



**SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA - PPGQUI**

ANIZIA FAUSTA FURTADO DURANS

**APLICAÇÃO DE MÉTODOS QUIMIOMÉTRICOS E ESPECTROMETRIA DE
MASSA PARA DETECTAR CÂNCER DE PRÓSTATA ATRAVÉS DA URINA**

Uberlândia
2022

ANIZIA FAUSTA FURTADO DURANS

**APLICAÇÃO DE MÉTODOS QUIMIOMÉTRICOS E ESPECTROMETRIA DE
MASSA PARA DETECTAR CÂNCER DE PRÓSTATA ATRAVÉS DA URINA**

Dissertação apresentada ao Programa de Pós-Graduação em Química do Instituto de Química da Universidade Federal de Uberlândia, como requisito para obtenção de título de Mestra em Química.

Área de Concentração: Química Analítica

Orientador: Prof. Dr. Waldomiro Borges Neto

Uberlândia
2022

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

D949a Durans, Anizia Fausta Furtado, 1979-
2022 Aplicação de métodos quimiométricos e espectrometria de massa para detectar câncer de próstata através da urina [recurso eletrônico] / Anizia Fausta Furtado Durans. - 2022.

Orientador: Waldomiro Borges Neto.
Dissertação (Mestrado) - Universidade Federal de Uberlândia.
Programa de Pós-Graduação em Química.
Modo de acesso: Internet.
Disponível em: <http://doi.org/10.14393/ufu.di.2022.5048>
Inclui bibliografia.
Inclui ilustrações.

1. Química. I. Borges Neto, Waldomiro, 1970-, (Orient.). II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Química. III. Título.

CDU: 54

André Carlos Francisco
Bibliotecário - CRB-6/3408



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
 Coordenação do Programa de Pós-Graduação em Química
 Av. João Naves de Ávila, 2121, Bloco 5I - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
 Telefone: (34) 3239-4385 - www.cpgquimica.iq.ufu.br - cpgquimica@ufu.br



ATA

Programa de Pós-Graduação em:	Química				
Defesa de:	Dissertação de Mestrado Acadêmico, 344, PPGQUI				
Data:	Vinte de junho de dois mil e vinte e dois	Hora de início:	9:00	Hora de encerramento:	12:25
Matrícula do Discente:	12022QMI002				
Nome do Discente:	Anizia Fausta Furtado Durans				
Título do Trabalho:	Aplicação de métodos quimiométricos e espectrometria de massa para detectar câncer de próstata através da urina				
Área de concentração:	Química				
Linha de pesquisa:	Instrumentação Analítica e Preparo de Amostras				
Projeto de Pesquisa de vinculação:	Aplicação de Métodos Quimiométricos				

Reuniu-se por meio de webconferência, Plataforma Google Meet, links <https://meet.google.com/tbm-hsmx-fhb>, <https://meet.google.com/cqb-ewtd-ios> e <https://meet.google.com/rqh-yxxf-axa>, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Química, assim composta: Professores Doutores: José Augusto da Col, da Universidade do Estado de Santa Catarina; Regina Massako Takeuchi, da Universidade Federal de Uberlândia e Waldomiro Borges Neto, orientador da candidata.

Iniciando os trabalhos o presidente da mesa, Dr. Waldomiro Borges Neto, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu à Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovada.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.



Documento assinado eletronicamente por **Waldomiro Borges Neto, Professor(a) do Magistério Superior**, em 20/06/2022, às 12:28, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Regina Massako Takeuchi, Professor(a) do Magistério Superior**, em 20/06/2022, às 12:28, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **José Augusto da Col, Usuário Externo**, em 20/06/2022, às 13:52, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3616659** e o código CRC **78DD7DB1**.

Dedico essa dissertação de mestrado a Deus, essa força maior, que me guia, a minha mãe Raimunda Furtado, essa mulher guerreira que me ensinou tudo que sei e estimula as minhas maiores virtudes e ao meu orientador, prof. Dr. Waldomiro Borges Neto que instigou a ser a melhor versão de mim, do início ao fim desse processo.

*Porque Dele, por Ele e para Ele são todas
as coisas. A Ele a glória por toda a eternidade! Amém.”*

Romanos 11, 36

AGRADECIMENTOS

Agradeço,

A Deus, pois sem a Sua graça não seria capaz de alcançar a conclusão deste trabalho e por ter me concedido a graça de estar diante de um grande milagre, que foi ver meu irmão acordar de um quadro de irreversibilidade.

À Universidade Federal de Uberlândia e aos seus funcionários por ter me proporcionado a estrutura necessária para que pudesse crescer academicamente e pessoalmente.

Ao meu orientador Dr. Waldomiro Borges Neto, a quem agradeço imensamente por ter despendido seu tempo para ensinar-me e por ter tido o privilégio de conviver com um doutor de profundo conhecimento.

Aos colegas do Laboratório de Quimiometria do Triângulo: Ademar Viagem Máquina, Lucas Gustavo da Costa, Carlos Alface, Rodrigo Santos, Thifanny Tiburcio, que compartilharam importantes conhecimentos comigo, prontos para me ajudar sempre que necessário.

Ao Dr. Frederico Garcia Pinto, da Universidade Federal de Viçosa Campus Rio Paranaíba pelo fornecimento dos dados obtidos em seu pós-doutorado.

Ao Dr. Timothy J. Garrett do departamento de Patologia e Imunologia do Laboratório de Medicina da Universidade da Flórida pela parceria que viabilizou este projeto.

A aluna de doutorado Ketolly Natanne da Silva Leal do Instituto de Química da UNICAMP pelas discussões sobre metabólica.

À minha mãe, Raimunda Furtado por ser meu pilar, estar ao meu lado e me fazer acreditar que tenho a força e as ferramentas necessárias para finalizar este ou qualquer outro trabalho que me disponha a iniciar.

Ao meu pai Joudã Brasil Durans *in memoriam*, por ter sido um exemplo de homem, meu herói que foi fundamental na formação do meu caráter.

Aos meus irmãos, Luce Furtado, Lice Furtado, Carlos Durans, Joelmy Furtado, Joelson Durans, Josuel Durans, Rodrigo Pinheiro que me incentivaram nos momentos difíceis e compreenderam a minha ausência enquanto eu me dedicava à realização deste trabalho.

Aos meus sobrinhos Ângela Durans, Adriano Furtado, Luciano Furtado, Rellyane Lopes e Emanuel Furtado para que eles possam ir muito além do que já foram entendendo que o futuro é feito a partir da constante dedicação no presente.

Às minhas cunhadas Elis Tavares e Jessica Carolina, obrigada por serem sempre tão solícitas.

Às minhas amigas, Jocelia Karina Menezes, Fabian Silva, Fabiola Menezes, Izailde Abreu, Luana Roberta por estarem sempre presente, pela amizade incondicional e pelo apoio demonstrado ao longo de todo o período de formação e dessa jornada chamada vida.

A Nagib Ribeiro, Valdenor Fontoura, Tácito Rêgo, José Almeida Junior, Marcelo Moraes, Karson Reis, Pablo Kennedy que estiveram ao meu lado durante esta jornada e por compreender minha dedicação a este projeto de pesquisa.

“[...] Porque a Deus nenhuma coisa é impossível.”

Lucas 1, 37

RESUMO

O câncer de próstata (PCa) é o segundo câncer mais diagnosticado em homens em todo o mundo, com uma estimativa de 358.989 mortes em 2020. A principal barreira para reduzir a mortalidade por PCa é o diagnóstico precoce da doença. Os testes clínicos mais utilizados atualmente aplicados ao diagnóstico do PCa são: o exame de sangue do antígeno prostático específico (PSA) e o exame retal digital (DRE). No entanto, esses testes apresentam inúmeras desvantagens. O exame de sangue do PSA tem baixa especificidade, apresentando uma alta taxa de falsos positivos ou negativos e as biópsias são procedimentos desagradáveis que podem produzir várias complicações de saúde, como infecções, incontinência e disfunção erétil. Além disso, as biópsias frequentemente falham a detecção do câncer devido à heterogeneidade do tumor. Assim, o desenvolvimento de um novo teste eficiente, utilizando uma abordagem não invasiva para o diagnóstico rápido e reprodutível da PCa, é uma necessidade clínica urgente. O campo emergente da metabolômica, no qual uma análise em larga escala de moléculas pequenas (metabólitos) de organismos biológicos é detectada qualitativa e quantitativamente em uma única etapa, tem imenso potencial para o diagnóstico precoce de muitas doenças. O perfil metabólico fornece uma visão geral da condição celular e fisiológico de um indivíduo em diagnósticos clínicos. A metabolômica tem sido cada vez mais usada para o diagnóstico de várias doenças em seres humanos, como Alzheimer, Parkinson e diabetes. Assim, a espectrometria de massa com ionização por spray de papel (PSI-MS) parece ser uma alternativa promissora para se aplicar a abordagem metabolômica em estudos em larga escala e, em seguida, estabelecer uma ferramenta para testes clínicos utilizados para o diagnóstico de doenças. Para esse fim, propõe-se a combinação do PSI – MS e de métodos quimiométricos, visando correlacionar as informações extraídas de espectros com as propriedades de interesse, no caso diagnosticar pacientes com câncer de próstata pela análise da urina. Assim, baseando nos resultados obtidos os métodos quimiométricos de análise discriminante por quadrados mínimos parciais (PLS-DA) e dados aplicados a modelagem independente e suave de analogia de classes (DD-SIMCA) apresentam-se como uma alternativa viável para detectar o câncer de próstata pela análise de amostras de urina, pois apresentaram índices de classificação de 100% na diferenciação entre as amostras de pacientes com câncer de próstata e pacientes saudáveis. A presente pesquisa foi divulgada no blog do Cláudio Lima, mostrando a importância da pesquisa para identificar diagnósticos de câncer da próstata. A presente pesquisa foi divulgada em vários sites, trabalhos acadêmicos, redes sociais, blog, e em jornais como: Comunica UFU.BR, G1 Triângulo e Alto Parnaíba, Diário de Uberlândia, FARMAT4H, Fiocruz.Br, PPGQUI.IQ.UFU/acontece, Conselho Regional de Química IX Região, Programa Manhã Total, Jornal da UFU, Programa Balanço Geral- TV Paranaíba. Em rádio como: Rádio Ciência-Universidade FM.

Palavras-chave: Câncer de próstata. Metabolômica. PSI-HRMS. Métodos de classificação. Quimiometria.

ABSTRACT

Prostate cancer (PCa) is the second most diagnosed type of cancer in men worldwide, it's estimated 358,989 deaths in 2020. The main barrier to reducing PCa mortality is early diagnosis of the disease. The most man commonly used clinical tests currently applied to the diagnosis of PCa are: the prostate-specific antigen (PSA) blood test and the digital rectal exam (DRE). However, these tests have numerous disadvantages. The PSA blood test has low specificity having a high rate of false positives or negatives, and biopsies are uncomfortable procedures that can produce several health complications such as infections, incontinence, and erectile dysfunction. Beside, biopsies often miss cancer detection due to tumor heterogeneity. It's necessary development of a new efficient test, using a non-invasive approach for the rapid and reproducible diagnosis of PCa, is an urgent clinical need. The emerging field of metabolomics, in which a large-scale analysis of small molecules (metabolites) from biological organisms is detected qualitatively and quantitatively in a single step, has immense potential for the early diagnosis of many diseases. The metabolic profile provides an overview of an individual's cellular and physiological status in clinical diagnoses. The metabolomics has been increasingly used for the diagnosis of several diseases in humans, such as Alzheimer's, Parkinson's and diabetes. So, paper spray ionization mass spectrometry (PSI-MS) appears to be a promising alternative to apply the metabolomics approach in large-scale studies and then establish a tool for clinical trials used for disease diagnosis. Therefore, it is essential to develop analytical methods to overcome the challenges proposed in this project using the analytical technique of mass spectrometry in combination with chemometric methods, aiming to correlate the information extracted from the spectra with the properties of interest, in this case, to diagnose patients with prostate cancer by urine analysis. Thus, based on the results obtained, the chemometric methods of discriminant analysis by partial least squares (PLS-DA) and data-driven-soft independent modelling of class analogy (DD-SIMCA) are presented as a viable alternative to detect cancer of prostate by the analysis of urine samples, as they presented classification rates of 100% in the differentiation between samples from patients with prostate cancer and healthy patients.

Keywords: prostate cancer. metabolomics. PSI-HRMS. classification methods. chemometrics.

LISTA DE FIGURAS

Figura 1 - Biópsia de próstata feita por ultrassom.....	19
Figura 2 - Formação da urina pelos rins	27
Figura 3 – Exame de PSA.....	30
Figura 4 - Imagem ilustrativa do toque retal	31
Figura 5 - Imagens de ressonância magnética do câncer de próstata	38
Figura 6 - Diagrama esquemático de um espectrômetro de massa.....	42
Figura 7 - Ilustração do processo de ionização por Electrospray	36
Figura 8 - Fonte com ionização por spray de papel.....	38
Figura 9 - Representação da construção da matriz X para modelagem multivariada	40
Figura 10 - Gráfico de Leverage versus Q Residuals	42
Figura 11 - Esquema da matriz de dados X e o vetor de classes y usados no PLS-DA.....	43
Figura 12 - Gráfico de estimativas para o modelo PLS-DA.	45
Figura 13 - Seleção do <i>threshold</i> pelo teorema da Bayes para o modelo PLS-DA.....	45
Figura 14 - Gráfico do resultado do modelo DD-SIMCA na etapa de treinamento.....	47
Figura 15 - Gráfico do resultado do modelo DD-SIMCA na etapa de teste	48
Figura 16 - Gráfico dos espectros de massa obtidos por PSI-HRMS, amostras da urina de 40 pacientes com câncer de próstata e 38 pacientes saudáveis.....	52
Figura 17 - Gráfico de Leverage versus Resíduos com 95% de confiança para o modelo PLS-DA	53
Figura 18 - Gráfico de Estimativas para amostras de urinas de pacientes com câncer de próstata e saudáveis.	54
Figura 19 - Gráfico dos pesos para as 3 VLs usadas para a construção do modelo PLS-DA, sendo (a) primeira variável latente; (b) segunda variável latente e (c) terceira variável latente	56
Figura 20 - Gráficos VIP escores do modelo PLS-DA	58

Figura 21 - Gráfico de aceitação de amostras com câncer de próstata pelo modelo DD-SIMCA no processo de treinamento. A área de aceitação é o limite para $\alpha= 0,05$, delimitação outlier é o limite para $\alpha= 0,01$	59
Figura 22 - Gráfico de verificação de outlier em amostras com câncer de próstata pelo modelo DD-SIMCA na etapa de treinamento.	59
Figura 23 - Gráfico de aceitação de amostras com e sem câncer de próstata pelo modelo DD-SIMCA na etapa de teste.	60

LISTA DE TABELAS

Tabela 1 - Esquema resumido do sistema de fases para estudos de busca de biomarcadores proposto em 2002 pelo EDRN, NCI, USA.....	25
Tabela 2 – Parâmetros de classificação obtidos para o modelo PLS-DA.....	54
Tabela 3 - Tabela de confusão do modelo PLS-DA nos conjuntos de treinamento e de teste..	55
Tabela 4 – Resultados do treinamento do modelo DD-SIMCA.....	59
Tabela 5 – Tabela de confusão do modelo DD-SIMCA nos conjuntos de treinamento e de teste.....	62

LISTA DE ABREVIATURAS E SIGLAS

INCA	Instituto Nacional de Câncer
PSA	Antígeno Prostático Específico (do inglês, Prostate Specific Antigen)
PCa	Câncer de Próstata
RM	Ressonância Magnética
LC-MS	Cromatografia Líquida acoplada a Espectrometria de Massa (do inglês, Liquid Chromatography – Mass Spectrometry)
PSI-MS	Espectrometria de Massa com Ionização por Spray de Papel
PSI-HRMS	Espectrometria de Massa de Alta Resolução com Ionização por Spray de Papel (do inglês, High Resolution Mass Spectrometry with Paper Spray Ionization)
EI	Ionização por Elétrons (do inglês, Electron Ionization)
CI	Ionização Química (do inglês, Chemical Ionization)
ESI	Ionização por Electrospray (do inglês, Ionization by Electrospray)
PLS-DA	Análise Discriminante por Quadrados Mínimos Parciais (do inglês, Discriminant Analysis with Partial Least Squares Method)
DD-SIMCA	Dados Aplicados a Modelagem Independente e Suave de Analogia de Classes (do inglês, data driven – soft independent modeling of class analogy)
PCA	Análise de Componentes Principais (do inglês, Principal Component Analysis)
PC	Componente Principal (do inglês, Principal Component)
PLS	Quadrados Mínimos Parciais (do inglês, Partial Least Squares)
VL	Variável Latente
RMSEC	Erro Quadrático Médio de Calibração (do inglês, Root Mean Squared Error of Calibration)
RMSECV	Erro Quadrático Médio de Validação Cruzada (do inglês, Root Mean Squared Error of Cross Validation)
RMSEP	Erro Quadrático Médio de Previsão (do inglês, Root Mean Squared Error of Prediction)

VIP	Variável Importante na Projeção
FP	Taxa de Falso Positivo
VP	Taxa de Verdadeiro Positivo
FN	Taxa de Falso Negativo
VN	Taxa de Verdadeiro Negativo
LOD	Limite de Detecção (do inglês, detection limit)

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos	17
1.2	Objetivos Específicos	17
2	Revisão Bibliográfica	18
2.1	Câncer de próstata	18
2.2	Biomarcadores	22
2.3	Urina como fonte de biomarcadores.....	26
2.4	Metabolômica	28
2.5	Antígeno Prostático Específico (PSA)	29
2.6	Toque retal.....	30
2.7	Ressonância Magnética (RM)	31
3	Técnicas Analíticas	32
3.1	Espectrometria de Massa (MS)	32
3.2	Espectrometria de massa com ionização por spray de papel (PSI-HRMS).....	37
4	Métodos Quimiométricos	39
4.1	Regressão por Quadrados Mínimos Parciais (PLS)	40
4.2	Detecção de Amostras Anômalas (Outlier)	41
4.3	Análises Discriminante por Quadrados Mínimos Parciais (PLS-DA)	42
4.4	Dados Aplicados a Modelagem Independente e Suave de Analogia de Classes (DD-SIMCA).....	46
4.5	Validação de modelos de classificação	48
5	Metodologia	50
6	Resultados e Discussões	52
6.1	Espectros de massa das amostras de urina analisadas por PSI-HRMS	52

6.3	Resultados do modelo DD-SIMCA.....	58
7	Conclusão	62
8	Referências Bibliográficas	63

1 INTRODUÇÃO

A próstata é uma glândula envolve a porção inicial da uretra, tubo pelo qual a urina armazenada na bexiga é eliminada, sendo responsável pela produção de parte do sêmen, líquido espesso que contém os espermatozoides, liberado durante o ato sexual. Em 2019, ocorreram quase 180 mil novos casos de câncer e 31.620 mil mortes por câncer nos Estados Unidos (SIEGEL *et al.* 2019). No Brasil, o câncer de próstata é o segundo mais comum entre os homens (atrás apenas do câncer de pele não-melanoma), assim como quando se considera ambos os sexos e valores absolutos. Em valores absolutos e considerando ambos os sexos, é o segundo tipo mais comum. O Instituto Nacional de Câncer (INCA) relata que no Brasil em 2019 ocorreram 15.983 mil mortes e em 2020 ocorreram 65.840 mil novos casos de câncer de próstata, (INCA, 2021).

A detecção precoce do câncer é uma estratégia para encontrar o tumor em fase inicial e, assim, possibilitar melhor chance de tratamento. No caso do câncer de próstata, esses exames são o toque retal, onde o médico palpa, com o dedo protegido por luva lubrificada, a próstata verificando se há nódulos (caroços) ou tecidos endurecidos (possível estágio inicial da doença) e o exame de sangue para avaliar a dosagem do PSA (antígeno prostático específico). Nenhum dos dois exames tem 100% de exatidão. Por isso, podem ser necessários exames complementares (INCA, 2021).

A biópsia é o único procedimento capaz de confirmar o câncer. A retirada de amostras de tecido da glândula para análise é feita com auxílio da ultrassonografia. Pode haver desconforto e presença de sangue na urina ou no sêmen nos dias seguintes ao procedimento, e há risco de infecção, o que é resolvido com o uso de antibióticos. Outros exames de imagem também podem ser solicitados, como tomografia computadorizada, ressonância magnética e cintilografia óssea (para verificar se os ossos foram atingidos) (INCA 2021).

Infelizmente, os procedimentos para diagnosticar este grave problema de saúde têm inúmeras desvantagens e são significativamente inadequados para o diagnóstico, sendo necessário pensar numa alternativa mais viável e se possível não invasiva para o diagnóstico de câncer de próstata.

Na metabolômica é possível detectar ou determinar diferentes metabólitos em organismos biológicos em uma única etapa de análise, o que permite uma resposta rápida como a requerida nos diagnósticos precoces. O perfil metabólico fornece uma visão geral do estado celular e fisiológico de um indivíduo, podendo ser usado em diagnósticos de várias doenças em humanos, como Alzheimer, Parkinson, diabetes e tuberculose (WISHART 2016).

A técnica de cromatografia líquida acoplada a espectrometria de massa (LC-MS) é uma das técnicas de alto potencial para a análise metabolômica, porém em grande escala apresenta algumas desvantagens como o tempo (20-30 min por injeção), preparação laboriosa de amostra e a quantidade usada de solvente. Assim, para superar tais desvantagens a espectrometria de massa com ionização por spray de papel (PSI-MS) parece ser uma alternativa promissora.

PSI é um método de ionização na espectrometria de massa que produz íons de fase gasosa no ar ambiente, separando a fase cromatográfica e minimizando a preparação prévia da amostra. O PSI-MS promove uma introdução e análise simples e direta de amostras com baixo custo e alta velocidade. A amostra é colocada diretamente em uma folha rígida de papel de cromatografia, deixada secar e, em seguida, analisada por MS após a aplicação de um solvente de spray para a eluição. As aplicações clínicas do PSI-MS estão em um estágio inicial, mas se mostram promissoras como uma técnica para o diagnóstico de doenças, tendo como vantagens o fato de usar um pequeno volume de amostra, sendo adequada para medição de muitos analitos clínicos em biofluidos em menos de 1 minuto em cada aquisição. Isso a torna uma ferramenta ideal para integração em um ensaio de perfil metabolômico rápido, (ALBERICE, 2014).

Devido a complexidade dos dados espectrais se faz necessário um tratamento multivariado visando extrair informações correlacionadas com a propriedade de interesse, no caso detectar a presença do câncer de próstata. Portanto, é imprescindível a utilização de métodos quimiométricos na obtenção de informações a partir desses espectros de massa. No presente trabalho a detecção do câncer de próstata é por meio da espectrometria de massa para rastrear os biomarcadores de câncer de próstata. O biomarcador é definido como uma substância encontrada no tecido, sangue ou outros fluidos podendo ser um sinal de uma determinada doença, em nosso caso o câncer de próstata.

A ampla utilização de LC-MS em análises metabolômicas deve-se, além das vantagens analíticas como robustez, reprodutibilidade e alta resolução, à versatilidade do método de separação intrínseco a técnica de HPLC. Esta permite a separação de compostos de uma vasta gama de polaridade, através de eluição isocrática ou de gradiente de eluição. Acetonitrila e metanol são os solventes orgânicos mais comumente empregados. A eluição isocrática é preferida para amostras simples (ou seja, menos de 10 componentes), enquanto o gradiente de eluição proporciona um conjunto mais rápido de análises, (ALBERICE, 2014).

A espectrometria de massa de alta resolução com ionização por spray de papel (PSI-HRMS) é um método que tem crescido rapidamente em popularidade nos últimos anos devido à sua simplicidade; é consolidado um método de amostragem e ionização direta, rápido e de baixo custo para análises de MS qualitativa e quantitativa. Portanto não existe método perfeito,

sempre tem limitações na análise de matrizes complexas como o sangue, plasma e urina, esperadas neste método, tais como: supressão iônica, pouca sensibilidade em níveis residuais e falta de especificidade (GÓMEZ-RÍOS et al. 2016).

Portanto, é interessante desenvolver modificações funcionais para que se possa vencer tais limitações e utilizar a PSI em análises de matrizes complexas, com o mínimo preparo de amostra. Propõe-se, para tanto, uma abordagem analítica com base no uso de polímeros impressos molecularmente (MIP) como substrato para o papel utilizado em espectrometria de massa de PSI (MIP-PSI) a fim de torná-lo mais seletivo para determinados tipos de analitos, permitindo, deste modo, a análise de metabólitos em urina.

O método de Análise Discriminante por Quadrados Mínimos Parciais - PLS-DA (*Partial Least Squares - Discriminant Analysis*) é um método multivariado utilizado para classificação de amostras onde é necessária a redução de variáveis e a variabilidade entre grupos seja maior que a dentro do grupo. O bloco **Y** em um modelo de PLS-DA indica a classe a qual uma amostra pertence. Quando se têm duas classes a serem discriminadas, existe uma variável dependente **y** que pode assumir os valores 0 ou 1. Idealmente, os valores previstos pelo modelo PLS-DA seriam 0 ou 1, mas na prática os valores previstos são próximos de 0 e 1. Tendo em conta que os valores previstos não são 0 ou 1, faz-se necessário o cálculo de um valor limite para separar as classes, denominado de *threshold*, (ALBERICE, 2014).

Atualmente, como alternativa ao método PLS-DA, existe métodos de modelagem de classe única ou classificadores de uma classe (*one class*), quando pretendem resolver problemas de classificação de amostras da mesma espécie. O método de dados aplicados a modelagem independente e suave de analogia de classes (DD-SIMCA) é um dos métodos *one class* que permite desenvolver modelos de classificação usando apenas informações das amostras da classe de interesse para posteriormente detectar se as informações de uma nova amostra se assemelham a essa classe ou não (ZONTOV et al. 2017). No entanto, até o presente momento não existem estudos publicados que relatam o seu uso para diagnosticar doenças a partir de espectros de massa. Assim, o presente trabalho visa inovar desenvolvendo um método multivariado baseado no uso do método DD-SIMCA para diagnóstico de câncer de próstata por análise de urina e comparar com o método PLS-DA, sendo ambos os modelos validados antes de tal comparação.

1.1 Objetivos

Esse trabalho tem como objetivo geral, desenvolver métodos analíticos para detectar o câncer de próstata pela análise da urina de pacientes usando espectrometria de massa com ionização por spray de papel (PSI-MS) e métodos quimiométricos de análise discriminante por quadrados mínimos parciais (PLS-DA) e dados aplicados a modelagem independente e suave de analogia de classes (DD-SIMCA).

1.2 Objetivos Específicos

- Construir modelos de PLS-DA e DD-SIMCA para detectar o câncer de próstata em amostras de urina analisadas por PSI-MS;
- Validar e comparar os modelos PLS-DA e DD-SIMCA construídos.

2 Revisão Bibliográfica

2.1 Câncer de próstata

De acordo com Ferlay (2012), o câncer de próstata (PCa) é considerado o segundo tipo mais comum em homens, abrange 87 países, que apresentam maior índice de diagnóstico de câncer de próstata, e é a quinta maior causa de morte por câncer. Existem alguns fatores de risco para desenvolver tal doença que estão relacionados à idade, à hereditariedade e à etnia, sendo 1,6 vezes mais comum em negros comparados aos brancos.

Para saber se tem indicação da existência da doença é necessário fazer o exame do toque retal, esse exame ainda é muito constrangedor e muitos homens não se sentem confortáveis. Mesmo fazendo tal exame, é preciso fazer um estudo histopatológico do tecido obtido pela biópsia da próstata para ter a confirmação do diagnóstico, (MOTTET *et al.* 2015).

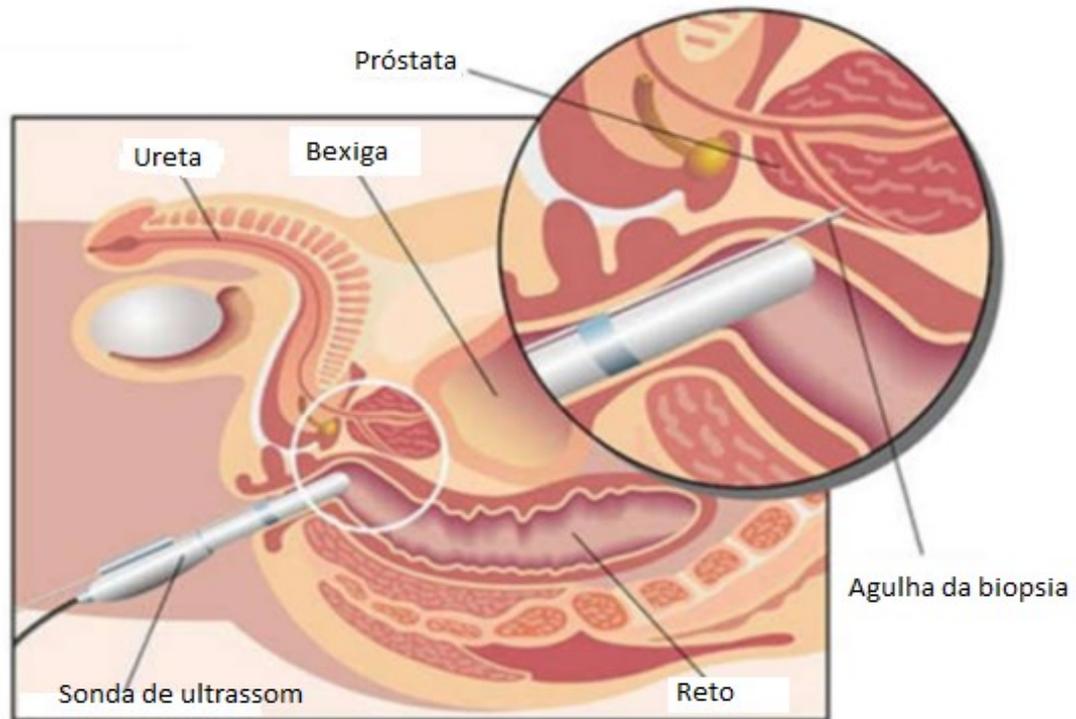
Sarris *et al.* (2018), o exame do toque retal para diagnosticar o câncer de próstata é preciso um médico urologista. Ao realizar o exame, o médico tenta definir algumas características da próstata: tamanho, consistência, superfície, forma, limites e sensibilidade. Isso permite que ele defina se há algum problema e estabeleça o próximo passo. Em geral, alterações no toque retal com PSA maior que 3 ng/mL já são suficientes para indicação de biópsia.

Conforme Paiva (2020), coloca a seguinte questão do diagnóstico sobre câncer de próstata:

Quando um indivíduo tem diagnóstico suspeito de PCa no rastreamento, necessita realizar a biópsia da glândula para confirmação histológica. A primeira biópsia da próstata foi realizada em 1937, com agulha e digitalmente dirigida ao nódulo da próstata. Com o advento do ultrassom e seu uso transretal, os nódulos não palpáveis passaram a ser vistos e biopsiados com agulhas apropriadas. O uso do ultrassom permite ao médico a localização do(s) nódulo(s) suspeito(s) inserindo a agulha no ponto exato para a coleta. Modificações na biópsia sextante padrão foram propostas tendo como foco a importância da lateralização das punções dirigidas, ou seja, obtenção de maior positividade quanto mais lateral na próstata for a biópsia. A biópsia é importante não só pela detecção do PCa como também no manejo clínico, ajudando a prever o prognóstico (PAIVA, 2020, p. 23).

O uso do ultrassom permite ao médico a localização do(s) nódulo(s) suspeito(s) inserindo a agulha no ponto exato para a coleta como mostra a Figura 1 como descrito por Paiva 2020.

Figura 1 - Biópsia de próstata feita por ultrassom



Fonte: Paiva (2020)

Feita a biópsia do câncer de próstata é necessário realizar a classificação do grau tumoral. Quando os pacientes são informados de que possuem um escore de Gleason 6 de 10, implica que seu prognóstico é intermediário, aumentando as chances de se ter uma neoplasia agressiva. Em tumores homogêneos a numeração é dobrada. Esta graduação é baseada na conformação arquitetural das glândulas, sua disposição e característica de crescimento e desconsidera características citológicas, sendo amplamente discutida e acordada na reunião da Sociedade Internacional de Uropatologia (*International Society of Urological Pathology – ISUP*), (EPSTEIN *et al.*, 2005).

Na Universidade Estadual de Campinas, no jornal da Unicamp, o professor Hernandes F. Carvalho, do Instituto de Biologia (IB) da Unicamp, fez um estudo, e revelou que a mulher também pode possuir uma glândula, cujas características se aproximam das do homem. A próstata feminina pode ser afetada pelas terapias de reposição hormonal ou pelo uso de anabolizantes, procedimentos que contribuiriam para o eventual desenvolvimento de tumores malignos.

A descoberta da próstata feminina promete fomentar intensos debates entre a comunidade científica e o professor Hernandes F. Carvalho, que também é coordenador do curso de Farmácia da Unicamp, os estudos preliminares indicam que o órgão estaria associado

com a chamada ejaculação feminina. Além disso, também teria ligação com o Ponto G, zona erógena que, graças à concentração de terminações nervosas e vasos sanguíneos, estaria relacionada à estimulação sexual da fêmea da espécie humana.

O professor Hernandes F. Carvalho explica que, dependendo do ambiente ou situação hormonal durante a formação do feto, a mulher pode ter um maior desenvolvimento do órgão. “No homem, o desenvolvimento da próstata depende da testosterona, hormônio produzido pelos testículos em dois momentos distintos: durante a embriogênese e, depois, na puberdade. Na mulher, a glândula não se desenvolve por causa da ausência dessa substância. No entanto, se o ambiente hormonal for alterado por algum motivo, a pessoa do sexo feminino pode, sim, vir a desenvolver o órgão na idade adulta”.

Essa hipótese foi confirmada por pesquisas realizadas com o auxílio de modelos animais. A equipe do IB usou nos estudos um roedor denominado gerbilo, comum no deserto da Mongólia, país da Ásia. Cerca de 50% das fêmeas dessa espécie apresentam próstata. Ao alterarem o ambiente hormonal dos animais, por meio da administração de testosterona, os cientistas fizeram três constatações importantes.

Lesões – Primeiro, identificaram um aumento da glândula. Segundo, confirmaram preliminarmente a atuação funcional do órgão. Por último, detectaram o desenvolvimento de lesões pré-malignas. O estudo rendeu artigo que foi publicado numa das mais prestigiosas revistas internacionais da área da reprodução, a *Biology of Reproduction*. Ao estabelecer uma associação desses resultados com a situação das mulheres, o professor Hernandes F. Carvalho lembra que algumas populações femininas têm o que os especialistas classificam de hiperandrogenismo. Mulheres com essa alteração hormonal apresentam naturalmente dosagens mais elevadas de testosterona. Uma das manifestações que acompanham o hiperandrogenismo é o hirsutismo, que consiste no crescimento de pêlos em áreas do rosto. Outra é o ovário policístico. “Mulheres com esses problemas devem merecer um cuidado especial por parte dos ginecologistas, pois podem vir a apresentar patologias ligadas ao desenvolvimento da próstata”, alerta o docente do IB.

Localizada logo abaixo da bexiga, envolvendo a uretra, a próstata masculina pesa entre 25 e 30 gramas e tem o formato de uma castanha. Sua consistência torna-se endurecida quando apresenta câncer. Tem por função produzir cerca de 70% do líquido seminal, substância fundamental para a vitalidade e transporte dos espermatozoides. No homem, o câncer de próstata atinge cerca de 10% dos indivíduos com mais de 50 anos e 50% dos que alcançam os 75 anos. O exame mais eficiente para a detecção da enfermidade é o toque retal. A possibilidade de ocorrência de uma glândula prostática feminina e sua dependência hormonal precisam ser

consideradas em mulheres com dosagens normalmente elevadas de andrógenos, que recebem tratamento baseado nessas drogas ou que apresentam ovário policístico, elementos relacionados ao quadro de hiperandrogenia.

A mesma atenção, prossegue o pesquisador, precisa ser dada às atletas que porventura façam uso de anabolizantes e aos transsexuais do tipo mulher-homem. Estes últimos, após a cirurgia para a mudança de sexo, normalmente são submetidos a tratamentos hormonais complementares.

O mesmo é válido em relação às mulheres que lançam mão de terapias de reposição hormonal, procedimentos adotados comumente após a menopausa. O professor Hernandes F. Carvalho esclarece que, nesse tipo de abordagem, a mulher recebe dosagens de um precursor do estrógeno, hormônio feminino.

Acontece que este precursor, depois de ser absorvido pelo organismo, tanto pode se transformar em hormônio feminino quanto em masculino. “Na hipótese de receber uma dosagem excessiva de testosterona, a mulher também pode ter manifestações patológicas relacionadas ao desenvolvimento da próstata. Em outras palavras, nesses casos os médicos devem fazer uma investigação mais profunda para identificar a possível presença da glândula e de eventuais doenças a ela relacionadas”, insiste o especialista.

O docente do IB afirma que é possível que cânceres de cavidade abdominal constatados em mulheres possam ter relação com problemas prostáticos, mas que essa associação dificilmente é feita. Por conta disso, em alguns casos a causa da doença é classificada como desconhecida. Hernandes F. Carvalho revela que uma diferença fundamental entre a próstata feminina e a masculina está na estrutura da glândula. No homem, o órgão conta com uma espécie de cápsula que restringe a disseminação das células. Na mulher, essa proteção não existe, o que permite a livre disseminação das células tumorais na cavidade abdominal.

As pesquisas foram conduzidas pela equipe do IB em conjunto com Fernanda C. A. Santos, que realizou seu doutorado no Instituto, e do professor Sebastião R. Taboga, da Unesp de São José de Rio Preto. De acordo com Hernandes F. Carvalho, os estudos estão tendo continuidade, desta vez com a investigação sobre o balanço dos hormônios masculino e feminino no desenvolvimento das próstatas feminina e masculina.

2.2 Biomarcadores

De acordo com Paiva (2020), a grande maioria dos biomarcadores tumorais é sintetizada por ambas as células, normais e cancerígenas, mas são produzidas em quantidades maiores por células cancerígenas. As principais características de um biomarcador tumoral ideal são sua especificidade para um determinado tipo de tumor e sua sensibilidade, sendo que seus níveis devem sinalizar com precisão a progressão e a regressão do tumor, (PAIVA, 2020).

Biomarcador é um termo abrangente para a investigação de processos biológicos e seu uso na detecção precoce, diagnóstico, monitoramento de doenças e escolha de tratamentos. Um biomarcador é definido como um indicador mensurável de um estado biológico específico, particularmente a presença ou o estágio de uma doença. De fato, um biomarcador refere-se a uma molécula, a qual pode pertencer a diversas classes de compostos e, como consequência, uma variedade de estratégias tem sido adotada para sua descoberta (ALBERICE, 2014).

A propagação de pesquisas com biomarcadores foi possível a partir do desenvolvimento de tecnologias para detecção, quantificação e identificação de biomoléculas. O crescente conhecimento dos mecanismos moleculares juntamente com a aplicação de técnicas de *high-throughput* (alto rendimento) nas análises *ômicas* vêm fornecendo a possibilidade de identificação de potenciais biomarcadores específicos para diferentes doenças. (ALBERICE, 2014).

De maneira geral os biomarcadores se correlacionam com suas respectivas neoplasias, e os biomarcadores de próstata mostram algumas ou todas as seguintes habilidades: diagnóstica, prognóstica, preditiva e farmacodinâmica. Para Barbieri *et al.*, (2011), são definidos assim:

Biomarcadores de diagnóstico representam uma estratégia de triagem (*screening*) ou confirmação do diagnóstico de uma neoplasia. Os biomarcadores prognósticos preveem o curso natural do câncer para distinguir o desfecho do tumor. Eles também ajudam a determinar quem tratar, a intensidade do tratamento, quais candidatos provavelmente responderão a um determinado medicamento e a dose mais eficaz. Biomarcadores preditivos avaliam o provável benefício de um tratamento específico. Já os biomarcadores farmacodinâmicos avaliam os efeitos iminentes do tratamento de um medicamento em um tumor e, possivelmente, podem determinar a dosagem nos estágios iniciais do desenvolvimento clínico de um novo medicamento anticâncer. A identificação de biomarcadores moleculares específicos no PCa é uma abordagem racional para enfrentar os desafios clínicos atuais de quem realiza biópsia prostática para quem oferece certas terapias intervencionistas, (BARBIERI *et al.*, 2011, p. 18).

Percebe-se que os biomarcadores possuem quatro habilidades, sendo que cada uma tem sua importância ao paciente, e podem ser usados visando um diagnóstico rápido e eficiente o que aumenta as chances de cura de um tratamento.

Conforme Alberice (2014), o objetivo de analisar a urina é para obter informações sobre as concentrações de metabolitos, de vias metabólicas, erros no metabolismo, efeitos dos medicamentos e monitorização da dosagem terapêutica, exposição a agentes químicos de fontes exógenas, consumo de drogas, controle de *doping*, nutrição e dieta, incluindo a ingestão de contaminantes de alimentos. Visto que a urina pode conter a substância alvo em conjunto com os metabolitos do diagnóstico e os compostos excretados podem ser concentrados pelos rins.

Apesar do reconhecimento dos benefícios que os biomarcadores podem trazer e dos esforços realizados para a descoberta de novos marcadores, poucos deles estão sendo utilizados na prática clínica. Desde 1998 a taxa de introdução de novos biomarcadores aprovado pelo FDA (*US Food and Drug Administration*) tem sido em média de um por ano. As razões para isso são várias e refletem o longo e difícil caminho desde a descoberta do candidato até a implementação do marcador (ALBERICE, 2014).

O desenvolvimento de um novo biomarcador segue um processo tão sistemático e rigoroso quanto a de um fármaco. Em 2002, a Rede de Pesquisa de Detecção Precoce (*Early Detection Research Network*, EDRN) do Instituto Nacional do Câncer (*National Cancer Institute*, NCI - USA) propuseram um sistema de cinco passos para estudos de busca de biomarcadores antes da sua implementação (ALBERICE, 2014).

Fase I – Esta etapa compreende essencialmente a fase de geração da hipótese, ou seja, a pergunta biológica envolvida com uma patologia particular. Nesta etapa, não só a pergunta biológica deve ser clara, mas também a escolha da matriz a ser utilizada na pesquisa, o que requer do pesquisador um estudo detalhado de tal. Dependendo da matriz alvo, testes *in vitro* ou em modelos animais, podem ser realizados.

Fase II – Desenvolvimento de ensaios preliminares em amostras de pacientes para determinação de potenciais biomarcadores e avaliação de sua reprodutibilidade e robustez. Nesta etapa, indivíduos com e sem câncer ou com tumores de diferentes estágios devem ser distinguidos pelo biomarcador, a fim de ser considerado promissor e passível de ser avaliado nas demais etapas do processo.

Fase III – O biomarcador é testado em um grupo de pacientes com relação a sua capacidade de discriminar entre os grupos avaliados. Se nesta etapa o biomarcador discrimina bem apenas determinada subpopulação, esta informação deve ser considerada para a seleção de populações apropriadas para nova triagem.

Fase IV – O biomarcador é testado em uma população relevante, representativa da população alvo, para determinar sua funcionalidade. Esta é, efetivamente, a fase de validação do marcador o qual deve demonstrar sua utilidade na prática clínica. Idealmente, o biomarcador deve ser testado em populações diferentes daquela da descoberta e as taxas de sensibilidade e especificidade devem ser determinadas.

Fase V – Finalmente, a última fase antes da implementação do marcador deve avaliar o impacto deste na sociedade. Devem ser avaliados não apenas a funcionalidade do biomarcador, mas também as dificuldades na sua implementação, custos e acesso pela comunidade, visando a viabilidade da sua implementação. Não é raro que nesta fase os interesses no marcador diminuam devido tanto à falta de avaliação dos parâmetros anteriormente mencionados como a não conformidade com eles.

O esquema de fases apresentado acima propõe a sistematização de estudos envolvendo biomarcadores, entretanto, como não há regulamentação para estes estudos, tal sistema não é seguido à risca.

Além da complexidade na sistemática, as pesquisas abrangendo busca de biomarcadores devem lidar com diversos desafios analíticos. Fundamentalmente, estudos que buscam por biomarcadores analisam a abundância relativa de moléculas por comparação, ou seja, moléculas de qualquer classe de compostos com abundância diferenciada entre os grupos pode ser um biomarcador em potencial. Por isso, determinar os parâmetros analíticos para todas as moléculas medidas nesse tipo de estudo é um desafio.

Parâmetros como eficiência de extração, recuperação e reprodutibilidade são determinadas em estudos analíticos que avaliam um único composto. Neste caso, tais parâmetros são determinados através de curva de calibração analítica e adicionando quantidades conhecidas do composto alvo na matriz de interesse. Entretanto, quando o objetivo é detectar o máximo número de compostos possíveis, estabelecer estes parâmetros para cada analito é muito difícil (ALBERICE, 2014).

A Tabela 1 mostra o exemplo de um resumo do sistema de fases usado por pesquisadores que visavam realizar estudos de possíveis biomarcadores.

Tabela 1 - Esquema resumido do sistema de fases para estudos de busca de biomarcadores proposto em 2002 pelo EDRN, NCI, USA

<i>Fase</i>	<i>Objetivo</i>	<i>Experimentação</i>	<i>Detalhes da amostra</i>
I	Geração da hipótese	Estudos pré-clínicos	Possibilidade de ensaios <i>in vitro</i> ou em modelos animais
II	Nomear e classificar os candidatos a biomarcadores	Reprodutibilidade e robustez; discriminação entre os grupos estudados	Ensaio preliminares em grupos reduzidos
III	Testar em uma população pequena	Avaliação da discriminação, estabelecimento de regras de predição	Ensaio em grupos de pacientes maiores que na fase anterior
IV	Validação do candidato a biomarcador	Avaliação da funcionalidade	Ensaio em população grande, multi-institucional
V	Avaliação do impacto do candidato a biomarcador	Relatório descrevendo sua funcionalidade e viabilidade de implementação	-

Fonte: Alberice (2014)

Outros processos importantes em análises de biomarcadores que devem ser avaliados são a manipulação e estocagem de amostras. Os compostos presentes nas amostras são suscetíveis a modificações quando estocados ou manuseados incorretamente e, assim, uma análise incorreta das moléculas diferencialmente expressas ocorreria. Até mesmo o número de ciclos de congelamento e descongelamento já foi relatado afetar compostos em soro e, por isso, é importante aderir a um procedimento padrão para evitar erros na avaliação dos marcadores (ALBERICE, 2014).

2.3 Urina como fonte de biomarcadores

A urina é considerada uma potencial fonte de biomarcadores voláteis. É influenciada por fatores intrínsecos e extrínsecos como a variação genética, a dieta, a ingestão de líquidos, hábitos e variações diárias. Em adição, os métodos de coleta das amostras voláteis e as condições cromatográficas podem alterar os perfis voláteis urinários. (CAVACO, 2011).

O perfil de biomoléculas de fluidos corporais tem sido bastante explorado na busca de biomarcadores para doenças. Neste cenário, a urina representa a matriz mais acessível e seu perfil biomolecular tem mostrado ser promissor na identificação de potenciais biomarcadores, principalmente relacionadas ao trato urinário. Particularmente, a disponibilidade de biomarcadores de tumor através da urina representa uma alternativa conveniente para sua detecção precoce e acompanhamento, devido ao contato direto da urina com o tumor. Assim, os compostos liberados das células tumorais enriquecidos na urina podem ser detectados nesse fluido corporal. (ALBERICE, 2014).

O uso de biomarcadores urinários não é uma prática recente. Durante muito tempo a glicose foi detectada em urina avaliando se formigas eram atraídas por ela. A presença de albumina na urina tem sido usada por séculos como indicador de doenças renais, avaliando sua capacidade de formar espuma após agitação. Assim, estudos para identificar biomarcadores de doenças na urina têm sido um componente fundamental da medicina investigativa ao longo dos séculos XX e XXI. Esses estudos têm se baseado no conhecimento da fisiopatologia da doença para identificar biomarcadores que podem ser testados em triagens clínicas (ALBERICE, 2014).

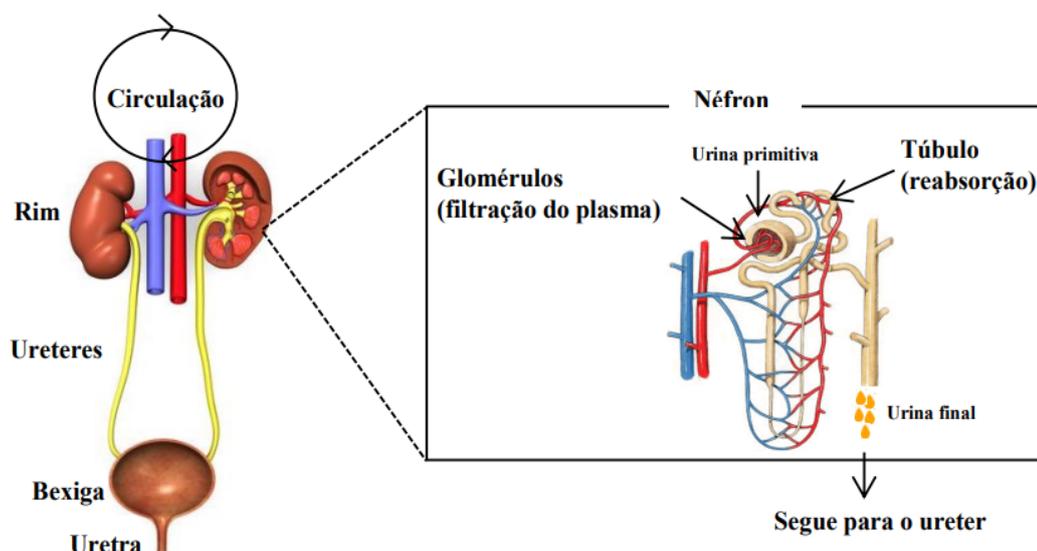
A urina é produzida pelo rim, o qual é formado por milhares de células funcionais chamadas néfrons. Os néfrons são divididos em glomérulos e tubo renal, sendo que os primeiros filtram o plasma originando a urina primitiva e o segundo absorve essa urina primitiva. A urina final chega através dos ureteres à bexiga, onde fica armazenada até eliminação mostrado na Figura 2. Destino de muitas das biomoléculas do plasma, o que lhe confere informações de órgãos mais distantes, a urina também é enriquecida com biomoléculas dos rins e trato urinário, o que a torna o fluido ideal para estudos relacionados a alterações nesses órgãos (ALBERICE, 2014).

Comparada com outros fluidos corpóreos, a urina tem características que a torna favorita para ser tomada como amostra biológica entre os demais. A principal vantagem do uso da urina nos estudos de biomarcadores, além de sua natureza não invasiva e fácil acessibilidade, é que

o paciente não é exposto a procedimentos desconfortáveis e não tem sua rotina de tratamento alterada.

Além disso, as biomoléculas contidas na urina são em geral, solúveis em água. O conteúdo biomolecular é relativamente estável, provavelmente devido ao fato da sua permanência estagnada na bexiga, o que é um fator positivo para a busca de biomarcadores (ALBERICE, 2014).

Figura 2 - Formação da urina pelos rins



Fonte: Alberice (2014)

Contudo, a urina também apresenta algumas desvantagens que devem ser superadas. A principal delas está relacionada ao fato que a concentração das biomoléculas urinárias varia conforme a ingestão de líquidos, variações na dieta, processos metabólicos ou catabólicos, ritmo cardíaco e exercícios, bem como nível circulatório. Entretanto, isso pode ser contornado normalizando-se a concentração de proteínas (e outras moléculas) pela razão de creatinina ou peptídeos presentes na mesma. Outra desvantagem está relacionada à baixa concentração de algumas biomoléculas na urina, como proteínas, por exemplo, tornando-se necessárias etapas adicionais para concentração delas, podendo levar a perdas. Ainda para o caso de proteínas, existem algumas delas que são abundantes na urina, como a albumina, que podem mascarar outras proteínas, e desta forma, uma etapa de depleção deve ser realizada. Por fim, a urina apresenta alto conteúdo de sal, o que é incompatível com algumas técnicas analíticas (ALBERICE, 2014).

2.4 Metabolômica

Rivera e Garrido (2019) definiram a metabolômica como o estudo sistemático de todos os processos químicos relativos a metabólitos, com massa molecular inferior a 1000 Da, fornecendo assim impressões digitais químicas e características que determinados processos intracelulares e extracelulares produzem.

Em vista disso, a metabolômica permite monitorar qualitativamente e quantitativamente pequenas moléculas da oxidação de ácidos graxos, aminoácidos orgânicos, açúcares e selecionados, que oportuniza a triagem de distúrbios metabólicos em organismos e tecidos (YANG et al., 2019). As concentrações metabólicas podem variar em resposta a perturbações na homeostase intracelulares e extracelulares, por meio de alterações genéticas, nutricionais e fisiopatológicas.

A metabolômica faz parte das ciências ômicas, que tem como objetivo analisar todas as pequenas moléculas (metabólitos) em sistema biológico de forma simultânea. Como os metabólitos representam um grupo diversificado de compostos com propriedades químicas e físicas variadas com ampla faixa de concentração, a análise metabolômica é tecnicamente desafiadora e a caracterização do metaboloma é difícil de ser alcançada (RIVERA E GARRIDO., 2019).

Nesse sentido, as técnicas analíticas estão em desenvolvimento para melhorar o rendimento, a sensibilidade principalmente para os metabólitos com menor abundância ou que tenham baixa eficiência de ionização, mínimo preparação de amostra, pequenos volumes e principalmente que permitam a quantificação de metabólitos.

Dentre as técnicas analíticas em destaque têm-se a espectrometria de massa e seus acoplamentos, espectroscopia de ressonância magnética, espectroscopia Raman, dentre outras. Cada técnica tem sua particularidade e estratégia analítica.

A alteração de metabólitos é notada em indivíduos doentes, quer como uma causa primária, quer como indicador de uma causa secundária. O conceito de que indivíduos têm um perfil metabólico que poderia ser refletido na composição de seus fluídos biológicos é bem conhecido. Os efeitos clínicos do metabolismo anormal incluem mal funcionamento de órgãos ou danos nos tecidos, distúrbios neurológicos, doenças cardiovasculares e morte prematura. A identificação e quantificação desses metabólitos anormais são importantes para o reconhecimento e a confirmação de um transtorno do metabolismo endógeno, (CHACE, 2001).

A metabolômica é útil para estudos na identificação de biomarcadores de doenças. A dopamina, a sarcosina e o ácido butírico são alguns exemplos de metabólitos biomarcadores

excretados pela urina. Os níveis urinários de dopamina podem ser correlacionados com várias condições patológicas relacionadas a doenças cardíacas, estresse, distúrbios neurológicos (doenças de Alzheimer e Parkinson e esquizofrenia) e tumores cancerígenos. A sarcosina é um biomarcador do câncer de próstata. Os níveis urinários de ácido butírico no organismo podem indicar coinfeção de bactérias em pacientes com Síndrome da Imunodeficiência Adquirida (SIDA/ AIDS), (IMAI *et al.*, 2012).

Atualmente o foco dos estudos em metabolômica tem sido para a descoberta de biomarcadores, que estão correlacionados com várias doenças (Yang *et al.*, 2019). Os biomarcadores têm característica definida que é medida como um indicador de processos biológicos normais, processos patogênicos ou respostas a uma exposição ou intervenção.

Portanto, a metabolômica permite abordagens de alto rendimento em pesquisas centradas em busca de biomarcadores, entendimento completo dos processos metabólitos durante as enfermidades e respostas fisiológica a estratégias de tratamento.

2.5 Antígeno Prostático Específico (PSA)

Conforme Castro (2011), os exames mais utilizados para o diagnóstico do câncer de próstata são o toque retal e a dosagem sérica do antígeno prostático específico (PSA). Porém o toque retal mostra algumas limitações quando a lesão prostática não é palpável, e quando alterado, a diferenciação entre lesão maligna e benigna é difícil. A dosagem sérica de PSA, em uso clínico desde 1986, é o marcador tumoral mais frequentemente utilizado para o diagnóstico do câncer de próstata.

As células epiteliais da zona de transição são as responsáveis pelos níveis séricos de PSA, e o aumento do volume prostático está diretamente relacionado com o aumento deste antígeno. Uma variedade de fatores pode afetar os níveis de PSA e deve ser considerada na interpretação dos resultados. As causas mais comuns de aumento do PSA são a prostatite, a hiperplasia prostática benigna e o câncer de próstata.

O diagnóstico com o exame de PSA tem a função de detecção da maior quantidade de casos de câncer de próstata. Apesar de possuir alta sensibilidade, este exame tem baixa especificidade para o câncer de próstata, ou seja, para manter taxas aceitáveis de diagnóstico, muitos pacientes são submetidos a biópsias desnecessárias, (CASTRO, 2011). O exame de PSA é realizado através do exame de sangue do paciente coletado em frascos específicos como mostra a Figura 3.

Figura 3 - Exame de PSA



Fonte: Urologia Integral (2022)

2.6 Toque retal

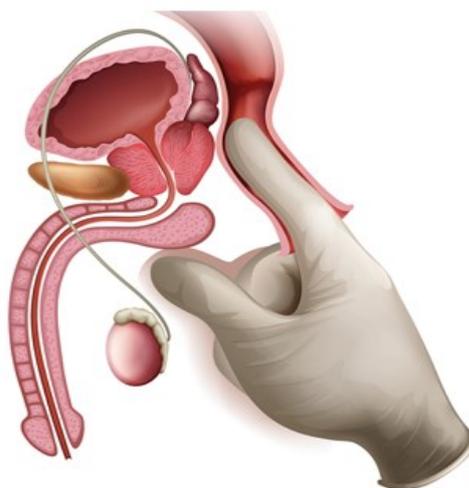
Para Souza (2011), uma das formas apontadas pela literatura para diagnóstico do câncer de próstata é o exame de toque retal, procedimento de baixo custo, rápido e que permite avaliar o tamanho, o formato e a consistência da próstata, embora não em sua total abrangência. Apesar das suas facilidades, acirra o imaginário masculino, sendo interpretado como uma afronta à masculinidade, o que pode influenciar na adesão ao exame.

Autores salientam que os homens podem apresentar resistência e constrangimento ao exame de toque retal, pois “viola” a sua masculinidade, no que se refere à condição de ser homem ativo. A resistência surge, então, porque veem o toque retal como algo que conspiraria contra a noção de masculino. Nesses casos, a masculinidade é usada como estrutura para a formação da identidade, ditando conceitos a serem seguidos para que sejam reconhecidos como “machos” e não serem questionados por aqueles que possuem as mesmas crenças. Nesse sentido, ser homem é um exercício contínuo de negação – mais do que de afirmação – negando atributos ditos femininos para se aproximar do que acredita ser a imagem ideal de homem, que foi construída ao longo da vida, (SOUZA, 2011).

A forma mais aceita atualmente de diagnóstico do câncer de próstata é a associação do toque retal à dosagem sérica do PSA. A chance de o indivíduo com toque retal alterado ter câncer de próstata aumenta conforme o valor do PSA. O PSA é uma protease produzida quase

que exclusivamente pelas células epiteliais prostáticas. Porém, alguns cuidados ao avaliar um exame de PSA devem ser tomados. O PSA é mais órgão-específico que câncer específico, (MARIA, 2008). Na Figura 4 tem-se uma representação do exame de toque retal.

Figura 4 - Imagem ilustrativa do toque retal



Fonte: Revista Saúde (2022)

2.7 Ressonância Magnética (RM)

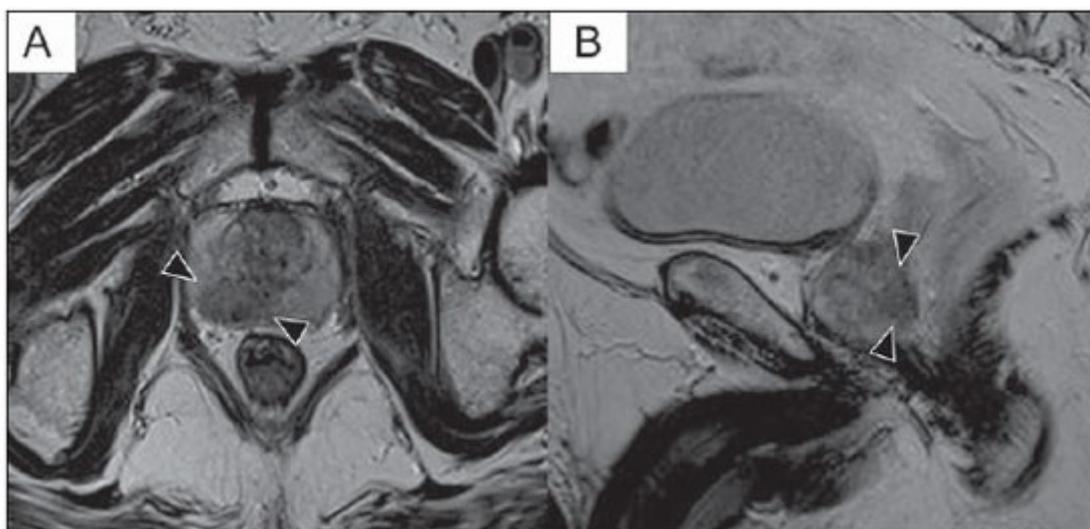
Para Melo (2009), em virtude do seu custo relativamente baixo, ao ser realizado com a sonda transretal oferece a melhor oportunidade de guiar a biópsia da glândula. Porém, limita-se ao estadiamento local, pela dificuldade no diagnóstico precoce da extensão extracapsular e a alta dependência do operador, o qual limita a reprodutibilidade da técnica.

Entre as demais técnicas radiológicas, a ressonância magnética é a ferramenta diagnóstica mais útil para avaliação dos estágios do tumor, principalmente quando se a utiliza com a bobina endorectal. A imagem por ressonância magnética possui sensibilidade significativamente maior (51–89%) na detecção do tumor quando comparada com a ultrassonografia transretal (USTR) (27–86%). Mas, da mesma forma, ambas têm baixa especificidade (58–94%), (MELO, 2009).

Recentemente, surgiu a espectroscopia por ressonância magnética (ERM) como nova esperança diagnóstica. Ela é capaz de, a partir das informações anatômicas provenientes das imagens de RM, demonstrar os indicadores metabólicos detectados na glândula prostática,

aprimorando a acurácia na provável localização do tumor. No Brasil, ainda são poucos os serviços de RM que possuem condições técnicas para realizar o exame da próstata com espectroscopia. Na Figura 5 temos imagens obtidas em um exame por ressonância magnética para diagnosticar o câncer de próstata.

Figura 5 – Imagens de ressonância magnética do câncer de próstata



Fonte: CBR (2022)

Sinais de extensão às vesículas seminais através de imagens pesadas em T2. Observam-se áreas hipointensas substituindo a vesícula seminal usualmente hiperintensa, seja de forma difusa (A) ou focal (B).

3 Técnicas Analíticas

3.1 Espectrometria de Massa (MS)

A espectrometria de massa em combinação com métodos de separação e métodos quimiométricos, torna-se a técnica analítica mais promissora a ser usada nas ciências ômicas. Tal técnica permite a análise de moléculas em fase gasosa, as quais são ionizadas e separadas de acordo com sua massa/carga (m/z) quando submetidas a condições específicas de campo elétrico e/ou magnético. A espectrometria de massa permite a determinação molecular, a caracterização estrutural, o estudo da reatividade de compostos, bem como a análise qualitativa e quantitativa de componentes de uma amostra ou mistura.

Um espectrômetro de massa é composto basicamente por uma fonte de ionização, um analisador de massa, um detector e um sistema de aquisição de dados. Atualmente, existem diversas fontes de ionização e analisadores de massa com características peculiares os quais favorecem a análise de determinada classe de compostos (ALBERICE, 2014).

A espectrometria de massa (MS - *Mass Spectrometry*) é uma técnica analítica de grande aplicabilidade capaz de fornecer informações sobre: (HARRIS, 2005)

- a composição elementar de amostras;
- as estruturas de moléculas inorgânicas, orgânicas e biológicas;
- a composição qualitativa e quantitativa de misturas complexas;
- a estrutura e a composição de superfícies sólidas;
- as razões isotópicas de átomos nas amostras.

Os princípios fundamentais da espectrometria de massa datam do final dos anos 1890, quando J. J. Thomson determinou a massa/carga do elétron, mas sua primeira aplicação em análises de rotina ocorreu no início da década de 40, quando a técnica começou a ser adotada pela indústria petrolífera para análise quantitativa de misturas constituídas por até nove hidrocarbonetos. Antes, as análises de misturas desse tipo eram feitas por destilação fracionada seguida por medidas de refração dos componentes separados, o que consumia, em média, 200 horas de trabalho do analista. Com o uso do espectrômetro de massa, informações semelhantes passaram a ser obtidas em poucas horas ou menos (HARRIS, 2005).

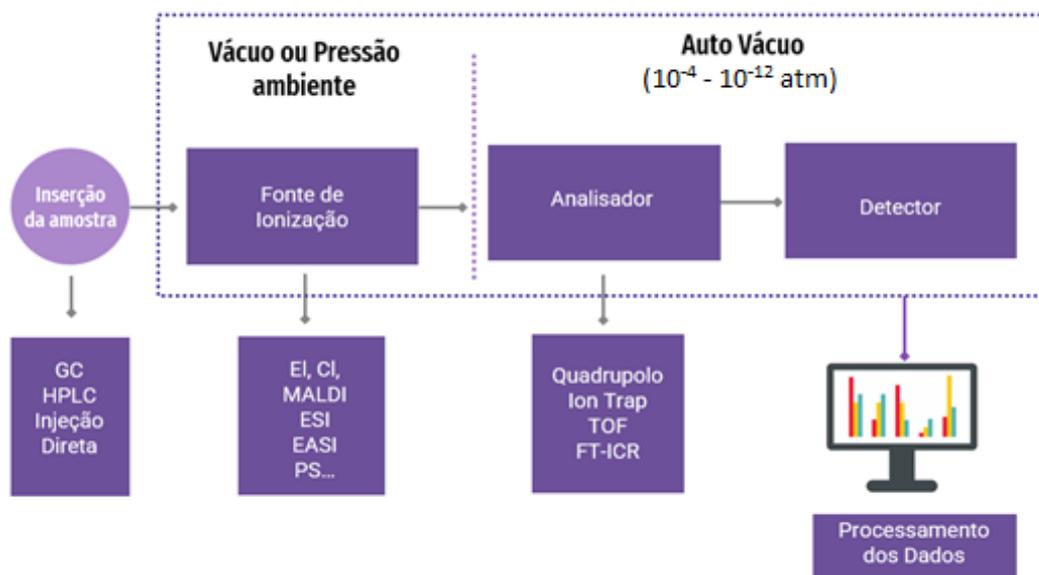
A partir de 1950, os espectrômetros de massa comerciais foram aperfeiçoados e adaptados para a identificação e elucidação estrutural de diversos compostos orgânicos. Essa aplicação da MS, combinada à invenção da ressonância magnética nuclear e ao desenvolvimento da espectroscopia no infravermelho, revolucionou o modo como os químicos orgânicos determinavam a estrutura das moléculas, sendo de fundamental importância (PAVIA, 2010).

Nas décadas de 1980 e 1990, as aplicações da espectrometria de massa tiveram mudanças significativas. O desenvolvimento de métodos de ionização à pressão atmosférica

permitiu que moléculas não-voláteis, com alto peso molecular ou termicamente instáveis como, por exemplo, a maioria das moléculas presentes nos alimentos ou amostras biológicas, fossem ionizadas e analisadas por MS. Em 2002, o pesquisador americano John B. Fenn, juntamente com Koichi Tanaka e Kurt Wüthrich, foi laureado com o prêmio Nobel de Química pelo desenvolvimento de métodos suaves de ionização e dessorção em análises espectrométricas de massa de macromoléculas biológicas (PAVIA, 2010).

A Figura 6 apresenta, de maneira esquemática, os componentes de um espectrômetro de massa. Em geral, a análise de um composto por MS compreende as seguintes etapas: (1) a introdução da amostra, (2) a ionização das moléculas, (3) a passagem por um analisador de massa, responsável por separar os íons formados de acordo com a sua massa/carga (m/z), (4) em seguida os íons são contados pelo detector que transforma o sinal em corrente elétrica e, por fim, (5) o processador converte a magnitude do sinal elétrico em função da razão m/z em dados, resultando no espectro de massa correspondente.

Figura 6 - Diagrama esquemático de um espectrômetro de massa



Fonte: a autora

O analisador de massa e o detector são sempre mantidos sob alto vácuo, enquanto a fonte, em diversas técnicas de ionização, opera à pressão atmosférica (API - *Atmospheric Pressure Ionization*). O tipo da fonte de ionização e do analisador de massa é o que determina a aplicabilidade da MS.

Inúmeros métodos de ionização, aplicáveis praticamente a todo tipo de analito, foram desenvolvidos ao longo da história da espectrometria de massa (HOFFMANN, 2007).

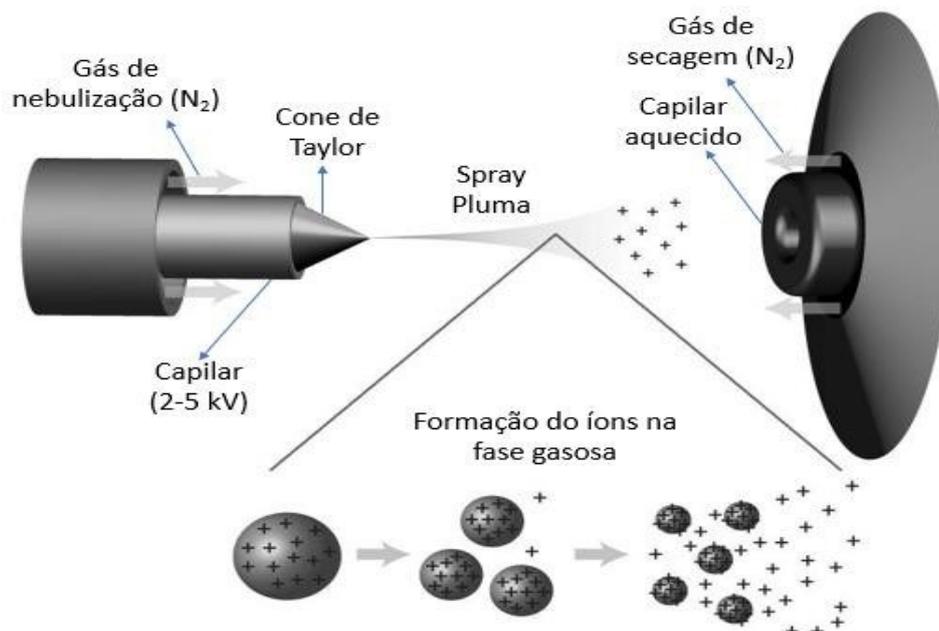
O método mais simples e comum de converter a amostra em íons é a Ionização por Elétrons (EI - *Electron Ionization*). Na EI, um feixe de elétrons de alta energia é emitido a partir de um filamento aquecido e colide com as moléculas do analito. Na colisão um ou mais elétrons são retirados das moléculas e cátions são criados. A energia transmitida à amostra durante o processo é alta e pode levar a uma grande quantidade de fragmentações do íon molecular. Uma desvantagem da EI é que a amostra deve ser relativamente volátil, o que impede a análise de compostos de alto peso molecular (PAVIA, 2010).

Na Ionização Química (CI - *Chemical Ionization*), as moléculas da amostra colidem com as moléculas de um gás reagente (ex.: H₂, CH₄, NH₃, CH₃OH) pré-ionizado. Com a colisão, algumas moléculas da amostra são ionizadas por vários mecanismos, como transferência de prótons, transferência de elétrons e formação de íons. Assim como a EI, a Ionização Química exige que a amostra seja volátil (baixo peso molecular) (PAVIA, 2010).

Em 1989, o trabalho publicado por (Apud Hoffmann, 2007) revolucionou o campo de estudo da espectrometria de massa ao divulgar uma nova técnica de ionização, menos “agressiva” e que opera à pressão atmosférica: a Fonte de Ionização *Electrospray* (ESI - *Electrospray Ionization*) (HOFFMANN, 2007).

Na ESI, os íons são formados fora do espectrômetro de massa. A amostra é diluída em um solvente volátil, em meio ácido ou básico, e as moléculas do analito são protonadas (modo positivo) ou desprotonadas (modo negativo). Os íons também podem ser observados na forma de íons, coordenados com cátions ou ânions. O processo de ionização por *Electrospray* mostrado na Figura 7 ocorre, basicamente em três etapas: (1) a nebulização da solução contendo a amostra em gotículas carregadas produzidas diretamente pela aplicação de voltagem no capilar (2-5 kV), (2) a liberação de íons a partir das gotículas e (3) o transporte dos íons da região de pressão atmosférica da fonte de ionização para a região de alto vácuo do analisador (FENN, 1989).

Figura 7 - Ilustração do processo de ionização por Electrospray



Fonte: Mendonça (2018)

A transferência dos íons presentes em solução para a fase gasosa se dá do seguinte modo: o tubo capilar expulsa pequenas gotículas carregadas (spray) para dentro da câmara de ionização. As gotículas carregadas enfrentam um contrafluxo de um gás de secagem (em geral, nitrogênio) que evapora as moléculas de solvente. Assim, a densidade de carga de cada gotícula aumenta até que as forças repulsivas excedam a sua tensão superficial (limite de Rayleigh), a gotícula então se divide em tamanhos menores (explosão coulômbica). O processo se repete até que os íons da amostra, livres de solvente, se encontrem em fase gasosa (FENN, 1989).

Com o surgimento da ESI, análises de moléculas não voláteis e de grande peso molecular, como as macromoléculas biológicas, se tornaram possíveis. Como a amostra a ser analisada deve ser introduzida em solução, é possível o acoplamento da ESI-MS com técnicas de separação, como a cromatografia em fase líquida (LC-ESI-MS). Outra peculiaridade da Ionização por *Electrospray* está no fato de ser um método de ionização brando, o que permite a análise da molécula intacta (sem ou com pouca fragmentação), mantendo inclusive as interações não covalentes (PRAMANIK, 2002).

Nas análises por MS, após a ionização da amostra, os íons são acelerados, por ação de um campo elétrico, na direção do analisador de massa, onde são separados de acordo com as suas razões m/z . Os analisadores de massa se diferenciam pelo seu modo de trabalho, pela sua exatidão e pela resolução. A exatidão é medida pelo erro calculado (em ppm) do desvio da medida experimental em relação ao valor teórico. A alta exatidão de massa oscila entre valores

de 0,1 a 50 ppm de erro. A resolução da massa pode ser definida como a capacidade de o analisador medir a razão m/z com a menor diferença possível da razão m/z teórica. Os analisadores de massa comerciais possuem resolução que variam de 1.000 (resolução unitária), até 100.000 (alta resolução) e de 100.000 até 1.000.000 (altíssima resolução) (HOFFMANN, 2007).

A espectrometria de massa (MS) é uma técnica instrumental que faz a mensuração das moléculas do analito por meio da separação de íons em fase gasosa, de acordo com suas diferentes razões massa/carga (m/z) – são cada vez mais utilizadas em metabolômica para fornecer uma visão instantânea do metabolismo. Por meio desta técnica é possível identificar compostos com grande exatidão. Em contrapartida mesmo utilizando a detecção fornecida pela MS, injeções diretas de amostras brutas, sem pré-tratamento, não são recomendadas, pois a presença de componentes da matriz pode interferir na análise do analito, dificultando, assim, a sua detecção e quantificação exata. (WEST *et al.*, 2016).

Para Martín Esteban (2013), a preparação da amostra é, ainda, considerada o obstáculo de todo o processo analítico e impacta em quase todos os passos posteriores, uma vez que é uma etapa crítica para a confirmação, identificação inequívoca e quantificação de analitos. A capacidade de analisar as amostras sem pré-tratamento ou com uma preparação mínima da amostra é o elemento-chave da espectrometria de massa ambiente (AMS).

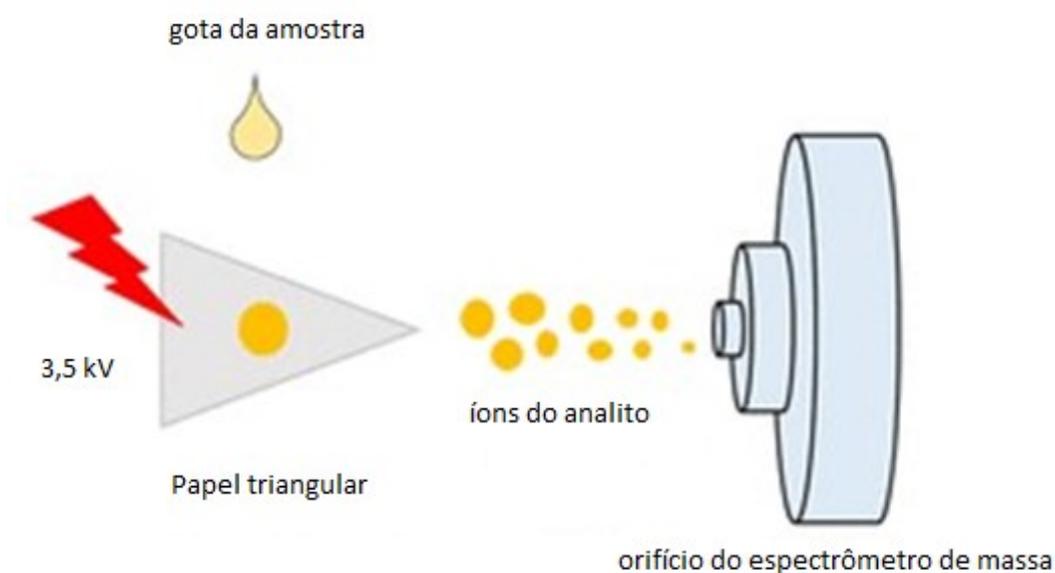
A espectrometria de massa ambiente (AMS) é uma família de técnicas de dessorção e ionização que revolucionaram o modo de obter íons na MS. Essas técnicas compartilham a capacidade de geração de íons na fase gasosa diretamente a partir de amostras não tratadas, reduzindo ou praticamente eliminando a extração de analito e a separação prévia. A simplicidade e a análise rápida são enfatizadas como características das técnicas ambientes, (WEST *et al.*, 2016).

3.2 Espectrometria de massa com ionização por spray de papel (PSI-HRMS)

Um dos métodos de ionização à pressão ambiente desenvolvidos recentemente é a Ionização por spray de papel (*Paper Spray*). Desenvolvido por Wang *et al.*, possui características similares ao ESI. Desde a sua divulgação, a PSI-MS tem sido amplamente utilizada devido à sua eficiência e capacidade de analisar amostras complexas com um mínimo ou nenhum preparo de amostras (LIU, 2010).

Na PSI-MS, um papel cromatográfico cortado em formato triangular é usado como substrato para a amostra e posicionado em frente à entrada do capilar do espectrômetro de massa. Após a adição da amostra, aplica-se uma diferença de potencial entre o papel e a entrada do equipamento, o que promove a formação de um spray de gotículas do solvente contendo os analitos ionizados devido ao acúmulo de cargas na ponta do papel. A geração de íons na ionização por spray de papel se dá pelo mesmo mecanismo do ESI, mostrado na Figura 7. A elevada diferença de potencial faz com que as gotículas formadas tenham um excesso de cargas (positivas ou negativas), e com a evaporação do solvente o volume da gota vai se reduzindo e se dividindo devido às forças de repulsão eletrostáticas, até restarem apenas moléculas do analito isoladas e carregadas (ESPY *et al.* 2012). Na Figura 8 é mostrado um esquema da ionização por spray de papel.

Figura 8 - Fonte com ionização por spray de papel



Fonte: Anal. Chem. (2021)

Por sua capacidade de analisar compostos em baixas concentrações em matrizes complexas, a PSI-MS vem sendo utilizada em diversas áreas e, dentre elas, a análise de alimentos, podendo ser aplicada na análise de uma grande variedade de alimentos, no controle de qualidade ou na detecção de fraudes e/ou adulterações (LIU, 2010).

O *Paper Spray Ionization* (PSI) é um método de AMS que tem crescido rapidamente em popularidade nos últimos anos; ele é consolidado como um método de amostragem e ionização direta, rápido e de baixo custo para a análise de MS qualitativa e quantitativa. Infelizmente não existe técnica perfeita, sempre tem limitações na análise de matrizes

complexas como o sangue, plasma e urina, esperadas nesta técnica, tais como: supressão iônica, pouca sensibilidade em níveis residuais e falta de especificidade (GÓMEZ-RÍOS et al., 2016).

Infelizmente, é interessante desenvolver modificações funcionais para que se possa vencer tais limitações e utilizar a PSI em análises de matrizes complexas, com o mínimo preparo de amostra. Propõe-se, para tanto, uma abordagem analítica com base no uso de polímeros impressos molecularmente (MIP) como substrato para o papel utilizado em espectrometria de massa de PSI (MIP-PSI) a fim de torná-lo mais seletivo para determinados tipos de analitos, permitindo, deste modo, a análise de metabólitos em urina.

4 Métodos Quimiométricos

A Quimiometria pode ser definida como a área da química que utiliza métodos matemáticos, estatísticos e de lógica formal visando extrair o máximo de informação química a partir de dados analíticos. Dentre os métodos quimiométricos podemos destacar três áreas: planejamento e otimização de experimentos, calibração multivariada e classificação, sendo que esta última ainda pode ser subdividida em métodos supervisionados e não supervisionados (MASSART et al., 1998).

Os equipamentos analíticos podem gerar grande quantidade de informação química (variáveis ou características) para muitas amostras. No entanto, a necessidade de extrair ao máximo a informação química útil que de outra forma estaria soterrada na avalanche de dados produzidos pela moderna instrumentação foi a causa principal do surgimento e rápido desenvolvimento da Quimiometria.

Muita ênfase tem sido dada aos sistemas multivariados, nos quais se pode medir muitas variáveis, simultaneamente (ou de forma sequencial, com grande eficiência) ao se estudar uma amostra qualquer. Nesses sistemas, a conversão da resposta instrumental no dado químico de interesse requer a utilização de métodos de estatística multivariada, álgebra matricial e análise numérica. Tais métodos constituem a melhor alternativa para a interpretação de dados para a aquisição do máximo de informações sobre o sistema (FERREIRA, 2015).

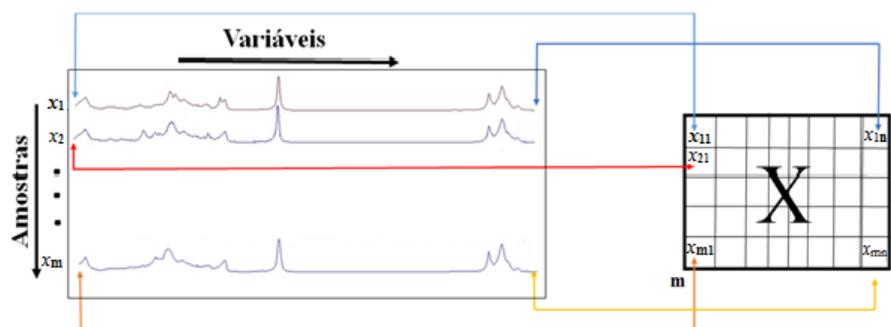
Os métodos espectrométricos comumente empregados em química orgânica e inorgânica tornam-se mais eficientes quando acoplados a métodos de matemática multivariada, o que permite a análise de grande quantidade de dados. A Quimiometria é também aplicada nas áreas de alimentos, saúde, medicamentos, combustíveis, meio ambiente, dentre outras, atuando na cadeia produtiva, controle de qualidade, estudos de adulterações, quantificações,

classificação de amostras, monitoramento do controle de qualidade, entre outras aplicações (TASIC et al., 2019).

Geralmente, durante a obtenção de informações quantitativas e qualitativas a partir de espectros complexos, e sendo dados multivariados, é indispensável a utilização de métodos quimiométricos, desenvolvidos e disponibilizados em programas computacionais.

Para realizar os procedimentos multivariados, os dados analíticos são organizados em uma matriz ordenada de linhas e colunas, onde cada linha corresponde a uma amostra “i” e cada coluna corresponde a uma variável “j”, em que $i = 1, 2, \dots, m$ e $j = 1, 2, \dots, n$ como representado na Figura 9. Em métodos supervisionados os dados são separados em uma matriz para a etapa de treinamento ou calibração e outra para a etapa teste ou previsão.

Figura 9 – Representação da construção da matriz X para modelagem multivariada



Fonte: a autora

4.1 Regressão por Quadrados Mínimos Parciais (PLS)

O método dos quadrados mínimos parciais (PLS) tem, em sua base, um conceito similar à PCA (Análise de Componentes Principais, do inglês *Principal Component Analysis*). Na PCA, ocorre uma decomposição de uma matriz X em um somatório de componentes principais (PCs) ortogonais entre si. Já para a construção dos modelos PLS, é feita uma decomposição simultânea entre duas matrizes: uma matriz X , de tamanho $m \times n$, que contém as n variáveis independentes (nessa dissertação, intensidade relativa em diferentes razões m/z) medidas para m amostras, e uma matriz Y , que contém os valores de referência, que podem ser obtidos através de uma análise química realizada por outras técnicas ou serem previamente conhecidos, quando as amostras são preparadas especificamente para a construção dos modelos (BRERETON, 2000).

A regressão por quadrados mínimos parciais é um método bastante usado nos dados de calibração multivariada que faz a análise de fatores. É um método que consiste na remoção de dados do conjunto de elementos da matriz \mathbf{X} e os correlaciona, por meio de operações matemáticas, com as informações retiradas do conjunto de dados de referência (vetor \mathbf{y}). Através da combinação da matriz \mathbf{X} com o vetor \mathbf{y} são obtidas as variáveis latentes (VL), as quais têm por finalidade alcançar a máxima covariância entre os dados analíticos e a propriedade de interesse. Para se ter a construção do modelo de calibração é utilizado um número de VL com o objetivo de ter menor erro de previsão, ou seja, que a diferença entre os valores de referência e os valores previstos pelo modelo seja o menor possível, BRERETON, 2000).

Conforme Wold *et al* (2001), na etapa de validação é preparado outro conjunto de amostras na mesma faixa de concentração das amostras de calibração e faz-se uma correlação entre essa nova matriz \mathbf{X} e o vetor de regressão, prevendo-se então a concentração dessas amostras.

Para determinar o número correto de VL é preciso realizar o método de validação cruzada, o qual é baseado na habilidade de previsão de um modelo construído por parte de um conjunto de dados seguido pela previsão do restante do conjunto de dados. A validação cruzada pode ser realizada em blocos, ou seja, amostras são deixadas de fora no processo de construção do modelo e a seguir essas amostras são previstas pelo modelo construído. Isso é semelhante ao método *leave-one-out*, em que uma amostra é deixada de fora no processo de construção do modelo e a seguir essa amostra é prevista pelo modelo construído. O processo é repetido até que todas as amostras tenham previsão e o Erro Quadrático Médio de Validação (RMSECV) é calculado.

4.2 Detecção de Amostras Anômalas (*Outlier*)

É preciso verificar a qualidade do conjunto de amostras, deve-se assegurar que essas amostras sejam homogêneas. Entretanto se o perfil de uma amostra for diferente das restantes, ela é considerada de amostra anômala ou outlier e poderá ter uma influência inadequada no modelo.

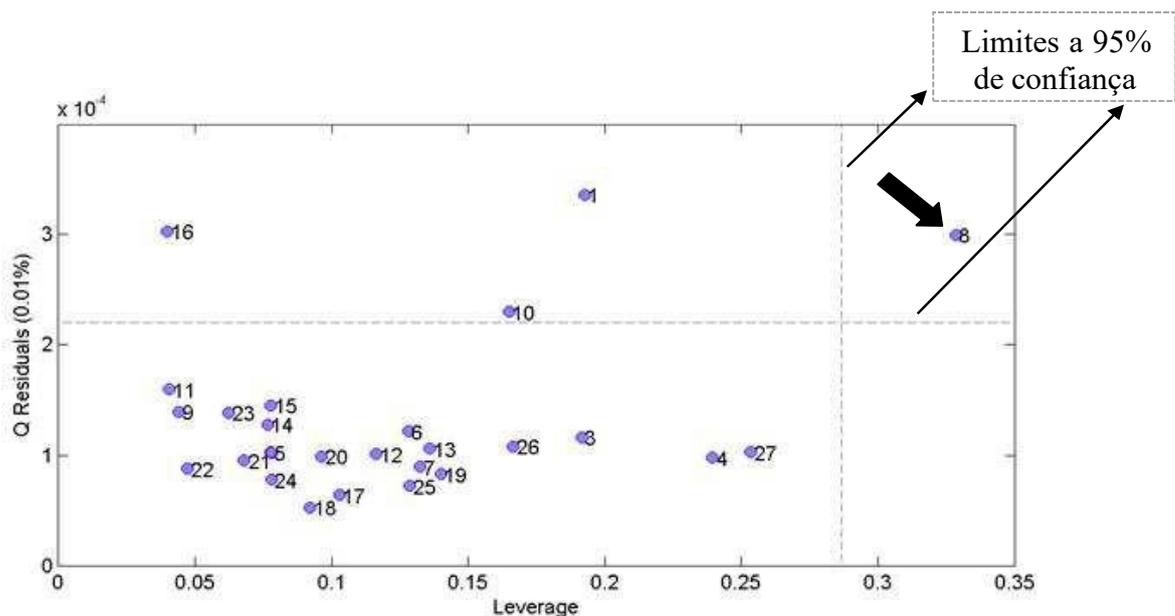
As amostras anômalas ocorrem por várias razões, como erros de digitação; erros de laboratório, como por exemplo, na determinação da propriedade de interesse por um método de referência; e os erros instrumentais, quando os sinais da amostra em questão são de baixa

qualidade. Uma vez que este tipo de amostra pode afetar a qualidade global do modelo quimiométrico, faz-se necessário à sua retirada do conjunto de calibração (VALDERRAMA et al., 2009).

A identificação de amostras anômalas é feita por meio do gráfico dos valores (o grau que uma amostra está distante da média do conjunto de dados) *versus* valores de resíduos (a parte não modelada nos dados espectrais ou nos valores de concentração analítica), conforme mostrado na Figura 10 (VANDEGINSTE et al., 1988).

Conforme mostra a Figura 10, a amostra de nº 8 está fora dos limites a 95% de confiança, representados pelas linhas tracejadas, isto é, apresenta um alto valor de leverage e alto valor de resíduos. Logo precisa ser retirada essa amostra 8 do conjunto de calibração, pois sua permanência implica no aumento de informação errônea, podendo diminuir a eficiência de previsão caso tal informação seja correlacionada com as amostras do conjunto de previsão.

Figura 10 - Gráfico de Leverage versus Q Resíduos



Fonte: a autora

4.3 Análises Discriminante por Quadrados Mínimos Parciais (PLS-DA)

O método da PLS-DA pode ser considerado uma adaptação do método PLS para a aplicação em análises qualitativas. Os princípios matemáticos utilizados nos dois métodos são

os mesmos, sendo que a diferença entre eles é a matriz \mathbf{Y} . No PLS, a matriz \mathbf{Y} é composta por valores numéricos que representam alguma propriedade da amostra, enquanto no PLS-DA, ela é composta por valores binários (normalmente 1 e 0), que indicam se uma amostra pertence ou não a uma classe pré-definida.

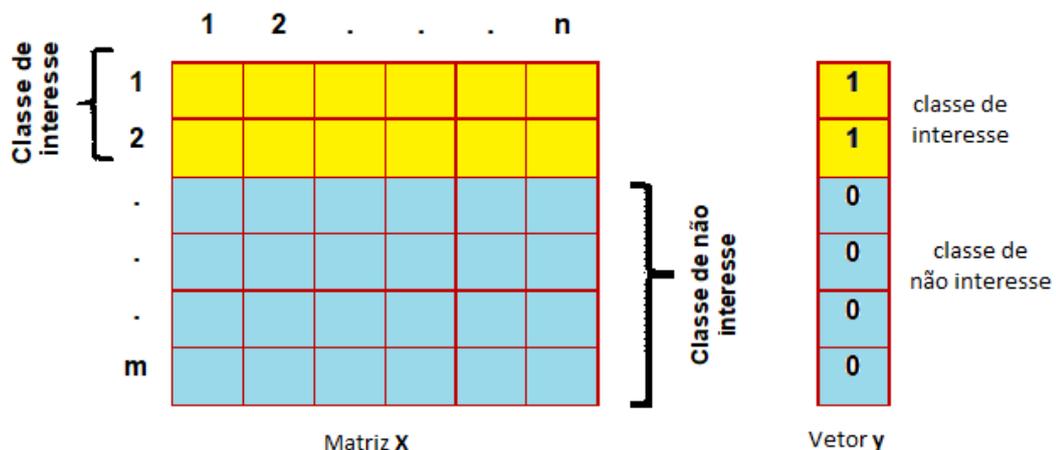
Partial Least Squares - Discriminant Analysis (PLS-DA) é um método multivariado utilizado para classificação de amostras onde é necessária a redução de variáveis e a variabilidade dentro do grupo é maior que a variabilidade entre grupos. O que diferencia o PLS-DA do PLS é que o bloco \mathbf{Y} em um modelo de PLS-DA indica a classe à qual a amostra pertence. Uma amostra tem valor igual a 1, se for um membro da classe de interesse e 0 se não pertencer a esta classe. Pode-se utilizar o PLS1 quando se tem apenas duas classes e o PLS2 quando se tem mais de duas classes. Em ambos os casos existe uma variável dependente y que pode assumir os valores 0 ou 1. Na Figura 11 temos um esquema da matriz de dados \mathbf{X} e o vetor de classes \mathbf{y} . Na construção do modelo PLS-DA o número de amostras dos conjuntos de calibração e previsão podem ser escolhidos segundo a norma ASTM E1655-05, realizando os cálculos conforme as Equações 1 e 2, respectivamente (ASTM STANDARD E1655-05, 2012).

$$n_{\text{cal}} = 6k + 1 \quad (1)$$

$$n_{\text{prev}} = 4k \quad (2)$$

Onde k é o número de variáveis latentes.

Figura 11 - Esquema da matriz de dados \mathbf{X} e o vetor de classes \mathbf{y} usados no PLS-DA



Fonte: a autora

O modelo PLS-DA fornece como resposta valores numéricos, que na maioria das vezes não são exatamente 0 ou 1. Para a classificação das amostras previstas é então estabelecido um limite de decisão (*threshold*) representado por uma linha tracejada no gráfico. Este limite pode ser definido arbitrariamente pelo analista (0,5 por exemplo) ou estimado a partir do Teorema

de Bayes, segundo o qual, o limiar assume que os valores de 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 y em que o número de falsos positivos e falsos negativos deve ser minimizado para previsões futuras (ALMEIDA et al., 2013). Quando a amostra apresentar um valor previsto acima do *threshold*, ela é considerada como pertencente à classe 1 e as amostras abaixo deste valor limite como pertencentes à classe 0.

O número de variáveis latentes (VL) do modelo PLS-DA é determinado no processo de validação cruzada onde geralmente se usa o método “*leave one out*” (deixar uma de fora por vez), onde na etapa de treinamento uma amostra é deixada de fora no processo de construção do modelo e a seguir essa amostra é prevista pelo modelo construído. Na sequência, esta amostra retorna ao conjunto de calibração, e outra é retirada e prevista pelo modelo construído pelas restantes, e esse processo é repetido até que todas as amostras tenham sido retiradas do modelo uma vez (MASSART et al., 1998). Quando se tem no conjunto de calibração acima de 20 amostras é recomendado o método de validação cruzada pelo critério da veneziana que é realizado em blocos de amostras, isto é, um número determinado de amostras é deixado de fora no processo de construção do modelo e a seguir essas amostras são previstas pelo modelo construído, seguindo o mesmo processo que o do método *leave-one-out*. Em ambos os casos, o processo possibilita a determinação do Erro Quadrático Médio de Validação Cruzada (RMSECV), de acordo com a Equação 3.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

Onde y_i , \hat{y}_i e n são, os valores reais, valores previstos pelo modelo PLS-DA e número de amostras de calibração, respectivamente.

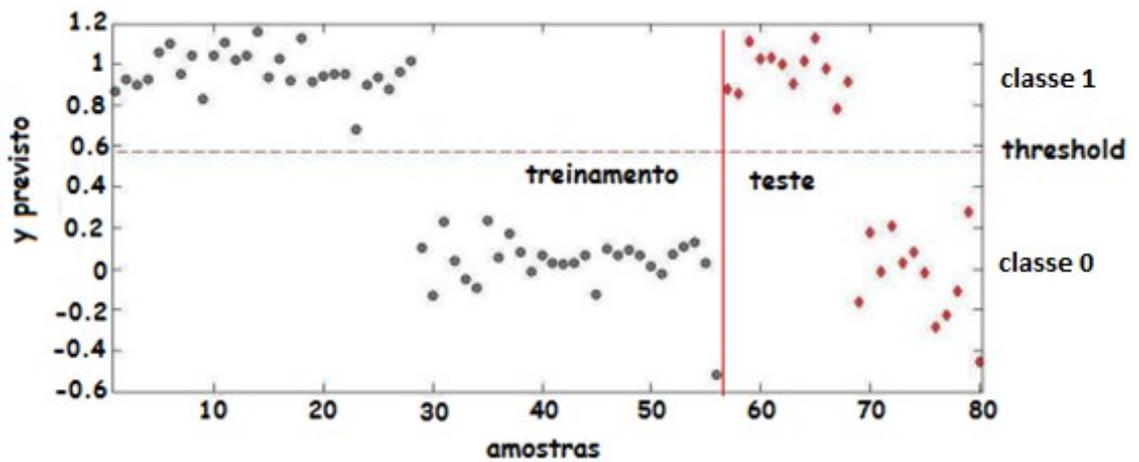
Além disso, as variáveis latentes devem descrever a covariância nas amostras das matrizes e ter a correlação máxima com a classe de valores conhecidos, dando menos peso para a variância do ruído (BARKER; RAYENS, 2003).

Quando se tem um modelo de classificação, dois parâmetros estatísticos são indispensáveis para avaliar o seu desempenho: a sensibilidade e a especificidade. A sensibilidade descreve a habilidade do teste em identificar resultados positivos, enquanto a

especificidade descreve a habilidade do teste em identificar resultados negativos (MENDES, 2012).

A Figura 12 mostra um exemplo do gráfico de estimativas para um modelo PLS-DA, em que nela pode se observar que as amostras do conjunto de treinamento (●) e de teste (◆) que pertencem à classe de interesse estão acima do *threshold*, separando-se das amostras do conjunto de treinamento (●) e de teste (◆) da classe não de interesse que estão abaixo do *threshold*.

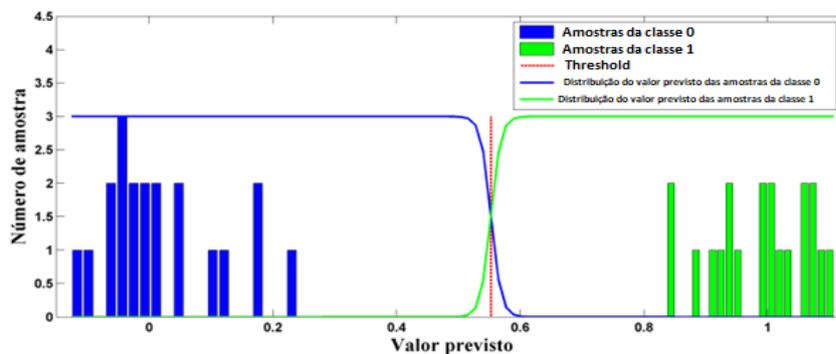
Figura 12 - Gráfico de estimativas para o modelo PLS-DA



Fonte: a autora

A seleção do valor limite de decisão (*threshold*) baseado no teorema de Bayes é representado no gráfico da Figura 13.

Figura 13 – Seleção do *threshold* pelo teorema da Bayes para o modelo PLS-DA



Fonte: a autora

Após a construção do modelo PLS-DA, pode-se fazer a análise da Variável Importante na Projeção (VIP) com intuito de identificar quais variáveis dos dados analíticos foram mais relevantes para separação das classes das amostras.

4.4 Dados Aplicados a Modelagem Independente e Suave de Analogia de Classes (DD-SIMCA)

O modelo DD-SIMCA (do inglês, *data driven – soft independent modeling of class analogy*) se diferencia do modelo SIMCA convencional por utilizar uma abordagem diferente para o cálculo dos limites de confiança que delimitam a classe alvo. Sua base matemática está na PCA, onde um modelo PCA independente é construído para modelar cada classe de interesse. Conforme sugerido por Zontov *et al.* (2017) e Rodionova *et al.* (2014), os passos utilizados para o desenvolvimento do DD-SIMCA estão descritos a seguir: (ZONTOV *et al.* 2017).

- i) Análise de componentes principais (PCA) do conjunto de treinamento contendo os dados instrumentais da classe alvo, devidamente pré-processados e centrados na média;
- ii) Escolha do número de componentes principais a ser utilizado;
- iii) Cálculo da distância de escores (área de *threshold* de aceitação – linha verde no gráfico), esta distância é calculada de forma análoga à dos valores de T2 de Hotelling;
- iv) Cálculo de distância *outlier* (área de *threshold* para amostras anômalas – linha vermelha no gráfico);
- v) Plotar o gráfico do conjunto de treinamento.

Os procedimentos de construção do modelo DD-SIMCA seguem os seguintes passos: Primeiro, a matriz de dados \mathbf{X} é decomposta pela análise de componentes principais (PCA) de acordo com a Equação 4.

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad (4)$$

Onde \mathbf{T} $[(i \times A), ex: 40 \times A]$ é a matriz dos escores, \mathbf{P} $[(j \times A), ex: (783 \times A)]$ é a matriz de pesos, \mathbf{E} $[(i \times j), ex: (40 \times 783)]$ é a matriz de resíduos e A é o número de componentes principais.

Na segunda etapa, para cada objeto $I = 1, \dots, I$ do conjunto de treinamento, duas distâncias são calculadas de acordo com as Equações 5 e 6, são elas a distância de escores (SD) e a distância *outlier* (OD) (POMERANTSEV, 2008).

$$h_i = t_i^t (\mathbf{T}^t \mathbf{T})^{-1} t_i = \sum_{i=1}^t \frac{t_{ia}^2}{\lambda_a} \quad (5)$$

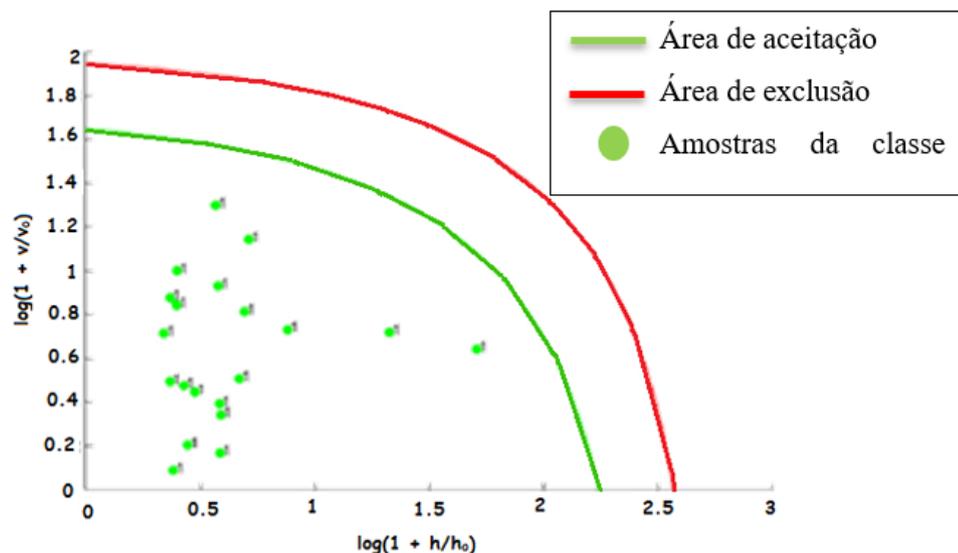
$$v_i = \sum_{j=1}^J e_{ij}^2 \quad (6)$$

onde λ_a , $a = 1, \dots, A$ são os elementos diagonais da matriz $\mathbf{T}^t\mathbf{T} = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_A)$

Para Pomerantsev (2008), SD representa a posição de uma amostra dentro do espaço dos escores, enquanto OD caracteriza uma distância para a amostra fora do espaço dos escores.

A Figura 14 mostra um exemplo do gráfico para um modelo DD-SIMCA na etapa de treinamento, em que nela pode se observar que as amostras do conjunto de treinamento que pertencem à classe de interesse (●) estão dentro dos dois limites estabelecidos pelas distâncias de escores e de *outlier*.

Figura 14 - Gráfico do resultado do modelo DD-SIMCA na etapa de treinamento

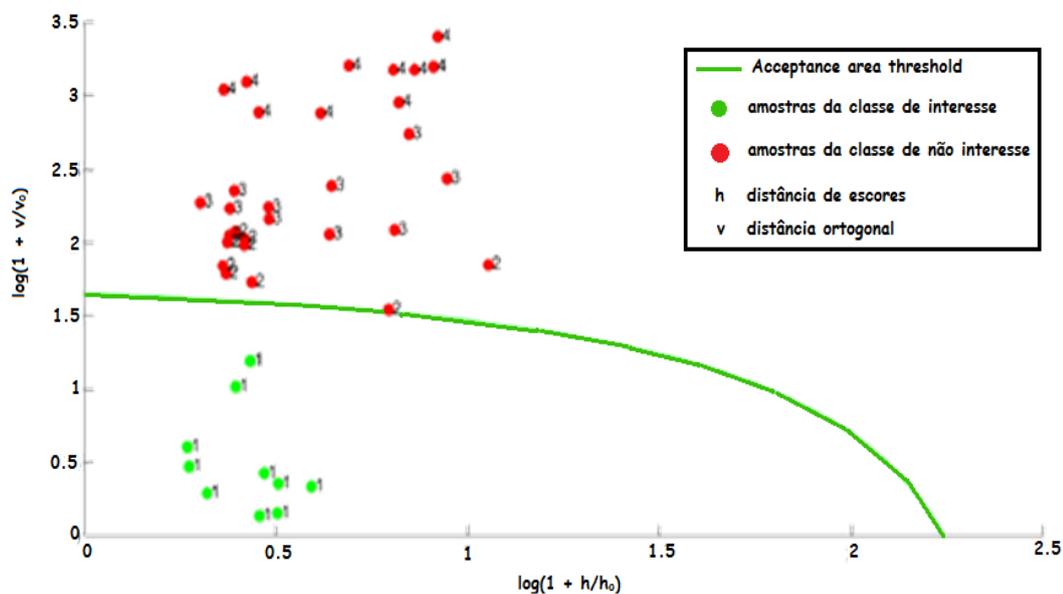


Fonte: a autora

Após esta etapa, a capacidade preditiva do modelo DD-SIMCA (representado por uma área de aceitação, na cor verde no gráfico) é avaliada por um conjunto teste ou previsão e caso pertençam a classe alvo (●) serão representadas dentro da área verde, enquanto as não pertencentes a classe de interesse (●) serão representadas na região fora da área verde.

A Figura 15 mostra a classificação das amostras do conjunto de teste, onde pode se observar que todas as amostras da classe de interesse estão dentro da área de aceitação (limite na cor verde), enquanto todas as amostras da classe de não interesse estão fora da área de aceitação, representando uma classificação 100% correta.

Figura 15 - Gráfico do resultado do modelo DD-SIMCA na etapa de teste



Fonte: a autora

4.5 Validação de modelos de classificação

A eficiência de um modelo de classificação representa sua capacidade preditiva tanto na etapa de treinamento quanto na de teste. Assim, devemos nos basear em critérios estatísticos como: taxa de falsos positivos e falsos negativos, sensibilidade, especificidade, eficiência e coeficiente de correlação de Matthew. A taxa de falsos positivos (FP) corresponde a probabilidade de uma amostra negativa ser classificada como positiva, calculada de acordo com a Equação 7 e a taxa falsos negativos (FN) corresponde a probabilidade de uma amostra positiva ser classificada como negativa, calculada de acordo com a Equação 8 (ALMEIDA et al., 2013).

$$FP = \frac{fp}{fp+vn} \times 100 \quad (7)$$

$$FN = \frac{fn}{fn+vp} \times 100 \quad (8)$$

Em que, 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.

Os parâmetros sensibilidade e especificidade do modelo são determinados com base nas Equações 9 e 10 para verificar a capacidade que o modelo tem em classificar corretamente as

amostras verdadeiras positivas e verdadeiras negativas, respectivamente (XU et al., 2012). O coeficiente de correlação de Matthew, calculado de acordo com a Equação 11, pode apresentar valores entre -1 e 1, onde o valor +1 representa uma classificação perfeita, zero representa uma classificação aleatória e -1 uma classificação inversa. A eficiência do modelo é calculada de acordo com a Equação 12. Tais parâmetros são apresentados na Tabela de Confusão.

$$\text{Sensibilidade} = \frac{vp}{vp+fn} \quad (9)$$

$$\text{Especificidade} = \frac{vn}{vn+fp} \quad (10)$$

$$\text{Coeficiente de Matthew} = \frac{(vp \times vn) - (fp \times fn)}{\sqrt{(vp+fp) \times (vp+fn) \times (vn+fp) \times (vn+fn)}} \quad (11)$$

$$\text{Acurácia} = \frac{vp+vn}{vp+vn+fp+fn} \quad (12)$$

No presente trabalho será feita a detecção do câncer de próstata do homem, por meio da espectrometria de massa com ionização por spray de papel e a partir dos resultados validados, principalmente a análise dos pesos das variáveis, encontrar os íons mais significativos na classificação das amostras de urina de pacientes com câncer de próstata. Tal informação será repassada aos colaboradores para identificar as moléculas que geraram os íons, assim esses compostos serão analisados visando identificar possíveis biomarcadores para o câncer de próstata, possibilitando o desenvolvimento de dispositivos para realizar exames do tipo rápido como o da Covid e teste de gravidez.

5 Metodologia

O presente trabalho foi desenvolvido em parceria com o Professor Doutor Frederico Garcia Pinto da Universidade Federal de Viçosa Campus Rio Paranaíba e o Professor Doutor Timothy J. Garrett da Universidade da Flórida nos Estados Unidos, onde foram realizadas as coletas das amostras de urina e realizadas as análises por espectrometria de massa de alta resolução com ionização por spray de papel (PSI-HRMS). Essas etapas foram realizadas durante o Pós-doutorado do Dr. Frederico Garcia Pinto supervisionado pelo Dr. Timothy J. Garrett quando foi adaptado o método com ionização por spray de papel. Então os dados dos espectros de massa foram encaminhados ao Laboratório de Quimiometria do Triângulo (LQT) para o tratamento matemático e modelagem. Mesmo a parte experimental não sendo realizada no LQT é apresentada uma breve abordagem das condições experimentais.

5.1 Amostras de urina

Amostras de urina de 40 pacientes com câncer de próstata foram obtidas no Biospecimen Core da rede SPORE em Prostate Cancer na Northwestern University (P50 CA180995). Pacientes com câncer de próstata apresentaram vários estágios de progressão da doença. Amostras de urina de 40 indivíduos saudáveis sem histórico médico anterior de câncer foram obtidas no Life Study (University of Florida, Gainesville, FL). O comitê de revisão institucional (IRB) do Florida Hospital aprovou o uso das amostras. As amostras de urina foram coletadas usando tubos de preservação de urina (Norgen Biotek, Thorold, ON, Canadá) e armazenadas a $-80\text{ }^{\circ}\text{C}$ até o uso para análise de espectrometria de massa de alta resolução com ionização por spray de papel (PSI-HRMS).

5.2 Análise da urina bruta por PSI-HRMS

As amostras de urina bruta foram usadas para análise PSI-HRMS. Amostras de urina congeladas ($-80\text{ }^{\circ}\text{C}$) foram descongeladas em temperatura ambiente. A urina bruta ($15\text{ }\mu\text{L}$) foi depositada em cartuchos Velox de papel triangular pré-cortados obtidos na Prosolia Inc. (Indianapolis, IN). Um dispositivo impresso em 3D foi usado (Prosolia Inc.) para reprodutibilidade na dispensação e posicionamento da amostra. Os cartuchos PSI foram carregados na unidade automatizada Prosolia Velox 360 PSI para análise de MS. Um volume de $80\text{ }\mu\text{L}$ de 3:7 MeOH:H₂O contendo 0,1% de ácido fórmico (v/v/v) foi usado como solvente

de umedecimento e spray. Amostras combinadas de cada grupo foram usadas para verificar a reprodutibilidade.

5.3 Análise PSI-HRMS

Os espectros de massa por PSI-HRMS foram obtidos usando um Thermo Scientific Q Exactive MS. A aquisição de dados foi realizada em modo de varredura completa (modo de ionização positiva) com uma faixa de massa m/z de 70-1000 com resolução de massa de 140 000. A temperatura capilar foi de 270 °C com um nível de lente S de 30. A voltagem do spray para ionização foi de 3,5 kV. A aquisição foi realizada em 0,5 min. Os arquivos de dados metabolômicos foram convertidos para o formato de arquivo mzXML usando Raw Converter. MZmine 2 foi empregado para a seleção e alinhamento de picos.

As espécies não detectadas com intensidade zero foram substituídas pelo valor do Limite de Detecção (LOD) de todas as características detectadas. Os dados foram filtrados por filtração de recurso em branco (BFF) para remover picos com $\geq 10\%$ de contribuição de intensidade do sinal de fundo. A normalização dos picos filtrados foi realizada para a corrente total de íons, log transformado e auto escalamento.

5.4 Análises Multivariadas

Os espectros de massa obtidos para as amostras de urina dos pacientes foram organizados numa matriz de dados X conforme descrito na Figura 9. Logo após foram tratados e pré-processados usando métodos adequados e rotinas computacionais desenvolvidas no Laboratório de Quimiometria do Triângulo (LQT). Os modelos PLS-DA e DD-SIMCA foram construídos usando os programas MATLAB versão R2018b (Mathworks, Inc.), PLS_Toolbox versão 9.0 (Eigenvetor Research) e um Toolbox desenvolvido pelos autores Zontov *et al.* (2017). Os dados foram pré-processados utilizando a normalização da soma, ou seja, dividindo cada região do espectro pela respectiva soma total da área de integração do espectro.

A partir dos espectros de massa de 80 amostras de urina, sendo 40 de pacientes com câncer de próstata (PCa) e 40 de pacientes sadios foi construído um modelo PLS-DA usando 48 amostras no conjunto de treinamento (24 com câncer e 24 sadios) e 32 amostras no conjunto teste (16 com câncer e 16 sadios). Já no modelo DD-SIMCA foram utilizadas as mesmas 80 amostras sendo 25 amostras no conjunto de treinamento (25 com câncer) e 55 amostras no

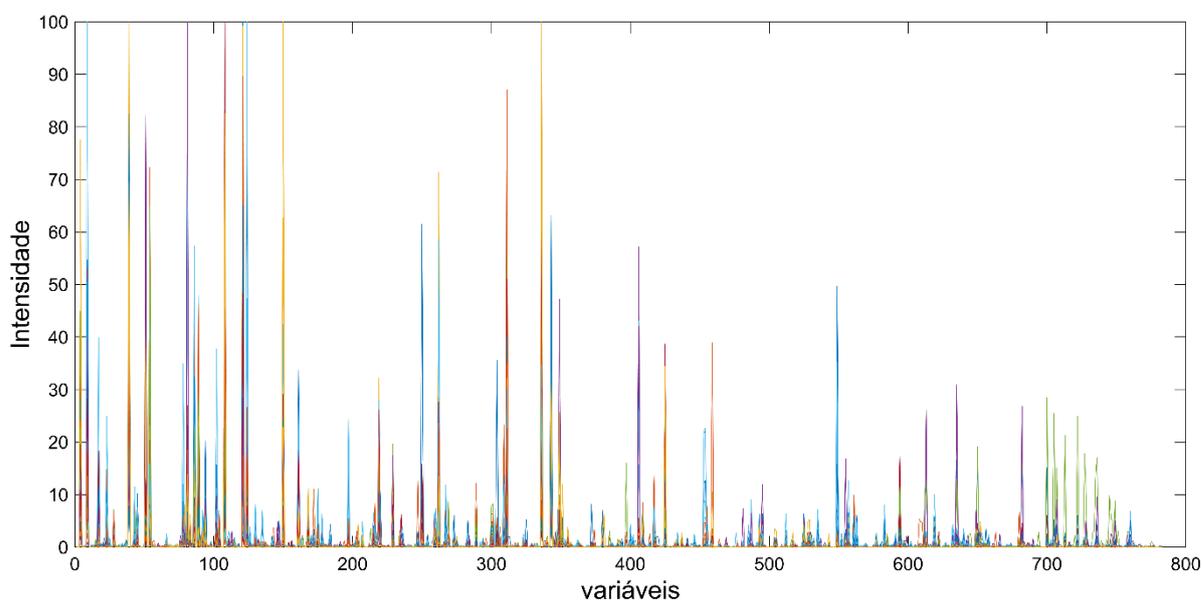
conjunto teste (15 com câncer e 40 sadios). Em ambos os casos os modelos foram validados conforme os parâmetros de mérito da tabela de confusão.

6 Resultados e Discussões

6.1 Espectros de massa das amostras de urina analisadas por PSI-HRMS

Na Figura 16 são mostrados os espectros de massa obtidos por PSI-MS para as 80 amostras, sendo 40 de urinas de pacientes com câncer e 40 de urinas de pacientes sadios, usadas para a construção dos modelos de calibração/treinamento e validação/teste. Os espectros foram pré-processados utilizando a normalização da soma, ou seja, dividindo cada região do espectro pela respectiva soma total da área de integração do espectro. Os espectros de cada amostra realizados em modo de ionização positiva mostraram os picos protonados em razão da sua massa pela carga (70-1.000 m/z).

Figura 16 - Gráfico dos espectros de massa obtidos por PSI-HRMS, amostras da urina de 40 pacientes com câncer de próstata e 40 pacientes sadios



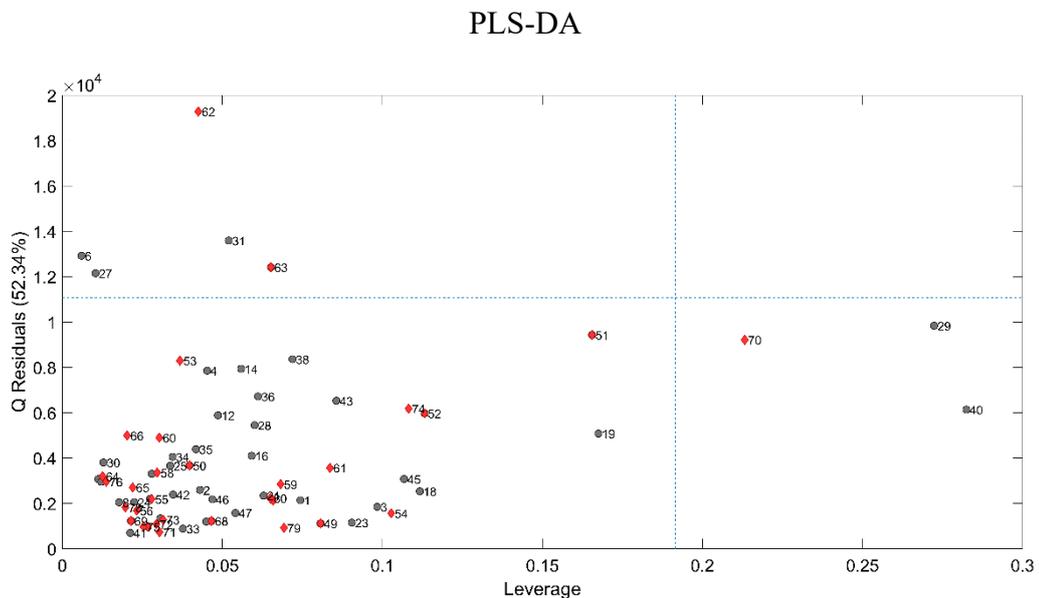
Ressalta-se que uma análise visual se torna quase impossível na diferenciação de um espectro do outro por conta das sobreposições de sinais, portanto, nesses casos, faz-se

necessária a utilização de métodos multivariados para a extração das informações para correlacionar com a propriedade de interesse, no caso a classe das amostras.

6.2 Resultados do modelo PLS-DA

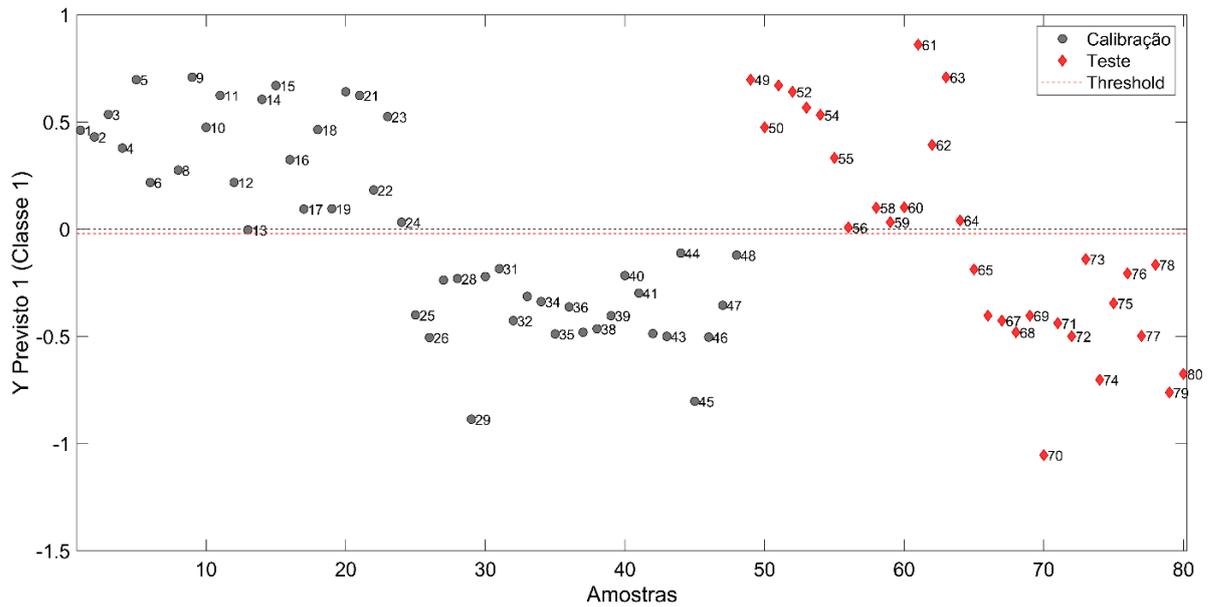
O modelo foi construído com 3 variáveis latentes (VL) e capturou 47,66 e 39,60% de variância explicada nos blocos X e y , respectivamente. A validação cruzada foi realizada por janelas venezianas (*venetian blinds*) utilizando 24 janelas com 2 amostras. Os dados foram centrados na média no bloco X . Nenhuma amostra anômala (*outlier*) foi detectada nesse modelo conforme mostra a Figura 17.

Figura 17 - Gráfico de Leverage versus Resíduos com 95% de confiança para o modelo



A Figura 18 apresenta o gráfico de estimativas por amostras obtido no modelo PLS-DA, onde observa-se que houve a separação das classes de forma eficiente. As amostras de 1 a 48 são as amostras de treinamento, isto é, as amostras que foram utilizadas para a construção do modelo. As amostras de 49 a 80 (totalizando 32) são as amostras do conjunto de teste para avaliar se o modelo construído consegue prever bem amostras desconhecidas. Observando o gráfico de estimativas verifica-se que: Todas as amostras de ambas as classes foram classificadas corretamente, sendo as amostras de urina dos pacientes com câncer de próstata com valores acima da linha *threshold*, enquanto as dos pacientes saudáveis abaixo.

Figura 18 - Gráfico de Estimativas para amostras de urinas de pacientes com câncer de próstata e saudáveis



O modelo PLS-DA foi validado usando as amostras do conjunto de teste e produziram os resultados dos parâmetros de classificação apresentados na Tabela 2.

Tabela 2 – Parâmetros de classificação obtidos para o modelo PLS-DA

Parâmetros	Valores do modelo
	PLS-DA
RMSECV	0,5739
RMSEC	0,5398
RMSEP	0,5862
Especificidade (Cal, CV, Prev)	1,0
Sensibilidade (Cal, CV, Prev)	1,0

Cal = calibração; CV = validação cruzada; Prev = previsão.

Os valores da sensibilidade e especificidade do modelo PLS-DA foram iguais a 1 significando que os modelos classificaram corretamente todas as amostras verdadeiras positivas e verdadeiras negativas, respectivamente.

A Tabela 3 apresenta a tabela de confusão do modelo PLS-DA na qual se observa que as taxas de *vp* e *vn* foi de 100% e as taxas de *fp* e *fn* foi de 0% para os conjuntos de treinamento e de teste das amostras de urina de pacientes com câncer e saudáveis. Isso significa que:

- Na etapa de treinamento do modelo, todas as 24 amostras de pacientes com câncer de próstata foram previstas na classe 1 e nenhuma amostra de sadios foi prevista na classe 1, assim como todas as 24 amostras de sadios foram previstas na classe 0 e nenhuma amostra com câncer de próstata foi prevista na classe 0;
- Na etapa de teste, todas as 16 amostras de pacientes com câncer foram previstas na classe 1 e nenhuma amostra de sadios foi prevista na classe 1, assim como todas as 16 amostras de sadios foram previstas na classe 0 e nenhuma amostra com câncer foi prevista na classe 0, como pode-se observar na tabela de confusão (Tabela 3);
- A Eficiência em ambas as etapas foi igual a 1 o que significa que o modelo PLS-DA apresenta um bom ajuste e alto desempenho;
- O Coeficiente de Correlação de Matthew's também resultou 1 para ambas as etapas o que representa uma classificação perfeita.

Tabela 3 – Tabela de confusão do modelo PLS-DA nos conjuntos de treinamento e de teste

		Tabela de confusão				
		VP	FP	VN	FN	N
Treinamento	Classe 1	100	0	100	0	24
	Classe 0	100	0	100	0	24
	Eficiência = 1 Coeficiente de Correlação Matthew's = 1					
Teste	Classe 1	100	0	100	0	16
	Classe 0	100	0	100	0	16
	Eficiência = 1 Coeficiente de Correlação Matthew's = 1					

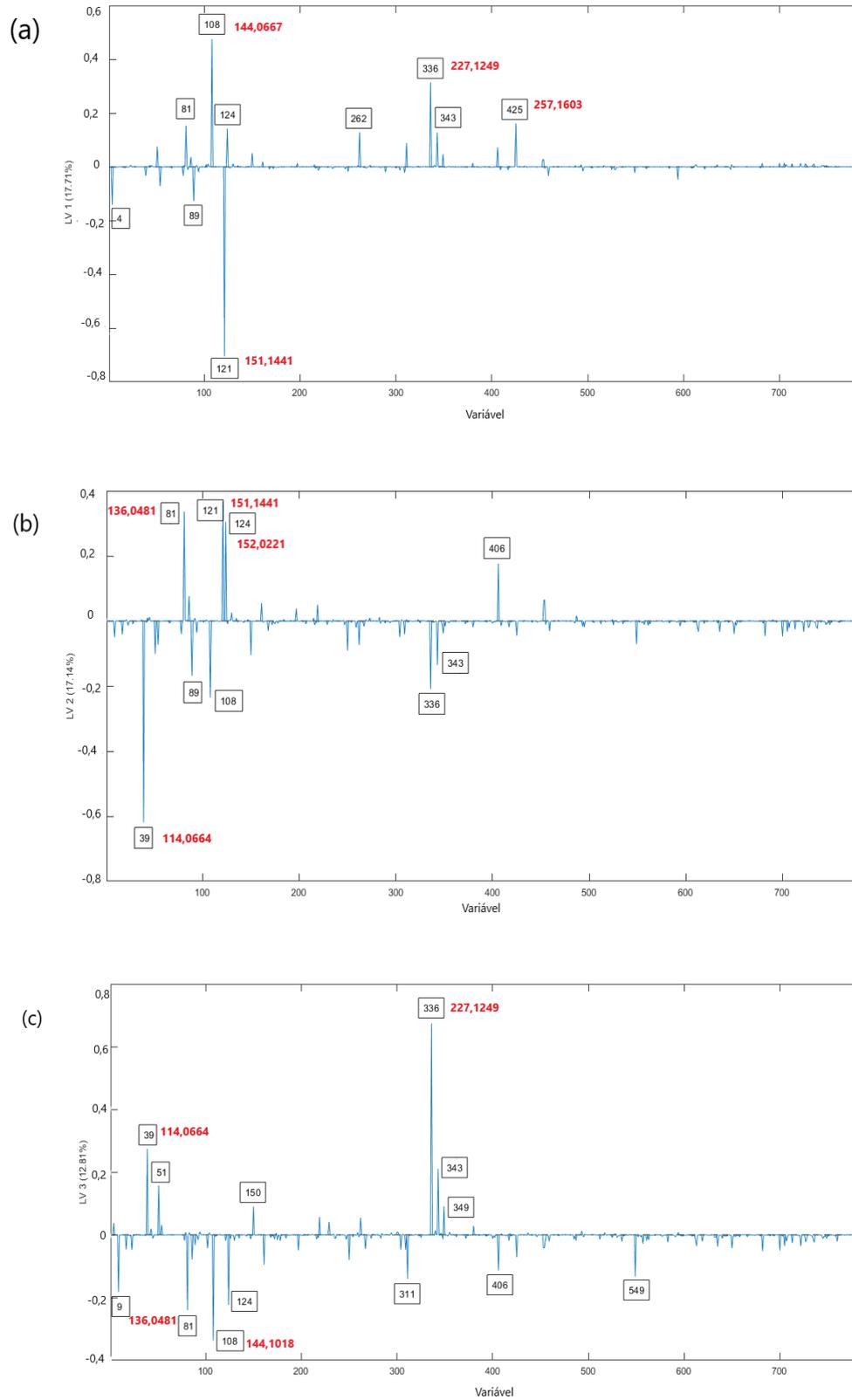
VP = taxa de verdadeiro positivo; VN = taxa de verdadeiro negativo;

FP = taxa de falso positivo; FN = taxa de falso negativo;

N = número de amostras.

A análise dos pesos das variáveis mostra quais variáveis foram mais significativas para a classificação das amostras de urina, onde se diferenciou as dos pacientes com câncer de próstata e as dos pacientes sadios. Como o modelo foi construído com 3 VL podemos visualizar nos gráficos da Figura 19 quais variáveis pesam mais em cada uma das variáveis latentes usadas no modelo. Se analisar os pesos das variáveis latentes, constata-se que na VL1 (Figura 19 (a)) as variáveis significativas foram (121, 108, 336, 425), na VL2 (Figura 19 (b)) foram (39, 81, 121, 124) e na VL3 (Figura 19 (c)) foram (39, 81, 108, 336), respectivamente.

Figura 19 - Gráfico dos pesos para as 3 VL usadas para a construção do modelo PLS-DA, sendo (a) primeira variável latente; (b) segunda variável latente e (c) terceira variável latente

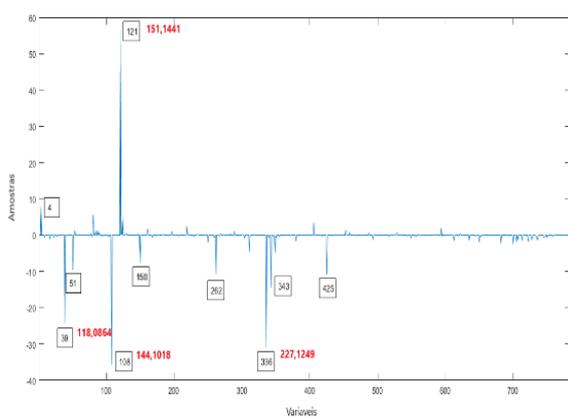


Também podemos plotar gráficos de VIP escores (representam as variáveis mais importantes nos escores das amostras) conforme mostra a Figura 20, assim escolhemos no conjunto de treinamento as amostras de números 1 (com câncer de próstata) (Figura 20 (a)) e 45 (sadio) (Figura 20 (b)) e no conjunto teste as amostras de números 49 (com câncer de próstata) (Figura 20 (c)) e 76 (sadio) (Figura 20 (d)).

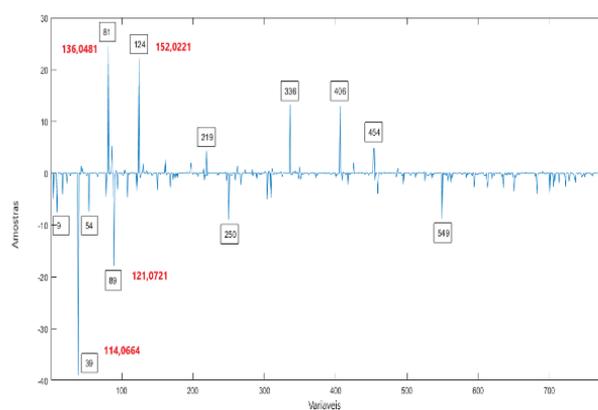
A partir da avaliação dos VIP escores, é possível pressupor que as razões m/z mais significativas para as amostras de urina com câncer de próstata foram 151,1441 (variável 121); 144,1018 (108); 227,1249 (336) e 257,1603 (425). Assim, esses íons podem levar as moléculas dos compostos a serem testados com biomarcadores para o câncer de próstata. Além deles, os demais em destaque nas Figuras 20 (a) e (c) podem contribuir neste sentido.

Figura 20 – Gráficos VIP escores do modelo PLS-DA

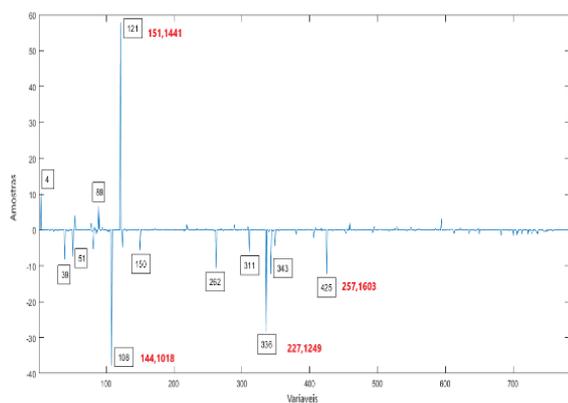
(a) Amostra do treinamento com câncer



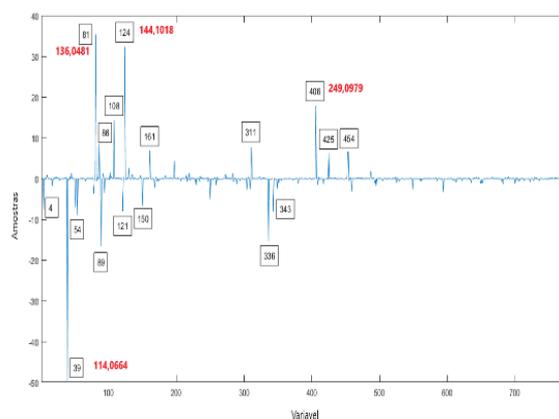
(b) Amostra do treinamento sadio



(c) Amostra do teste com câncer



(d) Amostra do teste sadio



6.3 Resultados do modelo DD-SIMCA

O modelo DD-SIMCA foi desenvolvido usando 4 componentes principais. Este número de componentes principais escolhido apresentou os melhores resultados, explicando 93,96% da variância total dos dados no conjunto de treinamento com $\alpha = 0,01$. A sensibilidade do modelo na etapa de treinamento foi igual a 1, o que equivale a dizer que nenhuma amostra da classe alvo (pacientes com câncer de próstata - PCa) foi classificada como extrema ou *outliers*. A soma da distância normalizada da pontuação e da distância ortogonal definiu o tipo de área de aceitação em «qui-quadrado», ou seja, uma área triangular (Figura 21). Esta figura mostra a classificação das amostras de PCa obtida pelo modelo DD-SIMCA, onde pode-se observar que nenhuma amostra de PCa foi encontrada na área dos extremos (espaço entre as linhas verde e vermelha) ou na área de *outlier* (espaço após a linha vermelha), que significa que todas as amostras de PCa foram classificadas corretamente porque estão dentro da área de aceitação, delimitadas pela linha verde com $\alpha = 0,01$.

A Tabela 4 apresenta os resultados obtidos na etapa de treinamento do modelo DD-SIMCA.

Figura 21 - Gráfico de aceitação de amostras com câncer de próstata pelo modelo DD-SIMCA no processo de treinamento. A área de aceitação é o limite para $\alpha = 0,05$, delimitação outlier é o limite para $\alpha = 0,01$

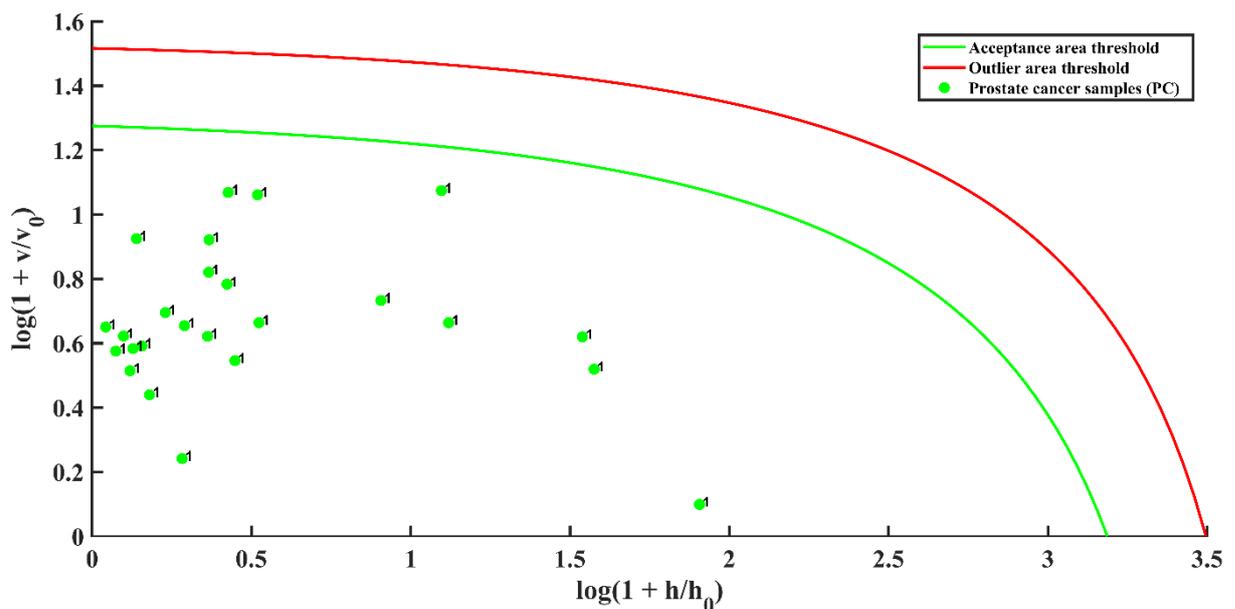
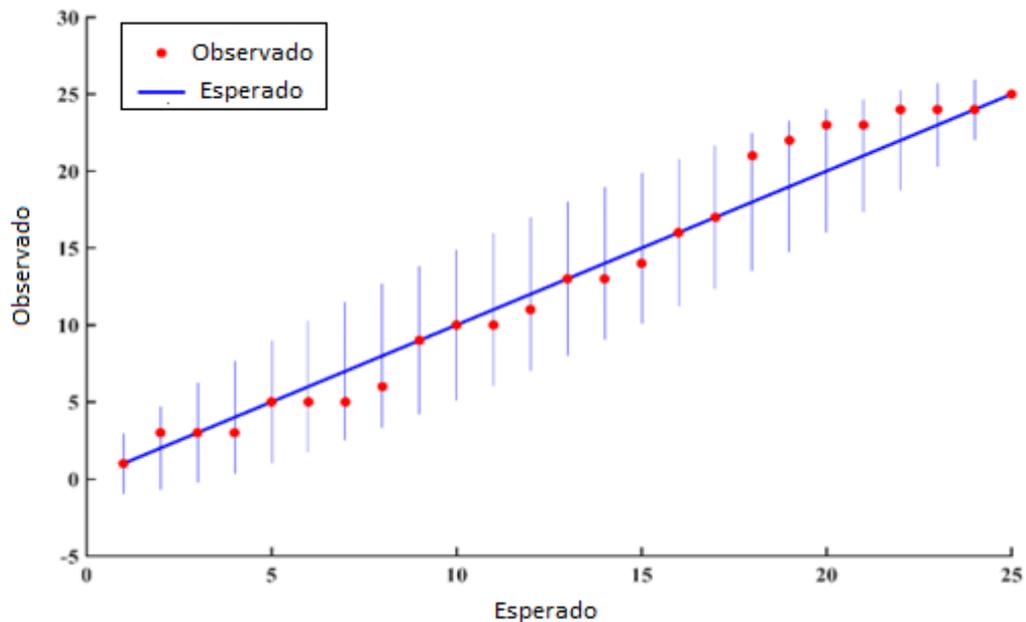


Tabela 4 - Resultados do treinamento do modelo DD-SIMCA

α	Componente Principal	Variância explicada	Amostras	Extremos	Ponto fora da curva	Sensibilidade
0,01	4	93,96%	25	0	0	1

A verificação das amostras *outliers* no conjunto de treinamento baseou-se na análise da dependência do número observado de amostras extremas versus os valores teoricamente esperados, apresentados no gráfico extremo especial (Figura 22). Nesta figura, observa-se que todas as amostras se enquadraram na área de tolerância (mostrada em linhas verticais), demonstrando que o conjunto de treinamento não contém *outliers*.

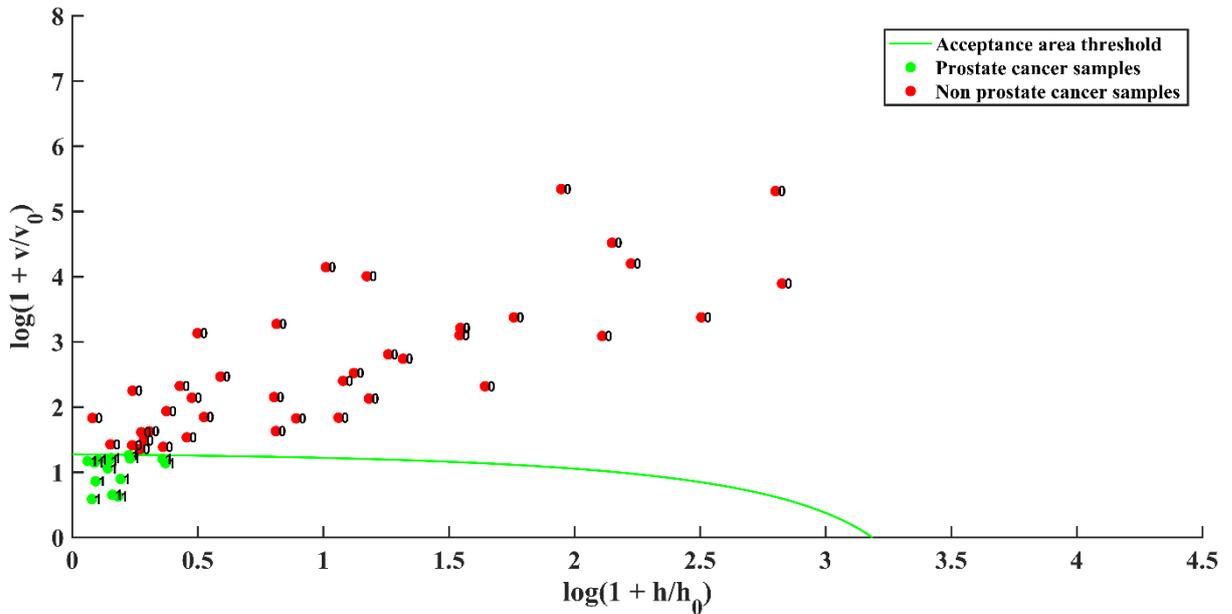
Figura 22 - Gráfico de verificação de *outlier* em amostras com câncer de próstata pelo modelo DD-SIMCA na etapa de treinamento



A capacidade preditiva do modelo DD-SIMCA foi avaliada com um conjunto de teste composto por 15 amostras da classe alvo (PCa) e 40 amostras de indivíduos saudáveis sem história médica prévia de câncer (NPCa). A especificidade de predição foi 1, o que equivale a dizer que todas as amostras do conjunto de teste foram classificadas corretamente em suas respectivas classes. A Figura 23 mostra a classificação das amostras do conjunto de teste, onde pode se observar que todas as amostras de PCa estão dentro da área de aceitação (área

delimitada pela curva de cor verde), representando uma classificação correta, enquanto que as amostras de urina de indivíduos saudáveis sem histórico médico anterior de câncer (NPCa) foram classificadas corretamente como de não interesse porque estão fora da área de aceitação, indicando que há diferenças entre os perfis espectrais dessas amostras com as amostras da classe alvo.

Figura 23 - Gráfico de aceitação de amostras com e sem câncer de próstata pelo modelo DD-SIMCA na etapa de teste



A Tabela 5 apresenta a tabela de confusão do modelo DD-SIMCA na qual se observa que as taxas de vp e vn foi de 100% e as taxas de fp e fn foi de 0% para os conjuntos de treinamento e de teste das amostras de urina de pacientes com câncer e sadios. Isso significa que:

- Na etapa de treinamento do modelo, todas as 25 amostras de pacientes com câncer de próstata foram previstas na classe alvo e nenhuma amostra ficou fora da área dos extremos;
- Na etapa de teste, todas as 15 amostras de pacientes com câncer foram previstas na classe alvo e nenhuma amostra de sadios foi prevista nessa classe, assim como todas as 40 amostras de sadios foram previstas fora da classe alvo e nenhuma amostra com câncer foi prevista fora da área de aceitação, como pode-se observar na tabela de confusão (Tabela 5);
- A Eficiência em ambas as etapas deu resultado 1 o que significa que o modelo DD-SIMCA apresenta um bom ajuste e alto desempenho;

- O Coeficiente de Correlação Matthew's também resultou 1 para ambas as etapas o que representa uma classificação perfeita.

Tabela 5 – Tabela de confusão do modelo DD-SIMCA nos conjuntos de treinamento e de teste

		Tabela de confusão				
		VP	FP	VN	FN	N
Treinamento	Classe 1	100	0	100	0	25
	Eficiência = 1					
	Coeficiente de Correlação Matthew's = 1					
Teste	Classe 1	100	0	100	0	15
	Classe 0	100	0	100	0	40
	Eficiência = 1					
Coeficiente de Correlação Matthew's = 1						

VP = taxa de verdadeiro positivo; VN = taxa de verdadeiro negativo;

FP = taxa de falso positivo; FN = taxa de falso negativo;

N = número de amostras.

A aplicação da espectrometria de massa com ionização por spray de papel associada aos métodos quimiométricos PLS-DA e DD-SIMCA permitiu o desenvolvimento de métodos de classificação para o diagnóstico não invasivo do câncer de próstata utilizando apenas informações relevantes dos espectros de massa das amostras de urina dos pacientes. A eficiência de tais métodos foi avaliada com base nos parâmetros de exatidão, sensibilidade e especificidade que apresentaram classificação correta das amostras do conjunto de treinamento e conjunto de teste de 100% para os dois modelos. Esses métodos analíticos propostos no trabalho são de fácil interpretação e eficientes, fazendo com que sejam alternativas viáveis para serem implementados visando o diagnóstico de pacientes com câncer de próstata através da análise não invasiva de suas urinas, contribuindo dessa forma para o diagnóstico precoce que leva uma maior chance de êxito quanto ao tratamento médico.

7 Conclusão

A utilização do (PSI-HRMS), associado aos métodos quimiométricos PLS-DA e DD-SIMCA, se mostrou uma alternativa viável na aplicação de um método para triagem de amostras de urinas a fim de detectar o câncer de próstata.

A espectrometria de massa com ionização por spray de papel mostrou-se ser um método rápido (em média um minuto por análise), não demanda pré-tratamento complexo da amostra e consome uma quantidade mínima de amostra e solvente.

A aplicação dos métodos PLS-DA e DD-SIMCA nos dados espectrais permitiu o desenvolvimento de dois métodos validados, estáveis, robustos e simples de interpretar para classificar amostras de pacientes com câncer de próstata e sadios.

Os resultados de classificação de ambos os modelos foram bastante satisfatórios sendo de 100% para os dois modelos.

De modo geral, os métodos PLS-DA e DD-SIMCA forneceram resultados promissores e concordantes entre si, de modo que os dois modelos podem ser considerados aptos para o desenvolvimento de métodos para detectar a presença de câncer de próstata através da análise de urina. Sendo promissor a possível identificação de biomarcadores, usando a análise de pesos, o que tornaria possível o desenvolvimento de exames do tipo rápido como o de Covid e do teste de gravidez, e até auxiliar no desenvolvimento de novos fármacos mais eficientes no tratamento do câncer de próstata.

A presente pesquisa foi divulgada em vários sites, trabalhos acadêmicos, redes sociais, blog e em jornais como: Comunica UFU.BR, G1 Triângulo e Alto Parnaíba, Diário de Uberlândia, FARMAT4H, Fiocruz.Br, PPGQUI.IQ.UFU/acontece, Conselho Federal de Química, Conselho Regional de Química IX Região, Programa Manhã Total, Jornal da UFU, Programa Balanço Geral- TV Parnaíba. Em rádio como: Radio Ciência- Universidade FM.

8 Referências Bibliográficas

ADENAN, M. N. H. et al. **Triagem de ninhos de pássaros comestíveis da Malásia para adulterantes estruturais e origem geográfica usando espectroscopia de infravermelho médio - refletância total atenuada (MIR-ATR) combinada com análise quimiométrica por Data-Driven - Soft Independent Modeling of Class Analogy.** *Forensic Chem*, v. 17, 2020. <https://doi.org/10.1016/j.foodcont.2021.108542>

ALBERICE, J. V. **Avaliação analítica de potenciais biomarcadores para câncer de bexiga em urina.** Tese apresentada ao Programa de Pós-Graduação em Química do Instituto de Química de São Carlos – USP, como requisito para obtenção de título de Doutor em Química, São Carlos, 2014. <https://www.teses.usp.br/teses/disponiveis/75/75135/tde-28082014-152517/pt-br.php>

ALMEIDA, M. R. et al. **Classification of Amazonian rosewood essential oil by Raman spectroscopy and PLS-DA with reliability estimation.** *Talanta*, v. 117, p. 305–311, 2013. <https://doi.org/10.1016/j.talanta.2013.09.025>

BARKER, M.; RAYENS, W. **Partial least squares for discrimination.** *Journal of Chemometrics*, Chichester, v. 17, p. 166-173, 2003. <https://doi.org/10.1002/cem.785>

BLOG DO CLÁUDIO LIMA. Disponível em: <<https://www.claudiolimablog.com.br/2022/03/maranhese-e-destaque-em-importante.html>>. Acesso em: 02 de jul de 2022.

BARBIERI, C. E. et al. **The mutational landscape of prostate cancer.** *European Urology*, v. 64, p. 567–576, 2013. <https://doi.org/10.1016/j.eururo.2013.05.029>

BRERETON, R. G. **Introduction to multivariate calibration in analytical chemistry** *Analyst*. v. 125. p. 2125-2154. 2000. <http://dx.doi.org/10.1039/b003805i>

Carvalho, Hernandes F. **Estudos desenvolvidos no IB e em outras instituições revelam que a mulher também pode ter a glândula.** Universidade Estadual de Campinas, 2007. <https://link.springer.com/article/10.1007/s43032-021-00475-9>

CHACE, D. H. **Mass Spectrometry in the Clinical Laboratory.** *Chemical Reviews*, v. 101, n. 2, p. 445-477, 2001. <https://pubs.acs.org/doi/10.1021/cr990077%2B>

DE SANTANA, F. B., BORGES NETO, W., POPPI, R. J. **Random forest as one-class classifier and infrared spectroscopy for food adulteration detection.** *Food Chemistry*, v. 293, p. 323-332, 2019. <https://doi.org/10.1016/j.foodchem.2019.04.073>

CHEN, T, MORRIS, J, MARTIN, E. **Response to the discussion of “Gaussian process regression for multivariate spectroscopic calibration”**. *Chemometrics and Intelligent Laboratory Systems*, v. 87, p. 69-71, 2007.

<https://doi.org/10.1016/J.CHEMOLAB.2006.09.007>

DEEV, V. et al. **Prostate cancer screening using chemometric processing of GC–MS profiles obtained in the headspace above urine samples**. *Journal of Chromatography B*, v. 1155, p. 122298, 2020.

<https://doi.org/10.1016/j.jchromb.2020.122298>

DRIVELOS, S. A. et al. **Geographical origin and botanical type honey authentication through elemental metabolomics via chemometrics**. *Food Chemistry*, v. 338, p. 127936, 2021.

<https://doi.org/10.1016/j.foodchem.2020.127936>

EPSTEIN, J. I. et al. **The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma**. *American Journal of Surgical Pathology*, v. 29, n. 9, p. 1228–1242, 2005.

<https://doi.org/10.1097/01.pas.0000173646.99337.b1>

ESPY, R. D.; MULIADI, A. R.; OUYANG, Z.; COOKS, R. G. **Spray mechanism in paper spray ionization**. *International Journal of Mass Spectrometry*. v. 167, p. 325-327, 2012.

<https://doi.org/10.1016/j.ijms.2012.06.017>

FENN, J. B.; MANN, M.; MENG, C. K.; WONG, S. F.; WHITEHOUSE, C. M. **Electrospray ionization for mass-spectrometry of large biomolecules**. *Science*. v. 246, p. 64-71, 1989.

<https://doi.org/10.1002/chin.199005359>

FERLAY, J. *et al.* **Cancer incidence and mortality worldwide: sources, methods and major patterns in Globocan**. *International Journal of Cancer*, v. 136, p. E359-E386, 2015.

<https://doi.org/10.1002/ijc.29210>

FERREIRA, M. M. C. **Quimiometria - Conceitos, Métodos e Aplicações**. Editora Unicamp. Campinas - SP. 2015.

<https://doi.org/10.7476/9788526814714>

GUPTA, K., BHAVSAR, A., SAO, A. K. **Detecting mitotic cells in HEp-2 images as anomalies via one class classifier**. *Computers in Biology and Medicine*, v. 111, p. 103328, 2019.

<https://doi.org/10.1016/j.compbiomed.2019.103328>

HARRIS, D. C. **Análise Química Quantitativa**. Editora LTC, 6ª Edição, São Paulo, 2005.

HOFFMANN, E.; STROOBANT, V. **Mass Spectrometry: Principles and Applications**, 3ª Edição, England, 2007.

INCA – Instituto Nacional de Câncer. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-prostata>>. Acesso em 10 de junho de 2022.

JORNAL DA UFU. Disponível em: <<https://www.youtube.com/watch?v=Ett4rCD3Tb0>>. Acesso em: 02 de jun 2022.

MÁQUINA, A. D. V. **Espectrometria no infravermelho médio e métodos quimiométricos PLS-DA e PLS: classificação e previsão do teor de biodiesel na mistura de biodiesel/diesel de mafurra, moringa e algodão.** Dissertação apresentada ao Programa de PósGraduação em Química do Instituto de Química da Universidade Federal de Uberlândia, como requisito para obtenção de título de Mestre em Química. Uberlândia, 2017.
<https://repositorio.ufu.br/handle/123456789/18098>

MÁQUINA, A. D. V., SITOIE, B. V., DE SANTANA, F. B., SANTOS, D. Q., BORGES NETO, W. **Determinação da Adulteração da Mistura B10 de Diesel e Biodiesel de Crambe Usando Espectroscopia de Ressonância Magnética Nuclear de Proton (RMN de 1H) com um Modelo de Modelagem Independente de Análise de Classe (DD-SIMCA) baseado em dados.** Anal Lett 2020; 0: 1-12. <https://doi.org/10.1080/00032719.2020.1782928>
<https://doi.org/10.1080/00032719.2020.1782928>

MASSART, D. L. et al. **Handbook of Chemometrics and Qualimetrics**, Volume 20B. Amsterdã: Elsevier Science, 1998.
<https://doi.org/10.14393/ufu.di.2017.115>

MAZIVILA, S. J., PÁSCOA, R. N. M.J. , CASTRO, R. C., RIBEIRO, D. S.M. , SANTOS, J. L. M. **Detection of melamine and sucrose as adulterants in milk powder using near-infrared spectroscopy with DD-SIMCA as one-class classifier and MCR-ALS as a means to provide pure profiles of milk and of both adulterants with forensic evidence: A short communication.** Talanta, v. 216, p. 120937, 2020.
<https://doi.org/10.1016/j.talanta.2020.120937>

MENDES, G. **Detecção de adulterações em gasolina e previsão de parâmetros físico-químicos a partir de curvas de destilação associadas a ferramentas quimiométricas.** 2012.
<http://hdl.handle.net/1843/SFSA-8ZKRMX>

MENDONÇA, K. H. **Espectrometria de massa com ionização por paper spray e métodos quimiométricos aplicados à identificação de adulterações em açúcar e óleo de coco.** 2018.
<http://hdl.handle.net/1843/SFSA-B23Q4X>

MOTTET, N. C. *et al.* **Guidelines on prostate cancer.** European Association of Urology. 2015; p. 1-137. https://uroweb.org/wp-content/uploads/09-ProstateCancer_LR.pdf

MOYA, M. M., KOCH, M. W., HOSTETLER, L. D. **Classifier networks of a class for target-aware applications.** I Int. Neural Netw. Soe, Portland, OR: Proc. Word Congr. Redes neurais; 1993, p. 797–801.

NEVES, M. D. G., POPPI, R. J. **Authentication and identification of adulterants in virgin coconut oil using ATR/FTIR in tandem with DD-SIMCA one class modeling.** Talanta, v. 219, p. 121338, 2020.
<https://doi.org/10.1016/j.talanta.2020.121338>

OLIVERI, P. **Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues – A tutorial.** *Analytica Chimica Acta*, v. 982, p. 9-19, 2017.

<https://doi.org/10.1016/j.aca.2017.05.013>

PANTAZI, X. E., MOSHOU, D., TAMOURIDOU, A. A. **Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers.** *Computers and Electronics in Agriculture*, v. 156, p. 96-104, 2019.

<https://doi.org/10.1016/j.compag.2018.11.005>

PAIVA, R. M. **Avaliação de micrornas como biomarcadores moleculares no câncer de próstata.** Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas: Fisiologia da Universidade Federal do Rio Grande do Sul, como requisito para obtenção de título de Doutor em Fisiologia. Porto Alegre, 2020.

<http://hdl.handle.net/10183/218110>

PAVIA, D. L., LAMPMAN, G. M., KRIZ, G. S., VYVYAN, J. R. **Introdução à Espectroscopia.** Cengage Learning. São Paulo. 2010.

POMERANTSEV, A. L. **Acceptance areas for multivariate classification derived by projection methods.** *Journal of Chemometrics*, v. 22, p. 601-609, 2008.

<https://doi.org/10.1002/cem.1147>

POMERANTSEV, A. L., RODIONOVA, O. Y. **On the type II error in SIMCA method.** *Journal of Chemometrics*, v. 28, p. 518-522, 2014.

<https://doi.org/10.1002/cem.2610>

POMERANTSEV, A. L., RODIONOVA, O. Y. **Concept and role of extreme objects in PCA/SIMCA.** *Journal of Chemometrics*, v. 28, p. 429-438, 2014.

<https://doi.org/10.1002/cem.2506>

PRAMANIK, B. N.; GANGULY, A. K.; GROSS, M. L. **Practical spectroscopy series, volume 32,** Applied Electrospray Mass Spectrometry. New York, 2002.

PROGRAMA BALANÇO GERAL- TV PARANAÍBA. Disponível em: <<https://tvparanaiba.com.br/arquivos/videos/estudo-da-ufu-propoe-diagnostico-do-cancer-de-prostata-atraves-da-urina-balanco-geral>>. Acesso em: 02 de jun 2022.

PROGRAMA MANHÃ TOTAL. Disponível em: <<https://www.youtube.com/watch?v=MdYtOfYfCuk>>. Acesso em: 02 de jun 2022.

RADIO CIENCIA- UNIVERSIDADE FM. Disponível em: <http://www.universidadefm.ufma.br/jornalismo106/radiociencia/engenheira-quimica-da-baixada-maranhense-se-destaca-em-pesquisa-que-identifica-diagnosticos-de-cancer-da-prostata>. Acesso em: 02 de jul de 2022.

RAJA, G., JUNG, Y., JUNG, S. H., KIM, T. J. **¹H-NMR-based metabolomics for cancer targeting and metabolic engineering –A review.** *Process Biochemistry*, v. 99, p. 112-122, 2020.

<https://doi.org/10.1016/j.procbio.2020.08.023>

RIVERA, R., GARRIDO, N. “Metabolomics.” *Metabolomics. Oxidants, Antioxidants and Impact of the Oxidative Status in Male Reproduction*: 277–85, 2019.

<https://doi.org/10.1039/c1mt00070e>

RODIONOVA, O. Y., TITOVA, A. V., POMERANTSEV, A. L. **Discriminant analysis is an inappropriate method of authentication.** *TrAC - Trends in Analytical Chemistry*, v. 78, p. 17-22, 2016.

<https://doi.org/10.1016/j.trac.2016.01.010>

RODIONOVA, O. Y., OLIVERI, P., POMERANTSEV, A. L. **Rigorous and compliant approaches to one-class classification.** *Chemometrics and Intelligent Laboratory Systems*, v. 159, p. 89-96, 2016.

<https://doi.org/10.1016/j.chemolab.2016.10.002>

SARRIS, A. B. et al. **Câncer de Próstata: uma breve revisão.** *Visão Acadêmica*, v. 19, p. 137-151, 2018.

<http://dx.doi.org/10.5380/acd.v19i1.57304>

SIEGEL, R. L., MILLER, K. D., JEMAL, A. **Cancer Statistics, Ca-Cancer.** *J. Clin.*, v. 69, p. 7-34, 2019.

WISHART, D. S. **Emerging Applications of Metabolomics in Drug Discovery and Precision Medicine.** *Nat. Rev. Drug Discov.*, v. 15, p. 473-484, 2016.

<https://doi.org/10.3322/caac.21332>

TV PARANAÍBA. Disponível em: <<https://tvparanaiba.com.br/arquivos/videos/estudo-da-ufu-propoe-diagnostico-do-cancer-de-prostata-atraves-da-urina-jornal-paranaiba>>. Acesso em: 02 de jun de 2022.

YANG, Q. et al. **“Metabolomics Biotechnology, Applications, and Future Trends: A Systematic Review”.** *Royal Society of Chemistry*, v. 9, p. 37245–37257, 2019.

<https://doi.org/10.1039/C9RA06697G>

ZONTOV, Y., RODIONOVA, O. Y., KUCHERYAVSHIY, S., POMERANTSEV, A. **“DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach,”** *Chemometrics and Intelligent Laboratory Systems*, v. 167, p. 23-28, 2017.

<https://doi.org/10.1016/j.chemolab.2017.05.010>