

## APLICAÇÃO DA ESPECTROMETRIA DE MASSAS COM IONIZAÇÃO POR PAPER SPRAY E MÉTODO QUIMIOMÉTRICO OPLS-DA PARA O DIAGNÓSTICO DO CÂNCER DE PRÓSTATA PELA ANÁLISE DE AMOSTRAS DE URINA

Thifanny T. P. da Silva<sup>a</sup>, Ademar D. V. Máquina<sup>a,b,\*</sup>, Anízia F. F. Durans<sup>a</sup>, Lucas G. da Costa<sup>a,c</sup>, Frederico G. Pinto<sup>c</sup>, Timothy J. Garrett<sup>d</sup> e Waldomiro Borges Neto<sup>a</sup>

<sup>a</sup>Instituto de Química, Universidade Federal de Uberlândia, 38408-100 Uberlândia – MG, Brasil

<sup>b</sup>Departamento de Ciências Exactas e Tecnológicas, Universidade Púnguè, PO Box 2301 Tete, Moçambique

<sup>c</sup>Instituto de Química, Universidade Federal de Viçosa, 36570-900 Rio Paranaíba – MG, Brasil

<sup>d</sup>Southeast Center for Integrated Metabolomics, Clinical and Translational Science Institute and Department of Pathology, Immunology, and Laboratory Medicine, University of Florida, 32610 Gainesville-FL, USA

Recebido: 23/09/2023; aceito: 19/03/2024; publicado online: 06/06/2024

APPLICATION OF MASS SPECTROMETRY WITH PAPER SPRAY IONIZATION AND THE OPLS-DA CHEMOMETRIC METHOD FOR THE DIAGNOSIS OF PROSTATE CANCER BY ANALYSIS OF URINE SAMPLES. In this study, paper spray ionization mass spectrometry (PSI-MS) allied to orthogonal projections on the latent structure-discrimination analysis (OPLS-DA) was employed to differentiate urine samples from patients with prostate cancer and control individuals (without prostate cancer). The parameters of sensitivity, specificity, efficiency and Matthew's correlation coefficient determined to analyze the efficiency and quality of the OPLS-DA model obtained a value equal to 1, which means, respectively, that the model has a good performance to correctly classify true positive samples and true negatives. These results showed 100% correct classification of the samples from the training and test set, demonstrating a high efficiency of the OPLS-DA model when classifying all samples considered as belonging to the class of interest (urine from control patient) into class 1 and all samples considered as belonging to the non-interest class (urine from patients with prostate cancer) in class 0. This efficiency resulting from the application of this model suggests that the PSI-MS combined to OPLS-DA is a promisor approach for a rapid diagnosis of prostate cancer, via chemical fingerprints of urine samples.

Keywords: urine; PSI-MS; differentiation; OPLS-DA; prostate cancer.

### INTRODUÇÃO

Dentre vários tipos de doenças cancerígenas, o câncer de próstata (PCa) é considerado como sendo a que mais afeta aos homens com idade acima dos 50 anos, no mundo inteiro.<sup>1</sup> Nos últimos anos, tem-se registrado um aumento de casos desta doença, como por exemplo, nos EUA se estima registrar 288300 novos casos em 2023, contra 268490 em 2022 e 248530 em 2021;<sup>2-4</sup> para o Brasil, se registrou uma estimativa de 71730 mil novos casos em 2022.<sup>5</sup> O aumento observado nas taxas de incidência de novos casos também é acompanhado com certa taxa de mortalidade dos pacientes, estimados em 34700 em 2023 e 15841 de mortes em 2020, nos EUA e no Brasil, respectivamente.<sup>4,5</sup>

Alguns estudos apontam o diagnóstico precoce do câncer como sendo a etapa primordial para o sucesso do seu tratamento, pois possibilita detecção desta doença na fase inicial. Para tal, é comumente usado procedimentos clínicos de toque retal e o exame de sangue do antígeno prostático específico (PSA), mas a sua confirmação é feita através de exames de biópsia.<sup>1,6</sup> Estes procedimentos apesar de proporcionarem vantagens de rastrear a dinâmica evolutiva e a heterogeneidade dos tumores e de detectar o surgimento muito precoce de resistência à terapia,<sup>7</sup> também apresentam algumas desvantagens por serem desconfortáveis para o paciente, deixa-o susceptível à complicações de saúde como infecções, incontinência e disfunção erétil; apresentam alto índice de amostras negativas que são classificadas como positivas e de amostras positivas classificadas como negativa, proporcionando uma baixa especificidade.<sup>8</sup> Com isso, torna-se necessário o desenvolvimento de metodologia eficiente, rápida, reprodutível e não invasiva para o diagnóstico de PCa.

Atualmente, o uso da metabolômica para detectar alterações fisiológicas que denunciam síndromes específicas tem sido bastante estudado por diversos pesquisadores. Duangkumpha *et al.*<sup>9</sup> identificaram níveis elevados de açúcares plasmáticos e álcoois de açúcar em pacientes diabéticos *mellitus* com insuficiência renal através da análise metabolômica de dados de cromatografia gasosa bidimensional e espectrometria de massa por tempo de voo (GC × GC-TOFMS); Wang *et al.*<sup>10</sup> desenvolveram uma pesquisa patológica integrada para o diagnóstico preciso da esquizofrenia combinando metabolômica e transcriptômica por cromatografia líquida acoplada à espectrometria de massas (LC-MS) e espectroscopia de ressonância magnética nuclear de hidrogênio (<sup>1</sup>H NMR). Gutiérrez *et al.*<sup>11</sup> desenvolveram um modelo de previsão de estágios iniciais da doença renal em indivíduos com diabetes tipo 2 usando a metabolômica da urina por ressonância magnética nuclear. Outras vantagens do uso de dados metabolômicos são: para testar a toxicidade de substâncias propostas para fármacos; para a determinação de fenótipos gerado por manipulação genética; identificação de fatores de risco para doenças que afetam o sistema nervoso central, como por exemplo de Parkinson, Alzheimer e Huntington; identificação de substratos endógenos para as enzimas e identificação de marcadores para descoberta de drogas.<sup>12-14</sup>

Os dados metabolômicos podem ser obtidos através de diferentes técnicas analíticas como as cromatografias gasosa (GC) e líquida (LC) ambas acopladas à espectrometria de massas (MS), a espectroscopia de ressonância magnética nuclear (RMN) e a espectrometria de massas com ionização por *paper spray* (PSI-MS).<sup>12</sup> A PSI-MS chama a atenção na obtenção de perfis metabólicos de amostras biológicas, como de urina por exemplo, pois não exige em muitos casos o preparo prévio das amostras para as análises. A ionização por *paper spray* (PSI) é uma importante técnica de ionização ambiente

\*e-mail: ademarmaquina@gmail.com

para aplicação em espectrometria de massas. As análises por PSI-MS são feitas depositando as amostras em um pedaço de papel triangular, conectado a uma fonte de alta tensão e posicionado em frente ao orifício de entrada do espectrômetro de massas (MS).<sup>15,16</sup> Quando um solvente é aplicado e uma alta voltagem (tipicamente de 3-5 kV) é fornecida ao papel, um *spray* de gotículas carregadas é induzido na ponta do triângulo de papel. Na PSI, os analitos presentes nas amostras são extraídos pelo solvente e transportados para a ponta do papel por capilaridade. Quando uma alta tensão é aplicada, ocorre a formação do cone de Taylor na ponta do papel (semelhante ao que ocorre na ionização *electrospray*, ESI), resultando na produção de íons em fase gasosa e subsequente detecção por MS.<sup>17-19</sup> Os analitos presentes na amostra sofrem um equilíbrio de ionização, onde sua extensão é definida pelo pKa do analito e pelo pH do meio. A aquisição dos resultados é rápida, demonstrando ser uma técnica promissora para o diagnóstico de doenças.<sup>20</sup>

Dada a complexidade dos perfis destes dados devido a complexidade das matrizes biológicas, vários pesquisadores têm recorrido ao uso ferramentas quimiométricas para o seu tratamento e posterior análise.<sup>21-23</sup> Para o caso de detecção e classificação são sugeridos os métodos supervisionados de reconhecimento de padrões, como por exemplo: análise discriminante por quadrados mínimos parciais (PLS-DA); K-ésimo vizinhos mais próximos (KNN); modelagem independente e flexível por analogia de classe (SIMCA) e dados aplicados a modelagem independente e suave de analogia de classes (DD-SIMCA).<sup>8,24</sup> O DD-SIMCA apesar de ser construído na base de dados da classe de interesse, ele pode apresentar falhas na detecção de novas amostras da classe de interesse que apresentam pequenas alterações, além disso, não proporcionam perfil das variáveis importantes (VIP) que mais contribuem para a detecção.<sup>25</sup> Como alternativa, pode ser usado o modelo PLS-DA porque a sua construção é feita com base nos dados do analito de interesse, neste caso de amostras de urina de pacientes controle e de não interesse, amostras de urina de pacientes com câncer de próstata.<sup>24</sup> Para melhorar os resultados do modelo PLS-DA, faz-se uma análise discriminante por projeções ortogonais à estrutura latente (OPLS-DA) que remove a variação ortogonal-Y na matriz de dados.<sup>26,27</sup>

Não há trabalhos publicados na literatura que relatem o emprego da PSI-MS com a ferramenta quimiométrica OPLS-DA aos dados de metabolômica para diagnosticar o câncer de próstata. Assim, o presente trabalho teve como objetivo o uso da técnica PSI-MS aliada a OPLS-DA para a diferenciação de amostras de urina de pacientes com câncer de próstata em relação a pacientes sem essa doença (grupo controle) pela análise do perfil metabólico dessas amostras.

## PARTE EXPERIMENTAL

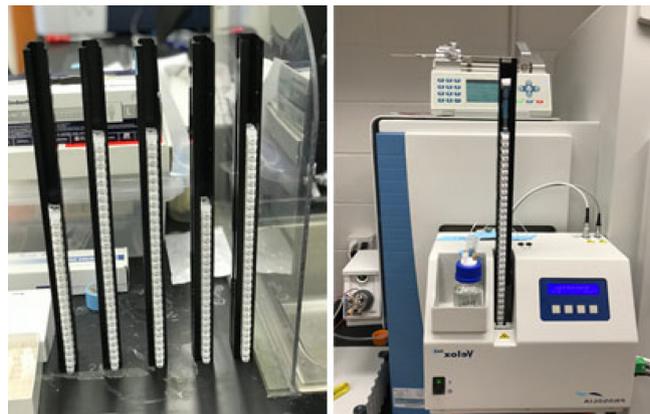
### Reagentes

Os reagentes utilizados, acetonitrila (ACN), metanol (MeOH), água (H<sub>2</sub>O) e ácido fórmico (FA), foram da marca Optima adquiridos da Thermo Fisher Scientific (Fairlawn, NJ, EUA). Todos os solventes utilizados neste trabalho foram grau HPLC 99,9%.

### Coleta e preparo de amostras

As redes SPORE de pesquisa em câncer de próstata (University of Northwestern, Chicago, EUA e Life Study da University of Florida, Gainesville, EUA) forneceram amostras de urina de pacientes com câncer de próstata (n = 40) e amostras de urina de indivíduos controle sem histórico médico de câncer (n = 40), respectivamente. Estas amostras foram coletadas usando tubos de preservação de urina (Norgen Bioteck, Thorold, Canadá) e armazenadas a -80 °C.

Para análise PSI-MS, as amostras foram descongeladas em temperatura ambiente e depositadas (15 µL) em cartuchos Velox de papel triangular pré-cortados, obtidos na Prosofia Inc. (Indianapolis, EUA). A reprodutibilidade na dispensação e posicionamento da amostra foi garantida com o uso de um dispositivo impresso em 3D (Prosofia Inc.). Então, 80 µL de MeOH:H<sub>2</sub>O 3:7 contendo 0,1% de ácido fórmico (v/v/v) foi usado como solvente de pulverização. A fonte de ionização *paper spray* utilizada neste trabalho é uma fonte comercial (Figura 1).



**Figura 1.** Ilustração dos racks com os cartuchos contendo o papel amostrador e o instrumento PSI-MS que foram utilizados neste trabalho

### Aquisição de dados espectrais e análise quimiométrica

Os dados espectrais foram obtidos em um espectrômetro de massas de alta resolução Thermo Scientific Q Exactive MS (Mundelein, EUA), usando o modo de varredura completa (ionização modo positivo) com uma faixa de massa de  $m/z$  70-1000 com resolução de massa de 140 000 em  $m/z$  200. A temperatura do capilar foi de 270 °C, lentes-S em 30 e uma voltagem para ionização por *paper spray* de 3,5 kV. O tempo de aquisição de sinais foi de 0,5 min. Os arquivos de dados metabolômicos foram convertidos para o formato de arquivo mzXML usando RawConverter<sup>28</sup> e o alinhamento dos picos foi realizado usando MZmine 2.<sup>29</sup> Espécies não detectadas com intensidade zero foram substituídas por metade do valor mínimo de todas as características detectadas. Os picos com contribuição de intensidade de sinal  $\geq 10\%$  do sinal de fundo foram removidos com base na filtração de recursos em branco (*blank feature filtration*, BFF). A normalização dos picos filtrados foi realizada para a corrente total de íons, log transformado e auto escalamento.<sup>8</sup>

Os softwares MATLAB<sup>30</sup> e PLS\_Toolbox<sup>31</sup> foram utilizados para pré-processar os dados e desenvolver o modelo OPLS-DA. Para realizar os procedimentos multivariados, os dados dos espectros de PSI-MS foram organizados em uma matriz ordenada de linhas e colunas, denominado de matriz **X**, onde cada linha corresponde a uma amostra “m” e cada coluna corresponde a uma variável “n”, na qual  $m = 1, 2, 3, \dots, 80$  e  $n = 1, 2, 3, \dots, 800$ . Um vetor **Y** foi criado contendo valores inteiros discretos 0 e 1 para indicar a classe não de interesse e de interesse, respectivamente. Os dados da matriz **X** foram normalizados e pré-processados usando OSC (*orthogonal signal correction*) para minimizar as variações sistemáticas indesejáveis e produzir modelos menos complexos com melhor poder preditivo.<sup>32,33</sup>

O OPLS-DA usa informações do vetor **y** para decompor a matriz **X** em três partes distintas, conforme descrito na Equação 1, onde **T<sub>p</sub>** é a matriz de escores preditiva para **X**, **P<sub>p</sub>** é a matriz de peso preditiva para **X**, **T<sub>o</sub>** é a matriz de escores Y-ortogonal correspondente, **P<sub>o</sub>** é a matriz de peso dos componentes Y-ortogonais, e **E** é a matriz residual de **X**.<sup>34,35</sup>

$$\mathbf{X} = \mathbf{T}_p \mathbf{P}_p^T + \mathbf{T}_o \mathbf{P}_o^T + \mathbf{E} \quad (1)$$

Para classificação, a variação ortogonal- $\mathbf{Y}$  é removida da matriz de dados  $\mathbf{X}$ , de acordo com a Equação 2, onde  $\mathbf{T}_o$  é a matriz de escores  $\mathbf{Y}$ -ortogonal para  $\mathbf{X}$  e  $\mathbf{P}_o$  é a matriz de pesos para  $\mathbf{Y}$ -ortogonal.

$$\mathbf{X}_p = \mathbf{X} - \mathbf{T}_o \mathbf{P}_o^T \quad (2)$$

A matriz  $\mathbf{Y}$  prevista é estimada usando o  $\mathbf{X}_p$  atualizado e os componentes preditivos do modelo OPLS-DA, que são estimados a partir do conjunto de dados de calibração.<sup>34</sup> O número de variáveis latentes (VLs) para o modelo OPLS-DA foi escolhido pelo critério de menor erro de previsão na validação cruzada *leave-7-out*. Na validação do modelo, uma nova matriz  $\mathbf{X}$  (construída com amostras que não foram usadas no treinamento) foi correlacionada com o vetor de regressão obtido no treinamento para prever as suas respectivas classes.

Idealmente, os valores previstos pelo modelo OPLS-DA seriam 0 ou 1, mas na prática os valores previstos são próximos de 0 e 1. Para separar as classes, calcula-se um valor de limiar (*threshold*) entre 0 e 1 de acordo com teorema de Bayes, segundo o qual, o limiar assume que os valores de  $\mathbf{Y}$ -previstos seguem uma distribuição semelhante ao que vai ser observado para as amostras futuras. Quando as duas distribuições estimadas se cruzam, é selecionado um limite, isto é, um valor de  $\mathbf{Y}$  em que o número de falsos positivos e falsos negativos deve ser minimizado para previsões futuras.<sup>36</sup> Assim, com o limiar calculado, a amostra é considerada como pertencente à classe 0 (não de interesse) quando apresenta valor previsto abaixo do limiar e considerada pertencente à classe 1 (de interesse) quando apresenta valor previsto acima do limiar. A probabilidade de uma amostra negativa ser classificada como positiva, designada de taxa de falso positivo (FP) é calculada de acordo com a Equação 3 e a probabilidade de uma amostra positiva ser classificada como negativa, designada de taxa de falso negativo (FN) é calculada de acordo com a Equação 4.<sup>37,38</sup>

$$FP = \frac{fp}{fp + vn} \times 100 \quad (3)$$

$$FN = \frac{fn}{fn + vp} \times 100 \quad (4)$$

onde,  $fp$  é o número de amostras falso positivas  $vn$  é o número de amostras verdadeira negativas,  $fn$  é o número de amostras falso negativas e  $vp$  é o número de amostras verdadeira positivas.

A capacidade que o modelo OPLS-DA tem em classificar corretamente as amostras verdadeira positivas e verdadeira negativas é verificada com a determinação dos parâmetros sensibilidade e especificidade, de acordo com as Equações 5 e 6, respectivamente.<sup>39</sup> Outros parâmetros utilizados para avaliar o desempenho do modelo são a eficiência e coeficiente de correlação de Matthews (*Matthews correlation coefficient*), calculados com base nas Equações 7 e 8, respectivamente. O coeficiente de correlação de Matthew pode ter valores entre  $-1$  e  $+1$ , onde o valor  $+1$  representa uma classificação perfeita, zero representa uma classificação aleatória e  $-1$  representa uma classificação inversa.

$$\text{Sensibilidade} = \frac{vp}{vp + fn} \quad (5)$$

$$\text{Especificidade} = \frac{vn}{vn + fp} \quad (6)$$

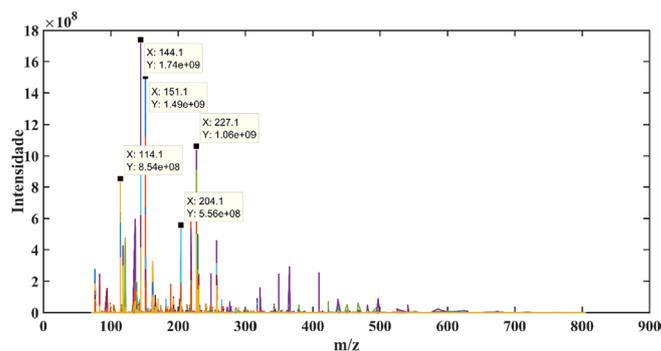
$$\text{Coeficiente de Matthews} = \frac{(vp \times vn) - (fp \times fn)}{\sqrt{(vp + fp) \times (vp + fn) \times (vn + fp) \times (vn + fn)}} \quad (7)$$

$$\text{Eficiência} = \frac{vp + vn}{vp + vn + fp + fn} \quad (8)$$

## RESULTADOS E DISCUSSÃO

### Espectros PSI-MS

A Figura 2 mostra os espectros PSI-MS obtidos pelas análises das amostras de urina dos pacientes controle e com câncer de próstata. Alguns picos foram identificados utilizando a massa exata e sua comparação com o banco de dados da plataforma CEU *Mass Mediator* seguindo a iniciativa de padrões metabolômicos (MSI).<sup>40-42</sup> Os compostos anotados estão apresentados na Tabela 1.



**Figura 2.** Espectros PSI-MS de 80 amostras de urina de pacientes controle ( $n = 40$ ) e com câncer de próstata ( $n = 40$ )

**Tabela 1.** Relação de compostos anotados na urina dos pacientes controle e com câncer de próstata pela análise por PSI-MS

$m/z$	Anotação	Íon/aduto	Erro / ppm
132,0767	creatina	[M + H] <sup>+</sup>	0
114,0663	creatinina	[M + H] <sup>+</sup>	1
215,1387	detiobiotina	[M + H] <sup>+</sup>	1
106,0866	dietanolamina	[M + H] <sup>+</sup>	3
145,0477	eritritol	[M + H] <sup>+</sup>	5
118,0612	guanidinoacetato	[M + H] <sup>+</sup>	1
137,0457	hipoxantina	[M + H] <sup>+</sup>	1
232,1543	isobutirilcarnitina	[M + H] <sup>+</sup>	0
175,1188	L-arginina	[M + H] <sup>+</sup>	1
162,1123	L-carnitina	[M + H] <sup>+</sup>	1
147,0764	L-glutamina	[M + H] <sup>+</sup>	0
156,0767	L-histidina	[M + H] <sup>+</sup>	0
132,1019	L-isoleucina	[M + H] <sup>+</sup>	0
106,0502	L-serina	[M + H] <sup>+</sup>	3
229,1545	Leu-pro	[M + H] <sup>+</sup>	1
217,1294	N-acetil-arginina	[M + H] <sup>+</sup>	1
131,1179	N-acetilputrescina	[M + H] <sup>+</sup>	0
137,0709	N-metilnicotinamida	[M + H] <sup>+</sup>	0
166,0861	fenilalanina	[M + H] <sup>+</sup>	1
168,0654	piridoxal	[M + H] <sup>+</sup>	1
173,0212	Sn-glicerol 3-fosfato	[M + H] <sup>+</sup>	1
126,0221	taurina	[M + H] <sup>+</sup>	1
138,0547	trigonelina	[M + H] <sup>+</sup>	2

## Análise quimiométrica

Para a discriminação das amostras de urina dos pacientes controle daqueles com câncer de próstata, o modelo OPLS-DA foi validado utilizando as amostras do conjunto de teste. Os resultados dos parâmetros de classificação obtidos por este modelo estão na Tabela 2, onde se pode verificar que os valores do erro quadrático médio de calibração (RMSEC) e erro quadrático médio de previsão (RMSEP) são relativamente baixos e considerados concordantes, sendo o RMSEP superior ao RMSEC. O modelo apresentou um bom ajuste com duas variáveis latentes, com o coeficiente de determinação ( $R^2$ ), observado na Figura 3, superior a 0,7, evidenciando que o modelo explica e prevê bem os valores atuais e previstos dos conjuntos de treinamento e teste. Nesta figura, também se observa que o parâmetro  $Q^2$  (coeficiente de correlação entre os dados originais e os dados previstos pela validação cruzada) foi maior que 0,5, o que equivale a dizer que o modelo OPLS-DA é robusto e apresenta boa capacidade preditiva.

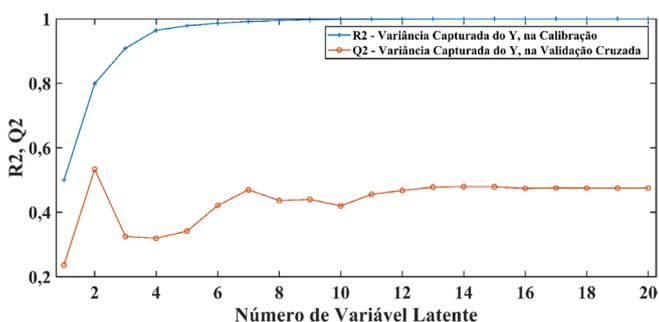


Figura 3. Parâmetros  $R^2$  e  $Q^2$  do modelo OPLS-DA desenvolvido com duas componentes

Tabela 2. Parâmetros de classificação obtidos pelo modelo OPLS-DA

Parâmetros	Valores
RMSECV / (% v/v)	0,3416
RMSEC / (% v/v)	0,2241
RMSEP / (% v/v)	0,3569

RMSECV: erro quadrático médio de validação cruzada; RMSEC: erro quadrático médio de calibração; RMSEP: erro quadrático médio de previsão.

A Tabela 3 apresenta uma matriz de confusão do modelo, onde pode-se observar valores das taxas de vp e vn igual a 1 e valores das taxas de fp e fn igual a 0 nos conjuntos de treinamento e de teste que contém amostras de urina de pacientes controle e com câncer de próstata. Isso significa que no processo de treinamento,

Tabela 3. Matriz de confusão e tabela de confusão do conjunto de treinamento e teste

	Matriz de confusão						Tabela de confusão			
	vp	fp	vn	fn	N	Erro	Classe 0	Classe 1		
Treinamento	Classe 0	1	0	1	0	24	0	Previsto como classe 0	24	0
	Classe 1	1	0	1	0	24	0	Previsto como classe 1	0	24
	Eficiência = 1 Coeficiente de correlação de Matthews = 1							Previsto como não atribuído	0	0
Teste	Classe 0	1	0	1	0	16	0	Previsto como classe 0	16	0
	Classe 1	1	0	1	0	16	0	Previsto como classe 1	0	16
	Eficiência = 1 Coeficiente de correlação de Matthews = 1							Previsto como não atribuído	0	0

vp: número de amostras verdadeira positivas; fp: número de amostras falso positivas; vn: número de amostras verdadeira negativas; fn: número de amostras falso negativas; N: número de amostras.

todas as 24 amostras de urina de pacientes controle foram previstas na classe 1 e nenhuma amostra de urina de pacientes com câncer foi prevista na classe 1, assim como todas as 24 amostras de urina de pacientes com câncer foram previstas na classe 0 e nenhuma amostra de urina de pacientes controle foi prevista na classe 0; no processo de teste, todas as 16 amostras de urina de pacientes controle foram previstas na classe 1 e nenhuma amostra de urina de pacientes com câncer foi prevista na classe 1, assim como todas as 16 amostras de urina de pacientes com câncer foram previstas na classe 0 e nenhuma amostra de urina de pacientes controle foi prevista na classe 0, como pode ser visto na tabela de confusão (Tabela 3).

A eficiência do modelo, calculada de acordo com a Equação 8, foi 1, significando um bom ajuste e desempenho do modelo. O coeficiente de correlação de Matthews resumiu a qualidade da tabela da matriz de confusão em um único valor numérico de 1 para os processos de treinamento e teste, representando uma classificação perfeita.

A Tabela 4 apresenta os resultados dos parâmetros sensibilidade e especificidade do modelo para o processo de treinamento, validação cruzada e teste, determinados com base nas Equações 5 e 6, respectivamente. Observa-se nesta tabela, que o valor da sensibilidade é igual a 1, o que significa que o modelo classificou corretamente todas as amostras verdadeiro positivo e o valor da especificidade também foi igual a 1, significando que o modelo classificou corretamente todas as amostras verdadeiro negativo. Esse valor de 1 também pode ser observado na curva de características operacionais do receptor (ROC), mostrada na Figura 4, que plota a sensibilidade e especificidade como funções de limite para separação de classe de interesse (amostra de urina de pacientes controle) e de não interesse (amostra de urina de pacientes com câncer de próstata).

Os resultados apresentados nas Tabelas 2, 3 e 4 mostram que o modelo OPLS-DA desenvolvido possui boa eficiência na classificação das amostras. Essa eficiência de classificação é observada na Figura 5, onde as amostras da classe de interesse (urina de pacientes controle) dos conjuntos de treinamento e teste estão acima do *threshold*, separando-se das amostras da classe não de interesse (urina de pacientes com câncer de próstata) do conjunto de treinamento e teste que estão abaixo do *threshold*.

A Figura 6 mostra a seleção do *threshold* em 0,4589 quando as distribuições estimadas da classe 1 (interesse) e classe 0 (não de interesse) se cruzam no  $Y$ -previsto, onde o número de falsos negativos e falsos positivos é minimizado para previsões futuras.

A Figura 7 mostra os pesos das duas variáveis latentes (VLs) escolhidas para o desenvolvimento do modelo OPLS-DA que selecionam as regiões espectrais que contribuíram para a diferenciação entre as classes. A primeira variável latente (VL-1) mostrada na Figura 7a, descreve 17,71% de toda variação definida por classe não de interesse,

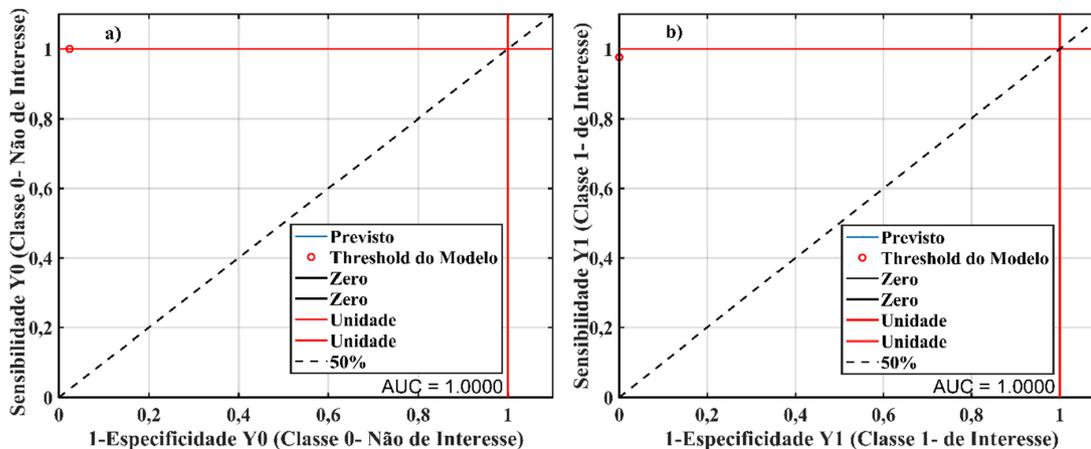


Figura 4. Curva ROC para (a) sensibilidade versus especificidade das amostras de urina de pacientes controle; (b) sensibilidade versus especificidade das amostras de urina de pacientes com câncer de próstata

Tabela 4. Valores dos parâmetros sensibilidade e especificidade obtidos pelo modelo OPLS-DA

Parâmetro	Valor		
	Calibração	Validação cruzada	Previsão
Sensibilidade	1	1	1
Especificidade	1	1	1

encontrada nas regiões de  $m/z$  200 a 230. A segunda variável latente (VL-2) mostrada na Figura 7b, descreve 17,14% de toda variação definida por classe de interesse, encontrada nas regiões de  $m/z$  110 a 155. No geral, as duas VLs aplicadas ao modelo descreveram 34,85% das variáveis que influenciaram a discriminação das classes de amostras de urina de pacientes controle e com câncer de próstata.

Os resultados dos testes de permutação usados para verificar o sobreajuste do modelo são mostrados na Tabela 5, na qual se observa que os valores do teste de classificação assinado em pares Wilcoxon (Wilcoxon), teste de classificação assinada em pares (*Sign test*) e Teste-*t* de randomização (*Random t-test*) são menores que 0,05, o que equivale a dizer que o modelo OPLS-DA não está sobreajustado.

CONCLUSÕES

A aplicação da ferramenta quimiométrica OPLS-DA para tratar os dados metabômicos obtidos por PSI-MS, permitiu a discriminação de amostras de urina de pacientes com câncer de próstata em relação a pacientes sem essa doença. A performance da abordagem proposta neste trabalho, foi analisada com base nos parâmetros de sensibilidade, especificidade, eficiência e coeficiente

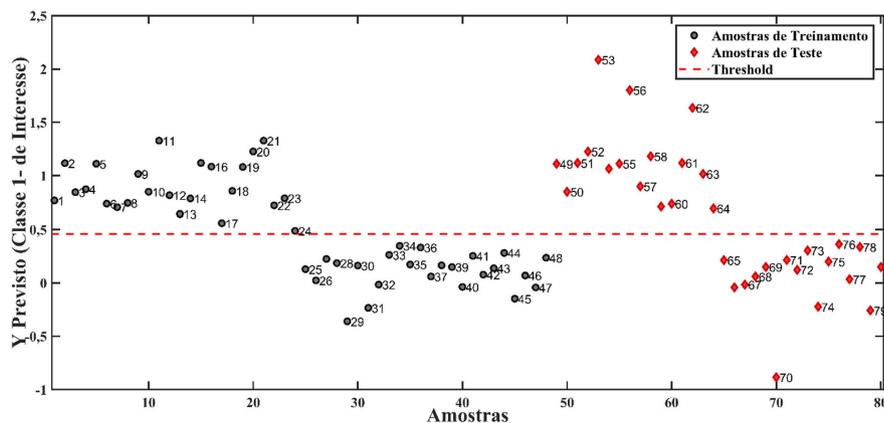


Figura 5. Discriminação entre amostras de urina de pacientes controle e com câncer de próstata dos conjuntos de treinamento e de teste

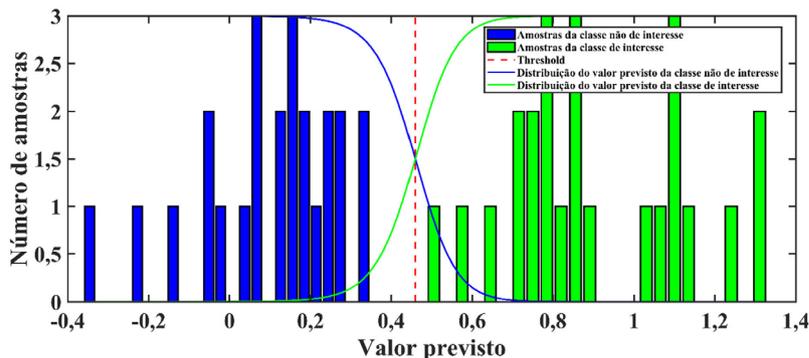


Figura 6. Seleção do valor de threshold que separa as classes de interesse e não de interesse

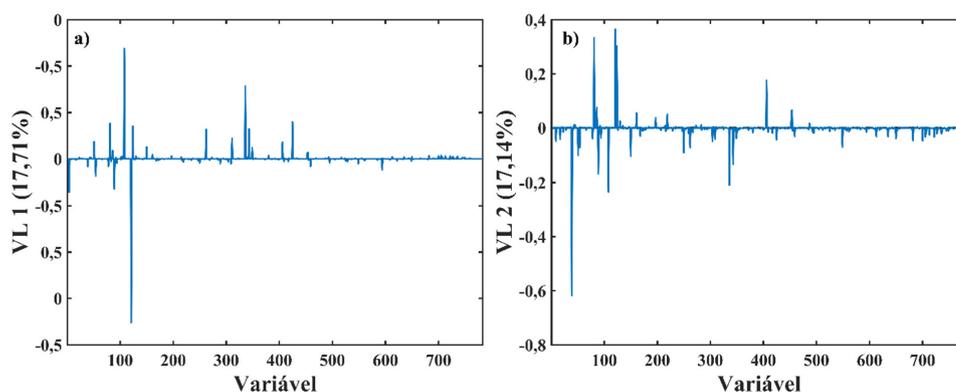


Figura 7. Pesos das VL-1 e VL-2 versus variáveis do modelo OPLS-DA

Tabela 5. Probabilidade de insignificância do modelo vs. amostras permutadas para o modelo com 2 componentes

	Wilcoxon	Sign test	Random t-test
Coluna-Y: 1			
Auto-previsão (RMSEC)	0,000	0,000	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,005
Coluna-Y: 2			
Auto-previsão (RMSEC)	0,000	0,000	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,005

RMSECV: erro quadrático médio de validação cruzada; RMSEC: erro quadrático médio de calibração.

de correlação de Matthews. Esses parâmetros levaram a 100% de acerto na classificação das amostras, tanto as de treinamento quanto as de teste, evidenciando o grande potencial dessa abordagem para discriminar o perfil metabólico de indivíduos acometidos pelo câncer de próstata pela análise da urina do paciente. Apesar do excelente resultado obtido na classificação das amostras, estudos posteriores a esse, utilizando um maior número de amostras, devem ser conduzidos para comprovar a eficiência dessa abordagem. Além disso, fica como sugestão a implementação dessa mesma abordagem em trabalhos posteriores para a diferenciação dos estágios do câncer, além do acompanhamento do tratamento dessa doença e do seu diagnóstico precoce pela análise da urina do paciente.

## AGRADECIMENTOS

Os autores agradecem ao National Institutes of Health (grant U24DK097209) e Programa de Convênio Estudantil PEC-PG (processo No. 190025/2017-4) pelo apoio financeiro e ao Prosolia pela doação da fonte de ionização *paper spray*. Agradecemos também ao CNPq (processo 308264/2022-3 e à FAPEMIG (processo APQ-03833-18) pelo apoio financeiro.

## REFERÊNCIAS

- Burns, D.; Anokian, E.; Saunders, E. J.; Bristow, R. G.; Fraser, M.; Reimand, J.; Schlomm, T.; Sauter, G.; Brors, B.; Korbel, J.; Weischenfeldt, J.; Waszak, S. M.; Corcoran, N. M.; Jung, C. H.; Pope, B. J.; Hovens, C. M.; Cancel-Tassin, G.; Cussenot, O.; Loda, M.; Sander, C.; Hayes, V. M.; Sorensen, K. D.; Lu, Y. J.; Hamdy, F. C.; Foster, C. S.; Gnanapragasam, V.; Butler, A.; Lynch, A. G.; Massie, C. E.; Woodcock, D. J.; Cooper, C. S.; Wedge, D. C.; Brewer, D. S.; Kote-Jarai, Z.; Eeles, R. A.; *Eur. Urol.* **2022**, *82*, 201. [Crossref]
- Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A.; *Ca-Cancer J. Clin.* **2021**, *71*, 7. [Crossref]
- Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A.; *Ca-Cancer J. Clin.* **2022**, *72*, 7. [Crossref]
- Siegel, R. L.; Miller, K. D.; Wagle, N. S.; Jemal, A.; *Ca-Cancer J. Clin.* **2023**, *73*, 17. [Crossref]
- Instituto Nacional de Câncer (INCA), *Estatísticas de Câncer*, <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros>, acessado em Maio 2024.
- Andersson, J.; Palsdottir, T.; Lantz, A.; Aly, M.; Grönberg, H.; Egevad, L.; Eklund, M.; Nordström, T.; *Eur. Urol., Suppl.* **2022**, *44*, 69. [Crossref]
- Heitzer, E.; Roberts, C. E. S.; Speicher, M. R.; *Nat. Rev. Genet.* **2019**, *20*, 71. [Crossref]
- Pinto, F. G.; Mahmud, I.; Rubio, V. Y.; Máquina, A. D. V.; Durans, A. F. F.; Borges Neto, W.; Garrett, T. J.; *Anal. Chem.* **2022**, *94*, 1925. [Crossref]
- Duangkumpha, K.; Jariyasopit, N.; Wanichthanarak, K.; Dhakal, E.; Wisanpitayakorn, P.; Thotsiri, S.; Sirivatanauksorn, Y.; Kitiyakara, C.; Sathirapongsasuti, N.; Khoomrung, S.; *J. Biol. Chem.* **2022**, *298*, 102445. [Crossref]
- Wang, T.; Li, P.; Meng, X.; Zhang, J.; Liu, Q.; Jia, C.; Meng, N.; Zhu, K.; Lv, D.; Sun, L.; Shang, T.; Lin, Y.; Niu, W.; Lin, S.; *Clin. Chim. Acta* **2022**, *524*, 84. [Crossref]
- Lucio-Gutiérrez, J. R.; Cordero-Pérez, P.; Fariás-Navarro, I. C.; Tijerina-Marquez, R.; Sánchez-Martínez, C.; Ávila-Velázquez, J. L.; García-Hernández, P. A.; Nández-Terreros, H.; Coello-Bonilla, J.; Pérez-Trujillo, M.; Parella, T.; Torres-González, L.; Waksman-Minsky, N. H.; Saucedo, A. L.; *J. Pharm. Biomed. Anal.* **2022**, *219*, 114885. [Crossref]
- Alarcon-Barrera, J. C.; Kostidis, S.; Ondo-Mendez, A.; Giera, M.; *Drug Discovery Today* **2022**, *27*, 1763. [Crossref]
- Robertson, D. G.; *Toxicol. Sci.* **2005**, *85*, 809. [Crossref]
- Quintero, M. E.; Pontes, J. G. M.; Tasic, L.; *Brain Res.* **2021**, *1773*, 147704. [Crossref]
- Su, Y.; Ma, X.; Ouyang, Z.; *Int. J. Mass Spectrom.* **2018**, *434*, 233. [Crossref]
- Vandergrift, G. W.; Gill, C. G.; *J. Mass Spectrom.* **2019**, *54*, 729. [Crossref]
- Frey, B. S.; Damon, D. E.; Tawiah, A. K. B.; *Mass Spectrom. Rev.* **2020**, *39*, 336. [Crossref]
- Maher, S.; Jjunju, F. P. M.; Damon, D. E.; Gorton, H.; Mah, Y. S.; Syed, S. U.; Heeren, R. M. A.; Young, I. S.; Taylor, S.; Badu-Tawiah, A. K.; *Sci. Rep.* **2016**, *6*, 35643. [Crossref]
- Wang, H.; Manicke, N. E.; Yang, Q.; Zheng, L.; Shi, R.; Cooks, R. G.; Ouyang, Z.; *Anal. Chem.* **2011**, *83*, 1197. [Crossref]
- Mahmud, I.; Pinto, F. G.; Rubio, V. Y.; Lee, B.; Pavlovich, C. P.; Perera, R. J.; Garrett, T. J.; *Anal. Chem.* **2021**, *93*, 7774. [Crossref]
- Raja, G.; Jung, Y.; Jung, S. H.; Kim, T. J.; *Process Biochem.* **2020**, *99*, 112. [Crossref]
- Deev, V.; Solovieva, S.; Andreev, E.; Protoshchak, V.; Karpushchenko, E.; Sleptsov, A.; Kartsova, L.; Bessonova, E.; Legin, A.; Kirsanov, D.;

- J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2020**, *1155*, 122298. [Crossref]
23. Drivelos, S. A.; Danezis, G. P.; Halagarda, M.; Popek, S.; Georgiou, C. A.; *Food Chem.* **2021**, *338*, 127936. [Crossref]
24. Ferreira, M. M. C.; *Quimiometria: Conceitos, Métodos e Aplicações*, 1ª ed.; Editora da Unicamp: Campinas, 2015.
25. Zontov, Y. V.; Rodionova, O. Y.; Kucheryavskiy, S. V.; Pomerantsev, A. L.; *Chemom. Intell. Lab. Syst.* **2017**, *167*, 23. [Crossref]
26. Máquina, A. D. V.; Siteo, B. V.; Cruz, W. O.; Santos, D. Q.; Borges Neto, W.; *Talanta* **2020**, *209*, 120590. [Crossref]
27. Máquina, A.; *Determinação do Teor de Biodiesel de Mafurra e Crambe em Misturas com Diesel por RMN de 1H e Regressão Multivariada por OPLS E OPLS-DA*; Tese de Doutorado, Universidade Federal de Uberlândia, Uberlândia, Brasil, 2020. [Crossref]
28. He, L.; Diedrich, J.; Chu, Y. Y.; Yates, J. R.; *Anal. Chem.* **2015**, *87*, 11361. [Crossref]
29. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M.; *BMC Bioinf.* **2010**, *11*, 395. [Crossref]
30. *Matlab®*, version R2018b; The MathWorks Inc., Natick, MA, USA, 2018.
31. *PLS\_Toolbox*, versão 9.2; Eigenvector Research, Inc., Manson, WA, USA, 2023.
32. Westerhuis, J. A.; de Jong, S.; Smilde, A. K.; *Chemom. Intell. Lab. Syst.* **2001**, *56*, 13. [Crossref]
33. Pers, T. H.; Martin, F. P.; Verdich, C.; Holst, C.; Johansen, J. V.; Astrup, A.; Polak, J.; Martinez, J. A.; Rezzi, S.; Blaak, E. E.; Saris, W. H. M.; Kochhar, S.; Macdonald, I. A.; Sørensen, T. I. A.; Ramadan, Z.; *Chemom. Intell. Lab. Syst.* **2008**, *93*, 34. [Crossref]
34. Max, B.; Mattias, R.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J.; *J. Chemom.* **2006**, *20*, 341. [Crossref]
35. Trygg, J.; Wold, S.; *J. Chemom.* **2002**, *16*, 119. [Crossref]
36. Wise, B. M.; Shaver, J. M.; Gallagher, N. B.; Rasmus, W. W.; Koch, B. R. S.; *PLS\_Toolbox 4.0 for use with MATLAB™*, Eigenvector Research, Inc., Wenatchee, WA, USA, 2006.
37. Wehrens, R.; Putter, H.; Buydens, L. M.; *Chemom. Intell. Lab. Syst.* **2000**, *54*, 35. [Crossref]
38. Rebafka, T.; Cléménçon, S.; Feinberg, M.; *Chemom. Intell. Lab. Syst.* **2007**, *89*, 69. [Crossref]
39. Xu, L.; Cai, C. B.; Cui, H. F.; Ye, Z. H.; Yu, X. P.; *Meat Sci.* **2012**, *92*, 506. [Crossref]
40. Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W. M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R.; *Metabolomics* **2007**, *3*, 211. [Crossref]
41. Mondul, A. M.; Moore, S. C.; Weinstein, S. J.; Karoly, E. D.; Sampson, J. N.; Albanes, D.; *Int. J. Cancer* **2015**, *137*, 2124. [Crossref]
42. Zeisel, S. H.; Warrier, M.; *Annu. Rev. Nutr.* **2017**, *37*, 157. [Crossref]