
HOSPITAL DE CLÍNICAS DE PORTO ALEGRE
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO COLABORADOR EM ÁLCOOL E DROGAS HCPA/SENAD
CENTRO DE PESQUISA EM ÁLCOOL E DROGAS

TERMO DE EXECUÇÃO DESCENTRALIZADA 03/2017 -
SENAD/HCPA PROCESSO SENAD: 08129.015636/2017-78

ARTIGO DE REVISÃO: Utilização da Inteligência Artificial em
Pesquisas sobre Álcool e Outras Drogas

Produto V

PROJETO: Análises Avançadas de Dados Brasileiros
Sobre Drogas com Inteligência Artificial e Translação
Para a Clínica

Investigador Principal: Prof. Dr. Flavio Pechansky
Coordenação do estudo: Prof.Dr. Felix Kessler e Profa. Dra. Lisia von Diemen

Porto Alegre, março de 2021

Utilização da Inteligência Artificial em Pesquisas sobre Álcool e Outras Drogas

Introdução

No intuito de atualizar o conhecimento da literatura recentemente publicada sobre Inteligência Artificial e uso de drogas, foi realizada uma pesquisa no EMBASE, WoS e PubMed. Foram selecionados artigos dos últimos dois anos (de fevereiro de 2019 a fevereiro de 2021). Na WoS foram encontrados 5 artigos; EMBASE foram 1462 e PubMed foram 634. Foi achado um total de 3281 artigos com 170 duplicatas. Após a sua remoção, totalizaram 3224. Estes foram ordenados por data de publicação e 39 foram previamente selecionados pelo título (buscando informações em relação a uso de técnicas de aprendizado de máquina aplicada ao uso de substâncias). Destes, foram retirados 13, porque o método não era especificamente de aprendizado de máquina ou porque o desfecho não era de interesse. Estes artigos foram resumidos pela nossa equipe de pesquisa, no texto que é apresentado nas próximas páginas. Também foram resumidas principais informações dos artigos no Quadro 1.

Este material complementa o manual pra Pesquisadores “Utilização da Inteligência Artificial em Pesquisas sobre Álcool e Outras Drogas” e alguns pontos do Capítulo 3 do manual são lembrados. As técnicas de aprendizado de máquina podem ser organizadas de acordo com o tipo do tipo de método de aprendizado: o supervisionado e não supervisionado.

Aprendizado supervisionado

O aprendizado supervisionado consiste na tarefa de aprender uma função que mapeia uma entrada para uma saída com base em exemplos de pares entrada-saída. Técnicas baseadas neste tipo de aprendizado inferem uma função a partir de dados de treinamento anotados, os quais consistem em um conjunto de exemplos em que é conhecida a saída desejada. O aprendizado é dito supervisionado porque um agente externo é quem rotula os exemplos de treinamento e age como um supervisor do processo de aprendizagem, fornecendo as saídas esperadas.

Os algoritmos supervisionados analisam os dados de treinamento e produzem uma função, que pode ser usada para inferir a saída de novos exemplos. As principais tarefas de mineração de dados associadas ao aprendizado supervisionado são:

- Classificação – método supervisionado que determina um modelo para um determinado atributo que é a função dos valores dos outros atributos.

- Regressão – predição do valor de uma variável contínua baseado no valor de outras variáveis, considerando um modelo de dependência linear ou não linear.

O objetivo da classificação é rotular, automaticamente, novas instâncias da base de dados com uma determinada classe aplicando a função ou modelo aprendido. Este modelo é baseado no valor dos atributos das instâncias de treinamento. Após a classificação, os dados de teste estarão categorizados em classes (rótulos de uma variável alvo categórica). Na regressão, a ideia é a mesma, mas a função ou modelo se ajusta a uma variável contínua.

Estes algoritmos de classificação e regressão podem ser organizados em diferentes tipos, de acordo com as técnicas que utilizam no aprendizado. Para cada técnica, diferentes algoritmos de aprendizado de máquina foram propostos e podem ser utilizados para as mais diversas análises, como indicado por extenso no manual supracitado.

- Classificadores baseados em regras (RIPPER - *repeated incremental pruning to produce error reduction*)
- Classificadores baseados em árvores (CART - *classification and regression tree*), floresta aleatório (*random forest*) que combina a saída de diferentes classificadores).
- Redes neurais artificiais (MLP – *Multilayer Perceptron*)
- *Support vector machines*(SVM)
- Modelos probabilísticos (*Naïve Bayes, Bayes Net*)
- Regressão linear
- Regressão logística

Aprendizado não-supervisionado

No aprendizado não-supervisionado, os dados de treinamento não estão rotulados, portanto a tarefa consiste em aprender uma função que organiza ou categoriza os dados de acordo com suas próprias características. As principais tarefas de mineração de dados associadas ao aprendizado não-supervisionado são:

- Associação – definição de regras do tipo A B, onde A e B são elementos que co-ocorrem em diferentes observações da base de dados. Uma das aplicações clássicas da

associação consiste na descoberta de produtos que são comprados juntos. A análise dos resultados pode determinar que ações devem ser tomadas para incrementar a venda de B, que produtos são afetados pelo produto A e que promoções podem incluir A para incrementar a venda de B. Analogamente, na área da medicina pode-se descobrir que sintomas estão relacionados.

- Descoberta de padrões sequenciais – definição de regras do tipo $(A) (C) \rightarrow B$ onde os pares AB e CB são elementos que co-ocorrem em diferentes observações da base de dados, sendo que CB ocorre após AB. Portanto, é possível descobrir que um paciente que apresentou o sintoma A e logo depois o sintoma C, possivelmente desenvolverá o sintoma B.

- Agrupamento (*clustering*) – classificação não supervisionada de registros em grupos. É realizado com base na similaridade entre os registros. Deve-se maximizar a similaridade intragrupo e minimizar a similaridade intergrupo. Exemplos de métodos de agrupamento são: particionamento, hierárquico e incremental.

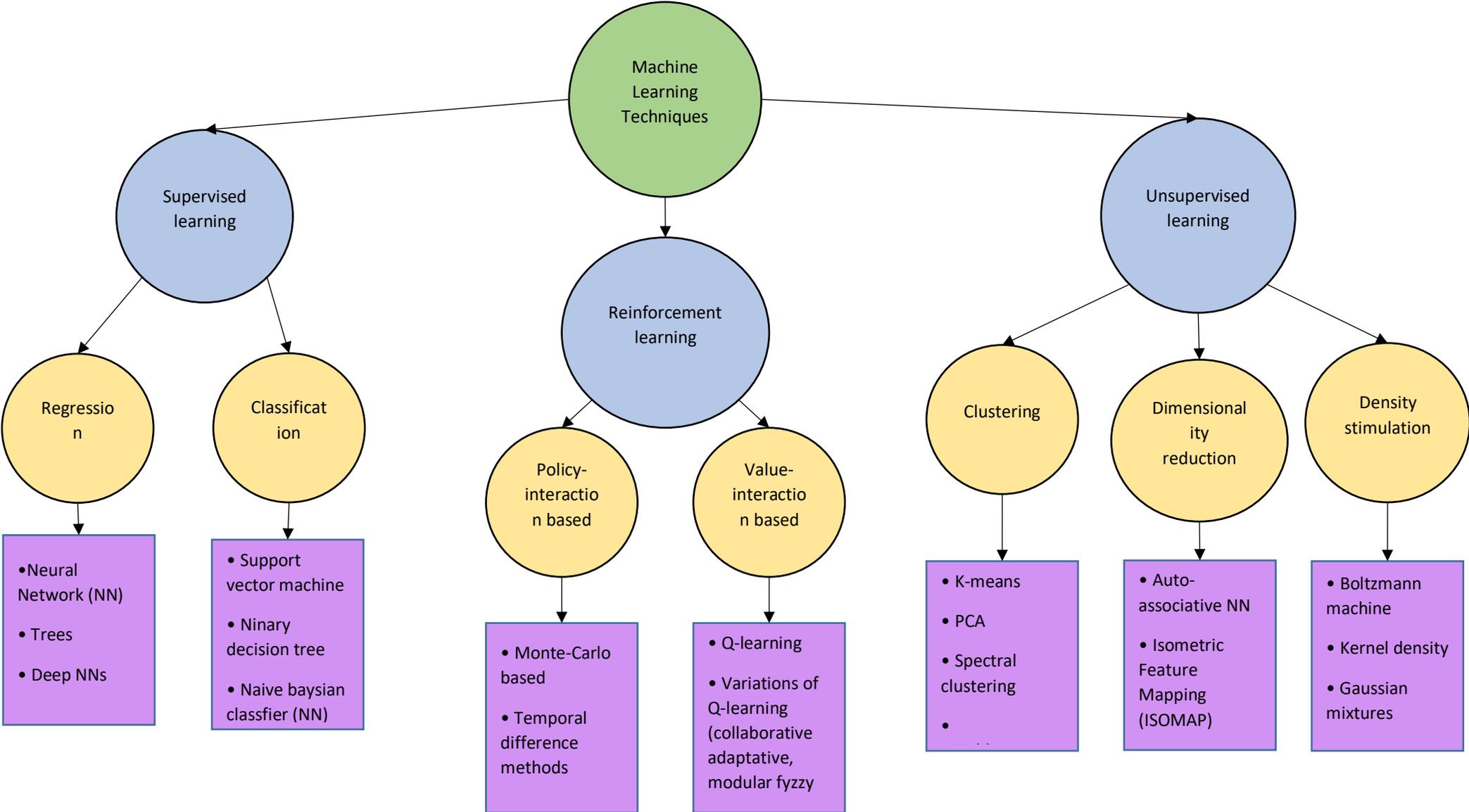
- Redução de dimensionalidade – método aplicado na etapa de pré-processamento dos dados com o objetivo de reduzir o número de variáveis a serem analisadas. Estes métodos podem simplesmente selecionar os atributos mais significativos ou criar novas variáveis que sejam compostas por conjuntos de variáveis originais, carregando o máximo de informação possível.

Alguns entre os principais algoritmos de aprendizado não-supervisionado estão citados no manual e são estes:

- *Partition Around Medoids* (PAM)
- *Density Based Spatial Clustering of Application with Noise* (DBSCAN)

A Figura 1 ilustra técnicas que compõem o aprendizado de máquina, mostrando a abrangência do assunto e sinalizando como revisões periódicas são importantes para a atualização deste conhecimento. Em relação à revisão realizada, é importante destacar que não se trata de uma revisão sistemática, e sim de uma revisão de parte da literatura científica publicada nestes últimos dois anos. No Quadro 1 são resumidas as principais informações dos artigos escolhidos

Figura 1. Figura do resumo das técnicas de Aprendizado de Máquina mais usadas.



Quadro 1. Quadro das informações principais dos artigos selecionados.

Ano	Autores	Título	Amostra	Desfecho	Técnica de Aprendizado de máquina utilizada
2021 Artigo 1	Wright, S. N., & Little, A. R.	NIDA vision for big data science to understand the biological underpinnings of substance use disorders. Neuropsychopharmacology, 46, 262.	-	-	Observação: carta ao editor que destaca a importância do big data, o aprendizado de máquina e ciência de dados como uma nova ferramenta para estudar os transtornos por uso de substância psicoativa.
2021 Artigo 2	Dugré, J. R., Potvin, S., Dellazizzo, L., & Dumais, A.	Aggression and delinquent behavior in a large representative sample of high school students: Cannabis use and victimization as key discriminating factors. Psychiatry Research, 296	63.196 estudantes de Ensino Médio	Problemas de conduta	K-means; método de Classificação Multinomial e Regression Tree (CART-aprendizagem por árvores de decisão)
2021 Artigo 3	Fojo, A. T., Lesko, C. R., Benke, K. S., Chander, G., Lau, B., Moore, R. D., Zandi, P. P., & Zeger, S. L.	A learning algorithm for predicting mental health symptoms and substance use. Journal of Psychiatric Research, 134, 22–29.	2444 participantes de coortes clínicas de pacientes em atendimento ambulatorial	Sintomas psiquiátricos futuros (depressão, ansiedade e mania) e uso de SPA (álcool, heroína e cocaína)	Modelo Tobit com abordagem bayesiana e hierárquica para resultados contínuos e um modelo Probit para resultados binários

<p>2020 Artigo 4</p>	<p>Sharma, B., Dligach, D., Swope, K., Salisbury-Afshar, E., Karnik, N. S., Joyce, C., & Afshar, M.</p>	<p>Publicly available Machine Learning Models for identifying opioid misuse from the clinical notes of hospitalized patients. BMC Medical Informatics and Decision Making, 20, 79.</p>	<p>161.520 pacientes de coorte observacional internados em um hospital terciário</p> <p>Aninhado: Caso-controle (n=1000) para construir um conjunto de dados para o uso indevido de opióides.</p>	<p>Uso indevido de opióides</p>	<p>Rede neural artificial e regressão logística</p>
<p>2020 Artigo 5</p>	<p>Panlilio, L. V, Stull, S. W., Bertz, J. W., Burgess-Hull, A. J., Kowalczyk, W. J., Phillips, K. A., Epstein, D. H., & Preston, K. L.</p>	<p>Beyond abstinence and relapse: cluster analysis of drug-use patterns during treatment as an outcome measure for clinical trials. Psychopharmacology, 237, 3369–3381.</p>	<p>Participantes tratados com metadona oriundos de três ensaios clínicos randomizados (n=166),(n=210),(n=50)</p>	<p>Padrões de uso de opióides e cocaína</p>	<p>Aprendizado de máquina não supervisionadas e os procedimentos utilizados foram agrupamento hierárquico de resultados categóricos e agrupamento longitudinal (k-means) de resultados quantitativos.</p>

2020 Artigo 6	Coughlin, L. N., Tegge, A. N., Sheffer, C. E., & Bickel, W. K.	A machine-learning approach to predicting smoking cessation treatment outcomes. <i>Nicotine and Tobacco Research</i> , 22, 415–422.	Duas coortes de fumantes n= 71 indivíduos na primeira coorte e n= 90 na segunda tratados com Terapia Cognitiva Comportamental (TCC) para dependência de tabaco	Abstinência de tabaco	Decision Trees algorithm
2020 Artigo 7	Grodin, E. N., Montoya, A. K., Bujarski, S., & Ray, L. A	Modeling motivation for alcohol in humans using traditional and Machine Learning approaches. <i>Addiction Biology</i> .	67 participantes bebedores pesados (“heavy drinkers”) que não buscavam tratamento.	Busca pelo álcool	Agrupamento de K-means e Random Forest
2020 Artigo 8	Karsinti, E., Labaeye, M., Piani, K., Fortias, M., Brousse, G., Bloch, V., Romo, L., & Vorspan, F.	Network analysis of psychotic manifestations among cocaine users. <i>Journal of Psychiatric Research</i> , 130, 300–305.	180 pacientes de um centro de assistência para adição	Manifestação de sintomas psicóticos em usuários de substâncias psicoativas	Análise de rede

2020 Artigo 9	Hajihosseini, M., Norozi, E., Azimi, T., & Ahmadinia, H.	Comparison of artificial neural networks and logistic regression for determination of social-cognitive factors impacting drug abstinence. <i>Journal of Substance Use</i> , 25, 295–300.	317 homens, amostra de conveniência de cinco programas de tratamento de uso de substâncias residenciais de curto prazo	Abstinência	Redes neurais artificiais
2020 Artigo 10	Cascini, F., De Giovanni, N., Inserra, I., Santaroni, F., & Laura, L.	A data-driven methodology to discover similarities between cocaine samples. <i>Scientific Reports</i> , 10, 15976.	Amostras de drogas de apreensões de cocaína de mais de 5 kg	Semelhanças entre perfis de amostras de cocaína	Algoritmos de aprendizado de máquina para classificar e comparar perfis de drogas
2020 Artigo 11	Yu, J. C., Fiore, V., Briggs, R., Braud, J., Rubia, K., Adinoff, B., & Gu, X.	Aberrant Neural Computation of Uncertainty Predicts Cocaine Use Relapse. <i>Biological Psychiatry</i> , 87, S22.	62 participantes humanos com transtorno por uso de cocaína e 25 controles	Tempo e status de recaída	Algoritmo de aprendizado de máquina
2020 Artigo 12	Han, D. H., Lee, S., & Seo, D. C.	Using Machine Learning to predict opioid misuse among U.S.	41.579 adolescentes Pesquisa Nacional	Uso indevido de opioides	Redes neurais artificiais, floresta aleatória distribuída e máquina de aumento de gradiente

		adolescents. Preventive Medicine, 130.	sobre Uso de Drogas e Saúde (2015-2017)		
2020 Artigo 13	Vázquez, A. L., Domenech Rodríguez, M. M., Barrett, T. S., Schwartz, S., Amador Buenabad, N. G., Bustos Gamiño, M. N., Gutiérrez López, M. L., & Villatoro Velázquez, J. A.	Innovative Identification of Substance Use Predictors: Machine Learning in a National Sample of Mexican Children. Prevention Science : The Official Journal of the Society for Prevention Research, 21, 171–181.	52.171 crianças mexicanas (amostra nacional)	Uso de substâncias ao longo da vida	Algoritmos de aprendizado de máquina
2019 Artigo 14	Jiménez, S., Angeles-Valdez, D., Villicaña, V., Reyes-Zamorano, E., Alcala-Lozano, R., Gonzalez-Olvera, J. J., & Garza-Villarreal, E. A.	Identifying cognitive deficits in cocaine dependence using standard tests and Machine Learning. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 95	2 amostras: a) 87 participantes (53 dependentes de cocaína e 34 não dependentes) b) 40 participantes (20 dependentes de cocaína e 20 não dependentes)	Dependência de cocaína	Algoritmos de Aprendizado de Máquina (ML): Modelo Linear Generalizado (Glm), Random forest e Elastic Net

2019 Artigo 15	Wilson, A. D., & Witkiewitz, K.	Predicting three-year recovery from end-of-treatment indicators: Machine Learning approach. Alcoholism: Clinical and Experimental Research, 43, 131A.	Braço ambulatorial do Project Match (N = 806) três anos após o tratamento	Perfil de recuperação	Abordagem de aprendizado de máquina exploratória Modelagem baseada em árvore
2019 Artigo 16	Garavan, H.	Data and theory driven investigations of cannabis's effects on adolescent neurodevelopment. Neuropsychopharmacology, 44, 62.	2.000 participantes recrutados aos 14 anos de idade (estudo IMAGEN)	Uso de cannabis	Análises de aprendizado de máquina
2019 Artigo 17	Kinreich, S., Meyers, J. L., Maron-Katz, A., Kamarajan, C., Pandey, A. K., Chorlian, D. B., Zhang, J., Pandey, G., Subbie-Saenz de Viteri, S., Pitti, D., Anokhin, A. P., Bauer, L., Hesselbrock, V., Schuckit, M. A., Edenberg, H. J., & Porjesz, B.	Predicting risk for Alcohol Use Disorder using longitudinal data with multimodal biomarkers and family history: Machine Learning study. Molecular Psychiatry.	656 participantes do Estudo Colaborativo da Genética do Alcoolismo (COGA) recrutados desde os 12 anos e não foram afetados em primeira avaliação e reavaliada anos depois como AUD	Transtorno do Uso de Álcool (AUD)	Linear Support Vector Machine (SVM)

			(DSM-5) (n = 328) ou não afetado (n = 328).		
2019 Artigo 18	Lichenstein, S., Scheinost, D., Potenza, M., Carroll, K., & Yip, S.	Distinct Neural Networks Predict Opioid Versus Cocaine Abstinence in Polysubstance Users. Biological Psychiatry.	53 pacientes poli usuários em tratamento de manutenção com metadona iniciando tratamento para cocaína.	Abstinência de cocaína e opioides	CPM (connectome-based predictive modelling)
2019 Artigo 19	Sekutowicz, M., Guggenmos, M., Kuitunen-Paul, S., Garbusow, M., Sebold, M., Pelz, P., Priller, J., Wittchen, H. U., Smolka, M. N., Zimmermann, U. S., Heinz, A., Sterzer, P., & Schmack, K.	Neural Response Patterns During Pavlovian-to-Instrumental Transfer Predict Alcohol Relapse and Young Adult Drinking. Biological Psychiatry.	Usuários de álcool (n=52), recém desintoxicados e jovens (n=136) que bebem socialmente sem transtorno por uso de substâncias.	Consumo de álcool e recaída	Esquema de classificação multivoxel baseado em máquinas de vetores de suporte linear.

<p>2019 Artigo 20</p>	<p>Erguzel, T. T., Noyan, C. O., Eryilmaz, G., Unsalver, B. O., Cebi, M., Tas, C., Dilbaz, N., & Tarhan, N.</p>	<p>Binomial Logistic Regression and Artificial Neural Network Methods to Classify Opioid-Dependent Subjects and Control Group Using Quantitative EEG Power Measures. Clinical EEG and Neuroscience.</p>	<p>75 pacientes dependentes de opioides e 59 controles</p>	<p>Dependência de opioides</p>	<p>Redes neurais artificiais</p>
<p>2019 Artigo 21</p>	<p>Lo-Ciganic, W.-H., Huang, J. L., Zhang, H. H., Weiss, J. C., Wu, Y., Kwoh, C. K., Donohue, J. M., Cochran, G., Gordon, A. J., & Malone, D. C.</p>	<p>Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions. JAMA Network Open.</p>	<p>560.057 participantes, sem câncer, com prescrição de opiáceos</p>	<p>Overdose de opioides</p>	<p>Algoritmos de aprendizado de máquina</p>
<p>2019 Artigo 22</p>	<p>Kim, N., McCarthy, D. E., Loh, W. Y., Cook, J. W., Piper, M. E., Schlam, T. R., & Baker, T. B.</p>	<p>Predictors of adherence to nicotine replacement therapy: Machine Learning evidence that perceived need predicts medication use. Drug and Alcohol Dependence.</p>	<p>623 participantes, dispostos a parar de fumar durante 30 dias</p>	<p>Adesão à terapia de reposição de nicotina</p>	<p>Algoritmo de aprendizado de máquina GUIDE (Generalized, Unbiased, InteractionDetection and Estimation)</p>

2019 Artigo 23	Ward, P. J., Rock, P. J., Slavova, S., Young, A. M., Bunn, T. L., & Kavuluru, R.	Enhancing timeliness of drug overdose mortality surveillance: A Machine Learning approach. PloS One.	Atestado de óbito de Kentucky	Mortalidade por overdose de drogas	Aprendizado de máquina usando processamento de linguagem natural
2019 Artigo 24	Wallace, G. T., Conner, B. T., & Shillington, A. M.	Using Classification Trees to Examine Predictors of Marijuana Use Among a Wide Range of Variables. Substance Use & Misuse.	4052 participantes estudantes universitários	Iniciação ao uso de maconha	Aprendizado de máquina exploratório (classification trees)
2019 Artigo 25	Vassileva, J., Shahidi, R., Taylor, B., Moeller, F. G., & Ahn, W. Y.	Machine Learning Identifies Common and Specific Markers of Addiction to Five Different Classes of Drugs. Biological Psychiatry.	595 participantes	Adição a diferentes classes de drogas	Modelos de aprendizado de máquina
2019 Artigo 26	Shen, Y. I., Lin, Z., Dziedzic, M., Harezlak, J., Halcomb, M. E., Yoder, K. K., & Oberlin, B. G.	Alcohol use disorder is accurately classified by behavioral and self-reported impulsivity using machine-learning methods.	32 participantes com transtorno por uso de álcool	Transtorno por uso de álcool	Regressão logística penalizada e random forest

		Alcoholism: Clinical and Experimental Research.			
--	--	--	--	--	--

A continuação os artigos selecionados foram resumidos pela nossa equipe de pesquisa, no texto que é apresentado nas próximas páginas.

Artigo 1

Wright, S. N., & Little, A. R. (2021). NIDA vision for big data science to understand the biological underpinnings of substance use disorders. Neuropsychopharmacology, 46, 262.

Resumo: Este artigo é uma carta ao editor que fala sobre a missão do Instituto Nacional de Abuso de Drogas (NIDA) de avançar a ciência sobre as causas e consequências do uso e da dependência de substâncias psicoativas. O NIDA se esforça para melhorar a saúde individual e pública através do apoio e condução estratégica de pesquisas básicas e clínicas sobre o uso de drogas, suas consequências e os mecanismos envolvidos nesta questão. As pesquisas biomédicas visam compreender como o abuso de drogas altera a biologia do cérebro e como gera o estado de dependência física e/ou comportamento compulsivo característico do vício. A visão da NIDA sobre como o uso de *big data* pode ser alavancada para revelar novos aspectos da biologia da dependência está estreitamente alinhada com o Plano Estratégico do NIH (*National Institute of Health*) para a Ciência de Dados. Nesta carta, é esclarecida uma abordagem com cinco medidas, a qual NIDA está focando para permitir maiores oportunidades de análise de dados, responder novas questões de investigação e examinar perspectivas alternativas sobre as questões originais de estudos anteriores. O NIDA também está trabalhando para que as capacidades computacionais mais recentes sejam combinadas com a investigação biomédica. A ciência de big data está bastante associada à inteligência artificial, particularmente em abordagens de aprendizado de máquina, e estes algoritmos têm desempenhado um papel importante nos últimos anos em relação aos dados de dependência química. Na carta, é colocada a necessidade de indivíduos com experiência em ciência da computação, bioinformática e matemática devido à incorporação da ciência de dados como uma nova ferramenta para estudar os transtornos por uso de substância psicoativa. A incorporação de diferentes dados permitirá descobertas científicas sobre as questões biológicas e comportamentais implícitas no vício de substâncias.

Artigo 2

Dugré, J. R., Potvin, S., Dellazizzo, L., & Dumais, A. (2021). Aggression and delinquent behavior in a large representative sample of high school students: Cannabis use and victimization as key discriminating factors. Psychiatry Research, 296.

Resumo: O estudo teve como objetivo utilizar um método de duas etapas para identificar melhor os diferentes padrões de problemas de conduta em uma amostra bastante representativa de jovens que estão no ensino médio, e também examinar as suas relações com diversos fatores de risco associados. Este estudo é uma pesquisa *survey* que foi realizada em 2651 aulas de francês e de inglês em 470 escolas de Quebec, Canadá. A amostra foi composta por 63.196 alunos de ensino médio. Do total, 50,7% eram homens e os jovens foram divididos de acordo com cinco níveis escolares: nível 1 (20%), nível 2 (20,9%), nível 3 (22,3%), nível 4 (19,3%), nível 5 (17,5%). O desfecho deste artigo foram os problemas de conduta (CP) e os subtipos dos mesmos, que variam muito entre os jovens. Os jovens foram comparados de acordo com variáveis sociodemográficas, psicológicas e uso de substância através de teste de Qui-quadrado e análise de variância (ANOVA). O método de agrupamento *K-means* foi utilizado para reduzir a heterogeneidade do CP. Também foi usado o método de Classificação Multinomial e *Regression Tree* (CART-aprendizagem por árvores de decisão) para examinar a especificidade dos fatores a fim de classificar os perfis de CP bem como suas interações hierárquicas. Como resultados das análises foram encontrados três padrões de CP: 1 - baixo AGG-RB (aggressive behaviors - non-aggressive/rule-breaking behaviors), 2 - alto AGG e baixo RB, e 3 - alto AGG-RB. A frequência de uso de cannabis, nível de vitimização e sintomas de hiperatividade foram os três fatores que melhor distinguiram os jovens dentre os padrões de CP. O modelo mostrou capacidade discriminatória de moderada a forte. Concluiu-se que o estudo contribui com novos conhecimentos para identificar jovens com padrões distintos de CP e que o uso de cannabis, vitimização e hiperatividade são fatores de risco chaves para distinguir os subtipos destes problemas de conduta.

Artigo 3

Fojo, A. T., Lesko, C. R., Benke, K. S., Chander, G., Lau, B., Moore, R. D., Zandi, P. P., & Zeger, S. L. (2021). A learning algorithm for predicting mental health symptoms and substance use. Journal of Psychiatric Research, 134, 22–29.

Resumo: O objetivo deste estudo foi desenvolver um algoritmo de predição Bayesiano que usa sintomas psiquiátricos de auto-relato e uso de substâncias psicoativas (SPA) dentro de uma população para prever comorbidades futuras de saúde mental e uso de SPA

para indivíduos daquela população. Esta abordagem foi validada a partir dos dados de duas coortes clínicas de pacientes em atendimento ambulatorial com o objetivo de demonstrar a possibilidade de generalizar em duas populações bem diferentes: The National Network of Depression Centers (NNDC, N = 1234) e The Johns Hopkins HIV Clinical Cohort (JHHCC, N = 1210). A amostra total foi composta por 2444 participantes sendo predominantemente do sexo feminino em NNDC (66,9%) e do sexo masculino em JHHCC (63,1%). As respectivas medianas de idade das coortes foram de 44 (31 a 57) e de 52 (45 a 58) anos. A primeira coorte (NNDC) compreende pacientes com transtornos de humor e a segunda (JHHCC), todas as pessoas de 18 anos ou mais infectadas com HIV que se inscreveram nesta clínica para cuidados com a doença e consentiram em compartilhar seus registros de saúde. O desfecho deste estudo é a presença de sintomas psiquiátricos futuros (depressão, ansiedade e mania) bem como o uso de SPA (álcool, heroína e cocaína) de pacientes. Para a análise, foi realizado um modelo Tobit com abordagem bayesiana e hierárquica para resultados contínuos e um modelo Probit para resultados binários. Os sintomas de saúde mental foram dicotomizados em moderado-grave e nenhum-leve. Previsões individuais produziram áreas sob a curva ROC(AUCs) de 0,84(95% CI, 0,80-0,88) e 0,85 (95% CI, 0,82-0,88) para sintomas de depressão nas duas coortes, AUCs de 0,84 (95% CI, 0,79-0,88) e 0,85 (95% CI, 0,82-0,88) para sintomas de ansiedade e uma AUC de 0,77 (95% CI, 0,72–0,82) para sintomas maníacos. Previsões de uso de substâncias produziram uma AUC de 0,92 (95% CI, 0,88–0,97) para uso de heroína, 0,90 (95% CI, 0,82–0,97) para uso de cocaína e 0,90 (95% CI, 0,88–0,92) para uso de álcool indevido. Conclui-se que esta abordagem teve um bom desempenho para fornecer previsões específicas do paciente no momento de cuidado sobre comorbidades de saúde mental e uso de substâncias. Ela pode ser aplicada a outros sintomas psiquiátricos e indicadores de uso de substâncias sendo também personalizável para sistemas de saúde específicos.

Artigo 4

Sharma, B., Dligach, D., Swope, K., Salisbury-Afshar, E., Karnik, N. S., Joyce, C., & Afshar, M. (2020). Publicly available Machine Learning models for identifying opioid misuse from the clinical notes of hospitalized patients. BMC Medical Informatics and Decision Making, 20, 79.

Resumo:

O objetivo do estudo é comparar o desempenho de abordagens de classificação de uso abusivo de opióides por meio de prontuário médico eletrônico comparando um modelo com informações sigilosas (PHI) e sem informações sigilosas dos pacientes. Este é um estudo de coorte observacional de pacientes internados em um hospital terciário composto por 161.520 observações entre 2007 e 2017. Também foi realizada uma amostra estratificada de caso-controle (n=1000) para construir um conjunto de dados para o uso indevido de opioides. Foi utilizado dois tipos de algoritmos de aprendizado de máquina: rede neural artificial e regressão logística. Os modelos de aprendizado de máquina sem PHI tiveram desempenho semelhante aos modelos com PHI. Os modelos sem PHI tiveram uma área sob a curva ROC maior que 0,90, principalmente usando algoritmo de rede neural convolucional. Os modelos sem PHI para construir fenótipos computáveis a partir de prontuários médicos eletrônicos são necessários para uma melhor vigilância e identificação de casos. O uso indevido de opioides é uma condição comportamental complexa que requer informações contidas em prontuários médicos eletrônicos. Além disso, as abordagens de aprendizado de máquina para classificação de texto podem ser uma solução viável para a identificação de pessoas com uso abusivo de opioides. Tais fenótipos computáveis para o uso indevido de opioides podem ser úteis para os sistemas de saúde para uma identificação e vigilância mais precisas de pacientes hospitalizados sem o risco de vazar qualquer PHI.

Artigo5

Panlilio, L. V, Stull, S. W., Bertz, J. W., Burgess-Hull, A. J., Kowalczyk, W. J., Phillips, K.

A., Epstein, D. H., & Preston, K. L. (2020). Beyond abstinence and relapse: cluster analysis of drug-use patterns during treatment as an outcome measure for clinical trials. Psychopharmacology, 237, 3369–3381.

Resumo: Este estudo foi uma análise de sequências de resultados de testes de drogas de três estudos publicados anteriormente pelo grupo de pesquisa. Cada participante dos estudos recebeu tratamento ambulatorial na clínica Archway em Baltimore. O objetivo foi analisar os padrões de uso de opióides e cocaína em três ensaios clínicos randomizados de gestão de contingências em participantes tratados com metadona. Os três estudos

analisados foram identificados como estudo 297(n=166), estudo 326 (n= 210) e estudo 390(n=50); em todos, foram feitas algumas exclusões e inserções de participantes na análise e, a partir de autorrelato e exame de urina, todos estavam usando opióides e cocaína no início do estudo. O desfecho deste estudo foi o padrão de uso associado ao tratamento de maior eficácia para saúde psicológica do usuário. As análises foram realizadas com técnicas de machine-learning não supervisionadas e os procedimentos utilizados foram agrupamento hierárquico de resultados categóricos e agrupamento longitudinal (k-means) de resultados quantitativos. A regressão logística multinomial foi utilizada para avaliar se as condições de contingência atribuídas aleatoriamente a cada um dos três estudos influenciaram a associação dos participantes com o grupo. Em cada ensaio clínico, foram identificados quatro grupos de padrões de uso, que podem ser descritos como uso de opióides, uso de cocaína, duplo uso (opióide e cocaína) e abstinência parcial/total. A gestão de contingências aumentou o número de membros nos grupos com níveis mais baixos de uso de drogas e menos sintomas de transtornos por uso de substâncias. Conclui-se que a análise de agrupamento fornece resultados mais interpretáveis e acionáveis do que os resultados de medidas tradicionais podendo oferecer uma resposta concreta à questão do que pode ser dito aos pacientes sobre as taxas de sucesso de novos tratamentos. De forma específica, a gestão de contingências tornou mais provável que uma pessoa adote um dos padrões de uso que foi associado à melhor saúde psicológica.

Artigo6

Coughlin, L. N., Tegge, A. N., Sheffer, C. E., & Bickel, W. K. (2020). A machine-learning approach to predicting smoking cessation treatment outcomes. Nicotine and Tobacco Research, 22, 415–422.

Resumo: O objetivo do estudo foi identificar características clínicas e psicossociais que predizem a abstinência duradoura em duas coortes de fumantes de baixa renda tratados com Terapia Cognitiva Comportamental (TCC) para dependência de tabaco avaliados em um estudo parental. Este estudo é uma análise de dados secundária de duas coortes de fumantes (uma de treino e a outra para validação) que fizeram terapia de grupo para parar de fumar. A amostra foi composta por 71 indivíduos na primeira coorte e 90 na segunda, todos sendo usuários de tabaco. Em relação a idade e sexo da amostra: a coorte de treino compôs 56,33% de homens com idade média de 46,55 anos e a coorte de validação

compôs 59,30% de homens com idade média de 51,71 anos. O estudo teve como desfecho a abstinência dentro do tratamento para parar de fumar. Para a análise estatística foram usadas Equações Estimativas Generalizadas (GEE) e uma abordagem de *Decision Trees* (abordagem de Aprendizado de máquina). Também foram utilizados os seguintes testes: teste t, teste de Mann-Whitney e teste de Qui-quadrado. Como resultados obteve-se que, na coorte de treino, a abordagem de *Decision Trees* melhorou significativamente na classificação de chance do status de fumante após o tratamento e no acompanhamento de 6 meses. A primeira divisão de todas as *Decision Trees*, que foi *Delay Discountig*, melhorou significativamente as taxas de classificação em ambas as coortes. “*Delay Discountig*” foi o melhor preditor na resposta do tratamento com TCC em grupo, prevendo status de tabagismo em 80% dos participantes no pós tratamento e de 81% dos participantes no acompanhamento. Concluindo, o estudo fornece um primeiro passo no tratamento personalizado para parar de fumar, no entanto são necessários mais trabalhos para esclarecer e avaliar esse tema.

Artigo7

Grodin, E. N., Montoya, A. K., Bujarski, S., & Ray, L. A. (2020). Modeling motivation for alcohol in humans using traditional and Machine Learning approaches. *Addiction Biology*.

Resumo: O objetivo do estudo é identificar preditores da motivação para a autoadministração do álcool utilizando um novo paradigma de programação de razão progressiva (PR). É uma pesquisa transversal que recrutou, através de folhetos e anúncios online em Los Angeles, bebedores pesados (“*heavy drinkers*”) que não buscavam tratamento. A amostra total foi composta por 67 participantes, sendo estes majoritariamente do sexo masculino (36 homens). A partir do agrupamento de *K-means*, foram identificados dois grupos na amostra: trajetórias de autoadministração motivadas (N = 41) e desmotivadas (N=26). As respectivas médias de idade destes grupos foram de 29.88(± 6.53) e 27.85 (± 6.68) e todos os participantes eram usuários de álcool sem sintomas significativos de abstinência e sem critérios de diagnóstico para outros transtornos por uso de substância, exceto de nicotina ou álcool. O desfecho deste estudo foi a busca pelo álcool. A curva de modelagem de crescimento foi usada para prever as trajetórias do álcool no ar expirado durante a autoadministração da substância. Os dados foram analisados usando duas abordagens: análise de regressão logística (abordagem

tradicional) e um modelo de *Random Forest* (abordagem de aprendizado de máquina). Em ambas as abordagens, “delay discounts” mais acentuados, que indicavam uma preferência por recompensas menores e mais rápidas, pode prever a motivação pela busca do álcool. Foi identificado ainda mais o desejo físico pelo álcool como um preditor para a motivação da autoadministração da substância. A aplicação adicional deste modelo para a ciência translacional sobre transtorno por uso de álcool e desenvolvimento de tratamento aparece de forma garantida.

Artigo8

Karsinti, E., Labaeye, M., Piani, K., Fortias, M., Brousse, G., Bloch, V., Romo, L., & Vorspan, F. (2020). Network analysis of psychotic manifestations among cocaine users. Journal of Psychiatric Research, 130, 300–305.

Resumo: Este estudo teve como objetivo adquirir melhor conhecimento, através de uma análise de rede, sobre a arquitetura dos fatores de risco (uso de cannabis, cocaína/crack, esquizofrenia) e como eles contribuíram para duas medidas diferentes de psicose (tendência à psicose e sintomas psicóticos transitórios induzidos pelo uso de cocaína) em usuários de cocaína. É um estudo quasi-experimental de dados transversais retirados de uma análise secundária. A amostra foi composta por 180 pacientes de um estudo multicêntrico descritivo sobre sintomas psicóticos transitórios induzidos por cocaína. Os pacientes foram recrutados de um centro de assistência para adição de drogas em Paris e subúrbio de Paris. A maioria da amostra (76%) era do sexo masculino, usava preferencialmente cocaína (72% *versus* crack) e descreveu ter tido pelo menos um sintoma psicótico induzido pela cocaína no seu histórico de uso ao longo da vida (91% da amostra). A idade média da amostra na época em que a entrevista foi realizada foi de 38 anos (± 8) e 88% desta população atendeu aos critérios do DSM IV para dependência de cocaína. De acordo com os registros médicos, 8,8% da amostra sofria de esquizofrenia. O desfecho deste estudo foi a manifestação de sintomas psicóticos em usuários de substâncias psicoativas. A partir da análise univariada foi observado que sintomas psicóticos transitórios induzidos por cocaína (TCIPS) são maiores em pacientes com alta tendência à psicose e esta última foi associada à dependência de cannabis ($p = 0,007$). As medidas da análise de rede mostraram que o produto preferencialmente utilizado (crack *versus* cocaína) estava associado com sintomas psicóticos transitórios induzidos por cocaína, tendência à psicose e, em menor grau, à esquizofrenia. Concluiu-se que, ao

utilizar a análise de rede, foi possível observar um fator de risco diferencial para o fenômeno da psicose em pacientes com transtorno por uso de cocaína. O uso intensivo de cocaína foi associado a TCIPS, estes correlacionados com tendência à psicose e esta última associada à esquizofrenia e ao uso pesado de cannabis. Sendo assim, se tem o conhecimento prévio de que o uso de cannabis é um fator de risco para TCIPS não de uma forma direta, mas mediada pela tendência à psicose.

Artigo9

Hajihosseini, M., Norozi, E., Azimi, T., & Ahmadinia, H. (2020). Comparison of artificial neural networks and logistic regression for determination of social-cognitive factors impacting drug abstinence. Journal of Substance Use, 25, 295–300.

Resumo: O objetivo deste estudo foi comparar o desempenho de modelos de regressão logística (LRM) e redes neurais artificiais (ANNs) na determinação de preditores de abstinência de drogas. É uma pesquisa prospectiva com amostra de conveniência de cinco programas de tratamento de uso de substâncias residenciais de curto prazo no sul de Khorasan, na província oriental do Irã. A amostra foi composta por 317 homens com média de idade de 28,5(± 6,6) anos. Todos tinham diagnóstico primário de transtorno de dependência de substância com base nos critérios diagnósticos do DSM-IV e precisavam estar em um estágio de desintoxicação. Neste estudo, os pacientes foram classificados como abstinentes se eles testaram negativo em seus resultados de toxicologia de urina e relataram nenhum uso de substância desde a avaliação inicial. O objetivo deste estudo foi avaliar se o modelo de ANNs tinha melhor desempenho em relação a outros métodos de análise de predição. Além das análises feitas com regressão logística e redes neurais, também foi utilizado o método de curva ROC para comparar o desempenho dos modelos. Ambos os métodos, LRM e ANNs, classificaram corretamente 82,2% e 84,4% dos participantes como grupo abstinente, respectivamente. As áreas sob a curva ROC foram 90 e 91,2 e a estatística *kappa* foi 0,643 e 0,688 para os modelos LRM e ANN, respectivamente. Os resultados mostraram que o desempenho da classificação não foi significativamente diferente entre os modelos. Conclui-se que a seleção de técnicas de predição para fatores de abstinência de drogas depende do conhecimento prévio sobre o padrão de dados e os objetivos do estudo.

Artigo 10

Cascini, F., De Giovanni, N., Inserra, I., Santaroni, F., & Laura, L. (2020). A data-driven methodology to discover similarities between cocaine samples. Scientific Reports, 10, 15976.

Resumo: O aprendizado de máquina tem sido usado para fins distintos no campo da ciência, mas nenhuma aplicação com drogas ilegais foi feita antes. Este estudo apresenta uma abordagem baseada em dados para descobrir semelhanças entre perfis de amostras de cocaína. Investigamos amostras de drogas de apreensões de cocaína de mais de 5 kg. Uma plataforma (PRIDE) continuamente atualizada com perfis de cocaína foi projetada para coletar informações ao longo do tempo, ela oferece várias vantagens para as ações de inteligência em toda a Europa. Assim, ele fornece uma metodologia ampla e padronizada que usa algoritmos de aprendizado de máquina para classificar e comparar perfis de drogas, destacar como as amostras de drogas são semelhantes e como é provável que compartilhem uma origem, lote ou processo de preparação comum. As atividades de pesquisa para criar a plataforma PRIDE para relacionar apreensões de cocaína incluíram análises laboratoriais, investigações em amostras de cocaína e o projeto e implementação da plataforma PRIDE. O estudo confirmou com sucesso a possibilidade de um aplicativo de aprendizado de máquina para classificação automática de perfis de cocaína.

Artigo 11

Yu, J. C., Fiore, V., Briggs, R., Braud, J., Rubia, K., Adinoff, B., & Gu, X. (2020). Aberrant Neural Computation of Uncertainty Predicts Cocaine Use Relapse. Biological Psychiatry, 87, S22.

Resumo: Os transtornos por uso de substâncias há muito tempo são associados a tomadas de decisão arriscadas. Nós investigamos a codificação da incerteza em 62 participantes humanos com transtorno do uso de cocaína e 25 controles saudáveis usando ressonância magnética funcional em duas tarefas de tomada de decisão. Descobrimos que a incerteza de decisão, quantificada por um modelo bayesiano, ativou partes cerebrais em ambas as tarefas em todos os participantes. A codificação da incerteza previu o tempo de recaída na tarefa do sinal de parada, mas não na tarefa de aprendizagem de recompensa. Um algoritmo de aprendizado de máquina previu o status de recaída com maior precisão ao usar respostas neurais da tarefa de sinal de parada, em comparação com o uso de sinais da aprendizagem reversa tarefa ou ambas as tarefas combinadas. Esses resultados

sugerem que a falha no rastreamento neural da incerteza de decisão pode ser um importante mecanismo e biomarcador para a recidiva de transtorno por uso de cocaína.

Artigo 12

Han, D. H., Lee, S., & Seo, D. C. (2020). Using Machine Learning to predict opioid misuse among U.S. adolescents. Preventive Medicine, 130.

Resumo: Este estudo avaliou o desempenho de três técnicas diferentes de aprendizado de máquina na previsão do uso indevido de opioides entre adolescentes norte-americanos. Os dados foram extraídos da Pesquisa Nacional sobre Uso de Drogas e Saúde 2015-2017 (N = 41.579 adolescentes, com idades entre 12–17 anos) e analisados em 2019. Os modelos de previsão foram desenvolvidos usando três algoritmos de ML, incluindo redes neurais artificiais, floresta aleatória distribuída e máquina de aumento de gradiente. A área sob a curva de característica de operação do receptor (AUROC) e a área sob a curva de recuperação de precisão (AUPRC) foram usadas como métricas de desempenho de previsão. A taxa geral de uso indevido de opióides entre adolescentes dos EUA foi de 3,7% (n = 1521). O desempenho da previsão foi semelhante nos quatro modelos. Em termos de AUPRC, a floresta aleatória distribuída apresentou o melhor desempenho em predição seguida de regressão logística penalizada, gradiente *boosting machine* e redes neurais artificiais. Os resultados sugerem que as técnicas de aprendizado de máquina podem ser uma técnica promissora.

Artigo 13

Vázquez, A. L., Domenech Rodríguez, M. M., Barrett, T. S., Schwartz, S., Amador Buenabad, N. G., Bustos Gamiño, M. N., Gutiérrez López, M. L., & Villatoro Velázquez, J. A. (2020). Innovative Identification of Substance Use Predictors: Machine Learning in a National Sample of Mexican Children. Prevention Science : The Official Journal of the Society for Prevention Research, 21, 171–181.

Resumo: O aprendizado de máquina fornece um método de identificação de fatores que discriminam entre usuários e não usuários de substâncias, melhorando potencialmente nossa capacidade de atender às necessidades dos serviços de prevenção disponíveis dentro do contexto com recursos limitados. O estudo aplicou algoritmos de aprendizado de máquina ao exame de uma ampla variedade de fatores para determinar sua importância na previsão do uso de substâncias ao longo da vida entre crianças mexicanas. O objetivo

era utilizar o aprendizado de máquina para identificar os fatores de alto impacto que melhor discriminam entre usuários e não usuários de substâncias em uma amostra nacional (N= 52.171) de crianças mexicanas (5ª, 6ª série; M idade = 10,40). Os participantes relataram informações sobre fatores individuais, fatores socioecológicos e uso de substâncias ao longo da vida. Os resultados sugerem que o uso de substâncias ilícitas do pai e do melhor amigo, e o sexo do entrevistado (masculino) parecem ser questões de triagem de alto impacto associadas ao início da substância durante a infância para jovens mexicanos. Esses dados ajudam os profissionais a restringir os esforços de prevenção, ajudando a identificar os jovens em maior risco.

Artigo 14

Jiménez, S., Angeles-Valdez, D., Villicaña, V., Reyes-Zamorano, E., Alcalá-Lozano, R., Gonzalez-Olvera, J. J., & Garza-Villarreal, E. A. (2019). Identifying cognitive deficits in cocaine dependence using standard tests and Machine Learning. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 95.

Resumo: Há uma necessidade crescente de abordar a variabilidade na detecção de déficits cognitivos com testes padrão de dependência de cocaína (DC). O objetivo do estudo foi identificar déficits cognitivos por meio de algoritmos de Aprendizado de Máquina (ML): Modelo Linear Generalizado (Glm), Floresta Random (Rf) e Rede Elástica (GlmNet), para permitir uma categorização mais eficaz de DC e controles não dependentes (NDC) e para resolver problemas metodológicos comuns. Para nossa validação, usamos dois conjuntos de dados independentes, o primeiro consistia de 87 participantes (53 CD e 34 NDC) e o segundo de 40 participantes (20 CD e 20 NDC). Todos os participantes foram avaliados com testes neuropsicológicos que incluíram 40 variáveis avaliando domínios cognitivos. Usando os resultados da avaliação cognitiva, os três algoritmos de ML foram treinados no primeiro conjunto de dados e testados no segundo para classificar os participantes em CD e NDC. Enquanto os três algoritmos tiveram um desempenho de curva operacional do receptor (ROC) acima de 50%, o GlmNet foi superior tanto no treinamento (ROC = 0,71) quanto nos conjuntos de dados de teste (ROC = 0,85) em comparação com Rf e Glm. Além disso, o GlmNet foi capaz de identificar os oito principais preditores de atribuição de grupo (DC ou NDC) de todos os domínios cognitivos avaliados. O estudo fornece evidências da eficácia das técnicas de ML para a

identificação de marcadores cognitivos potenciais e para mostrar seções de testes cognitivos padrão, que parecem interrompidos em pacientes dependentes de cocaína.

Artigo 15

Wilson, A. D., & Witkiewitz, K. (2019). Predicting three-year recovery from end-of-treatment indicators: Machine Learning approach. Alcoholism: Clinical and Experimental Research, 43, 131A.

Resumo: Análises recentes identificaram quatro perfis latentes de recuperação do transtorno por uso de álcool no braço ambulatorial do Project Match (N = 806) três anos após o tratamento. Os perfis foram identificados como 1) bebedores pesados frequentes de mau funcionamento, 2) bebedores pesados infrequentes de mau funcionamento, 3) bebedores pesados ocasionais de alto funcionamento e 4) bebedores pesados não pesados de alto funcionamento. A investigação atual usou uma abordagem de aprendizado de máquina exploratória, particionamento recursivo, para examinar quais das 28 variáveis medidas no final do tratamento no Project Match eram mais preditivas de associação nos dois perfis de alto funcionamento, 3 anos após o tratamento. As interações hierárquicas desses preditores foram examinadas com modelagem baseada em árvore, e pontos de corte com potencial utilidade clínica foram estabelecidos para certas medidas. A validação cruzada K-fold foi usada para evitar o sobreajuste e para reduzir o modelo para a solução mais estável. Pontuações do Inventário de Depressão de Beck inferiores a 7, ao final do tratamento, maximizam as chances de estar em um dos perfis de recuperação de alto funcionamento três anos após o tratamento. Outros fortes preditores de associação com perfil de recuperação de alto funcionamento incluíram escores mais altos de propósito de vida e um número médio de bebidas consumidas em todos os dias observados no último mês de tratamento que não excedeu 3,093 bebidas padrão por dia.

Artigo 16

Garavan, H. (2019). Data and theory driven investigations of cannabis's effects on adolescent neurodevelopment. Neuropsychopharmacology, 44, 62.

Resumo: Os recentes aumentos na disponibilidade e potência da cannabis levantam preocupações de que ela possa ter um impacto negativo no desenvolvimento do adolescente. Estudos longitudinais bem desenvolvidos oferecem percepções valiosas: ao

ajudar a identificar diferenças cerebrais que precedem ou seguem o uso de cannabis, eles podem sugerir possíveis causas e consequências. Nós relatamos dados do estudo IMAGEN, um estudo de dez anos com 2.000 participantes recrutados aos 14 anos de idade. Para identificar possíveis causas para o uso de cannabis, realizamos análises de aprendizado de máquina avaliando milhares de variáveis multimodais adquiridas aos 14 anos por seu valor na previsão do uso de cannabis dois anos depois. Para avaliar as consequências, comparamos os volumes de massa cinzenta de participantes que, aos 14 anos, relataram apenas um ou dois usos na vida de cannabis contra participantes controles. Também avaliamos as mudanças na reatividade da amígdala após cinco anos nos participantes que estavam usando cannabis aos 14 anos e continuaram a usar até os 19 anos. A abordagem exploratória para identificar os preditores do uso de cannabis identificou um número esparsa de características preditivas. Os resultados sugerem que mudanças cerebrais podem surgir desde os primeiros usos de cannabis em um adolescente. O estudo sugere que o uso de cannabis pode ser previsto usando métodos exploratórios baseados em dados e por diferenças hipotéticas no funcionamento da amígdala. Os resultados sugerem que um perfil de futuro usuário é possível, enquanto a “normalização” da reatividade da amígdala pode sugerir um sistema biológico relacionado ao uso reforçado e sustentado de cannabis ao longo da adolescência.

Artigo 17

Kinreich, S., Meyers, J. L., Maron-Katz, A., Kamarajan, C., Pandey, A. K., Chorlian, D. B., Zhang, J., Pandey, G., Subbie-Saenz de Viteri, S., Pitti, D., Anokhin, A. P., Bauer, L., Hesselbrock, V., Schuckit, M. A., Edenberg, H. J., & Porjesz, B. (2019). Predicting risk for Alcohol Use Disorder using longitudinal data with multimodal biomarkers and family history: Machine Learning study. Molecular Psychiatry.

Resumo: Os modelos preditivos conseguiram distinguir entre indivíduos com Transtorno do Uso de Álcool (AUD) e controles. No entanto, os modelos preditivos que identificam quem é propenso a desenvolver AUD e os biomarcadores que indicam uma predisposição para AUD ainda não estão claros. Nossa amostra (n = 656; 376M e 280F) incluiu descendentes e não descendentes de ancestrais europeus americanos (EA) e afro-americanos (AA) do Estudo Colaborativo da Genética do Alcoolismo (COGA) que foram recrutados desde os 12 anos e não foram afetados em primeira avaliação e reavaliada anos depois como AUD (DSM-5) (n = 328) ou não afetado (n = 328). A análise de aprendizado

de máquina foi realizada para 220 medidas de EEG, 149 polimorfismos de nucleotídeo único (SNPs) relacionados ao álcool de um grande estudo de associação ampla do genoma (GWAS) de uso / abuso de álcool e dois históricos familiares (mãe DSM-5 AUD e pai DSM -5 AUD). Foi usado o classificador supervisionado, Linear Support Vector Machine (SVM) para testar quais recursos avaliados prevêm o desenvolvimento de AUD. Foram realizadas análises estratificadas por idade, gênero e ancestralidade. Os resultados indicam taxas de precisão significativas e mais altas para o AA em comparação com os modelos de predição da EA e uma tendência de precisão do modelo mais alta entre as mulheres em comparação com os homens para ambos os ancestrais. Os recursos combinados de EEG e SNP modelam modelos com desempenho superior com base apenas em recursos de EEG ou apenas em recursos de SNP para amostras EA e AA. Esta superioridade multidimensional foi confirmada em uma análise de acompanhamento nas faixas etárias de AA (12-15, 16-19, 20-30) e na faixa etária de AA (16-19). Em ambas as amostras de ancestralidade, o grupo de idade mais jovem obteve pontuação de precisão mais alta do que os outros dois grupos de idade mais avançada. O AUD materno aumentou a precisão do modelo nas amostras de ambas as linhagens. Várias medidas discriminativas de EEG e características SNPs foram identificadas, incluindo gama posterior inferior, conectividade de onda lenta superior (delta, teta, alfa), relação gama frontal superior, correlação beta superior na área parietal e 5 SNPs: rs4780836, rs2605140, rs11690265, rs692854 e rs13380649. Os resultados destacam a importância da uniformidade de amostragem seguida por análise estratificada (por exemplo, ancestralidade, gênero, período de desenvolvimento) e uma seleção mais ampla de características.

Artigo 18

Lichenstein, S., Scheinost, D., Potenza, M., Carroll, K., & Yip, S. (2019). Distinct Neural Networks Predict Opioid Versus Cocaine Abstinence in Polysubstance Users. Biological Psychiatry.

Resumo: Usando uma nova abordagem de aprendizado de máquina, o objetivo do estudo é identificar previamente padrões de conectividade da rede neural, que podem prever a abstinência de cocaína durante o tratamento, buscando facilitar o desenvolvimento do tratamento personalizado e inovador. Com base na coleta de urina quinzenal, de 53 pacientes poli usuários em tratamento de manutenção com metadona iniciando tratamento

para cocaína. O CPM (*connectome-basedpredictivemodelling*), foi aplicado para identificar redes neurais de abstinência de opióides, avaliando a especificidade das redes de abstinência de cocaína e opióides para prever diferentes resultados, em diferentes estados cerebrais. Os resultados mostraram que redes neurais distintas, predizem opióides, na abstinência de cocaína durante o tratamento.

Artigo 19

Sekutowicz, M., Guggenmos, M., Kuitunen-Paul, S., Garbusow, M., Sebold, M., Pelz, P., Priller, J., Wittchen, H. U., Smolka, M. N., Zimmermann, U. S., Heinz, A., Sterzer, P., & Schmack, K. (2019). Neural Response Patterns During Pavlovian-to-Instrumental Transfer Predict Alcohol Relapse and Young Adult Drinking. Biological Psychiatry.

Resumo: O objetivo do estudo, através da transferência instrumental pavloviana, é testar se as respostas neurais, podem prever recaídas futuras em pacientes dependentes de álcool e o consumo futuro em adolescentes. Os dados foram coletados em uma universidade de medicina e hospital universitário em Berlim. A amostra contém um grupo de usuários de álcool (n=52), recém desintoxicados e um grupo de jovens (n=136), do sexo masculino, que bebem socialmente e não possuem diagnóstico de TUS. Para prever o consumo futuro, foi usado um esquema de classificação multivoxel baseado em máquinas de vetores de suporte linear. Testando o esquema de classificação dos participantes, houve 71,2% de precisão, na previsibilidade de recaídas, podendo também prever mudanças longitudinais no consumo de álcool. Os resultados sugerem que respostas neurais, durante a transferência instrumental pavloviana, podem constituir um prognóstico generalizado marcador para um futuro consumo abusivo de álcool.

Artigo 20

Erguzel, T. T., Noyan, C. O., Eryilmaz, G., Unsalver, B. O., Cebi, M., Tas, C., Dilbaz, N., Tarhan, N. (2019). Binomial Logistic Regression and Artificial Neural Network Methods to Classify Opioid-Dependent Subjects and Control Group Using Quantitative EEG Power Measures. Clinical EEG and Neuroscience.

Resumo: A regressão logística, e redes neurais artificiais, são métodos usados em estudos, nas classificações de dados médicos. A regressão logística, é uma sugestão para problemas médicos, devido à sua metodologia bem estabelecida e coeficientes que contribuem para avaliação das interpretações clínicas. As redes neurais artificiais, são

modelos gráficos, que possuem capacidade de modelar quaisquer dados. O objetivo do estudo, é gerar um modelo dicotomizado, em uma amostra de 75 pacientes dependentes de opioides e 59 indivíduos controle. Esses pacientes foram recrutados de um Hospital Psiquiátrico de Istambul, entre 2013 e 2016. Os critérios de exclusão foram: uso de múltiplas drogas, comorbidades psiquiátricas, transtornos neurológicos e pacientes tratados por eletroconvulsoterapia nos últimos 3 meses. As mudanças neurais, causadas pelo uso de drogas, podem ser detectadas com método de eletroencefalografia quantitativa. Observamos nos resultados, que a metodologia é potencialmente útil para médicos descobrirem marcadores biológicos e suas capacidades discriminativas, para vários tipos de transtorno.

Artigo 21

Lo-Ciganic, W.-H., Huang, J. L., Zhang, H. H., Weiss, J. C., Wu, Y., Kwoh, C. K.,

Donohue, J. M., Cochran, G., Gordon, A. J., & Malone, D. C. (2019). Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions. JAMA Network Open.

Resumo: O objetivo do estudo é desenvolver um algoritmo capaz de prever o risco de overdose de opioides, com pelo menos uma prescrição de opiáceos. A amostra possui um n de 560.057 participantes, sem câncer, com prescrição de opiáceos. Divididos em três grupos: treinamento (n=186 686), teste (n=186 685) e validação (n=186 686), de acordo com os resultados a amostra é predominantemente branca do sexo feminino. Foi usado uma regressão logística multivariada, para prever o risco de overdose nos três primeiros meses, após o início do tratamento com opioides prescritos. Os resultados mostram que os algoritmos de aprendizado de máquina, parecem funcionar para o risco, previsão e estratificação de overdose de opioides.

Artigo 22

Kim, N., McCarthy, D. E., Loh, W. Y., Cook, J. W., Piper, M. E., Schlam, T. R., & Baker, T. B. (2019). Predictors of adherence to nicotine replacement therapy: Machine Learning evidence that perceived need predicts medication use. Drug and Alcohol Dependence.

Resumo: A não adesão à medicação, para cessar o uso de nicotina, é um problema frequente. Identificar preditores do abandono, antes que ele ocorra, pode ajudar a explicar a não adesão, e sugerir intervenções personalizadas para resolvê-lo. Portanto, o objetivo do estudo, é identificar e caracterizar subgrupos de fumantes, com base na adesão à terapia de reposição de nicotina. A amostra teve um n de 623 participantes, dispostos a parar de fumar durante 30 dias, com pelo menos 18 anos de idade, tendo feito uso mínimo de 5 cigarros por dia, pelos últimos 6 meses. Os critérios de exclusão foram: uso de medicações para fumo, histórico de derrame, ataque cardíaco, ataque isquêmico transitório ou eletrocardiograma anormal, diagnóstico de psicose, gravidez ou falta de interesse

em usar um método contraceptivo, durante o tratamento. Os participantes foram randomizados em R-UC ou A-OT, R-UC usou adesivos de nicotina e A-OT pastilhas combinadas com adesivos. O algoritmo de aprendizado de máquina GUIDE foi usado e em cada partição o teste estatístico Qui-quadrado para avaliar as associações entre cada preditor de resultado, e selecionar as variáveis preditoras mais importantes. Os resultados demonstraram maior motivação e confiança para parar, e acreditar que a medicação para parar de fumar é segura e fácil de usar foi associada a uma maior adesão, que foi especialmente alta naqueles que eram mais dependentes e mais expostos ao fumo.

Artigo 23

Ward, P. J., Rock, P. J., Slavova, S., Young, A. M., Bunn, T. L., & Kavuluru, R. (2019). Enhancing timeliness of drug overdose mortality surveillance: A Machine Learning approach. PloS One.

Resumo: As mortes por overdose de drogas são identificadas nos sistemas de vigilância por meio dos códigos CID-10 presentes nas certidões de óbito, a codificação CID-10 leva tempo, mas as informações em texto livre estão disponíveis nas certidões de óbito, antes da

codificação. O objetivo do estudo, foi desenvolver um método de aprendizado de máquina, que classificasse certificados de óbito de texto livre, como overdose de drogas, para fornecer uma vigilância mais rápida acerca da mortalidade por overdose de drogas. Usando os dados do atestado de óbito de Kentucky, os campos de texto livre foram *tokenizados* e os recursos foram criados a partir desses tokens, usando processamento de linguagem natural.

O modelo de aprendizado de máquina com maior pontuação alcançou 0,96 valor preditivo positivo (PPV) e sensibilidade de 0,98 para um escore F de 0,97 na identificação de overdoses fatais de medicamentos em dados de teste. Este modelo de aprendizado de máquina obteve desempenho significativamente maior para sensibilidade ($p < 0,001$) do que a abordagem baseada em regras. A engenharia de recursos adicionais pode melhorar a previsão do modelo. Esse modelo pode ser implantado nas certidões de óbito assim que o

texto livre estiver disponível, eliminando o tempo necessário para codificar as certidões de óbito. O aprendizado de máquina usando processamento de linguagem natural é uma abordagem relativamente nova no contexto da vigilância de condições de saúde.

Este método apresenta uma aplicação acessível de aprendizado de máquina que melhora a oportunidade da vigilância da mortalidade por overdose de drogas. Como tal, pode ser empregado para informar as respostas de saúde pública à epidemia de overdose de drogas em tempo quase real, ao contrário de várias semanas após os eventos.

Artigo 24

Wallace, G. T., Conner, B. T., & Shillington, A. M. (2019). Using Classification Trees to Examine Predictors of Marijuana Use Among a Wide Range of Variables. Substance Use & Misuse.

Resumo: Jovens adultos possuem um risco elevado, de consequências negativas relacionadas ao uso de maconha. O estudo busca identificar possíveis preditores

de iniciação ao uso de maconha em jovens adultos, em diferentes populações, para prever quais deles têm mais experiências negativas. O estudo foi transversal, seus dados foram coletados em uma Universidade de Colorado. A amostra teve um n de 4052 estudantes universitários, sendo predominantemente brancos, do sexo feminino e com média de idade de 22,77 anos. Divididos entre um grupo controle (nunca usaram maconha) e um grupo de participantes que alegaram terem usado durante a participação da pesquisa. Análise de aprendizado de máquina exploratória (*classification trees*) foi utilizada. O uso de álcool e comportamento sexual, foram identificados como potenciais preditores para a iniciação do uso de maconha. Foi concluído que a prevenção de comportamentos de risco à saúde em geral, podem ser importantes para a prevenção da iniciação ao uso de maconha.

Artigo 25

Vassileva, J., Shahidi, R., Taylor, B., Moeller, F. G., & Ahn, W. Y. (2019). Machine Learning Identifies Common and Specific Markers of Addiction to Five Different Classes of Drugs. Biological Psychiatry.

Resumo: O objetivo do estudo, foi identificar marcadores comportamentais que classificam a dependência de opiáceos (n=159), dependência de estimulantes (n=168), dependência de álcool (n=42), dependência de maconha (n=92) e dependência de nicotina (n=187), usando aprendizado de máquina. A amostra inclui 595 participantes, as análises foram feitas por modelos de regressão e validação cruzadas, para prever as dependências, com base em dados demográficos, psiquiátricos, de personalidade e neurocognitivos. Os resultados mostraram preditores comuns entre substâncias específicas, na dependência de 5 diferentes drogas, a variável de personalidade possui uma maior previsibilidade, em comparação a variável neurocognitiva. Sendo a psicopatia o único preditor comum de dependência, sugerindo que pode ser um importante marcador de diagnóstico para o vício, independentemente da classe de drogas.

Artigo 26

Shen, Y. I., Lin, Z., Dziedzic, M., Harezlak, J., Halcomb, M. E., Yoder, K. K., & Oberlin, B. G. (2019). Alcohol use disorder is accurately classified by behavioral and self-reported impulsivity using machine-learning methods. Alcoholism: Clinical and Experimental Research.

Resumo: A identificação de fatores de risco é altamente importante para o desenvolvimento de transtornos por uso de álcool, tendo utilidade clínica e etiológica.

Fatores de risco que preveem o desenvolvimento posterior do transtorno, se agrupam amplamente em torno da impulsividade, busca de sensação e traços

relacionados à ansiedade. Foi realizada uma análise retrospectiva de indivíduos com o diagnóstico de TUA e bebedores sociais, usando dois métodos diferentes de aprendizado de máquina com variáveis demográficas, auto-relato e comportamentais para classificar o transtorno por uso de álcool versus bebedores sociais. A amostra contém um grupo de indivíduos que possuem diagnóstico de TUA (transtorno por uso de álcool), com um n de

32 participantes, sendo predominantemente homens. Contendo informações demográficas, de personalidade e realizando um teste monetário. Nos resultados a impulsividade auto-relatada, DD (*delay discounting*) e educação foram identificados como os preditores mais influentes por ambos os métodos. Além disso, gênero e raça foram identificados como preditores em regressão logística penalizada. A regressão logística penalizada descartou o uso de maconha, histórico familiar, idade e IMC como preditores, enquanto a *random forest* encontrou histórico familiar, raça, gênero e maconha como preditores menos influentes. DD foi um preditor influente em ambos métodos, apoiando a hipótese do estudo. Comportamentos impulsivos foram classificadores fortes e consistentes de transtorno por uso de álcool, identificando a impulsividade como um fenótipo de vício importante. As técnicas de classificação, podem detectar fatores de risco importantes, sendo uma abordagem promissora para melhor compreensão da etiologia e subtipos de TUA, aumentando potencialmente a detecção precoce de risco e, em última análise, aumentando a eficácia do tratamento.

Conclusão

Nas últimas décadas, tem ocorrido um enorme avanço da Psiquiatria devido à aplicação das técnicas de inteligência artificial e aos grandes consórcios internacionais de compartilhamento de dados. Anteriormente, os dados eram escassos com poucos pacientes e informações sobre eles. Entretanto, este cenário está mudando progressivamente, principalmente, na área da Dependência Química, haja vista a utilização de coleta de dados em tempo real através de uso de aplicativos de *smartphone* associado com dados genéticos e de neuroimagem. Além disso, o avanço da inteligência artificial dentro da Psiquiatria das Adições permite uma mudança de paradigma, movendo-se das evidências baseadas no grupo para as evidências baseadas no sujeito. Dessa maneira, é possível desenvolver tratamentos personalizados, prever desfechos e identificar subtipos de pacientes com o mesmo diagnóstico, mas que diferem em sua sintomatologia, tornando os diagnósticos mais precisos.

A partir dessa breve revisão narrativa da literatura apresentada, percebe-se que é necessário continuar as investigações em Psiquiatria da Adição usando técnicas de inteligência artificial com o intuito de desenvolver abordagens mais eficazes e eficientes no tratamento de pessoas com transtorno por uso de substâncias.

Referências

1. Wright, S. N., & Little, A. R. (2021). NIDA vision for big data science to understand the biological underpinnings of substance use disorders. *Neuropsychopharmacology*, 46, 262.
2. Dugré, J. R., Potvin, S., Dellazizzo, L., & Dumais, A. (2021). Aggression and delinquent behavior in a large representative sample of high school students: Cannabis use and victimization as key discriminating factors. *Psychiatry Research*, 296.
3. Fojo, A. T., Lesko, C. R., Benke, K. S., Chander, G., Lau, B., Moore, R. D., Zandi, P. P., & Zeger, S. L. (2021). A learning algorithm for predicting mental health symptoms and substance use. *Journal of Psychiatric Research*, 134, 22–29.
4. Sharma, B., Dligach, D., Swope, K., Salisbury-Afshar, E., Karnik, N. S., Joyce, C., & Afshar, M. (2020). Publicly available Machine Learning models for identifying opioid misuse from the clinical notes of hospitalized patients. *BMC Medical Informatics and Decision Making*, 20, 79.
5. Panlilio, L. V., Stull, S. W., Bertz, J. W., Burgess-Hull, A. J., Kowalczyk, W. J., Phillips, K. A., Epstein, D. H., & Preston, K. L. (2020). Beyond abstinence and relapse: cluster analysis of drug-use patterns during treatment as an outcome measure for clinical trials. *Psychopharmacology*, 237, 3369–3381.
6. Coughlin, L. N., Tegge, A. N., Sheffer, C. E., & Bickel, W. K. (2020). A machine-learning approach to predicting smoking cessation treatment outcomes. *Nicotine and Tobacco Research*, 22, 415–422.
7. Grodin, E. N., Montoya, A. K., Bujarski, S., & Ray, L. A. (2020). Modeling motivation for alcohol in humans using traditional and Machine Learning approaches. *Addiction Biology*.
8. Karsinti, E., Labaeye, M., Piani, K., Fortias, M., Brousse, G., Bloch, V., Romo, L., & Vorspan, F. (2020). Network analysis of psychotic manifestations among cocaine users. *Journal of Psychiatric Research*, 130, 300–305.
9. Hajhosseini, M., Norozi, E., Azimi, T., & Ahmadiania, H. (2020). Comparison of artificial neural networks and logistic regression for determination of social-cognitive factors impacting drug abstinence. *Journal of Substance Use*, 25, 295–300.

10. Cascini, F., De Giovanni, N., Inserra, I., Santaroni, F., & Laura, L. (2020). A data-driven methodology to discover similarities between cocaine samples. *Scientific Reports*, 10, 15976.
11. Yu, J. C., Fiore, V., Briggs, R., Braud, J., Rubia, K., Adinoff, B., & Gu, X. (2020). Aberrant Neural Computation of Uncertainty Predicts Cocaine Use Relapse. *Biological Psychiatry*, 87, S22.
12. Han, D. H., Lee, S., & Seo, D. C. (2020). Using Machine Learning to predict opioid misuse among U.S. adolescents. *Preventive Medicine*, 130.
13. Vázquez, A. L., Domenech Rodríguez, M. M., Barrett, T. S., Schwartz, S., Amador Buenabad, N. G., Bustos Gamiño, M. N., Gutiérrez López, M. L., & Villatoro Velázquez, J. A. (2020). Innovative Identification of Substance Use Predictors: Machine Learning in a National Sample of Mexican Children. *Prevention Science : The Official Journal of the Society for Prevention Research*, 21, 171–181.
14. Jiménez, S., Angeles-Valdez, D., Villicaña, V., Reyes-Zamorano, E., Alcalá-Lozano, R., Gonzalez-Olvera, J. J., & Garza-Villarreal, E. A. (2019). Identifying cognitive deficits in cocaine dependence using standard tests and Machine Learning. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 95.
15. Wilson, A. D., & Witkiewitz, K. (2019). Predicting three-year recovery from end-of-treatment indicators: Machine Learning approach. *Alcoholism: Clinical and Experimental Research*, 43, 131A.
16. Garavan, H. (2019). Data and theory driven investigations of cannabis's effects on adolescent neurodevelopment. *Neuropsychopharmacology*, 44, 62.
17. Kinreich, S., Meyers, J. L., Maron-Katz, A., Kamarajan, C., Pandey, A. K., Chorlian, D. B., Zhang, J., Pandey, G., Subbie-Saenz de Viteri, S., Pitti, D., Anokhin, A. P., Bauer, L., Hesselbrock, V., Schuckit, M. A., Edenberg, H. J., & Porjesz, B. (2019). Predicting risk for Alcohol Use Disorder using longitudinal data with multimodal biomarkers and family history: Machine Learning study. *Molecular Psychiatry*.
18. Lichenstein, S., Scheinost, D., Potenza, M., Carroll, K., & Yip, S. (2019). Distinct Neural Networks Predict Opioid Versus Cocaine Abstinence in Polysubstance Users. *Biological Psychiatry*.
19. Sekutowicz, M., Guggenmos, M., Kuitunen-Paul, S., Garbusow, M., Sebold, M., Pelz, P., Priller, J., Wittchen, H. U., Smolka, M. N., Zimmermann, U. S., Heinz,

- A., Sterzer, P., & Schmack, K. (2019). Neural Response Patterns During Pavlovian-to-Instrumental Transfer Predict Alcohol Relapse and Young Adult Drinking. *Biological Psychiatry*.
20. Erguzel, T. T., Noyan, C. O., Eryilmaz, G., Unsalver, B. O., Cebi, M., Tas, C., Dilbaz, N., Tarhan, N. (2019). Binomial Logistic Regression and Artificial Neural Network Methods to Classify Opioid-Dependent Subjects and Control Group Using Quantitative EEG Power Measures. *Clinical EEG and Neuroscience*.
21. Lo-Ciganic, W.-H., Huang, J. L., Zhang, H. H., Weiss, J. C., Wu, Y., Kwok, C. K., Donohue, J. M., Cochran, G., Gordon, A. J., & Malone, D. C. (2019). Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions. *JAMA Network Open*.
22. Kim, N., McCarthy, D. E., Loh, W. Y., Cook, J. W., Piper, M. E., Schlam, T. R., & Baker, T. B. (2019). Predictors of adherence to nicotine replacement therapy: Machine Learning evidence that perceived need predicts medication use. *Drug and Alcohol Dependence*.
23. Ward, P. J., Rock, P. J., Slavova, S., Young, A. M., Bunn, T. L., & Kavuluru, R. (2019). Enhancing timeliness of drug overdose mortality surveillance: A Machine Learning approach. *PloS One*.
24. Wallace, G. T., Conner, B. T., & Shillington, A. M. (2019). Using Classification Trees to Examine Predictors of Marijuana Use Among a Wide Range of Variables. *Substance Use & Misuse*.
25. G. T., Conner, B. T., & Shillington, A. M. (2019). Using Classification Trees to Examine Predictors of Marijuana Use Among a Wide Range of Variables. *Substance Use & Misuse*.
26. Vassileva, J., Shahidi, R., Taylor, B., Moeller, F. G., & Ahn, W. Y. (2019). Machine Learning Identifies Common and Specific Markers of Addiction to Five Different Classes of Drugs. *Biological Psychiatry*.
27. Shen, Y. I., Lin, Z., Dziedzic, M., Harezlak, J., Halcomb, M. E., Yoder, K. K., & Oberlin, B. G. (2019). Alcohol use disorder is accurately classified by behavioral and self-reported impulsivity using machine-learning methods. *Alcoholism: Clinical and Experimental Research*.