57º SBPO
SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
Gramado • RS
DE 5 A 9 DE OUTUBRO DE 2025

UM MODELO BASEADO EM DADOS CLÍNICOS PARA RASTREIO DE RETINOPATIA DIABÉTICA NA TRIAGEM DE PACIENTES

Victor Augusto do Carmo Duarte, Erito Marques de Souza Filho
Universidade Federal Rural do Rio de Janeiro
Av. Gov. Roberto Silveira, s/n – Moquetá, Nova Iguaçu – RJ, 26020-740
victoradocduarte@gmail.com, proferitomarques@gmail.com

Igor Melo de Almeida
Universidade Federal de São Paulo
R. Botucatu, 740 – Vila Clementino, São Paulo – SP, 04023-062
igorm.dml@gmail.com

Luís Felipe Leite da Silva, Ronaldo Altenburg Odebrecht Curi Gismondi
Universidade Federal Fluminense
R. Des. Athayde Parreiras, 100 – Fátima, Niterói – RJ, 24070-090
luis_leite@id.uff.br, ronaldogismondi@gmail.com

RESUMO

Apresentamos uma aplicação de algoritmos de *machine learning* e de *Linear Programming Discriminant Analysis* (LPDA) no rastreio de retinopatia diabética (RD) para a triagem de pacientes a partir de dados clínicos e demográficos. Os algoritmos foram treinados e testados em um conjunto de dados de 1460 indivíduos selecionados do banco de dados da *National Health and Nutrition Examination Survey* (NHANES) referente aos anos de 2005-2020. Os métodos se mostraram eficazes por diversas métricas (precisão, sensibilidade, F1-Score, acurácia e área sob a curva ROC), alcançando resultados superiores a 90%. As variáveis utilizadas no modelo foram a idade do paciente à época do diagnóstico de diabetes (real e normalizada), uso de insulina (binária), uso de antidiabéticos orais (binária) e pressão arterial sistólica e diastólica (reais e normalizadas).

PALAVRAS-CHAVE. Retinopatia diabética; Programação Linear; *Machine Learning*; NHANES.

Tópicos: PO na Área de Saúde.

ABSTRACT

We present an application of machine learning algorithms and Linear Programming Discriminant Analysis (LPDA) in diabetic retinopathy (DR) screening for patient triage based on clinical and demographic data. The algorithms were trained and tested on a dataset of 1460 individuals selected from the National Health and Nutrition Examination Survey (NHANES) database for the years 2005-2020. The methods proved effective across various metrics (precision, sensibility, F1-Score, accuracy, and area under the ROC curve), achieving results exceeding 90%. The variables used in the model included the patient's age at the time of diabetes diagnosis (real and normalized), insulin usage (binary), oral antidiabetic medication usage (binary), and systolic and diastolic blood pressure (real and normalized).

KEYWORDS. Diabetic retinopathy; Linear Programming; Machine Learning; NHANES.

Paper topics: OR in Health.

DOI: 10.59254/sbpo-2025-212499
https://proceedings.science/p/212499?lang=pt-br

1. Introdução

A retinopatia diabética (RD) é uma complicação microvascular crônica e progressiva do diabetes mellitus (DM) que pode levar a graus variados de perda visual [Brasil, 2023]. A deficiência visual e a cegueira decorrentes da RD são uma das principais causas de cegueira na população economicamente ativa [Brasil, 2023]; estima-se que atingiram, em 2020, cerca de 3,2 milhões de pessoas globalmente [Oh et al., 2021], acometendo até um terço de todos pacientes com DM [Brasil, 2023]. A detecção precoce da RD é essencial para o tratamento e prevenção da perda visual, uma vez que a doença pode ser assintomática até fases avançadas.

Por essas razões, tem havido um interesse crescente em desenvolver modelos preditivos para a RD usando abordagens de aprendizado de máquina, análise estatística e outros [Oh et al., 2021]. Embora existam diversas estratégias para prever o risco de RD, a aplicação da *Linear Programming Discriminant Analysis* (LPDA), proposta neste trabalho, tem vantagens significativas. A programação linear é uma técnica amplamente utilizada em otimização restrita e gerenciamento de recursos [Maia et al., 2021], podendo ser adaptada para problemas de previsão. Além disso, ela permite a inclusão de restrições específicas, como restrições de recursos limitados ou restrições de risco aceitável [Nueda; Gandía; Molina, 2022] e estas podem ser importantes na previsão de RD. Neste trabalho, propõe-se utilizar modelos de *machine learning* e de LDPA para classificação de pacientes segundo o risco de desenvolver RD, usando dados clínicos e demográficos públicos da NHANES dos anos de 2005-2020. É possível que este modelo seja útil na triagem de indivíduos com risco de RD e permita a implementação de medidas preventivas.

2. Modelo de LPDA

Considere um conjunto C de pontos no espaço p-dimensional $\mathbb{R}^p$. Sejam X e Y partições de C. Tem-se por objetivo encontrar um hiperplano H de dimensão $p-1$ tal que H separe C em X e Y (ou se aproxime disso, conforme será abordado adiante). Conforme a proposta por Nueda, Gandía e Molina [2022], "*a hyperplane H in $\mathbb{R}^p$ is an $(p-1)$–affine set and can be represented as $H = \{x \in \mathbb{R}^p | a^t \cdot x = b\}$, where $b \in \mathbb{R}$ and $a \in \mathbb{R}^p$, $a \neq 0_p$, and they are unique up a common non-zero multiple*". Inicialmente, busca-se um hiperplano H que possa separar estritamente X e Y. A propriedade que garante que o hiperplano H de fato é um hiperplano separador foi demonstrada por Nueda, Gandía e Molina [2022], e afirma que "*X and Y are strictly separable if and only if there exists a hyperplane $H = \{x \in \mathbb{R}^p | a^t \cdot x = b\}$ such that the following system is consistente. Such hyperplane is named separator hyperplane*". Considere o sistema mencionado:

$$\sigma = \begin{cases} a^t \cdot x_i \geq b + 1, & i = 1, \dots, n_1 \\ a^t \cdot y_j \leq b - 1, & j = 1, \dots, n_2 \end{cases}$$

Entretanto, caso esse hiperplano que separa estritamente X e Y não exista, "*we focus on a hyperplane that minimizes a measure of the deviation os this goal, called separation error*" [Nueda; Gandía; Molina, 2022]. Considere que σ é inconsistente. Isso implica a existência de $i \in \{1, \dots, n_1\}$ tal que $b + 1 - a^t \cdot x_i < 0$ ou a existência de $j \in \{1, \dots, n_2\}$ tal que $a^t \cdot y_j - b + 1 > 0$. A partir disso, o *separation error* associado a um hiperplano $H = \{x \in \mathbb{R}^p | a^t \cdot x = b\}$ pode ser definido a partir da função $f(a, b)$ em (I), chamada *separation error function*. Sejam m_i e w_j os pesos associados a cada medida.

$$f(a,b) = \sum_{i=1}^{n_1} m_i \cdot max\{b + 1 - a^t \cdot x_i, 0\} + \sum_{j=1}^{n_2} w_j \cdot max\{-b + 1 + a^t \cdot y_j, 0\} \quad \text{(I)}$$

Conforme Nueda, Gandía e Molina [2022], "*the separation error function is a non-negative, convex and non-differenciable function and the aim is to solve the problem min $f(a, b)$*". Ademais, esse problema pode ser reescrito de forma equivalente como em (II).

DOI: 10.59254/sbpo-2025-212499
https://proceedings.science/p/212499?lang=pt-br

$$Minimizar \sum_{i=1}^{n_1} m_i \cdot u_i + \sum_{j=1}^{n_2} w_j \cdot v_j$$

$$sujeito\ a \begin{cases} u_i \geq -a^t \cdot x_i + b + 1, & i = 1, \ldots, n_1 \\ u_i \geq 0, & i = 1, \ldots, n_1 \\ v_j \geq a^t \cdot y_j - b + 1, & j = 1, \ldots, n_2 \\ v_j \geq 0, & j = 1, \ldots, n_2 \end{cases} \quad \text{(II)}$$

Assim, busca-se a equação do hiperplano $H = \{x \in \mathbb{R}^p | a^t \cdot x = b\}$ que satisfaça (II) na otimalidade. As variáveis de decisão correspondentes às coordenadas do vetor *a* e ao escalar *b* são de grande interesse para o modelo desenvolvido, pois são coeficientes da equação do hiperplano que se deseja encontrar. As demais variáveis u_i e v_j, $i = 1, \ldots, n_1$ e $j = 1, \ldots, n_2$, correspondem ao *separation error* associado a cada ponto x_i e y_j em relação ao hiperplano H. Em suma, a solução ótima de (II) aponta a combinação de coeficientes que gera o hiperplano que minimiza a *separation error function* de H, e que da forma mais eficiente possível divide o conjunto em dois grupos X e Y pelo critério de ter ou não RD. O Algoritmo 1, no Apêndice I, apresenta a implementação na linguagem de programação Python da função responsável pela modelagem e pela busca da solução ótima do problema associado à LPDA, de acordo com (II), utilizando a biblioteca *gurobipy*.

3. Métodos

Para a coleta de dados, foram utilizados arquivos resultantes da NHANES dos anos de 2005-2020, que incluem dados demográficos e sobre diabetes em uma amostra representativa da população dos Estados Unidos, contendo 82.189 pessoas. No entanto, limitando-se a amostra aos indivíduos dos quais se pode obter todas as variáveis de interesse, e fixando tamanhos iguais para indivíduos com e sem RD, tem-se uma amostra de 1460 pessoas. Dentre as variáveis disponíveis na base de dados, foram selecionadas por critério de disponibilidade, juízo clínico dos autores, semelhança com outros trabalhos [Tsao; Chan; Su, 2018] e melhora dos resultados um total de seis variáveis para compor o modelo e serem utilizadas pelos algoritmos de previsão. Uma discussão acerca de seus funcionamentos e aplicações ao modelo de previsão de RD encontra-se em Das, Biswas e Bandyopadhyay [2022]. Entre eles estão algoritmos de *machine learning* como *Random Forest* (RF), *Support Vector Machine* (SVM), *K-Nearest Neighbors Classifier* (KNN), *Naive Bayes Classifier* (NB), *Gradient Boosting* (GB) e *Extreme Gradient Boosting* (XGB), e de rede neural como *Multilayer Perceptron* (MLP).

Em contrapartida, a LPDA permite encontrar a equação de um hiperplano ótimo para separar a amostra nos grupos de treino (80% da amostra) de quem tem e não tem RD. A equação do hiperplano é então utilizada para prever no grupo de teste (20% da amostra) se um indivíduo tem ou não a doença. As variáveis pertinentes foram: idade à época do diagnóstico de diabetes, uso ou não de insulina, uso ou não de antidiabéticos orais, pressão arterial sistólica, pressão arterial diastólica (variáveis independentes) e ter ou não ter RD (variável dependente) [Reddy; Sethi; Rajender, 2020]. A Tabela 1 apresenta uma síntese do perfil das variáveis na população.

Variáveis		Valores
Age of diagnosis for diabetes (DID040)		47.81±16.14 (1–80)
	1-20	87 (5.96%)
	21-40	394 (26.99%)
	41-60	691 (47.33%)
	60+	288 (19.73%)

DOI: 10.59254/sbpo-2025-212499
https://proceedings.science/p/212499?lang=pt-br

Variáveis		Valores
Insulin usage (DIQ050)		
	Sim	649 (44.45%)
	Não	811 (55.55%)
Oral antidiabetic medication usage (DIQ070)		
	Sim	605 (41.44%)
	Não	855 (58.56%)
Systolic blood pressure (DIQ300S)		131.31±19.45 (70–212)
Diastolic blood pressure (DIQ300D)		76.64±13.77 (20–150)
Diabetic retinopathy diagnosis (DIQ080)		
	Sim	730 (50.00%)
	Não	730 (50.00%)

Tabela 1. Perfil da amostra estudada em função das variáveis utilizadas.

Uma descrição do modelo matemático desenvolvido e baseado em Nueda, Gandía e Molina [2022] encontra-se na Seção 2. Para implementar tal modelo, no contexto do objetivo deste trabalho, foi utilizada a biblioteca *gurobipy* da linguagem de programação Python, e esta foi responsável por encontrar a solução ótima do problema de programação linear. A modelagem foi feita como apontado por Nueda, Gandía e Molina [2022]. As variáveis selecionadas para o modelo foram adaptadas a partir da base de dados da NHANES, incluindo informações clínicas e demográficas facilmente acessíveis, como a idade do diagnóstico de diabetes, uso de insulina, uso de antidiabéticos orais e a pressão arterial sistólica e diastólica. Da amostra de 1460 indivíduos, metade com e metade sem RD, 80% foi separado para o grupo de treino e 20% alocado no grupo de teste. Ao conjunto X pertencem os indivíduos do grupo de treino com RD, e ao Y os sem RD. Uma vez estabelecidos tais conjuntos, é possível otimizar o *separation error* [Nueda; Gandía; Molina, 2022]. O Apêndice 1 apresenta uma implementação em Python da função responsável pela modelagem e pela busca pela solução ótima do problema de programação linear utilizando a *gurobipy*. Além disso, uma vez encontrada a classificação dos indivíduos do grupo de teste a partir do hiperplano, pode-se comparar o estado de saúde real com o previsto a fim de avaliar o modelo. Para tal, utilizamos métricas de precisão, sensibilidade, F1-score, acurácia e área sob a curva ROC.

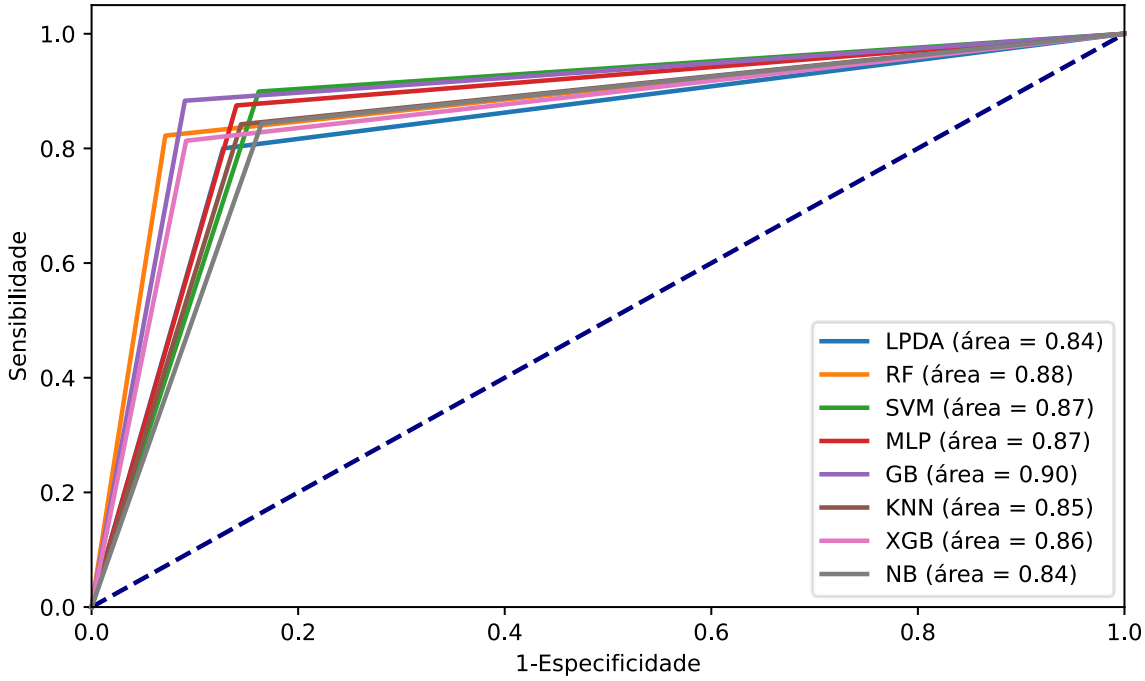
4. Resultados

Foram incluídos neste estudo dados referentes a uma amostra de 1460 indivíduos submetidos à NHANES nos anos de 2005-2020. Em relação à idade à época do diagnóstico de diabetes, 87 indivíduos (5,96%) tinham até 20 anos, 394 (26,99%) tinham entre 21 e 40 anos, 691 (47,33%) tinham entre 41 e 60, e 288 (19,73%) tinham acima de 60 anos. Ressalta-se que não foram obtidos os dados relativos à classificação de DM e nem ao tempo da doença. A qualidade dos modelos foi avaliada com diferentes métricas, a saber: acurácia, precisão, sensibilidade, F1-score, e área sob a curva ROC.

A Figura 1 representa graficamente a métrica da curva ROC, com a sensibilidade dada pela taxa de verdadeiros positivos e a especificidade dada pela taxa de verdadeiros negativos, no caso LPDA, mostrando a relação entre essas duas medidas para todos os valores de corte possíveis. Para além disso, segue dos resultados da amostra neste método 37,0% de verdadeiros positivos (VP), ou seja, indivíduos com RD que foram corretamente identificados, 46,9% de verdadeiros negativos (VN), ou seja, indivíduos sem RD que foram assim classificados, 6,8% de falsos positivos (FP), ou indivíduos sem RD aos quais o algoritmo atribuiu um diagnóstico da doença, e 9,2% de falsos negativos (FN), com indivíduos com RD que erroneamente foram alocados no grupo dos que não receberam o diagnóstico. Os demais métodos utilizados também são representados abaixo.

DOI: 10.59254/sbpo-2025-212499
https://proceedings.science/p/212499?lang=pt-br

Figura 1. Desempenho segundo a curva ROC na previsão do diagnóstico de RD.



.

Como detalhado na Tabela 2, podem ser comparados os resultados obtidos com LPDA, RF, SVM, MLP, GB, KNN, XGB e NB. De modo geral, nota-se que a LPDA obteve um desempenho competitivo, com acurácia de 83,9% (indicando sua proporção de classificações corretas), precisão de 84,38% (indicando a proporção de VP dentre os classificados como positivos) e sensibilidade de 80% (indicando a proporção dos VP dentre os reais positivos). Certas métricas de alguns métodos, como a sensibilidade do RF (91,85%), superam o desempenho da LPDA (80%), enquanto que outras, como a precisão do XGB (81,33%), são superadas (84,38%). Com área sob a curva ROC de 0,84, valor que, embora inferior ao dos principais métodos de *machine learning* e redes neurais, ainda é competitivo, nossos resultados sugerem que o modelo linear proposto é capaz de realizar uma separação razoável entre as classes, mesmo sem os mecanismos mais complexos de modelagem não-linear presentes nos outros algoritmos. É notória a similaridade da performance dos métodos de *machine learning* e redes neurais em relação ao método de programação linear explorado neste trabalho. Embora técnicas como GB, MLP e RF apresentem vantagens marginais em algumas métricas, a LPDA demonstra ser uma abordagem viável e competitiva, com uma vantagem em sua formulação interpretável.

Método	**Acurácia** $\frac{VP+VN}{VP+VN+FP+FN}$	**Precisão** $\frac{VP}{VP+FP}$	**Sensibilidade** $\frac{VP}{VP+FN}$	**F1-Score** $\frac{2 \cdot \textbf{Precisão} \cdot \textbf{Sensibilidade}}{\textbf{Precisão} + \textbf{Sensibilidade}}$	**ROC**
LPDA	83,90%	84,38%	80,00%	82,13%	0,84
RF	86,64%	81,58%	**91,85%**	86,41%	0,87
SVM	86,30%	**89,92%**	79,26%	84,25%	0,87
MLP	87,67%	88,37%	84,44%	86,36%	0,88
GB	**89,73%**	88,32%	89,63%	**88,97%**	**0,89**
KNN	84,93%	84,21%	82,96%	83,58%	0,84
XGB	85,96%	81,33%	90,37%	85,61%	0,86
NB	83,90%	84,38%	80,00%	82,13%	0,84

Tabela 2. Valores das métricas para avaliação dos algoritmos de classificação.

DOI: 10.59254/sbpo-2025-212499
https://proceedings.science/p/212499?lang=pt-br

5. Discussão

Segundo as métricas discriminadas, os algoritmos de predição de RD apresentaram desempenho consistente entre si se comparados aos resultados uns dos outros. Esses resultados indicam que o modelo é capaz de identificar corretamente a maioria dos pacientes com risco de desenvolver retinopatia diabética, bem como de excluir aqueles que não apresentam risco significativo. Além disso, o modelo apresentou um bom desempenho em termos de tempo de execução e facilidade de interpretação, tornando-o uma ferramenta prática para a triagem e gerenciamento de pacientes com diabetes. Em particular, uma vantagem significativa do uso da LPDA é permitir a inclusão de restrições específicas, como restrições de recursos limitados ou restrições de risco aceitável, que podem ser importantes na previsão de risco de RD.

Comparando os resultados deste trabalho com a literatura científica recente, um estudo de Reddy, Sethi e Rajender [2020] propôs o método de aprendizado de máquina SVM (*Support Vector Machine*) com kernel gaussiano para detectar RD em suas fases iniciais. Outros métodos testados pelos autores foram *random forest*, *decision tree*, *adaptative boosting* e afins. O modelo utilizou medidas de glicemia, pressão arterial, e outros dados clínicos. Com o SVM, foram obtidos melhores resultados que em todos os outros métodos comparados no estudo, atingindo uma acurácia de 81,3% (inferior a nossos resultados). O método de aprendizado de máquina SVM desenvolvido por Tsao, Chan e Su [2018] obteve como resultado as métricas acurácia e área sob a curva ROC nos valores 83,9% (igual ao pior de nossos resultados) e 0,933 (superior a nossos resultados), respectivamente. O modelo levou em consideração a pressão arterial, o Índice de Massa Corporal, a idade, o sexo, o tempo de diagnóstico da doença, uso de insulina e outros parâmetros clínicos. Esses trabalhos indicam a eficácia de abordagens de aprendizado de máquina na previsão de RD. Uma abordagem alternativa, proposta por Ghoushchi *et al*. [2021], foi a utilização das meta-heurísticas *Genetic Algorithm* (GA) e *Ant Colony Optimization* (ACO) para prever a RD. A base de dados utilizada continha imagens de 224 angiografias de retina entre pacientes com e sem diabetes. Analisando a acurácia e a precisão, GA e ACO obtiveram 95% e 96,4% (acima de nossos resultados), e 84,3% e 87,5% (abaixo de nossos resultados), na ordem.

A literatura destaca a importância da detecção precoce da retinopatia diabética para prevenir a perda de visão em pacientes com diabetes [Maia et al., 2021]. A triagem regular de pacientes diabéticos para detectar sinais de RD é recomendada por diretrizes clínicas, como as da *American Diabetes Association* [2021] e o *Protocolo Clínico e Diretrizes Terapêuticas da Retinopatia Diabética* do Ministério da Saúde [Brasil, 2021]. No entanto, o rastreio populacional pode ser custoso, demandar recursos não disponíveis para ser aplicada em todos pacientes e exige profissionais capacitados munidos destes recursos [Brasil, 2021; Vujosevic et al., 2020]. Isso pode dificultar o acesso à triagem e atrasar diagnósticos precoces e de casos assintomáticos em estágios iniciais da doença. Portanto, a previsão do risco de desenvolver RD usando abordagens de aprendizado de máquina, análise estatística, meta-heurísticas, programação linear, e quaisquer outros métodos factíveis e confiáveis, pode ser uma alternativa promissora, facilitando a seleção efetiva dos pacientes que melhor se beneficiariam da triagem. O modelo de programação linear proposto neste artigo contribui para essa linha de pesquisa, fornecendo uma abordagem adicional para prever o risco de RD usando variáveis clínicas e demográficas facilmente acessíveis.

Dentre as limitações do nosso estudo, destacam-se o relativamente pequeno tamanho da amostra que pôde ser analisada, a dificuldade de extrapolar os achados para outras populações além da estudada (considerando as peculiaridades genéticas, culturais e sociais de cada população), a quantidade de bases de dados consultadas (apenas dados da NHANES), e a falta de outras variáveis conhecidamente relevantes no desfecho clínico de RD, como o tempo total de DM do paciente, o Índice de Massa Corporal e o sexo [Tsao; Chan; Su, 2018]. Estudos futuros devem considerar a ampliação da amostra, a aplicação em diferentes populações e a inclusão de mais variáveis para avaliar a eficácia e a validade externa do modelo proposto.

DOI: 10.59254/sbpo-2025-212499
https://proceedings.science/p/212499?lang=pt-br

6. Conclusão

Os modelos de predição baseados em programação linear ou em *machine learning* propostos neste estudo demonstraram sua capacidade de prever satisfatoriamente a RD em pacientes com diabetes. Com precisão, sensibilidade, F1-score e acurácia superiores a 80% (salvo a sensibilidade do SVM), encontramos resultados promissores na predição de RD através de dados clínicos e demográficos facilmente acessíveis. Os resultados deste estudo corroboram que *machine learning* e sugerem que a programação linear podem ser técnicas úteis e eficientes para predição RD, auxiliando na triagem de pacientes diabéticos. Ademais, levantamos a hipótese de que o modelo poderia ser adaptado para outras doenças, utilizando variáveis e populações representativas da condição médica em questão.

A dependência do modelo de apenas variáveis simples, como a idade na época do diagnóstico de diabetes e a pressão arterial, pode facilitar a triagem de pacientes para a detecção precoce de RD, principalmente em cenários de recursos limitados, comumente presentes na gestão de saúde pública. Adicionalmente, a abordagem de programação linear proposta neste trabalho tem vantagens significativas em relação a outras técnicas de previsão, especialmente no que se refere à capacidade de incluir restrições específicas no modelo e ao baixo tempo de execução. Em resumo, a aplicação da LPDA aqui proposta se mostra uma ferramenta útil para prever o risco de RD em pacientes com diabetes, bem como os métodos de *machine learning* consolidados na literatura científica recente. Os resultados obtidos fornecem evidências preliminares da utilidade clínica do modelo e, por isso, destacamos a importância de continuar a pesquisa nessa área como suporte à gestão pública no rastreio de RD para a triagem de pacientes.

Referências

Brasil. Ministério da Saúde. Secretaria de Atenção Especializada à Saúde. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Portaria Conjunta nº 17, de 1º de outubro de 2021. Aprova o Protocolo Clínico e Diretrizes Terapêuticas da Retinopatia Diabética. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/protocolos-clinicos-e-diretrizes-terapeuticas-pcdt/arquivos/2021/portal-portaria-conjunta_pcdt_retinopatia-diabetica_.pdf>. Acesso em: 13 nov. 2023

Oh K, Kang HM, Leem D, Lee H, Seo, KY, Yoon S. Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images. Scientific Reports, 11(1897), 2021. DOI: 10.1038/s41598-021-81539-3

Maia EG, Passos CM, Granado FS, Levy RB, Claro RM. Replacing ultra-processed foods with fresh foods to meet the dietary recomendations: a matter of cost? Cadernos de Saúde Pública, 37(13), 2021. DOI: 10.1590/0102-311x00107220

Nueda MJ, Gandía C, Molina MD. LPDA: A new classification method based on linear programming. PLoS ONE, 17(7), 2022. DOI: 10.1371/journal.pone.0270403

Tsao H, Chan P, Su EC. Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. BMC Bioinformatics, 19(283), 2018. DOI: 10.1186/s12859-018-2277-0

Das D, Biswas SK, Bandyopadhyay S. A critical review on diagnosis of diabetic retinopathy using machine learning and deep learning. Multimed Tools Applications, 81(18), 25613-25655, 2022. DOI: 10.1007/s11042-022-12642-4

DOI: 10.59254/sbpo-2025-212499
https://proceedings.science/p/212499?lang=pt-br

Reddy SS, Sethi N, Rajender R. Discovering optimal algorithm do predict diabetic retinopathy using novel assessment methods. EAI Endorsed Transactions on Scalable Information Systems, 8(29), 2020. DOI: 10.4108/eai.13-7-2018.165505

Ghoushchi SJ, Ranjbarzadeh R, Dadkhah AH, Pourasad Y, Bendechache M. An Extended Approach to Predict Retinopathy in Diabetic Patients Using the Genetic Algorithm and Fuzzy C-Means. *Biomed Res Int*., 2021:5597222, 2021. DOI: 10.1155/2021/5597222

American Diabetes Association. Standards of Medical Care in Diabetes — 2021 Abridged for Primary Care Providers. *Clinical Diabetes*, 3TSAO(1), 14-43, 2021. DOI: 10.2337/cd21-as01

Brasil. Ministério da Saúde. Secretaria de Atenção Especializada à Saúde. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Portaria Conjunta nº 17, de 1º de outubro de 2021. Aprova o Protocolo Clínico e Diretrizes Terapêuticas da Retinopatia Diabética. Brasília: Ministério da Saúde, 2021. 10 p.

Vujosevic S, Aldington SJ, Silva P, Hernández C, Scanlon P, Peto T, Simó R. Screening for diabetic retinopathy: new perspectives and challenges. Lancet Diabetes Endocrinol. 2020 Apr;8(4):337-347. doi: 10.1016/S2213-8587(19)30411-5. Epub 2020 Feb 27. PMID: 32113513.

DOI: 10.59254/sbpo-2025-212499
https://proceedings.science/p/212499?lang=pt-br

Apêndice I

```python
def gurobi(x, y): #recebe conjuntos X e Y como parâmetro
    import gurobipy as grb #biblioteca para resolver o PPL

    ui, vj, a = [], [], [] #vetores para variáveis de decisão
    modelo = grb.Model() #criando modelo
    b = modelo.addVar(vtype = grb.GRB.CONTINUOUS)
        #variável do termo b do hiperplano

    for _ in range(len(x[0])): #variáveis do vetor a do hiperplano
        a.append(modelo.addVar(vtype = grb.GRB.CONTINUOUS))

    for _ in x: #percorrendo cada indivíduo do grupo X
        ui.append(modelo.addVar(vtype = grb.GRB.CONTINUOUS))
            #separation error de cada indivíduos com RD
        modelo.addConstr(ui[-1] >= 0)
            #restrição de não negatividade
        modelo.addConstr(ui[-1] >= -np.dot(a, _) + b + 1)
            #restrição do modelo proposto

    for _ in y: #percorrendo cada indivíduo do grupo Y
        vj.append(modelo.addVar(vtype = grb.GRB.CONTINUOUS))
            # separation error de cada indivíduo sem RD
        modelo.addConstr(vj[-1] >= 0)
            #restrição de não negatividade
        modelo.addConstr(vj[-1] >= np.dot(a, _) - b + 1)
            #restrição do modelo

    sum_ui = modelo.addVar(vtype = grb.GRB.CONTINUOUS)
        #variável do somatório dos ui
    sum_vj = modelo.addVar(vtype = grb.GRB.CONTINUOUS)
        #variável do somatório dos vj
    modelo.addConstr(sum_ui == grb.quicksum(ui))
        #variável do somatório dos ui
    modelo.addConstr(sum_vj == grb.quicksum(vj))
        #variável do somatório dos vj
    modelo.setObjective(sum_ui + sum_vj, grb.GRB.MINIMIZE)
        #objetivo de minimizar a separation error function
    modelo.optimize()
        #chamada à função para otimização na biblioteca
    return [i.x for i in modelo.getVars()[:1+len(x[0])]]
        #coeficientes do hiperplano
```

Algoritmo 1. Código em Python para modelagem e solução de (II) na Seção 2.

DOI: 10.59254/sbpo-2025-212499
https://proceedings.science/p/212499?lang=pt-br