriamente escolhido e é alterado para outro valor possível, entre verdadeiro, falso ou nulo. Após essa alteração, é gerada uma explicação local para cada caso e, assim, se compara o atributo mais importante nos dados originais com o atributo mais relevante para o conjunto gerado, esta comparação é usada para computar a métrica de estabilidade.

4. RESULTADOS EXPERIMENTAIS

Nessa seção apresentamos os resultados experimentais deste trabalho, incluindo a parametrização, apresentada na Tabela II, a análise dos modelos, cujos resultados estão descritos na Tabela III, assim como os resultados de interpretabilidade, descritos na subseção 4.1.

Em geral, a métrica de acurácia é um melhor indicador em casos que temos dados bem distribuídos. Como a base está mal distribuída, a métrica F1 é melhor para indicação da qualidade do modelo.

Deste modo, vemos que, no intervalo de confiança de 95%, não há diferença estatisticamente signifi-

	Min Sample Leaf		Max Leaves		Interactions	
EBM	5		3		3	
	Subsample	Estimators	Max Depth	Learning Rate	Col bytree	Col bylevel
XGBoost	0,7	500	6	0,01	0,4	0,4
	Bootstrap	Min Sample Leaf	Estimators	Max Depth	Min Samples Split	
RF	false	2	1400	50	2	
	Kernel		Gamma		C	
SVM	RBF		0,05		2	

Table II. Valor dos hiperparâmetros otimizados de cada modelo.

Modelo	F1	Sensibilidade	Especificidade	Precisão	Acurácia
EBM	0,81 - 0,92 (0,87)	0,94 - 1,00 (0,98)	0,03 - 0,29 (0,14)	0,69 - 0,85 (0,78)	0,69 - 0,85 (0,77)
XGBoost	0,82 - 0,91 (0,87)	0,94 - 1,00 (0,98)	0,05 - 0,32 (0,17)	0,71 - 0,86 (0,79)	0,70 - 0,84 (0,78)
RF	0,83 - 0,92 (0,88)	1,00 - 1,00 (1,00)	0,03 - 0,27 (0,14)	0,70 - 0,85 (0,78)	0,71 - 0,86 (0,79)
SVM	0,82 - 0,92 (0,87)	0,97 - 1,00 (0,99)	0,03 - 0,29 (0,14)	0,71 - 0,85 (0,78)	0,71 - 0,85 (0,78)

Table III. Resultado da classificação dos algoritmos. Para todas as métricas é dado o limite inferior e superior em um intervalo de confiança de 95%. Entre parênteses temos o valor médio das métricas.

ficativa entre os modelos quanto à métrica F1. Como mencionado, o modelo EBM foi escolhido por ser um modelo auto explicável.

4.1 Resultados da interpretabilidade

Conforme descrito na metodologia, para avaliar a interpretação foi calculada a métrica de estabilidade da interpretação e também foi feita a análise com especialista.

A estabilidade da interpretabilidade foi de 98,5%, mostrando que, para pequenas mudanças nos dados de entrada, a interpretabilidade do modelo é bem estável. Ademais, as interpretações se mostraram relevantes quando apresentadas para especialistas, ressaltando atributos que tradicionalmente não são utilizados como critério para diagnósticos, tendo se mostrado relevantes tanto como ferramenta de auxílio no processo de triagem, quanto na auditoria da classificação. Em um dos exemplos em que o especialista concordou com a classificação da máquina, foi possível observar que o modelo descreve como indicio de TDAH que “*o informante relata que a criança tem dificuldade de brincar ou envolver-se em atividades de forma calma*”, que é um atributo utilizado por ferramentas psicométricas que avaliam o TDAH. Todavia o modelo também capturou atributos latentes, como “*o informante relata que a criança tem o desempenho inferior ao das outras crianças na terceira atividade extracurricular que ela faz (violão, pipa, ler, jogar video-game)*” e a considerou negativa para o diagnóstico, destacando que a informação da criança praticar ao menos três atividades extracurriculares é relevante para o contexto do exame, de acordo com o especialista.

Por outro lado, em um exame em que o profissional discordou do diagnóstico da máquina, foi possível capturar a utilização de diversos atributos considerados por ele irrelevantes, como “*O informante não respondeu se a criança tem alergia a comidas*” e utilizou diversas cruzamentos com a informação “*o informante relata que a criança não vai bem na escola (tira notas baixas)*” indicando que a máquina estava atribuindo um peso desproporcional a essa característica. Destarte, os resultados da avaliação da interpretação por especialistas corroboram com o nosso objetivo de tornar a utilização do classificador mais segura, compreensível e auditável.

5. TRABALHOS RELACIONADOS

O estudo de técnicas de inteligência artificial vem sendo amplamente investigado como ferramenta de auxílio em diagnósticos médicos, sendo utilizadas inclusive na psiquiatria [Tamilarsan Ramasamy

2022; 2022; Itani et al. 2019]. Nosso foco na utilização de um modelo interpretável para fazer a classificação se baseia na demanda por modelos auditáveis na área médica [Lipton 2017; Oliveira et al. 2020].

A utilização do aprendizado de máquina na detecção de TDAH nos artigos estudados segue seis caminhos distintos [Loh et al. 2022] dos quais três serão destacados nessa seção: (i) a utilização de neurofisiologia e neuroimagem [Itani et al. 2019], (ii) a análise de aspectos fisiológicos do paciente (eletroencefalogramas, eletrocardiogramas, sensores de movimento etc.) [Slobodin et al. 2020] e, finalmente, (iii) o estudo de questionários e escalas de auto/heteroavaliação do paciente [Rosi et al. 2022; da Silva Jandre 2021], este sendo o mais fiel à metodologia de diagnóstico de déficit estipulada pelo DSM-V [Association et al. 2014].

Modelos computacionais interpretáveis por profissionais de saúde facilitam ao médico escrutinar o diagnóstico para um paciente específico, tendo em mente todas as particularidades que podem passar despercebidas pelo modelo. Dito isso, os dois primeiros caminhos (i.e., (i) análises baseadas em neuroimagem e (ii) análises baseadas em neurofisiologia tendem a utilizar uma combinação de modelos de aprendizado de máquina e *deep learning* (DL). Este último, opaco, utiliza modelos de interpretação post-hoc, ou seja, interpretações posteriores ao treinamento, mas limitadas quanto à importância dos atributos contextuais apresentados para o especialista [Lipton 2017; 2018]. O modelo adotado nesse trabalho, em contrapartida a (i) e (ii), se baseia primariamente em dados tabulares coletados por psicométrias elaboradas a partir do DSM-V.

Dessa forma, os modelos baseados nos dados psicométricos, tendem a utilizar modelos otimizados para a classificação de dados tabulares, como o XGBoost e árvores de decisão [Gorishniy et al. 2021; Loh et al. 2022]. Todavia, ainda que a estrutura básica destes modelos seja interpretável, uma limitação comum é combinar a estes modelos que não são interpretáveis, ou não apresentar uma avaliação adequada a respeito da utilização de modelos voltados para a interpretação do especialista.

6. CONCLUSÕES

O objetivo deste trabalho foi propor um classificador interpretável para diagnóstico de TDAH. O modelo foi aplicado em uma base de dados real de TDAH e conseguiu um bom resultado tanto em acurácia (entre outras medidas de desempenho) como em interpretabilidade, dado em conta o viés existente nos dados. Esse resultado demonstra que, para esse tipo de tarefa, os modelos interpretáveis são uma alternativa possível. A interpretabilidade é um importante fator para aplicabilidade desses modelos em contextos ecológicos de saúde e educação, condicionado a uma auditoria do modelo por parte dos profissionais de saúde e minimização de associações espúrias. Ademais, durante o processo de tratamento e elaboração ficou evidente a importância da introdução do profissional médico no meio de pesquisa computacional, tanto pela perspectiva da compreensão do conteúdo abordado nas bases de dados, quanto para elaboração de interpretações acessíveis que podem ser avaliadas por médicos. Em contrapartida, é importante também que o profissional de computação esteja inserido no meio médico para que a produção de dados seja otimizada para os modelos que serão construídos a partir deles.

Deste modo, o próximo passo desse projeto é aplicar a avaliação do modelo junto a um grupo de especialistas que possa avaliar o quão relevante é essa interpretação do modelo em seus contextos de atuação e o quanto os padrões reconhecidos pelo modelo se aproximam daqueles utilizados pelos psiquiatras.

REFERENCES

- ANDRADÓTTIR, S. Chapter 20 an overview of simulation optimization via random search. In *Simulation*, S. G. Henderson and B. L. Nelson (Eds.). Handbooks in Operations Research and Management Science, vol. 13. Elsevier, pp. 617–631, 2006.

- ARNOLD, L. E., HODGKINS, P., KAHLE, J., MADHOO, M., AND KEWLEY, G. Long-term outcomes of adhd: academic achievement and performance. *Journal of attention disorders* 24 (1): 73–85, 2020.
- ASSOCIATION, A. P. ET AL. *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*. Artmed Editora, 2014.
- BERRAR, D. Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology* vol. 1, pp. 542–545, 2018.
- BRASIL, H. H. AND BORDIN, I. A. Convergent validity of k-sads-pl by comparison with cbcl in a portuguese speaking outpatient population. *BMC psychiatry* 10 (1): 1–11, 2010.
- CAPIK, C. AND GOZUM, S. Psychometric features of an assessment instrument with likert and dichotomous response formats. *Public Health Nursing* 32 (1): 81–86, 2015.
- CHEN, T., HE, T., BENESTY, M., KHOTILOVICH, V., TANG, Y., CHO, H., CHEN, K., ET AL. Xgboost: extreme gradient boosting. *R package version 0.4-2* 1 (4): 1–4, 2015.
- CORTES, C. AND VAPNIK, V. Support-vector networks. *Machine learning* 20 (3): 273–297, 1995.
- CUTLER, A., CUTLER, D. R., AND STEVENS, J. R. Random forests. In *Ensemble machine learning*. Springer, pp. 157–175, 2012.
- DA SILVA JANDRE, C. R. O uso de personas e aprendizado de máquina na análise do perfil de crianças e adolescentes com transtorno de déficit de atenção/hiperatividade, 2021.
- GILPIN, L. H., BAU, D., YUAN, B. Z., BAJWA, A., SPECTER, M., AND KAGAL, L. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*. IEEE, pp. 80–89, 2018.
- GORISHNIY, Y., RUBACHEV, I., KHRULKOV, V., AND BABENKO, A. Revisiting deep learning models for tabular data. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan (Eds.). Vol. 34. Curran Associates, Inc., pp. 18932–18943, 2021.
- ITANI, S., ROSSIGNOL, M., LECRON, F., AND FORTEMPS, P. Towards interpretable machine learning models for diagnosis aid: a case study on attention deficit/hyperactivity disorder. *PloS one* 14 (4): e0215720, 2019.
- LIPTON, Z. C. The doctor just won't accept that! *arXiv preprint arXiv:1711.08037*, 2017.
- LIPTON, Z. C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16 (3): 31–57, 2018.
- LOH, H. W., OOI, C. P., BARUA, P. D., PALMER, E. E., MOLINARI, F., AND ACHARYA, U. R. Automated detection of adhd: Current trends and future perspective. *Computers in Biology and Medicine* vol. 146, pp. 105525, 2022.
- MATTOS, P., ROHDE, L. A., AND POLANCZYK, G. V. Adhd is undertreated in brazil. *Brazilian Journal of Psychiatry* vol. 34, pp. 513–514, 2012.
- MOLNAR, C. *Interpretable machine learning*. Lulu Press, Inc, 2020.
- NORI, H., JENKINS, S., KOCH, P., AND CARUANA, R. Interpretml: A unified framework for machine learning interpretability. *arXiv preprint arXiv:1909.09223*, 2019.
- OLIVEIRA, D. M., RIBEIRO, A. H., PEDROSA, J. A., PAIXÃO, G. M., RIBEIRO, A. L. P., AND MEIRA, W. Explaining end-to-end ecg automated diagnosis using contextual features. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 204–219, 2020.
- RIBEIRO, M. T., SINGH, S., AND GUESTRIN, C. Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*, 2016.
- ROBNIK-ŠIKONJA, M. AND BOHANEK, M. Perturbation-based explanations of prediction models. In *Human and machine learning*. Springer, pp. 159–175, 2018.
- ROSI, E., GRAZIOLI, S., CRIPPA, A., CANDELIERI, A., CECCARELLI, S. B., MAURI, M., MANZONI, M., MAURI, V., TRABATTONI, S., MOLteni, M., ET AL. Towards tediagnostic procedures in child neuropsychiatry: addressing adhd diagnosis and autism symptoms through supervised machine learning, 2022.
- SIMMS, L. J., ZELAZNY, K., WILLIAMS, T. F., AND BERNSTEIN, L. Does the number of response options matter? psychometric perspectives using personality questionnaire data. *Psychological assessment* 31 (4): 557, 2019.
- SLOBODIN, O., YAHAV, I., AND BERGER, I. A machine-based prediction model of adhd using cpt data. *Frontiers in Human Neuroscience* vol. 14, 2020.
- SPEARMAN, C. The proof and measurement of association between two things. *The American Journal of Psychology* 15 (1): 72–101, 1904.
- TAMILARASAN RAMASAMY, D. J. J. Early risk detection of depression from social media posts using hierarchical attention networks. *JOURNAL OF ALGEBRAIC STATISTICS* 13 (1): 483–489, 2022.
- WINKLER, J. K., FINK, C., TOBERER, F., ENK, A., DEINLEIN, T., HOFMANN-WELLENHOF, R., THOMAS, L., LALLAS, A., BLUM, A., STOLZ, W., ET AL. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *JAMA dermatology* 155 (10): 1135–1141, 2019.