



Máquina de vetores de suporte para predição de ansiedade em dependentes químicos em reabilitação

Support vector machine for predicting anxiety in chemical dependency rehabilitation patients

Máquina de vectores de soporte para la predicción de ansiedad en pacientes de rehabilitación de dependencia química

Pedro Elias Patente Freire ¹, Ana Clara Borges Silva ², Lucas Magalhaes Portilho Carrara ¹, Chrystian Araujo Pereira ³

1 Discente de Medicina, Departamento de Medicina, Universidade Federal de Lavras (UFLA), Lavras (MG), Brasil.

2 Mestre em Nutrição e Saúde, Doutoranda do Programa de Pós-Graduação em Plantas Medicinais, Aromáticas e Condimentares, Departamento de Agricultura, Universidade Federal de Lavras (UFLA), Lavras (MG), Brasil.

3 Doutorado em Agroquímica, Professor do Departamento de Medicina, Departamento de Medicina, Universidade Federal de Lavras (UFLA), Lavras (MG), Brasil.

Autor correspondente: Pedro Elias Patente Freire

E-mail: pedro.freire@estudante.ufla.br

Resumo

Objetivo: Relacionar variáveis clínicas de internos em reabilitação química com a ansiedade, através do método de aprendizado de máquina. **Método:** Estudo em campo, realizado em uma Comunidade Terapêutica, onde considerou-se dados de 25 internos. Dentre os parâmetros, têm-se as substâncias psicoativas de dependência, tempo de uso e abstinência, idade e o questionário GAD-7. O algoritmo utilizado foi a Máquina de Vetores de Suporte (MVS). As métricas de análise de desempenho foram matriz de confusão e AUC. **Resultados:** A prevalência de reabilitação em cocaína ou crack foi de 92% dos internos seguida de álcool em 76%. As maiores métricas foram acurácia de 68%, sensibilidade 89%, especificidade 88%, F1 score 59% e AUC de 0,91. **Conclusão:** O algoritmo MVS se mostrou promissor para ser utilizado na predição de ansiedade em internos em processo recuperação por substâncias psicoativas.

Descritores: Aprendizado de Máquina; Dependência Química; Ansiedade.



Abstract

Objective: Relate clinical variables of inpatients in chemical rehabilitation to anxiety, using the machine learning method. **Method:** A field study conducted in a Therapeutic Community, considering data from 25 inpatients. Among the parameters are the psychoactive substances of dependence, duration of use and abstinence, age, and the GAD-7 questionnaire. The algorithm used was the Support Vector Machine (SVM). The performance analysis metrics were the confusion matrix and AUC. **Results:** The prevalence of rehabilitation for cocaine or crack was 92% of the inpatients, followed by alcohol at 76%. The highest metrics were an accuracy of 68%, sensitivity of 89%, specificity of 88%, F1 score of 59%, and AUC of 0.91. **Conclusion:** The SVM algorithm proved to be promising for use in predicting anxiety in inpatients undergoing recovery from psychoactive substances.

Keywords: Machine Learning; Dependences, Chemical; Anxiety.

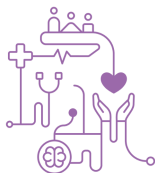
Resumen

Objetivo: Relacionar variables clínicas de internos en rehabilitación química con la ansiedad, mediante el método de aprendizaje automático. **Método:** Estudio de campo realizado en una Comunidad Terapéutica, considerando datos de 25 internos. Entre los parámetros se encuentran las sustancias psicoactivas de dependencia, tiempo de uso y abstinencia, edad y el cuestionario GAD-7. El algoritmo utilizado fue la Máquina de Vectores de Soporte (SVM). Las métricas de análisis de rendimiento fueron: matriz de confusión y el AUC. **Resultados:** La prevalencia de rehabilitación en cocaína o crack fue del 92% de los internos, seguida por alcohol en un 76%. Las métricas más altas fueron una precisión del 68%, sensibilidad del 89%, especificidad del 88%, puntaje F1 del 59% y un AUC de 0.91. **Conclusión:** El SVM demostró ser prometedor para su uso en la predicción de ansiedad en internos en proceso de recuperación de sustancias psicoactivas.

Descriptores: Aprendizaje Automático; Dependencia Química; Ansiedad.

Introdução

O uso de drogas e de substâncias psicoativas tornou-se um grande problema de saúde pública, que vem crescendo significativamente na sociedade, acometendo



pessoas de diversos gêneros e faixas etárias, afetando o convívio social e gerando impactos graves à saúde física e mental desses indivíduos. De acordo com o Ministério da Saúde (MS), no ano de 2021, o Sistema Único de Saúde (SUS) realizou um total de 400,3 mil atendimentos a indivíduos com transtornos mentais e comportamentais decorrentes do uso de substâncias psicoativas. A maioria dos pacientes eram do sexo masculino, com idades situadas predominantemente entre 25 e 29 anos ⁽¹⁾.

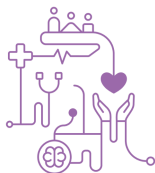
Desse modo, é comum observar uma forte associação entre o uso de drogas estimulantes ou depressivas e o desenvolvimento de ansiedade, apesar de que a relação entre cada tipo de transtorno de ansiedade desenvolvido e a substância em abuso não é bem estabelecida ⁽²⁾. Alguns estudos demonstram uma proporção de cerca de 17.7% no que se refere ao diagnóstico simultâneo de Transtorno por Uso de Substância (TUS) e desordens de ansiedade ⁽³⁾.

Para isso, ao longo dos últimos anos, foi possível observar um maior emprego de métodos de aprendizado de máquina (MAM) para relação entre variáveis clínicas e previsão de parâmetros que sejam úteis para o manejo do paciente ⁽⁴⁾.

Dentre esses algoritmos, é possível citar a Máquina de Vetores de Suporte (MVS) que têm demonstrado resultados promissores quando utilizada para prever condições clínicas a partir de dados de pacientes. Em particular, pesquisas recentes têm focado na utilização de valores clínicos como preditores e na aplicação do questionário Generalized Anxiety Disorder-7 (GAD-7) como uma métrica para avaliar a classe, ou seja, a presença e a gravidade da ansiedade. Esses estudos indicam que a MVS pode ser uma ferramenta eficaz na previsão precoce de níveis de ansiedade, permitindo intervenções mais rápidas e personalizadas para os pacientes ^(5,6). Portanto, esta pesquisa teve como objetivo relacionar variáveis clínicas de internos em reabilitação química com a ansiedade através do modelo preditivo MVS.

Métodos

Este trabalho consiste em um estudo de campo, realizado em uma Comunidade Terapêutica no município de Lavras, Minas Gerais, Brasil, abrangendo um espaço amostral de 25 internos. A pesquisa foi aprovada pelo Comitê de Ética



em Pesquisa com Seres Humanos (COEP) da Universidade Federal de Lavras (UFLA) sob o parecer 6803699/CAAE: 77207923.9.0000.5148. Os dados selecionados para análise foram as substâncias psicoativas (SPA) citadas como causa de dependência pelos indivíduos em reabilitação neuropsicossocial, tempo de uso da SPA, idade, tempo de abstinência e o score final do questionário GAD-7. Tais dados compuseram o *dataset* para aplicação do método de aprendizado de máquina (MAM). O modelo selecionado foi a Máquina de Vetores de Suporte (MVS), implementado na linguagem Python, através da biblioteca Scikit-learn.

Coleta dos dados e formação do dataset inicial

A coleta de dados para pesquisa foi realizada através de anamnese e da aplicação de um questionário a cada um dos internos, sendo as respostas tabuladas em um arquivo separado por vírgula (CSV). Para identificação das SPA mais utilizadas pelos indivíduos em reabilitação, foi perguntado sobre a dependência prévia à álcool, cocaína ou crack, tabaco e maconha. Os internos que relataram o uso de crack foram considerados no mesmo grupo de usuários de cocaína, devido ao fato de o crack ser derivado da pasta base, portanto, possuindo o mesmo princípio ativo ⁽⁷⁾. Para cada uma das afirmações, foi atribuído um valor binário. Já para o tempo de uso, foi solicitado o valor em anos do primeiro uso que levou a dependência àquela substância. Além disso, o tempo, em meses, de abstinência (último uso) e a idade, em anos.

Para avaliar o grau de ansiedade, foi utilizado o questionário GAD-7. Este instrumento é composto por sete perguntas que avaliam a gravidade dos sintomas de ansiedade experimentados nas últimas duas semanas. Cada item é avaliado em uma escala de 0 a 3, totalizando uma pontuação máxima de 21 pontos. As pontuações mais altas indicam maior gravidade dos sintomas de ansiedade. Uma pontuação de 10 ou mais no GAD-7 representa um ponto de corte razoável para identificar casos de ansiedade de grau moderado a grave ⁽⁸⁾. Portanto, para tal parâmetro, também foram atribuídos valores binários, onde internos que tiveram pontuação maior ou igual a 10 foram considerados com pontuação diferente de zero.

Pré-processamento dos dados e formação do dataset transformado



Devido à utilização de dados binários e valores não binários, foi necessário usar a técnica de padronização dos dados. Assim, foi aplicado o Standard Scaler removendo a média e o dimensionamento para a variância unitária da base de dados ⁽⁹⁾. Assim, formou-se o *dataset* transformado, de modo a seguir uma melhor proporção entre os dados.

Máquina de vetores de suporte

A Máquina de Vetores de Suporte (MVS) é uma ferramenta que propõe um algoritmo linear ou não-linear, sendo um MAM supervisionado. A principal vantagem do MVS é a sua capacidade de superar problemas de dimensionalidade e operar com sucesso, sendo, por isso, frequentemente aplicado para classificação de classes binárias ⁽¹⁰⁾. Para isso, determina um hiperplano que separa o espaço amostral em duas populações, ou classes, funcionando como um bom preditor para recursos multidimensionais ⁽¹¹⁾.

Seleção de hiperparâmetros

A seleção dos melhores hiperparâmetros na MVS tem impacto direto em seu desempenho e otimização ⁽¹²⁾. Portanto, sua configuração se faz necessária para melhor abordagem no que tange ao *dataset* transformado e suas limitações. Para nosso MAM, foram analisados diferentes hiperparâmetros de escolha do *kernel*, sendo eles sigmoide, linear e polinomial. Para cada *kernel*, obteve-se as métricas de análise de desempenho, além da matriz de confusão agregada. Assim, teve-se como objetivo demonstrar as diferentes abordagens do algoritmo MVS.

Atributos e Métricas de análise de desempenho

Dentre as métricas para avaliar o desempenho do nosso modelo, a validação cruzada, ou *cross-validation*, é uma abordagem segura. Para isso, o espaço amostral é dividido em subamostras que serão usados tanto para treinamento, quanto para teste. Isso permite que todos dados do dataset sejam usados, obtendo a taxa de erro estimada (isto é, o erro médio dessas subamostras) e a acurácia



média. A validação cruzada fornece uma estimativa melhor da taxa de erro do que treinar/testar ao custo de mais computação ⁽¹³⁾. Em nosso modelo, usamos o *cross-validation k-fold*, que consiste em dividir o espaço amostral em k subamostras de tamanhos iguais, treinando em k-1, testando em k e repetindo k vezes esse procedimento. Para nosso trabalho, utilizamos k = 5.

Além disso, foi gerada a matriz de confusão, que mostra a taxa de verdadeiros positivos, verdadeiros negativos, verdadeiros previstos e falsos previstos. Portanto, foi possível montar as seguintes métricas: acurácia, especificidade, sensibilidade e o escore F1 ^(14, 15). A matriz de confusão foi montada a partir da agregação de cada *k-fold*, sendo a soma dos valores individuais (soma das 5 matrizes criadas).

Por fim, foi definida a curva ROC (Receiver Operating Characteristic), que é uma medida bidimensional de desempenho de classificação, representando a taxa de verdadeiros positivos (sensibilidade) em função da taxa de falsos positivos (1 - especificidade). Portanto, pode-se interpretar a curva como uma comparação do desempenho do classificador em todo o *dataset* de distribuições de classes. Frequentemente, a principal métrica extraída da curva ROC é a AUC (Area Under Curve). Quando AUC é igual a 1, o classificador representa um modelo perfeito, já uma AUC de 0.5 diz respeito a um modelo de classificação aleatória. Um ponto de interesse da AUC é que ela demonstra um comportamento geral das formas de classificação, pois é independente do limite usado para obter um rótulo de classe, promovendo uma boa forma de comparação entre diferentes MAM ⁽¹⁶⁾.

Resultado e discussão

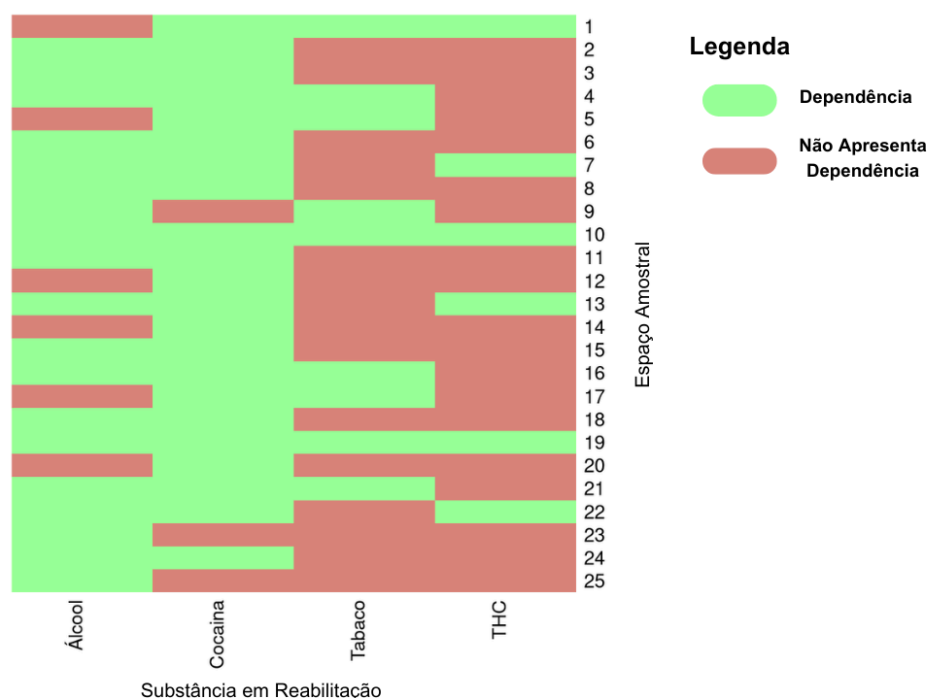
Análise da população

Para nosso espaço amostral, o primeiro resultado apresentado é sobre a distribuição das substâncias psicoativas citadas pelos indivíduos em reabilitação



(Figura 1). A prevalência de reabilitação em cocaína ou crack foi de 92% dos internos seguida de álcool em 76%. Já o tabaco seguiu em 36% e maconha (THC) 24%. O mais frequente co-uso de SPA foi o de cocaína ou crack e álcool, sendo relatado em 64% da população. Já o co-uso de tabaco juntamente à cocaína ou álcool foi de 24% da população.

Figura 1: Distribuição do uso de diferentes substâncias psicoativas.



A figura 1 representa a distribuição binária de cada indivíduo, na qual a área verde corresponde a reabilitação a SPA (valor binário positivo de 1), e a vermelha a ausência (valor binário de 0). Apenas 2 indivíduos apresentaram dependência às 4 substâncias analisadas, em contrapartida, 5 indivíduos estavam em reabilitação de apenas uma droga.

Dataset e aplicação do modelo

Em seguida, após a etapa de coleta dos dados, o *dataset* foi formado de acordo com a Tabela 1. Portanto, teve-se uma divisão entre os previsores (x) e classe (y). De modo que x teve 7 variáveis (valores binários para cada SPA (álcool, cocaína, tabaco e maconha), tempo de uso, tempo de abstinência e idade) e y apenas 1 classe (presença ou não de ansiedade: GAD-7 maior ou igual a 10).

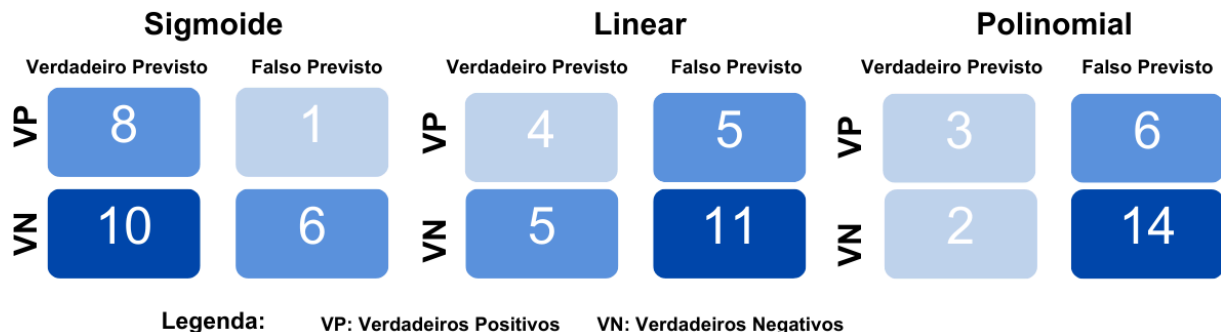


Tabela 1: Divisão de previsores e classe.

Previsores				Classe
Substâncias Psicoativas em Reabilitação (vício prévio)	Tempo de Uso	Tempo de Abstinência	Idade	SCORE GAD-7 ≥ 10
Valores binários para cada SPA	16.56 ± 6.88 anos	4.19 ± 4.79 meses	41.31 ± 10.77 anos	Valores binários

O limiar para considerar positivo o valor do score GAD-7 foi de 10 pontos. Isso aconteceu em 36% da população analisada. Após isso, o modelo de MVS foi aplicado sobre o banco de dados, tendo em consideração os hiperparâmetros de otimização. Como resultado, formaram-se as seguintes matrizes de confusão representada pela Figura 2, sendo consequência da soma de cada dobra durante o processo de validação cruzada para cada *kernel* individualmente.

Figura 2: Matriz de confusão agregada para cada *kernel*.



Na figura 2 temos que VN representa os verdadeiros negativos (True Negative) e VP os verdadeiros positivos (True Positive). A partir da interpretação numérica, podemos extrair as seguintes métricas para avaliação do modelo de MVS de acordo com a Tabela 2.

Tabela 2: Valores médios de acurácia, sensibilidade, especificidade e F1 para o modelo MVS em diferentes *kernels*.

Acurácia (<i>accuracy</i> , ACC)	Sensibilidade (<i>recall</i> , REC)	Especificidade (<i>specificity</i> , SPEC)	F1 (F1 <i>score</i>)	AUC
---	---	--	-----------------------	-----



Sigmoide c = 0.2 tol = 0.0001	56%	89%*	38%	59%*	0.76
Linear c = 10 tol = 0.0001	60%	44%	69%	44%	0.76
Polinomial c = 1 tol = 0.0001	68%*	33%	88%*	43%	0.91*

*Maiores valores obtidos para cada métrica.

Dentre as métricas usadas para avaliação do modelo, a curva ROC e a AUC são bons parâmetros de comparação de modelos. Para nossos resultados, a AUC foi extraída das diferentes “plotagens” ROC no gráfico.

Assim o modelo de Máquina de Vetores de Suporte (MVS) teve acurácia (ACC) de 68%, especificidade (SPEC) de 88% e AUC de 0.91 com os hiperparâmetros: regularização (c) de 1, tolerância para o critério de parada (tol) de 0.0001 no modelo *kernel* polinomial (observando-se os maiores valores obtidos). Já no modelo sigmoide, com os hiperparâmetros $c = 0.2$ e $tol = 0.0001$, obteve-se sensibilidade (REC) de 89% e F1 de 59%. As métricas mais baixas foram para o *kernel* linear, conforme mostra a Tabela 2.

A ACC é uma métrica frequentemente usada em aplicações de MAM na medicina, em que representa a razão entre as amostras corretamente classificadas e o número total de amostras no conjunto de dados de avaliação. Já a REC mostra a correta identificação de classes positivas (uma REC alta tem baixa influência de falsos positivos), e a SPEC das classes negativas (uma SPEC alta tem baixa influência de falsos negativos), sendo importantes mensuradores do impacto de valores falsos (positivos ou negativos) no desempenho do modelo. Já para a métrica F1, nosso modelo obteve como máxima 59%. Tal avaliação representa uma medida sob influência da precisão e sensibilidade, portanto, um alto valor de F1 traz uma boa classificação de instâncias positivas, minimizando os casos falsos positivos e negativos ⁽¹⁷⁾.

Um estudo feito por Tabares et al.⁽¹⁸⁾ com o objetivo de prever ansiedade em jovens utilizando MAM teve como previsores valores clínicos, dentre eles o uso de



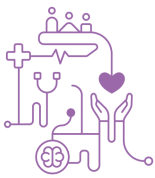
drogas, e sociodemográficos e como classe o questionário GAD-7 e dentre outros. Comparativamente, obtiveram para o modelo MVS uma ACC, REC e F1 score de 89%. De modo semelhante, Sau et al.⁽¹⁹⁾ compararam diferentes algoritmos para predição de ansiedade em jovens não dependentes, utilizando-se de outro questionário para avaliação do score. Como resultado, obtiveram para MVS, durante a validação cruzada, uma ACC de 76,1% e AUC 0.75. Outro trabalho realizado nessa população foi de Park et al.⁽²⁰⁾, em que desenvolveu-se um modelo para prever o risco de pacientes com dependência a álcool abandonarem o tratamento ambulatorial, tendo no algoritmo MVS uma ACC de 70%, REC 64%, SPEC 70% e AUC de 0.67.

Na literatura, não foram encontrados trabalhos que tivessem como previsores somente valores clínicos de internos em reabilitação química e a classe, o questionário GAD-7. Também foram escassos os trabalhos de *machine learning* que avaliaram ansiedade em indivíduos em reabilitação neuropsicossocial no período de abstinência. Destaca-se a importância de avaliar a dependência química a mais de uma substância psicoativa, já que frequentemente o multi-uso é observado como demonstrado na Figura 1, em que 80% dos internos encontram-se em recuperação de mais de uma SPA.

Aplicações terapêuticas para comunidade médica

Os transtornos por uso de substâncias (TUS) apresentam uma forte relação com o Transtorno de Ansiedade Generalizada (TAG)⁽²¹⁾. Pontuações acima de 10 pontos no questionário GAD-7 tem demonstrado um bom ponto de corte para identificar TAG⁽⁸⁾, apesar dos critérios do DSM-V serem pautados nos últimos 6 meses, e não nas últimas 2 semanas como apresenta-se no questionário⁽²¹⁾.

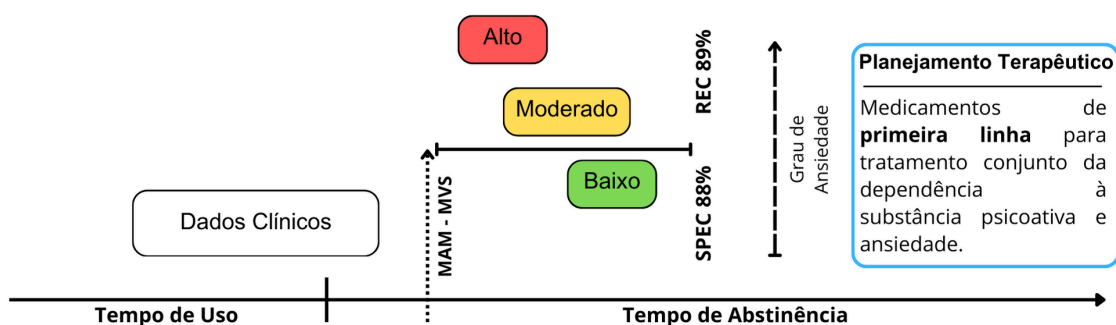
Para a associação entre desordens de ansiedade e abstinência alcoólica, o uso de antidepressivos inibidores seletivos de recaptção de serotonina (ISRS), como a paroxetina, apresentaram nível de evidência Ia. Já o uso de benzodiazepínicos foi recomendado em até 4 semanas, sob monitoramento clínico, também com evidência Ia. Medicamentos como gabapentina, pregabalina e antagonistas de opióides atuaram tanto na remissão da ansiedade como no processo de abstinência, apresentando evidência Ia^(21, 22, 23).



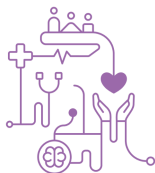
Em relação a abstinência à retirada de cocaína, o beta-bloqueador propranolol se mostrou promissor ao tratamento em pacientes em reabilitação. Durante o período de abstinência, esses indivíduos demonstram sensibilidade aos efeitos da adrenalina e noradrenalina, o que pode desencadear o surgimento de ansiedade. Portanto, o uso de betabloqueadores pode ser uma estratégia eficaz para mitigar a ansiedade relacionada à cocaína, além de aliviar alguns dos sintomas mais incômodos do desejo por essa substância ⁽²⁴⁾. De modo geral, os antidepressivos não apresentam efeitos consistentes em qualquer medida para tratamento, ao invés disso, tem sido relacionado a maiores taxas de abandono e efeitos adversos. Estudos demonstraram que os psicoestimulantes, agonistas da dopamina e antipsicóticos não tiveram efeitos positivos sobre o placebo, de modo análogo, os antipsicóticos apresentam resultados conflitantes ⁽²⁵⁾.

Assim, é comum observar a associação dos TUS com diferentes níveis de ansiedade. Em nosso modelo, é possível propor um esquema de aplicação do MAM-MVS para prática clínica (Figura 3).

Figura 3: Esquema para prática do modelo de previsão.



Legenda: A figura demonstra um modelo de aplicação do MAM-MVS (Modelo de Aprendizado de Máquina - Máquina de Vetores de Suporte) para a comunidade médica, tendo como alicerce o desenvolvimento de ansiedade ao longo do tempo de abstinência. Primeira linha, refere-se aos medicamentos com nível de evidência Ia. REC (sensibilidade), SPEC (especificidade).



O planejamento terapêutico envolve amplas abordagens, dentre elas, a farmacológica. A partir dos resultados que obtivemos em diferentes configurações de *kernel*, ressaltando as máximas de REC e SPEC, é possível ter uma visão ampliada da janela terapêutica. Assim, a aplicação de medicamentos de primeira linha, tendo em vista as métricas de desempenho estabelecidas, pode ser uma opção para implementação de uma terapêutica combinada. Assim, espera-se um impacto positivo da identificação precoce dos níveis de ansiedade no processo de reabilitação neuropsicossocial.

Limitações

Fundamentalmente, o estudo teve uma população restrita para o modelo, de modo que o *dataset* foi composto por dados específicos de usuários em processo de reabilitação.

Cabe ressaltar a importância de um grande espaço amostral para desenvolvimento de um MAM robusto. Paralelamente, o estudo de Tabares et al.⁽¹⁸⁾ contou com 234 jovens entre 18 e 28 anos, com mais de 30 parâmetros previsores. Já o de Park et al.⁽²⁰⁾ contou com 839 pacientes hospitalizados.

Conclusão

O desenvolvimento de ferramentas pautadas nos MAM torna-se evidente para auxílio na conduta médica, essencialmente, modelos que sirvam de alicerce de integração e confiabilidade.

Tendo em vista o *dataset* utilizado, e suas limitações, o algoritmo MVS se mostrou promissor para ser utilizado na predição de ansiedade em internos em processo de reabilitação química, considerando diferentes variáveis, tendo potencial para auxílio na conduta terapêutica.

Referências

1. Atendimento a pessoas com transtornos mentais por uso de álcool e drogas aumenta 12% no SUS [Internet]. Ministério da Saúde. 2022.
2. de Matos MB, de Mola CL, Trettim JP, Jansen K, da Silva RA, Souza LD de M, et al. Psychoactive substance abuse and dependence and its association with anxiety



- disorders: a population-based study of young adults in Brazil. *Revista Brasileira de Psiquiatria*. 2018 Feb 15;40(4):349–53.
3. Soraya S, Mahdavi M, Saeidi M, Seddigh R, Nooraeen S, Sadri M, et al. Prevalence of anxiety disorders and its co-occurrence with substance use disorder: a clinical study. *Middle East Current Psychiatry*. 2022 Apr 20;29(1).
 4. Chhetri B, Goyal LM, Mittal M. How machine learning is used to study addiction in digital healthcare: A systematic review. *International Journal of Information Management Data Insights*. 2023 Nov;3(2):100175.
 5. Albagmi, Faisal Mashel, et al. Prediction of generalized anxiety levels during the Covid-19 pandemic: A machine learning-based modeling approach. *Informatics in Medicine Unlocked*. 28 (2022): 100854.
 6. Pintelas EG, Kotsilieris T, Livieris IE, Pintelas P. A review of machine learning prediction methods for anxiety disorders. *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*. 2018 Jun 20;
 7. Castro RA de, Ruas RN, Abreu RC, Rocha RB, Ferreira R de F, Lasmar RC, et al. Crack: pharmacokinetics, pharmacodynamics, and clinical and toxic effects. *Revista Médica de Minas Gerais [Internet]*. 2015;25(2). Available from: <http://rmmg.org/exportar-pdf/1782/v25n2a17.pdf>
 8. Spitzer R, Kroenke K, Williams J, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7 [Internet]. *Archives of internal medicine*. 2006. Available from: <https://pubmed.ncbi.nlm.nih.gov/16717171/>
 9. de Amorim LBV, Cavalcanti GDC, Cruz RMO. The choice of scaling technique matters for classification performance. *Applied Soft Computing [Internet]*. 2023 Jan;133:109924. Available from: <https://arxiv.org/pdf/2212.12343>
 10. Brereton RG, Lloyd GR. Support Vector Machines for classification and regression. *The Analyst*. 2010;135(2):230–67.
 11. Li Y, Cui Z, Liao Q, Dong H, Zhang J, Shen W, et al. Support vector machine-based multivariate pattern classification of methamphetamine dependence using arterial spin labeling. *Addiction Biology*. 2019 Jan 9;24(6):1254–62.
 12. Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*. 2020 Nov;415:295–316.
 13. King RD, Orhobor OI, Taylor CC. Cross-validation is safe to use. *Nature Machine Intelligence [Internet]*. 2021 Apr 1;3(4):276–6. Available from: <https://www.nature.com/articles/s42256-021-00332-z>
 14. Rocha J de L, Salles EOT, Andreão RV. Detecção da Apneia Obstrutiva do Sono Através da Variabilidade da Frequência Cardíaca. *J Health Inform [Internet]*. 20º de julho de 2023 [citado 27º de maio de 2024];15(Especial). Disponível em: <https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/1084>



15. Yates LA, Aandahl Z, Richards SA, Brook BW. Cross validation for model selection: a review with examples from ecology. *Ecological Monographs*. 2022 Nov 13;93(1).
16. Marzban C. The ROC Curve and the Area under It as Performance Measures. *Weather and Forecasting*. 2004 Dec;19(6):1106–14.
17. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports* [Internet]. 2022 Apr 8;12(1):5979. Available from: <https://www.nature.com/articles/s41598-022-09954-8>
18. Tabares T, Vélez Álvarez, Consuelo, Salcedo B, Murillo Rendón, Santiago. Anxiety in Young People: Analysis from a Machine Learning Model. 2024 Jan 1 [cited 2024 May 27]; Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4791415
19. Sau A, Bhakta I. Screening of anxiety and depression among seafarers using machine learning technology. *Informatics in Medicine Unlocked*. 2019;16:100228.
20. Park SJ, Lee SJ, Kim H, Kim JK, Chun JW, Lee SJ, et al. Machine learning prediction of dropping out of outpatients with alcohol use disorders. Le KNQ, editor. *PLOS ONE*. 2021 Aug 2;16(8):e0255626.
21. Back SE, Brady KT. Anxiety Disorders with Comorbid Substance Use Disorders: Diagnostic and Treatment Considerations. *Psychiatric Annals*. 2008 Nov 1;38(11):724–9.
22. Smith JP, Book SW. Anxiety and Substance Use Disorders: A Review. *The Psychiatric times* [Internet]. 2008;25(10):19–23. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2904966/>
23. Gimeno C, Dorado ML, Roncero C, Szerman N, Vega P, Balanzá-Martínez V, et al. Treatment of Comorbid Alcohol Dependence and Anxiety Disorder: Review of the Scientific Evidence and Recommendations for Treatment. *Frontiers in Psychiatry* [Internet]. 2017 Sep 22;8(173). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5614930/>
24. Kampman KM. New Medications for the Treatment of Cocaine Dependence. *Psychiatry (Edgmont)* [Internet]. 2005 Dec 1;2(12):44–8. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2994240/>
25. Schwartz EKC, Wolkowicz NR, De Aquino JP, MacLean RR, Sofuoglu M. Cocaine Use Disorder (CUD): Current Clinical Perspectives. *Substance Abuse and Rehabilitation*. 2022 Sep;Volume 13:25–46.