



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA
UNIVERSIDADE FEDERAL DOS VALES
DE JEQUITINHONHA E MUCURI
PROGRAMA DE PÓS-GRADUAÇÃO
EM BIOCOMBUSTÍVEIS**



LÍGIA RODRIGUES DE OLIVEIRA

**CLASSIFICAÇÃO DO TEOR DE BIODIESEL
METÁLICO DE MACAÚBA E NÍGER NA MISTURA
COM DIESEL, POR MEIO DA TÉCNICA DE
ESPECTROSCOPIA MIR E DOS MÉTODOS
CLASSIFICADORES PLS-DA E SVM**

UBERLÂNDIA

2021

LÍGIA RODRIGUES DE OLIVEIRA

**CLASSIFICAÇÃO DO TEOR DE BIODIESEL
METÁLICO DE MACAÚBA E NÍGER NA MISTURA
COM DIESEL, POR MEIO DA COMBINAÇÃO DA
TÉCNICA DE ESPECTROSCOPIA MIR E DOS
MÉTODOS CLASSIFICADORES PLS-DA E SVM**

Tese apresentada ao Programa de Pós-Graduação em Biocombustíveis, da Universidade Federal de Uberlândia, como exigência parcial para a obtenção do Título de Doutora em Ciência e Tecnologia de Biocombustíveis.

Área de Concentração/Linha de pesquisa: Biocombustíveis/Produtos e coprodutos.

Orientador: Prof. Dr. Douglas Queiroz Santos

UBERLÂNDIA

2021

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

O48 2021	<p>Oliveira, Lígia Rodrigues de, 1985- Classificação do teor de biodiesel metílico de macaúba e níger na mistura com diesel, por meio da combinação da técnica de espectroscopia MIR e dos métodos classificadores PLS-DA e SVM [recurso eletrônico] / Lígia Rodrigues de Oliveira. - 2021.</p> <p>Orientador: Douglas Queiroz Santos . Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-graduação em Biocombustíveis. Modo de acesso: Internet. Disponível em: http://doi.org/10.14393/ufu.te.2021.175 Inclui bibliografia. Inclui ilustrações.</p> <p>1. Biocombustível. I. , Douglas Queiroz Santos, 1980-, (Orient.). II. Universidade Federal de Uberlândia. Pós- graduação em Biocombustíveis. III. Título.</p> <p style="text-align: right;">CDU: 662.756</p>
-------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em
Biocombustíveis

Av. João Naves de Ávila, 2121 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4385 / 4208 - www.iq.ufu.br - ppbic@iqufu.ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Biocombustíveis			
Defesa de:	Tese de Doutorado, 14, PPBIC			
Data:	24 de fevereiro de 2021	Hora de início:	14:00	Hora de encerramento: 16:30
Matrícula do Discente:	11713PGB001			
Nome do Discente:	Lígia Rodrigues de Oliveira			
Título do Trabalho:	CLASSIFICAÇÃO DO TEOR DE BIODIESEL METÍLICO DE MACAÚBA E NÍGER NA MISTURA COM DIESEL, POR MEIO DA COMBINAÇÃO DA TÉCNICA DE ESPECTROSCOPIA MIR E DOS MÉTODOS CLASSIFICADORES PLS-DA E SVM			
Área de concentração:	Biocombustíveis			
Linha de pesquisa:	Produtos e coprodutos			
Projeto de Pesquisa de vinculação:				

Reuniu-se em ambiente virtual, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Biocombustíveis, assim composta: Professores Doutores: João Paulo Victorino Santos, do Instituto Federal de Educação Ciência e Tecnologia de Goiás - Campus Itumbiara; Eduardo Chagas de Oliveira, da UNIUBE; Antônio Carlos Ferreira Batista e Anízio Márcio de Faria, da Universidade Federal de Uberlândia; e Douglas Queiroz Santos, orientador(a) do(a) candidato(a).

Iniciando os trabalhos o(a) presidente da mesa, Dr(a). Douglas Queiroz Santos, apresentou a Comissão Examinadora e o candidato(a), agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do 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):

[A]provado(a).

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

O competente diploma será expedido após cumprimento dos demais requisitos,

conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Anizio Marcio De Faria, Professor(a) do Magistério Superior**, em 26/02/2021, às 16:08, 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 **Douglas Queiroz Santos, Professor(a) do Ensino Básico, Técnico e Tecnológico**, em 27/02/2021, às 10:27, 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 **Antonio Carlos Ferreira Batista, Professor(a) do Magistério Superior**, em 01/03/2021, às 14:55, 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 **JOÃO PAULO VICTORINO SANTOS, Usuário Externo**, em 04/03/2021, às 14:50, 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 **Eduardo Chagas de Oliveira, Usuário Externo**, em 09/03/2021, às 17:21, 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 **2594663** e o código CRC **ACC6D384**.

Dedico este trabalho em especial a Deus. Aos meus pais Geraldo Gomes de Oliveira e Anita Rodrigues de Brito. Também, aos meus irmãos e amigos que compartilharam desse momento, me incentivando e confiando no meu potencial.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, por ser sempre o meu horizonte, o Pai protetor com quem posso contar e a Nossa Mãe por me acolher sempre em seu colo nos momentos de atribulação.

À Universidade Federal de Uberlândia (UFU), e ao Programa de Pós Graduação em Biocombustíveis por me ter proporcionado a obtenção de conhecimento e formação na área de Ciência e Tecnologia de Biocombustíveis, além do presente trabalho de doutoramento.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao suporte financeiro para a realização desta pesquisa.

Ao meu orientador, Prof. Dr. Douglas Queiróz Santos, por me receber como sua orientanda, pelo desafio que me propôs na condução deste trabalho, me levando a ter determinação, mais confiança e foco para execução da tese. Obrigada pela orientação neste período.

Agradeço ao professor Ubirajara, com quem eu tive o prazer de trabalhar, me apoiando e dando suporte estatístico necessário. Obrigada pela dedicação e esforço em tempos tão difíceis que estamos passando, mostrando a sua generosidade e gentileza.

Ao coordenador Ricardo Reis e aos professores do Programa de Pós-Graduação de Biocombustíveis e do Instituto de Química da Universidade Federal de Uberlândia (UFU), pelo tratamento, conselhos e conhecimentos transmitidos.

Aos funcionários da UFU, ao Gabriel Fonseca Rezende da secretária do Programa de Pós Graduação em Biocombustíveis, pela atenção requeridas em todos os processos técnico-administrativos e aos técnicos do multiusuário pela assistência prestadas.

Aos meus pais, Geraldo Gomes e Anita Rodrigues de Brito, pelo amor incondicional a que me dedicaram, por proporcionarem esse caminho de estudo e me tornado quem eu sou.

Aos meus irmãos Lânio, Josman, Jusa, cunhadas, amigos e tios que me incentivaram e colaboraram nesse processo.

A todos, entre outros, amigos e colegas, que fizeram-se importantes, muito obrigada.

“A Sabedoria é sempre resplandecente. Ela é facilmente encontrada por aqueles que a amam e a procuram. Ela até se antecipa, dando-se a conhecer aos que a desejam. Quem por ela madruga não se cansa, pois a encontra sentada à sua porta. Meditar sobre ela é a perfeição da prudência; e quem fica acordado por causa dela, em breve há de viver despreocupado. Pois ela mesma sai à procura dos que a merecem, cheia de bondade, aparece-lhes nas estradas e vai ao seu encontro em todos os seus projetos” (LÍVRO DA SABEDORIA, Sb 6,12-16).

RESUMO

Nas últimas décadas, foram desenvolvidas políticas públicas de incentivo ao uso de biocombustíveis no Brasil, tais como o estabelecimento de um teor mínimo obrigatório de adição de biodiesel ao diesel comercializado. Visando evitar adulterações na composição dessa blenda estabelecida por lei, faz-se necessário um rigoroso controle de qualidade ao longo de toda cadeia produtiva e de distribuição. Nesta temática, pesquisas podem ser fomentadas, no sentido de desenvolver metodologias que forneçam respostas rápidas e confiáveis para o controle qualitativo dessa mistura binária. Neste trabalho, foram desenvolvidos modelos analíticos para classificar amostras de biodiesel metílico de macaúba e níger em mistura com diesel, usando a combinação das técnicas analíticas/quimiométricas de Espectroscopia MIR e dos métodos preditivos, como de Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e de Máquinas de Vetores de Suporte (SVM). Para a construção dos modelos discriminativos, produziu-se um total de 74 combinações de diesel BX, geradas pela mistura do biodiesel metílico de macaúba e níger ao diesel puro, em um intervalo de concentração de 0,25% a 30,00% (v/v). O primeiro grupo de amostras foi preparado utilizando a proporção de 10% de biodiesel para 90% de diesel puro (B10), e o segundo de acordo com as diferentes percentagens BX (B3 à B9 e B11 à B30 contendo, respectivamente, de 3 a 9% e 11 a 30% de biodiesel). Foram utilizadas em cada modelo PLS-DA, 50 amostras para os conjuntos de treinamento e de 24 amostras para teste. Na análise de qualidade dos modelos, os erros quadráticos médios obtidos foram baixos, com percentagens dentro da reprodutibilidade do método normatizado. Os critérios de desempenho obtidos para os modelos desenvolvidos, tais como sensibilidade, especificidade, precisão e eficiência, foram iguais a 100%, os quais também apresentaram excelente capacidade em discriminar todas as amostras em suas respectivas blends, com uma previsão de 100%. O algoritmo de SVM também mostrou ser uma boa opção para discriminação entre as classes B10 e BX do biodiesel metílico de níger e macaúba. Quanto à análise de desempenho dos modelos preditivos, a base de dados foi dividida em três subconjuntos: treinamento, teste e validação, compreendendo 50/12/12 amostras, respectivamente. Diferentes funções do *Kernel*, como a linear, polinomial, sigmoide e a função de base radial (RBF) foram testadas, visando avaliar a capacidade de generalização do classificador. Se tratando dos modelos de níger, o *kernel* RBF foi o que apresentou a melhor acurácia (0,92) em relação aos demais, com os hiperparâmetros definidos $C = 10$ e $\gamma = 10^{-4}$. Já para os modelos de macaúba, foram obtidas excelentes medidas de acurácia para todos os *Kernel* avaliados, com os hiperparâmetros definidos $C = 10$ e $\gamma = 10^{-5}$. A precisão e eficiência de classificação da SVM foi superior a 70,0% nas análises de teste e validação, em ambos os modelos desenvolvidos. Os resultados também revelaram valor igual a 100% para os dados de treino, o que reforça a capacidade de discriminação dos modelos SVM. Em geral, a boa performance obtida neste estudo, sugere que as metodologias analíticas são viáveis. A combinação das técnicas analíticas/quimiométricas permitiu discriminar, de maneira rápida e eficiente, as amostras de B10 e BX dos dois biodieseis metílicos na mistura com o diesel, oferecendo também a possibilidade de minimizar o uso de solventes e reagentes. Desta maneira, podem ser indicadas para o controle de qualidade destes combustíveis.

Palavras-chave: Biocombustíveis; Oleaginosas; FT-MIR; Controle de Qualidade; Quimiometria.

ABSTRACT

In recent decades, public policies have been developed to encourage the use of biofuels in Brazil, such as the establishment of a minimum mandatory content of biodiesel to commercialized diesel. In order to avoid adulterations in the composition of this blend established by law, a strict quality control is necessary throughout the entire production and distribution chain. Regarding this topic, many researches can be promoted, in order to develop methodologies that could give us quick and reliable answers for the quality control of this binary mix. In this work, analytic methods were developed to classify methyl biodiesel of macaúba and níger samples (both combined as a mixture) combined with diesel, by the combination of analytical/chemometric techniques of medium infrared spectroscopy (MIR), Discriminant Analysis by Partial Least Squares (PLS-DA), and by the use of support vector machines (SVM). For the construction of the discriminative models, a total of 74 combinations of BX diesel were produced, generated by the addition of methyl biodiesel of macaúba and níger to the pure diesel (B100), in a concentration range from 0,25% to 30,00% (v/v). The first group of samples was prepared using a proportion of 10% biofuel and 90% of pure diesel (B10), and the second according to different BX percentages (B3 to B9 and B11 to B30 containing, each of them, from 3 to 9% and from 11 to 30% of biodiesel). For each PLS-DA model were used 50 samples for the training sets and 24 samples for the testing. During the quality analysis of the models, the root mean squared errors of prediction were low, with percentages within the reproducibility of the standardized method. The performance criteria obtained for the developed models, such as sensitivity, specificity, precision and efficiency were of 100%. They also presented an excellent capacity to discriminate all the samples by its respective blends, with a prediction 100% accurate. The SVM algorithm also proved to be a great option to discriminate between the two classes of methyl biodiesel of macaúba B10 and BX. Regarding the performance analysis of the predictive models, the data base was divided in three subsets: training, test and validation, including 50/12/12 samples respectively. Different functions of the *Kernel*, such as the linear, polynomial, sigmoid and the radial base function (RBF) were tested, aiming to evaluate the generalization capacity of the classifier. Regarding the níger model, the RBF *Kernel* was the one that presented the best accuracy (0,92) in relation to the rest, with the hyperparameters defined as $C = 10$ and $\gamma = 10^{-4}$. For the macaúba models, a great precision was obtained for all the tested *Kernels*, with the hyperparameters defined as $C = 10$ and $\gamma = 10^{-5}$. The precision and the efficiency of classification of the SVM were higher than 70,0% in the test analysis and validation in both models. The results also revealed a value equal to 100% for the training data, which reinforces the discrimination capacity of the SVM models. In general, the good performance obtained in this study, suggests that the analytical methodologies are viable. The combination of analytical / chemometric techniques allowed to discriminate, quickly and efficiently, the samples of B10 and BX from the two methyl biodiesels in the mixture with diesel, also offering the possibility of minimizing the use of solvents and reagents. The good performance obtained in this study, indicates that this kind of methodology is viable, and could be used in the quality control of this fuel.

Keywords: Biofuels; Oilseeds; Oleaginous; FT-MIR; Quality control; Chemometrics.

LISTA DE FIGURAS

Figura 3.1 – Reação de transesterificação para a produção de Biodiesel.	22
Figura 3.2 – Principais características não conformes para o óleo diesel.....	30
Figura 3.3 – Matérias-primas utilizadas na produção de biodiesel (B 100) do ano de 2008 a 2020.....	33
Figura 3.4 – Produção Nacional de biodiesel em 2020.....	34
Figura 3.5 – Características morfológicas de <i>Acrocomia Aculeata</i> . (A) palmeira; (B) inflorescência; (C) frutos (D) estrutura e composição média do fruto da macaúba.	35
Figura 3.6 – Características morfológicas de <i>Guizotia abyssinica</i> Cass. (A) flor; (B) folhas, flores e pecíolo; (C) sementes.	38
Figura 4.1 – Espectrofotômetro com Transformada de Fourier, equipado com um acessório de Refletância Total Atenuada (ATR) de Cristal de ZnSe.	41
Figura 4.2 – Processo de reflexão no ATR horizontal.....	42
Figura 4.3 – Organização dos dados para a construção da matriz X.	42
Figura 4.4 – Gráfico <i>de leverage</i> versus Q Residual	43
Figura 4.5 – Representação da matriz de dados X e o vetor y usados no PLS-DA.....	45
Figura 4.6 – Representação gráfica do resultado do modelo PLS-DA.	46
Figura 4.7 – Esquema de indução de um classificador em aprendizado supervisionado.	50
Figura 4.8 – Representação gráfica de um conjunto treinamento com separação ótima, hiperplano e vetores de suporte (a). Margem de separação entre fronteiras das classes positivas e negativas (b).....	53
Figura 4.9 – Figura exemplificativa de inserção da variável ξ_i em um modelo de classificação binária não linearmente separável.....	56
Figura 4.10 – (a) Mapeamento de características pela função de <i>Kernel</i> no espaço original. (b) Projeção das amostras de entrada no espaço de características de alta dimensão. (c) Representação ilustrativa de uma função não linear de classificação.....	57
Figura 5.1 – Fluxograma da produção de biodiesel.....	61
Figura 5.2 – Exemplo gráfico de Curva ROC.	67
Figura 6.1 – Espectros MIR originais dos biodieseis metílicos em mistura com diesel: (a) Macaúba; (b) Níger, sem correção da linha de base e sem cortes.	69

Figura 6.2 – Espectros MIR corrigidos dos biodieseis metílicos em mistura com diesel: (a) Macaúba; (b) Níger.	70
Figura 6.3 – Gráfico dos pesos de VL1, VL2, VL3, VL4, VL5 e VL6 versus número de onda do modelo PLS-DA de: (a) Macaúba e (b) Níger.	73
Figura 6.4 – Gráfico de <i>leverage</i> versus <i>Q residuals</i> com 95% de confiança para os modelos PLS-DA dos biodieseis metílicos: (a) Macaúba e (b) Níger.	74
Figura 6.5 – Valores de classe estimados para conjuntos de treinamento (●) e de teste (◆) pelo modelo PLS-DA na discriminação de amostras de B10 e BX do biodiesel metílico: (a) Macaúba e (b) Níger.	75
Figura 6.6 – Relação entre número de onda (cm^{-1}) e absorvância entre as classes B10 e BX para o biodiesel metílico de (a) Níger e (b) Macaúba.	77
Figura 6.7 – Curvas ROC e AUC referentes ao modelo SVM de <i>kernel</i> RBF aplicado às amostras de (a) treino, (b) teste e (c) validação para o biodiesel metílico de níger.....	80
Figura 6.8 – Curvas ROC e AUC referentes ao modelo SVM de <i>kernel</i> Linear aplicado às amostras de (a) treino, (b) teste e (c) validação para o biodiesel metílico de macaúba.....	81

LISTA DE TABELAS

Tabela 5.1 – Número de amostras aplicadas na construção dos modelos PLS-DA.	62
Tabela 5.2 – Parâmetros de ajuste das funções <i>Kernel</i>	65
Tabela 5.3 – Parâmetros da matriz de confusão para análise SVM.	66
Tabela 6.1 – Bandas de absorções correspondentes aos modos vibracionais comuns aos biodieseis metílico de macaúba e níger em mistura com o diesel.	71
Tabela 6.2 – Parâmetros de Exatidão dos modelos PLS-DA.	72
Tabela 6.3 – Parâmetros de classificação dos modelos PLS-DA obtidos dos biodieseis metílico de Macaúba e Níger.	76
Tabela 6.4 – Resultados de performance de acordo com os hiperparâmetros γ e C para os dados do biodiesel metílico de níger.	77
Tabela 6.5 – Resultados de performance de acordo com os hiperparâmetros γ e C para os dados do biodiesel metílico de macaúba.	78
Tabela 6.6 – Treino dos modelos segundo diferentes <i>kernel</i> para os dados do biodiesel metílico de níger.	79
Tabela 6.7 – Treino dos modelos segundo diferentes <i>kernel</i> para os dados do biodiesel metílico de macaúba.	79
Tabela 6.8 – Resumo da matriz de confusão e principais medidas de acurácia do modelo.	82

LISTA DE ABREVIATURAS E SIGLAS

ABNT NBR	Norma Brasileira da Associação Brasileira de Normas Técnicas
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
AUC	Área sob a curva (do inglês, <i>area Under the ROC Curve</i>)
ANVISA	Agência Nacional de Vigilância Sanitária
ASTM	Sociedade Americana de Testes e Materiais (do inglês, <i>American Society of Testing and Materials</i>)
ATR	Refletância Total Atenuada (do inglês, <i>Attenuated Total Reflectance</i>)
B10	Mistura Biodiesel/Diesel com 10% de Biodiesel
B100	Biodiesel Puro
BX	Mistura Biodiesel/Diesel com X% de Biodiesel
CCM	Coefficiente de Correlação de Matthew's
CEN	Comitê Europeu de Normalização (em inglês, <i>Comité Européen de Normalisation</i>)
CNPE	Conselho Nacional de Política Energética
CV	Validação Cruzada (do inglês, <i>Cross Validation</i>)
EPE	Empresa de Pesquisa Energética
FT-MIR	Infravermelho Médio com Transformada de Fourier (do inglês, <i>Fourier Transform Midle Infrared</i>)
HATR	Refletância Total Atenuada Horizontal
IEA	Agência Internacional de Energia (do inglês, <i>International Energy Agency</i>)
ISO	Organização Internacional de Regulamentação (do inglês, <i>International Organization for Standardization</i>)
LIBSVM	Biblioteca para Máquinas de Vetores de Suporte (do inglês, <i>Library for Support Vector Machines</i>)
OSH	Hiperplano de separação ótima (do inglês, <i>Optimal Separating Hiperplane</i>)
PLS	Quadrados Mínimos Parciais (do inglês, <i>Partial Least Squares</i>)
PLS-DA	Análise Discriminante por Quadrados Mínimos Parciais (do inglês, <i>Partial Least Squares – Discriminant Analysis</i>)

PMQC	Programa de Monitoramento da Qualidade dos Combustíveis
RFA	Associação de Biocombustíveis Renováveis (do inglês, <i>Renewable Fuels Association</i>)
RMSEC	Erro Quadrático Médio de Calibração (do inglês, <i>Root Mean Square Error of Calibration</i>)
RMSECV	Erro Quadrático Médio de Validação Cruzada (do inglês, <i>Root Mean Square Error of Cross-Validation</i>)
RMSEP	Erro Quadrático Médio de Previsão (do inglês, <i>Root Mean Square Error of Prediction</i>)
ROC	<i>Receiver Operating Characteristic</i>
SOx	Óxidos de enxofre
SRM	Risco de minimização estrutural (do inglês, <i>Structural Risk Minimization</i>)
SVM	Máquinas de Vetores de Suporte (do inglês, <i>Support Vector Machine</i>)
USDA	Departamento de Agricultura dos Estados Unidos- DAEU (do inglês, <i>United States Department of Agriculture- USDA</i>)
VL	Variável Latente

SUMÁRIO

RESUMO	8
ABSTRACT	9
1 INTRODUÇÃO.....	17
1.1 Estrutura da Tese.....	20
2 OBJETIVOS	21
2.1 Objetivo Geral.....	21
2.2 Objetivos Específicos.....	21
3 REFERENCIAL TEÓRICO	22
3.1 Produção de Biodiesel.....	22
3.2 Importância da produção de biodiesel.....	27
3.3 Mistura Biodiesel/Diesel: Normas e controle de qualidade	28
3.4 Óleos vegetais	32
3.4.1 Macaúba.....	34
3.4.2 Níger.....	37
4 MÉTODOS QUIMIOMÉTRICOS	39
4.1 Quimiometria.....	39
4.2 Instrumentação analítica.....	40
4.2.1 Espectroscopia MIR aliada a Refletância Total Atenuada (ATR).....	40
4.2.2 Organização e pré-processamento dos dados espectrais	42
4.3 Métodos analíticos	44
4.3.1 Análise Discriminante por Mínimos Quadrados Parciais	44
4.3.2 Avaliação de desempenho dos modelos PLS-DA.....	47
4.3.3 Conceitos sobre aprendizado de máquina e fundamentação em Máquina de Vetores de Suporte	49
4.3.4 Fundamentação em Máquina de Vetores de Suporte.....	51

5 PROCEDIMENTO EXPERIMENTAL.....	60
5.1 Produção do Biodiesel.....	60
5.2 Preparo das amostras.....	61
5.3 Obtenção dos Espectros MIR	62
5.4 Análise Quimiométrica.....	62
5.4.1 Modelos PLS-DA.....	62
5.4.2 Modelos SVM.....	64
6 RESULTADOS E DISCUSSÃO.....	68
6.1 Caracterização dos Espectros MIR	68
6.2 Modelos PLS-DA.....	71
6.3 Modelos SVM.....	76
7 CONCLUSÕES.....	83
7.1 Conclusões gerais.....	83
7.2 Recomendações e perspectivas futuras	84
REFERÊNCIAS.....	85

1 INTRODUÇÃO

Nas últimas décadas, foram desenvolvidas políticas públicas de incentivo ao uso de biocombustíveis no Brasil, com o desenvolvimento dos programas de incentivo à produção de novas matrizes energéticas, incluindo as de origem na biomassa vegetal, entre outras opções (gordura animal), que poderiam substituir total ou parcialmente os derivados petroquímicos (AZEVEDO; LIMA, 2016).

Os biocombustíveis são uma das formas quando se pensa em reduzir a queima de combustíveis fósseis. A forma mais comum de obtenção deste biocombustível é por meio da reação óleos vegetais com metanol ou etanol, na presença de um catalisador, em processo químico ou biológico conhecido como transesterificação. Os produtos desta reação são as misturas de ésteres etílicos ou metílicos de ácidos graxos, que compõem o biodiesel, e glicerina, cujo maior constituinte é o glicerol (GUARIEIRO *et al.*, 2008; RUSCHEL *et al.*, 2014).

A sua cadeia produtiva, apresenta diversas vantagens em relação ao uso do diesel, como no caso, são menos tóxicos pela ausência de compostos sulfurados e aromáticos voláteis e sua combustão gera menores teores de gases poluentes que o óleo diesel, pois retiram CO₂ da atmosfera para crescimento da planta e a quantidade de CO₂ liberada na combustão desses óleos é menor que a gasta na sua produção (HUAN; ZHOU; LIN, 2012). Além disso, a reação de transesterificação produz o biodiesel com estruturas moleculares mais simples que o seu triglicerídeo precursor, diminuindo a sua viscosidade comparativamente ao óleo ou gordura, resultando numa queima eficiente que reduz significativamente a deposição de resíduos no motor (LÔBO; FERREIRA; CRUZ, 2009).

A maior parte do biodiesel produzido e comercializado no Brasil utiliza como matéria-prima o óleo de soja refinado por se tratar de um processo bem estabelecido, mas em termos estratégicos, a dependência desta oleaginosa, pode influenciar tanto o valor final do combustível quanto na sua utilização como alimento, causando um desequilíbrio comercial nestes setores (ANP, 2018). Logo, outras pesquisas, podem ser realizadas no sentido de viabilizar o uso de outras fontes oleaginosas, preferencialmente espécies que apresentam maior potencial de

aproveitamento e consideráveis rendimentos nos teores de óleo (ARANSIOLA *et al.* 2014; HOCHSCHEIDT *et al.*, 2020).

Na literatura referenciam como exemplo a macaúba (*Acrocomia aculeata*), uma palmeira que produz cerca de 25 toneladas frutos ha⁻¹·ano⁻¹ com o teor de 50 a 70% de óleo (COPPEL *et al.*, 2018), e cujos coprodutos podem ser aplicados na produção de cosméticos, tortas forrageiras, além de combustível de alto poder calorífico (CICONINNI *et al.*, 2013; COLOMBO *et al.*, 2018).

Outra oleaginosa, ainda não explorada na literatura para a produção de biodiesel é a níger (*Guizotia abyssinica*). Essa espécie, de origem no continente africano, apresenta potencial energético promissor. Algumas delas, como o crambe e a mafurra, já estão em fase de estudos e adaptação no Brasil (AMBROSANO, 2012; FRANCHINI, 2014).

O biodiesel frente ao seu valor de mercado e elevado consumo, pode ser alvo de adulterações durante sua produção ou transporte, sendo necessário um rigoroso controle de qualidade ao longo de toda a cadeia produtiva e de distribuição. No Brasil, essa fiscalização é feita pela Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP, 2016). A ANP estabelece o teor mínimo obrigatório da adição de biodiesel ao diesel comercializado no país. Este teor vem sendo incrementado gradativamente com objetivo de atingir a proporção de 15% (v/v) biodiesel/diesel até maio de 2023, segundo o cronograma do Conselho Nacional de Política Energética (ANP, 2020). Nesta perspectiva, é relevante o desenvolvimento de metodologias que possam fornecer respostas rápidas e eficientes para o controle qualitativo dessa mistura binária.

A Espectroscopia de Infravermelho Médio (MIR) com Reflectância Total Atenuada (ATR) associada a métodos quimiométricos pode, de maneira eficiente, atender tal demanda, com a vantagem da amostragem ser direta, não destrutiva e usar pequeno volume de amostras, minimizando a geração de resíduos químicos (SUAREZ, 2014).

Em geral, a combinação das técnicas analíticas/quimiométricas, permite gerar excelentes resultados de classificação dos óleos vegetais empregando o perfil espectral como impressão digital da amostra (SANTANA *et al.*, 2020). A aquisição direta de medidas de reflectância detectam diversos grupos funcionais presentes nos materiais, contudo os espectros complexos requerem o estabelecimento de modelos quimiométricos para serem relacionados com as propriedades de interesse, seja ela qualitativa ou quantitativa (GUARIEIRO *et al.*, 2008).

Algumas pesquisas, nesta temática, têm sido realizadas para a análise de muitas matrizes de dados complexas, envolvendo diferentes concentrações da mistura biodiesel/diesel (FILGUEIRAS *et al.*, 2016; RUSCHEL, 2017; MAHMUDI, MOSTAFAEI; MIRZAEI-GHALEH, 2019; SITOIE *et al.*, 2020; TOSATO *et al.*, 2020), embora, poucos estudos sejam aplicados para a análise de dados qualitativos. Da mesma forma que o aspecto quantitativo é indispensável, uma análise qualitativa pode a princípio fornecer uma resposta mais rápida, para em seguida, se preciso, realizar a análise quantitativa, podendo ser um recurso estratégico para o controle de qualidade da mistura BX pelos órgãos de fiscalização.

Diante do exposto, buscou-se desenvolver e validar modelos analíticos para classificar amostras de biodiesel metílico de macaúba e níger quanto à mistura biodiesel/diesel, usando a combinação das técnicas analíticas/quimiométricas de espectroscopia MIR e dos métodos preditivos de Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e de Máquinas de Vetores de Suporte (SVM).

1.1 Estrutura da Tese

Esta tese encontra-se dividida em cinco capítulos.

O capítulo 1 apresenta uma revisão bibliográfica sobre os principais tópicos abordando o biodiesel: a produção, características, o monitoramento e o controle de qualidade da mistura binária biodiesel/diesel; incluindo as biomassas mais utilizadas; além das considerações relativas às espécies macaúba e níger utilizadas para a produção de biodiesel neste trabalho.

O capítulo 2 traz uma explanação sobre a Quimiometria, como de sua importância na análise de dados químicos multivariados, além das vantagens da técnica de Espectroscopia MIR, incluindo a fundamentação teórica quanto aos métodos discriminativos de Análise Discriminante por Quadrados Mínimos Parciais e de Máquinas de Vetores de Suporte.

O capítulo 3 apresenta os materiais e métodos utilizados para o desenvolvimento da pesquisa. Foram descritas a produção de biodiesel, a obtenção dos espectros MIR, a organização e divisão dos dados; as métricas de avaliação utilizadas e a otimização e o desenvolvimento dos modelos preditivos.

O capítulo 4 contém a caracterização dos espectros MIR dos biodieseis metílicos de macaúba e níger, além dos resultados e discussão dos modelos desenvolvidos. Parte dos resultados compõem o artigo intitulado “Classificação do teor de biodiesel metílico de macaúba na mistura com diesel, por meio da combinação das técnicas analíticas/quimiométricas de espectroscopia MIR e PLS-DA” que foi submetido à Revista Virtual de Química em outubro de 2020.

Finalmente, no capítulo 5, são apresentadas as conclusões e recomendações de propostas para futuras pesquisas a partir da experiência obtida no desenvolvimento da presente pesquisa.

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver e validar métodos analíticos para classificar amostras de biodiesel metílico de óleos de níger e macaúba quando misturados com diesel usando a combinação das técnicas analíticas/quimiométricas de Espectrometria no Infravermelho Médio e os métodos discriminativos PLS-DA e SVM.

2.2 Objetivos Específicos

- Construir modelos de PLS-DA e SVM para classificar amostras de misturas biodiesel/diesel dentro da conformidade em relação ao teor de biodiesel.
- Identificar as variáveis mais importantes e sua contribuição na construção dos modelos PLS-DA e SVM.
- Investigar a contribuição da seleção dos parâmetros *kernel* na acurácia produzida pelo processo de classificação em SVM.
- Verificar os tipos de função *Kernel* de SVM com maior capacidade de generalização do modelo classificador.
- Validar os modelos PLS-DA e SVM construídos tendo como base os parâmetros do método.

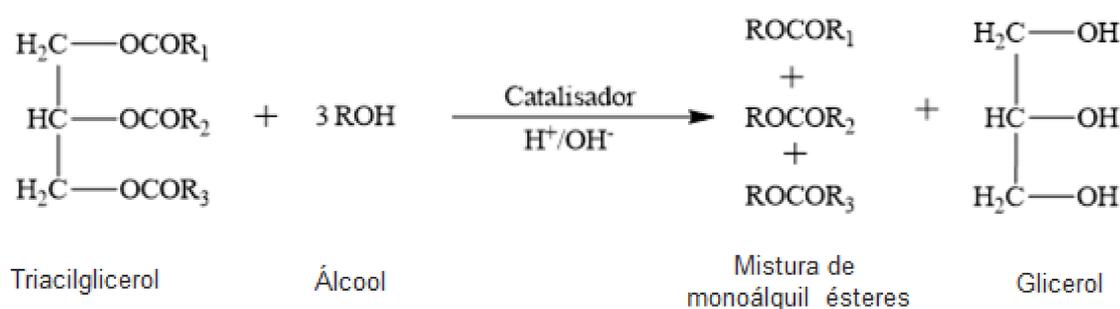
3 REFERENCIAL TEÓRICO

3.1 Produção de Biodiesel

O biodiesel é definido como uma mistura de ésteres monoalquílicos de ácidos carboxílicos de cadeia longa, derivado de fontes renováveis, como óleos vegetais e gordura animal, que pode ser obtido a partir do processo de transesterificação e/ou esterificação, entre outros processos, obedecendo a uma dada especificação (ANP, 2016). Das rotas tecnológicas empregadas para a produção de biodiesel, a transesterificação é amplamente utilizada, sobretudo em óleos de baixa acidez (MARX, 2016).

Como mostra na Figura 3.1, essa rota consiste numa reação química em que as moléculas de triacilgliceróis (matéria prima) reagem com o álcool, de cadeia curta, metanol ou etanol, na presença de um catalisador ácido, básico ou enzimático, produzindo aqui ésteres (biodiesel) e glicerol (INTARAPONG *et al.*, 2013).

Figura 3.1 – Reação de transesterificação de um triacilglicerídeo



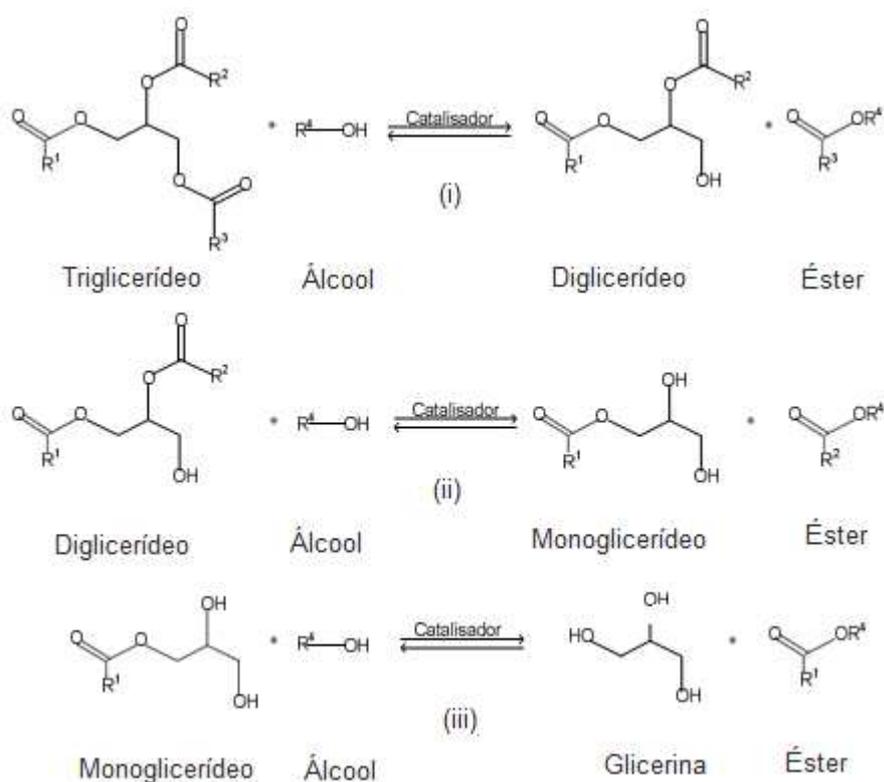
Fonte: MARCHETTI *et al.*, 2007.

Este processo tem por objetivo modificar a estrutura molecular de óleos vegetais tornando-o com propriedades mais adequadas ao seu uso como combustível (CONCEIÇÃO *et al.*, 2005). A inserção direta de óleos vegetais nos motores causa entupimento dos bicos devido a sua alta viscosidade, quando comparado com o diesel fóssil, e pode provocar a combustão incompleta nos motores por causa de algumas características físico-químicas (elevada massa

molecular, alto ponto de ignição, e seu baixo poder calorífico), e o caráter poli-insaturado, o que justifica a necessidade da transesterificação (DEMIRBAS, 2008).

O mecanismo geral é composto por 3 reações consecutivas e reversíveis, nas quais são formados diglicerídeos e monoglicerídeos como intermediários na equação. Sua estequiometria requer 1 mol de triglicerídeo para 3 moles de álcool, mas devido ao fato desta reação ser reversível, um excesso de álcool se faz necessário para aumentar o rendimento de produção dos ésteres alquílicos e permitir a sua fácil separação da glicerina formada (SUAREZ *et al.*, 2007). A Figura 3.2 indica as três reações consecutivas e reversíveis, onde há a formação de diglicerídeos e éster (i) através do ataque do álcool ao triglicerídeo, em seguida ocorre a formação de monoglicerídeos e éster (ii) como produtos intermediários e, por fim, a formação dos produtos éster e glicerina (iii).

Figura 3.2 – Mecanismo da reação de transesterificação, onde R^1 , R^2 e R^3 são as cadeias do ácido graxo, e R^4 é o grupo alquil do álcool.



Fonte: ALVES, 2012.

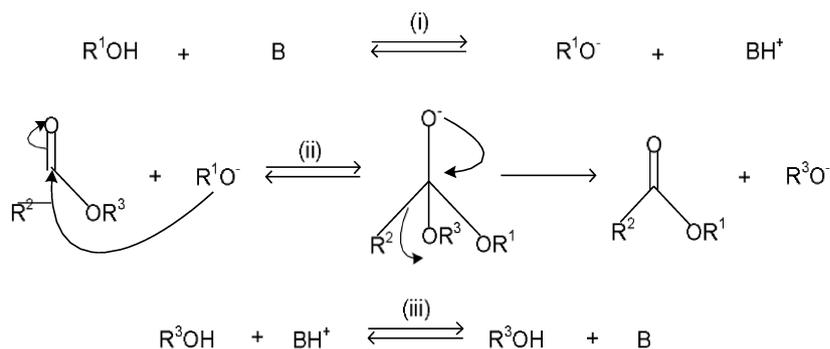
Geralmente, a transesterificação alcalina e a rota metílica, denominada metanólise, são os processos químicos mais aplicados à produção de biodiesel. Em virtude da reversibilidade das etapas de reação, o álcool deve ser adicionado em excesso em relação à quantidade de triacilgliceróis, a fim de se alcançar a máxima conversão possível. Além disso, embora a reação se processe pela simples mistura entre os reagentes, o uso de um catalisador se faz necessário para maximizar a conversão (JOTHIRAMALINGAM; WANG, 2009; TUBINO; ROCHA JUNIOR; BAUERFELDT, 2014).

O álcool mais comum neste processo é o metanol, haja visto que: (i) permite a separação espontânea do glicerol; (ii) o processo empregado é mais simples; (iii) de alto rendimento (empregando tanto metóxidos, quanto hidróxidos de sódio ou potássio), (iv) de baixo custo e menor quantidade de água comparado ao etanol. No entanto, o etanol é menos tóxico, renovável, e produz um biodiesel com maior número de cetano e lubricidade (LÔBO; FERREIRA; CRUZ, 2009). Mas, ainda precisam ser superadas algumas limitações para se alcançar um processo viável pela rota etílica: é ligeiramente menos reativo do que o metanol e requer mais etapas para a produção de biodiesel, o que representa um custo maior para o processo (RICHARD *et al.*, 2013). Desta forma, a opção de rota a ser utilizada na produção de biodiesel, seja ela, metílica ou etílica, irá depender da matéria-prima empregada, dos preços dos produtos utilizados na produção, do tempo de duração e do consumo de energia (MARCHETTI *et al.*, 2007, SINGH *et al.*, 2006).

Os ácidos e bases de Brønsted largamente empregados como catalisadores na reação de transesterificação de óleos vegetais e gorduras animais são: ácidos sulfúrico, fosfórico e clorídrico; e os hidróxidos e alcóxidos de sódio ou potássio (MARCHETTI *et al.*, 2007; SINGH *et al.*, 2006). As transesterificações catalisadas por bases ocorrem mais rapidamente que as transesterificações catalisadas por ácidos. Estas, além de serem as mais comuns comercialmente, ocorrem sob condições mais brandas e apresentam um alto rendimento (na maioria das vezes superior a 90%). Por esse motivo, junto ao fato de que os catalisadores alcalinos são menos corrosivos que os compostos ácidos, os processos industriais em geral preferem catalisadores básicos. Contudo, se o óleo possuir um alto teor de ácidos graxos livres ou uma grande quantidade de água presente, o emprego de catalisadores ácidos é desejável (KNOTHE *et al.*, 2005; MENDONÇA, 2008).

A catálise básica requer condições anidras mais rígidas comparativamente a ácida, já que a presença de água conduz à hidrólise irreversível dos lipídeos, podendo formar emulsão ou sabão (saponificação) se a concentração de catalisador for maior do que a necessária (LU *et al.*, 2016). Na Figura 3.3 é indicado o mecanismo para a reação de transesterificação de triglicerídeos com álcoois utilizando um catalisador básico.

Figura 3.3 - Mecanismo para a reação de transesterificação alcalina.



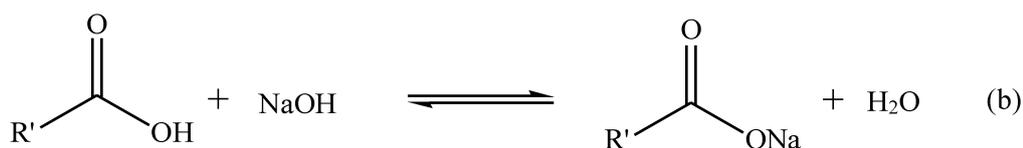
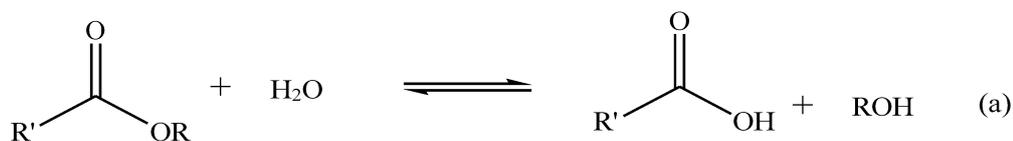
R^1 = grupo alquil do álcool,

R^2 e R^3 = cadeia carbônica do ácido graxo

Fonte: Adaptado de Suarez *et al.*, 2007.

Na reação alcalina, primeiramente, a base reage com o álcool formando um alcóxido (i). Em seguida, uma carbonila do triglicerídeo sofre um ataque nucleofílico do alcóxido, conforme indicado na reação (ii), formando um intermediário tetraédrico onde são formados os diglicerídeos e o éster. Este composto intermediário desprotona o catalisador e regenera o álcool do diglicerídeo, que se torna apto a reagir novamente, agora com a segunda molécula de álcool. Os diglicerídeos e os monoglicerídeos são formados através deste mesmo método formando na reação final glicerina e éster (ALVES, 2012; SUAREZ *et al.*, 2007). Por outro lado, a presença de água dá origem à hidrólise de alguns ésteres produzidos pelo ataque do íon hidróxido (Figura 3.4a), com a consequente formação de sabão (Figura 3.4b). Os sabões formados durante a reação são removidos com o glicerol ou são eliminados durante a etapa de lavagem aquosa. No entanto, essa reação de saponificação é indesejável, pois reduz o rendimento da reação, dificultando consideravelmente a purificação dos ésteres e do glicerol devido à formação de emulsões (MARCHETTI *et al.*, 2007; SOLOMONS; FRYHLE, 2002).

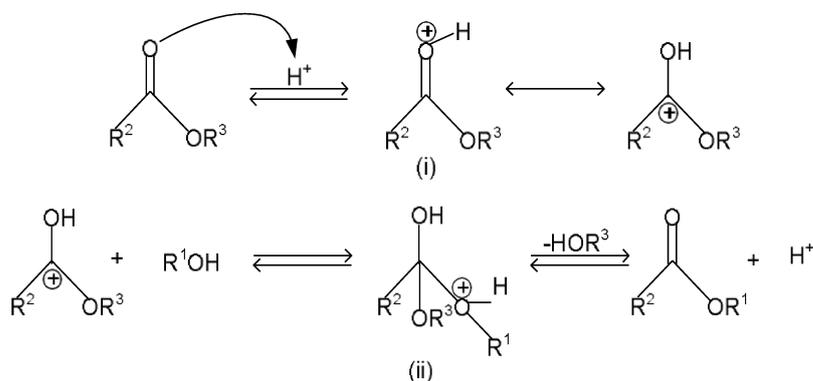
Figura 3.4 - Reação de hidrólise de ésteres de ácidos graxos (a); Reação de saponificação de ácidos graxos (b).



Fonte: SCHUCHARDT *et al.*, 2004.

Tendo-se a catálise ácida como alternativa, seu uso inclui a combinação de reações reversíveis. A sua elevada conversão é devida à capacidade de reagir tanto com ácidos graxos quanto com sais de ácidos graxos presentes no sistema (MENEGHETTI; MENEGHETTI; BRITO, 2013). O ácido sulfúrico é o mais empregado, embora apresente uma cinética muito lenta de reação comparada ao uso de catalisadores básicos. Outra desvantagem do uso desse catalisador está associada à necessidade de sua remoção após a transesterificação, visando prevenir possíveis danos às partes integrantes dos motores (CANAKSI; GERPEN, 2001). Porém a sua utilização se faz necessária quando altos teores de ácidos graxos livres estão presentes nos glicerídeos.

Figura 3.5 - Mecanismo da reação de transesterificação catalisada por ácido



Na Figura 3.5 indica-se o mecanismo para a reação de transesterificação de triglicerídeos com álcoois de cadeia curta em meio ácido. Uma carbonila, do triglicerídeo, sofre um ataque eletrofílico do H^+ , indicado na reação (i), formando um carbocátion. Em seguida, este carbocátion sofre um ataque nucleofílico do álcool, formando um intermediário tetraédrico (diglicerídeos e éster) indicado na reação (ii). Este composto intermediário regenera o álcool

do diglicerídeo e está apto a reagir novamente, agora com a segunda molécula de álcool. Os diglicerídeos e os moniglicerídeos são formados através deste mesmo método formando na reação final glicerina e éster (SUAREZ *et al.*, 2007; MILINSKI, 2007).

3.2 Importância da produção de biodiesel

O biodiesel foi inserido na matriz energética brasileira por meio do Programa Nacional de Produção e Uso de Biodiesel (PNPB), um programa interministerial do Governo Federal que tem como diretrizes: implantar um programa sustentável, promovendo inclusão social; garantir preços competitivos, qualidade e suprimento; e produzir o biodiesel a partir de diferentes fontes oleaginosas e em regiões diversas (EMPRESA DE PESQUISA ENERGÉTICA- EPE, 2019).

Desde o início do Programa em 2005, já foram produzidos, até dezembro de 2019, mais de 40,6 bilhões de litros deste biocombustível. Comparativamente, o Brasil continua entre os três maiores produtores e consumidores de biodiesel no ranking internacional, junto a EUA e Indonésia, segundo o Departamento de Agricultura dos Estados Unidos (DAEU; em inglês: *United States Department of Agriculture- USDA*) (USDA, 2019). Os dois principais países, Brasil e Estados Unidos, continuam com alta participação neste mercado, com 85% da produção e comercialização (RENEWABLE FUELS ASSOCIATION- RFA, 2020).

Acompanhando essa tendência, a possibilidade de inserir o biodiesel na matriz energética brasileira permitiu atrair grandes investimentos na construção de usinas de produção de biodiesel no País. Em dezembro de 2019 a capacidade instalada correspondeu a 9,3 bilhões de litros, dividida entre 51 usinas produtoras autorizadas (AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCMBUSTÍVEIS-ANP, 2020). Pioneiro mundial no uso de biocombustíveis, o Brasil alcançou uma posição almejada por muitos países que buscam desenvolver fontes renováveis de energia como alternativas estratégicas ao petróleo (EMPRESA DE PESQUISA ENERGÉTICA- EPE, 2017). Além disso, o biodiesel também apresenta vantagens consideradas relevantes, sob ponto de vista ambiental quando comparado aos combustíveis convencionais: São provenientes de fontes renováveis, biodegradáveis, com menor emissão de compostos sulfurados (SOx) e aromáticos voláteis (benzeno, tolueno, xilenos) (CORDEIRO *et al.*, 2011). Também possui um elevado teor de oxigênio, o que melhora a lubrificidade e a combustão do óleo, e ainda promove a redução de emissões gasosas

(monóxido de carbono (CO), dióxido de nitrogênio (NO₂) e ozônio (O₃), assim como de materiais particulados (JOTHIRAMALINGAM; WANG, 2009; HUANG; ZHOU; LIN, 2012). Outro aspecto importante é que a produção de energias renováveis gera oportunidades de emprego e renda para a população rural na agricultura familiar; diminui a importação de combustíveis fósseis; diversifica a matriz energética por meio do uso de novas tecnologias de produção (GEISSDOERFER; VLADIMIROVA; EVANS, 2018; JAIN, 2019).

3.3 Mistura Biodiesel/Diesel: Normas e controle de qualidade

O biodiesel substitui total ou parcialmente o óleo diesel de petróleo em motores ciclodiesel automotivos (de caminhões, tratores, camionetas, automóveis, etc) ou estacionários (geradores de eletricidade, calor, etc). Pode ser usado puro ou misturado ao óleo diesel em diversas proporções (RIBEIRO, 2013). As misturas binárias de biodiesel e óleo diesel são designadas pela abreviação BX, onde X é a percentagem de biodiesel adicionada à mistura. Como exemplo, o combustível B5 possui 95% de diesel e 5% de biodiesel. Já o B100 é um combustível preparado a partir do biodiesel puro (AZEVEDO; LIMA, 2016).

O óleo diesel é o principal produto do fracionamento do petróleo no Brasil. Esse combustível é constituído por hidrocarbonetos parafínicos, olefínicos e aromáticos com cadeia carbônica de 6 a 30 átomos. Alguns compostos apresentam também enxofre, oxigênio e nitrogênio. Enquanto o biodiesel é constituído por ésteres alquílicos obtidos, principalmente, a partir da transesterificação de óleos vegetais ou gordura animal, provenientes de algum álcool de cadeia curta, como o metanol, e de ácidos de cadeia longa, entre 12 e 22 átomos de carbono (CANCELA; CESCÓN, 2020).

O incremento gradual na mistura segue a Resolução do Conselho Nacional de Política Energética, CNPE nº 16, de 29 de outubro de 2018 (CNPE, 2018), que visa o aumento do percentual de biodiesel na mistura de 1% ao ano, a partir de 1º de junho de 2019, até atingir a proporção de 15% (v/v) em 2023. Assim, desde que sejam obedecidas as condicionantes de aprovação de testes nos motores para esse teor (EMPRESA DE PESQUISA ENERGÉTICA-EPE, 2019).

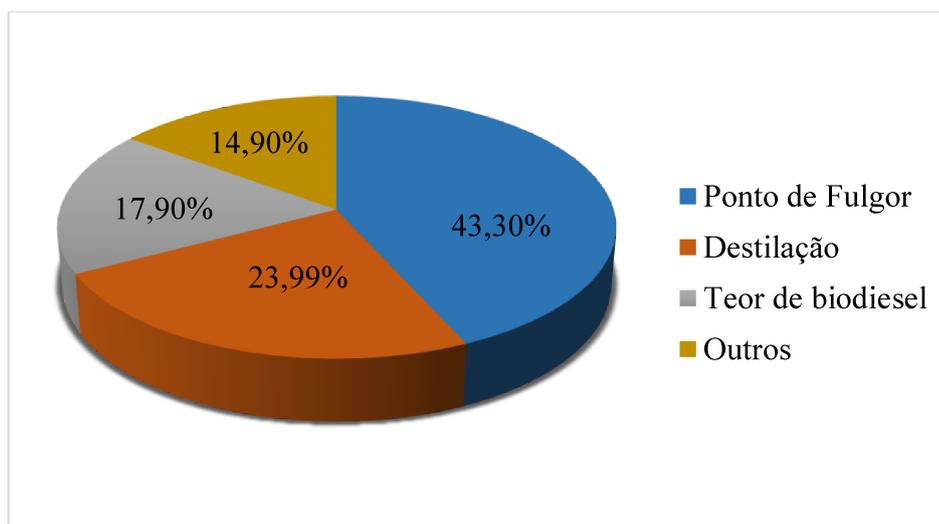
Durante boa parte do tempo, teve-se o uso da mistura B10, com a aprovação em 1º de março de 2018, de acordo com a Resolução nº 23 do CNPE, de 09 de novembro de 2017, justificando neste trabalho, a premissa da adição dessa mistura binária, conforme previsto em lei. Após isso já foram aprovados o uso da mistura B11 e B12. Entretanto, como medida salutar, feita pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), foi autorizada a redução excepcional e temporária, dos atuais 12% para 11% no bimestre de novembro e dezembro de 2020, com o objetivo de garantir o abastecimento interno de diesel B no país, em razão da pequena oferta no período de Pandemia (AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS-ANP, 2020).

A mistura óleo diesel/biodiesel tem sua especificação estabelecida pela ANP (Resolução ANP 15/2006). O Programa de Monitoramento da Qualidade dos Combustíveis (PMQC) é o instrumento utilizado pela Agência para verificar os indicadores gerais da qualidade dos combustíveis comercializados no País. Por meio deste programa identificam-se focos de não conformidade, visando eliminar assimetrias de informações, além de orientar e aperfeiçoar a atuação da área de fiscalização de sua competência. O PMQC realiza as coletas das amostras de combustíveis, notadamente, em postos revendedores. O plano de coleta baseia-se em amostragem estratificada simples, a partir do estabelecimento de subpopulações (subárea) de postos em cada estado, por meio de sorteio, a cada mês. Essas amostras, são analisadas no Centro de Pesquisas e Análises Tecnológicas da ANP, assim como nos laboratórios de universidades e instituições de pesquisa contratados para esta finalidade. As especificações dos padrões de qualidade do biodiesel seguem de acordo à aplicação das normas: a) Associação Brasileira de Normas Técnicas (ABNT), b) *American Society for Testing and Materials* (ASTM) e c) *Comité Européen de Normalisation* (CEN) ou *International Organization for Standardization* (ISO).

Na

Figura 3.2 é apresentado os resultados consolidados das análises da qualidade do diesel comercializado no País. Das 6.728 amostras avaliadas em setembro/2020, foram observados 67 resultados de ensaios não conformes. Entre eles, destacaram-se ponto de fulgor, destilação, teor de biodiesel e outros, que corresponderam, respectivamente, a 43,30%, 23,99%, 17,90% e 14,90%, do total de não conformidades. Na categoria “outros”, foram englobadas as não conformidades de cor, aspecto, teor de água e enxofre.

Figura 3.2 – Principais características não conformes para o óleo diesel.



Fonte: ANP - Mês de Referência: setembro/2020.

Apesar desse cenário, os resultados do monitoramento da ANP em relação à média nacional, no mesmo mês/ano, esternalizou índices de conformidade a partir de 96%, indicando uma tendência de equilíbrio das conformidades deste combustível em percentuais elevados (AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCMBUSTÍVEIS-ANP, 2020).

A técnica adotada pela ANP para determinar o teor de biodiesel em misturas BX é a espectroscopia na região do infravermelho médio (MIR), via norma ABNT NBR 15568 (ABNT, 2008). Os procedimentos químicos/analíticos são de determinação quantitativa, que consiste em modelos de calibração, com base na regressão por Mínimos Quadrados Parciais (PLS). O método da ABNT NBR estabelece o desenvolvimento de modelos para faixas analíticas de 0-8% (v/v) e 8-30% (v/v) do conteúdo de biodiesel no óleo diesel e esses modelos devem ter raiz quadrática média dos erros de predição (RMSEP) que não pode ser superior a 0,1% (v/v) e 1% (v/v), respectivamente (ALVES; POPPI, 2013).

A respeito da norma, ainda existem críticas e discussões sobre a total eficiência de sua aplicação, quando implementadas em análises de rotina. Devido à dificuldade em obter a precisão exigida, além da necessidade de usar uma grande quantidade de amostras, equipamentos avançados e reagentes específicos (iso-octano, por exemplo), que tornam o processo demorado e dispendioso (SUAREZ, 2014; SANTANA *et al.*, 2020). Diante disso, faz-se necessário o desenvolvimento de metodologias/técnicas capazes de fornecer respostas confiáveis em reduzido intervalo de tempo e com menores custos envolvidos; e que isso passe

a ser estratégico no controle de qualidade da mistura BX a ser disponibilizado diretamente nas saídas de distribuidoras, facilitando a verificação de não conformidades dos combustíveis (SOUZA, 2015; RIBEIRO, 2017).

Da mesma forma que o aspecto quantitativo é indispensável, uma análise qualitativa pode a princípio fornecer uma resposta mais rápida, para em seguida, se preciso, realizar a análise quantitativa. A Espectroscopia de Infravermelho Médio (MIR) com Reflectância Total Atenuada (ATR), aliada a métodos quimiométricos, pode, de maneira eficiente, atender tal demanda, com a vantagem da amostragem ser direta, não destrutiva e usar pequeno volume de amostras, minimizando a geração de resíduos químicos. No contexto geral, a combinação de técnicas analíticas/quimiométricas mostra excelentes resultados de classificação dos óleos vegetais, podendo ser utilizadas para a análise de muitas matrizes de dados complexas, envolvendo diferentes concentrações da mistura biodiesel/ diesel (FILGUEIRAS *et al.*, 2016).

Guariero *et al.* (2008) enfatizaram que a espectroscopia no infravermelho é uma ferramenta analítica que pode ser utilizada para quantificar misturas biodiesel/diesel, principalmente pela existência de bandas distintas no espectro no infravermelho do diesel mineral e do biodiesel permitindo assim, que esse tipo de análise seja aplicado para quantificar o percentual de biodiesel misturado ao diesel.

Mueller *et al.* (2013) empregaram a espectroscopia MIR com o sensor de reflectância total atenuada (ATR), juntamente com análise estatística multivariada para classificar amostras de biodiesel, produzidos a partir de seis variedades distintas de óleos vegetais. A análise de componente principal (PCA), análise de agrupamento hierárquico (HCA), e modelagem independente de analogias de classe (SIMCA) foram aplicadas. Os resultados indicaram que é possível desenvolver metodologias confiáveis e rápidas pelo uso da ATR e também por aplicação dos métodos de análise multivariada aos dados de infravermelho.

Ruschel (2017) analisou conjuntos de amostras de gasolina e etanol, adulteradas e não-adulteradas, que foram caracterizados através da espectroscopia no infravermelho com transformada de Fourier com acessório de reflectância total atenuada (FTIR-ATR), para posterior aplicação de ferramentas quimiométricas aos dados espectroscópicos gerados. Os dados foram aplicados para os modelos discriminativos PLS-DA e SVM. No geral, com base nas técnicas utilizadas, todos os resultados mostraram que tais modelos são ótimas alternativas para distinguir as amostras adulteradas das não-adulteradas para os combustíveis estudados.

Ambos modelos de classificação PLS-DA e SVM-DA apresentaram sensibilidade e especificidade máximas.

Gelinski *et al.* (2018), desenvolveram modelos para monitoramento da reação transesterificação catalisada por base, usando também as metodologias empregadas. Os modelos foram criados usando regressão por Mínimos Quadrados Parciais (*Partial Least Squares*, PLS), feita com amostras sintéticas simulando a composição da reação para diferentes razões molares etanol/óleo, obtida da soja. Os resultados demonstraram técnicas rápidas e precisas de quantificação, permitindo controle efetivo da reação, como também reduzindo tempo e custos de processo.

3.4 Óleos vegetais

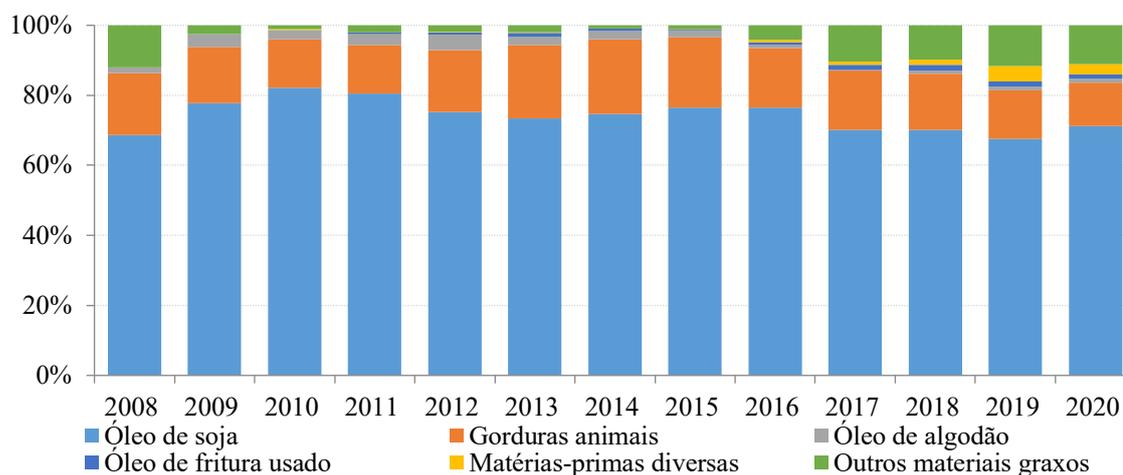
As projeções mundiais previstas para 2020 pela *International Energy Agency* (IEA)- Agência Internacional de Energia, assinalam crescente substituição das fontes de combustível de origem fóssil pelas fontes renováveis de origem em biomassa, dentre elas as derivadas da cana-de açúcar e do milho, para a produção de etanol, e as provenientes dos óleos vegetais de canola, de soja, de algodão, entre outras, para a produção de biodiesel (INTERNATIONAL ENERGY AGENCY - IEA, 2020).

O aumento da demanda dos mercados de alimentos e oleoquímicos, tende a criar oportunidades para diversificar a matriz fornecedora de óleos vegetais (LEVERMANN, 2014). Além disso, a grande extensão territorial coloca o Brasil em uma posição vantajosa no que diz respeito à diversidade de oleaginosas a serem utilizadas como matérias-primas na produção de biodiesel. As diferentes características ambientais das cinco regiões brasileiras facilitam o aparecimento de espécies distintas em cada área. Aliado a isso estão a boa qualidade do solo e a disponibilidade de água doce para irrigação em escala industrial (CÉSAR *et al.*, 2015).

Segundo o boletim da ANP (2020), a soja ainda mantém-se como principal insumo para a produção de biodiesel (B100), correspondendo ao equivalente a 71% do total, com um aumento de 3% em relação a 2019. A segunda fonte no ranking de produção das usinas foi a gordura animal (12,0%), seguida por outros materiais graxos (11,0%). Ao longo dos anos, outras matérias-primas, vêm sendo inseridas nesse mercado, incluindo, o óleo de palma, óleo de amendoim, óleo de nabo-forrageiro, óleo de mamona, óleo de sésamo, óleo de canola e óleo

de milho, equivalente a 3% do total. Em sequência, temos o óleo de algodão e o óleo de fritura usado, relativo a 1% para ambos, conforme a Figura 3.3.

Figura 3.3 – Matérias-primas utilizadas na produção de biodiesel (B 100) do ano de 2008 a 2020



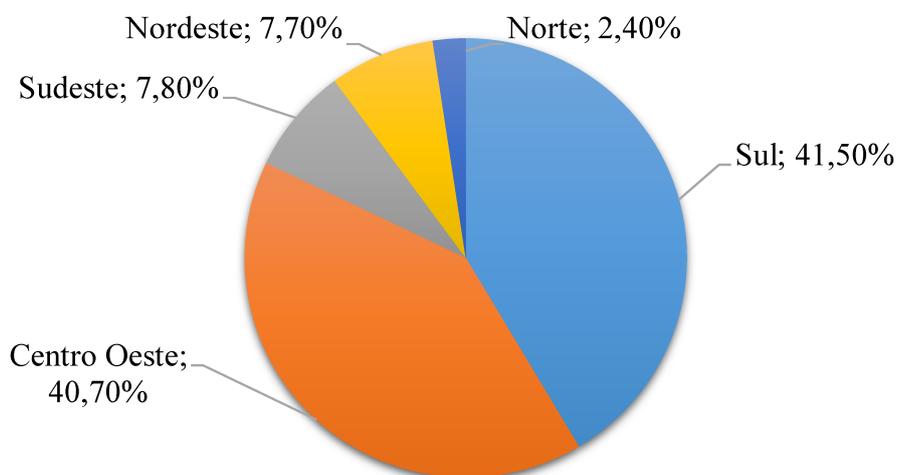
*Gordura animal inclui: gordura bovina, gordura de frango e gordura de porco

Fonte: ANP- Mês de Referência: setembro/2020.

Apesar de a soja ainda ser matéria-prima predominante para produção brasileira de biodiesel, há uma grande preocupação a respeito da extensão de terras cultiváveis que ela pode ocupar, já que além de ser usada no abastecimento energético também tem papel importante no mercado alimentício. Além disso, em termos estratégicos, não é interessante que um país dependa de apenas um produto, fazendo com que seja crescente a busca por novas oleaginosas com potencial para obtenção de biodiesel (RIBEIRO, 2017).

Acredita-se que para estimar a capacidade instalada para 2029, que a distribuição regional das plantas seja mantida, de forma que o mercado continue com uma capacidade de produção cerca de 20% acima da demanda projetada para cada região (EMPRESA DE PESQUISA ENERGÉTICA-EPE, 2020). Como demonstrado na Figura 3.4, as regiões Sul (41,5%) e Centro-Oeste (40,7%) continuam mantendo sua liderança na produção deste biocombustível, embora a demanda esteja mais concentrada na região Sudeste, que com o Nordeste apresentaram o equivalente, a 7,8% da produção. Com estímulo à inserção de novas culturas, adaptadas às condições edafoclimáticas das regiões Norte e Nordeste, estas também podem vir a apresentar grande potencial de crescimento na produção de biodiesel (AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCMBUSTÍVEIS-ANP, 2020).

Figura 3.4 – Produção Nacional de biodiesel em 2020.



Fonte: ANP- Mês de Referência: setembro/2020.

A partir desses dados, é importante enfatizar também que a viabilidade de cada oleaginosa dependerá da sua competitividade técnica, econômica e ambiental, além de aspectos como teor em óleo, produtividade por unidade de área, adequação a diferentes sistemas produtivos, sazonalidade e demais condições relacionadas com o ciclo de vida da planta (ATABANI *et al.*, 2013; SAMADI *et al.*, 2014).

3.4.1 Macaúba

A *Acrocomia aculeata* (Jacq.) Lood. ex. Martius, é uma espécie de palmeira nativa, popularmente conhecida como Macaúba na região central do Brasil (Minas Gerais, Goiás e Tocantins); Macaíba no Nordeste; ou ainda Bocaiúva nos estados de Mato Grosso e Mato Grosso do Sul, dentre outras várias denominações intrarregionais (LIMA *et al.*, 2018). Minas Gerais é um dos principais estados de ocorrência desta espécie, tendo, inclusive, uma lei de incentivo ao cultivo, extração, comercialização, consumo e transformação da Macaúba, o Pró-Macaúba (Lei nº 19.485). Evidenciam ampla distribuição no território brasileiro, ocorrendo desde os estados de São Paulo e Rio de Janeiro, passando por toda a região Centro-Oeste, pelo Nordeste e Norte do Brasil, sendo também encontrada na região Sul (CICONINI *et al.*, 2013).

De acordo com Coppel *et al.* (2018), a macaúba apresenta uma produtividade interessante que a coloca entre as principais culturas agroenergéticas: produz cerca de 25 toneladas frutos ha⁻¹·ano⁻¹, com alto teor de óleo (50-75%), ou seja, 6200 kg de óleo ha⁻¹·ano⁻¹,

comparável às produções obtidas de dendê africano (*Elaeis guineensis*). Além do alto rendimento em óleo, a macaúba agrega outros requisitos para o desenvolvimento sustentável, destacando-se: o uso integral dos coprodutos como fonte de cogeração de energia, e/ou obtenção de produtos de alto valor agregado, como carvões ativados (EVARISTO *et al.*, 2016) e a aptidão para os sistemas de produção em Integração Lavoura Pecuária Floresta (CONCEIÇÃO *et al.*, 2015; CARDOSO *et al.*, 2017).

Na Figura 3.5 pode ser visualizada a palmeira, os cachos com frutos, as inflorescências, e um corte de um fruto de macaúba.

Figura 3.5 – Características morfológicas de *Acrocomia Aculeata*. (A) palmeira; (B) inflorescência; (C) frutos (D) estrutura e composição média do fruto da macaúba.



Fonte: CICONINI, 2013; NUCCI, 2007.

O fruto da macaúba (Figura 3.5-C e D) é considerada, como parte mais importante desta palmeira, uma vez que todo ele pode ser aproveitado: casca, polpa, castanha e amêndoa (NUNES *et al.*, 2015). Empregado para fins energéticos, o óleo da polpa de macaúba é destinado à produção de biocombustíveis (biodiesel e bioquerosene), apresentando cerca de 78% de ácidos graxos insaturados, especialmente ácido oleico. O óleo da amêndoa é utilizado na fabricação de cosméticos, e é composto por aproximadamente 71% de ácidos graxos saturados que compreendem principalmente os ácidos láurico, mirístico, oleico e palmítico, além de um alto teor de compostos bioativos como carotenoides e tocoferóis (CÉSAR *et al.*, 2015; SILVA *et al.*, 2020). Os resíduos de ambas as partes, polpa e amêndoa também são aproveitados para ração animal, a torta de macaúba apresenta uma excelente composição nutritiva, que inclui proteínas, fibras, ferro, cálcio e vitaminas essenciais (PIMENTA, 2010). O endocarpo é utilizado para a obtenção de carvão ativado com ótimos rendimentos em

gasogênicos, operações metalúrgicas e siderúrgicas, pelo alto poder calorífico (EVARISTO *et al.*, 2016). A madeira é usada na confecção de ripas e calhas de água e as folhas como forragem e fibras têxteis, praticamente não deixando resíduos inaproveitáveis (LOBATO, 2014).

A palmeira da macaúba (Figura 3.5-A) é denominada estipe, essa estipe é cilíndrica, sem ramificações e com folhas aglomeradas em seu ápice (LORENZI *et al.*, 2010). Possui também inflorescências de coloração amarelada, agrupadas em cachos pendentes com comprimento de 50 a 80 cm, conforme a Figura 3.5-B (CICONINI, 2012; COSTA *et al.*, 2014). Entre outras características, a planta é de fácil adaptação, se desenvolve bem em ambientes mais secos, sendo capaz de resistir a doenças e suportar grandes variações climáticas (CÉSAR *et al.*, 2015; NAVARRO-DÍAZ *et al.*, 2014).

Os frutos são esféricos ou ligeiramente achatados, lisos, marrom-amarelados, com diâmetro que varia de 2,5 a 5 cm e apresentam-se agrupados em cachos que podem pesar até 60 kg (Figura 3.5-C e 3.5-D) (IHA *et al.*, 2014). Contém uma amêndoa oleaginosa, envolvida por um endocarpo rígido, por um mesocarpo rico em óleo, fibra e mucilagem; e um epicarpo lenhoso. O mesocarpo (polpa) é, fibroso, de sabor muitas vezes adocicado, de coloração amarela ou esbranquiçada. A amêndoa é mais proteica e com maior teor de óleo, sendo também comestível. De acordo com Cardoso *et al.*, (2020) o teor de óleo destas duas partes, dependem da região de estudo.

Pesquisadores da Empresa Brasileira de Pesquisa energética- EMBRAPA, têm realizado estudos de melhoramento genético da espécie, de desenvolvimento dos sistemas de produção, da qualidade da matéria-prima e de processamento do óleo. Os trabalhos, além disso, visam a adequação dos sistemas de cultivo que tornem a macaúba uma fonte viável de matéria-prima para a produção de biodiesel (FAVARO *et al.*, 2018; JUNQUEIRA *et al.*, 2019).

Estudos realizados pela Embrapa Cerrados (DF) indicam que a *Acrocomia aculeata* tem potencial para produzir até (8 ton/ha/ano) de óleo. O potencial produtivo dessa espécie pode ser comparado ao rendimento anual das oleaginosas comumente utilizadas na produção de biodiesel como a soja (6 ton/ha.ano), o algodão (4 ton/ha/ano) e o girassol (7 ton/ha/ano) (CÉSAR *et al.*, 2015; NOBRE *et al.*, 2014).

Evaristo *et al.*, (2016) após avaliarem 100 genótipos provenientes de Mirabela, região do Norte de Minas Gerais, estimaram o potencial de produção de óleo das dez plantas mais produtivas, encontrando 5,71 ton/ha/ano de óleo de polpa e 1,00 ton/ha/ano de óleo da amêndoa. Genótipos de macaúbas das cidades de Montes Claros, Coração de Jesus, e de São Gotardo, também apresentaram esta tendência. Para a mesma amostragem, na região de Lavras, Sul de Minas, foram verificados 2,3 ton/ha/ano de óleo de polpa, 0,4 ton/ha/ano de óleo de amêndoa, a cada 400 plantas/ha.

Berton *et al.* (2013) estimaram produções entre 5.38 ton/ha/ano de óleo da polpa a 1,00 de óleo da amêndoa, entre plantas de populações nativas de macaúbas de São Paulo e Minas Gerais, para densidade de 400 plantas/ha. Na região do Estado de Mato Grosso do Sul, valores estimados com base em observações de genótipos de macaúba de Campo Grande chegaram a 1,05 ton/ha/ano de rendimento de óleo na polpa e 0,42 ton/ha/ano de rendimento de óleo na amêndoa, considerando a mesma densidade de 400 plantas/ha e eficiência de extração de 70% (CICONINI *et al.*, 2013). Para ambos os biomas, Cerrado e Pantanal, os rendimentos estimados por Ciconini *et al.* (2013) indicaram que plantas dessa região do Mato Grosso do Sul tem menor potencial comparado às plantas de populações dos estados de Minas Gerais e São Paulo.

Quanto a isso, mais resultados promissores estão surgindo em periódicos de pesquisa qualificados nos últimos cinco anos (ANTONIASSI *et al.*, 2020; CARDOSO *et al.*, 2020; JUNQUEIRA *et al.*, 2019; FAVARO *et al.*, 2018). Portanto, esses e demais estudos podem contribuir como importante referência do conhecimento da espécie na produção de biocombustíveis.

3.4.2 Níger

Níger (*Guizotia abyssinica* Cass.), é um grão oleaginoso, com significativo valor comercial, destacando-se como matéria-prima na ração de pássaros, como insumo na produção de sabonetes, tintas, lubrificantes e, potencial uso como biodiesel (GORDIN *et al.*, 2015).

O gênero *Guizotia*, pertence à mesma família botânica do girassol (*Asteraceae*) e cártamo (*Compositae*) (Figura 3.6-A), compreendendo seis espécies, das quais *Guizotia abyssinica* Cass. é o único cultivado, com maior concentração de plantas na Etiópia, África tropical, onde é nativa (GELETA *et al.*, 2010). Sua semente é importante para a produção de óleo na Etiópia e em certas partes da Índia, (WEISS, 2000), em rotação com cereais e

leguminosas (BOTTEGA, 2012). Todas as partes da planta podem ser visualizadas na Figura 3.6.

Figura 3.6 – Características morfológicas de *Guizotia abyssinica* Cass. (A) flor; (B) folhas, flores e pecíolo; (C) sementes.



Fonte: Adaptado de <https://www.agefotostock.com/age/es/Stock-Images>.

Essa é uma espécie herbácea anual, rústica, cujas sementes (Figura 3.6 --C) apresentam em torno de 30% a 45% de teor de óleo (GELETA *et al.*, 2011). Destes, quanto ao perfil de ácidos graxos, são verificadas porcentagens de 9,2% de ácido palmítico, 10,1% de ácido esteárico, 9,0% de ácido oléico e 71,7% de ácido linoléico, tornando-os adequados em termos de composição, para o preparo e comercialização de biodiesel (SARIN *et al.*, 2009; YADAV *et al.*, 2012). Além disso, o óleo das sementes se assemelha ao do cártamo e do girassol, com alto teor de ácido linoléico (C18:2n-6), que pode ser de até 85%, dependendo da origem (RAMADAN; MÖRSEL, 2004). Os grãos ainda possuem cerca de 28% de proteínas em sua composição (FATIMA *et al.*, 2015).

Em relação às características agrônômicas, a planta pode atingir um porte de 0,5-1,5 m de altura, apresentando folhas opostas, serrilhadas, que chegam à aproximadamente 22 cm de comprimento (Figura 3.6 --B). As sementes são negras brilhantes (Figura 3.6 --C) e muito leves, com peso de 1.000 sementes variando de 3 a 5 gramas (RAMADAN; MÖRSEL, 2002). Sua população é em média de 400.000 plantas por/ha⁻¹, sendo os níveis de rendimento de óleo em torno de 200-300 kg ha⁻¹, embora possam chegar a 500-600 kg ha⁻¹, quando bem manejadas (RAMADAN; MÖRSEL, 2003). Seu ciclo é em torno de 90 a 110 dias, o que faz de níger também uma boa opção para cultivo entressafra ou cultura de rotação, podendo ser cultivada com milho ou trigo (AMBROSANO, 2012; FRANCHINI, 2014).

A espécie é de média exigência, a qual possui condições de crescer em zonas temperadas e tropicais, assim como de se desenvolver em solos ácidos e pobres (GETINET; SHARMA, 1996). Dessa maneira, permite-se adaptar bem às condições edafoclimáticas brasileiras (CARNEIRO *et al.*, 2008), embora ainda não é explorada para a produção de biodiesel.

4 MÉTODOS QUIMIOMÉTRICOS

4.1 Quimiometria

Na química, frequentemente, as amostras contêm muitas informações complexas a serem analisadas simultaneamente, e/ou poucos analitos na presença de muitos interferentes. A necessidade, de uma instrumentação mais eficiente para fazer face a esses sistemas exige, assim, ferramentas computacionais mais rápidas e confiáveis para manipular e interpretar as informações obtidas (MALIK; TAULER, 2013; FERREIRA, 2015). Neste caso, a Quimiometria é o campo da química utilizada para o planejamento e otimização das condições experimentais, e para a extração de informações relevantes dos dados produzidos. Em geral, envolvem a aplicação de métodos matemáticos e estatísticos na análise de dados químicos multivariados (MASSART *et al.*, 1998; SANTOS; TOGNOLLI; OLIVEIRA, 2010).

As técnicas de classificação multivariada (também conhecidas como reconhecimento supervisionado de padrões) lidam com respostas qualitativas, ou seja, definem relações matemáticas entre um conjunto de variáveis descritivas (por exemplo, medições químicas) em uma variável qualitativa, isto é, a associação a uma categoria definida. Uma vez treinado o modelo de classificação, é possível prever a associação de amostras desconhecidas a uma das classes definidas (BALLABIO; CONSONNI, 2013).

Alguns dos métodos mais aplicados em quimiometria são: PCA (*Principal Component Analysis*), PCR (*Principal Component Regression*), PLS (*Partial Least Squares*), SIMCA (*Soft Independent Modeling of Class Analogy*), entre outros. Têm sido conhecidas como técnicas mais robustas as redes neurais e a SVM (*Support Vector Machine*).

4.2 Instrumentação analítica

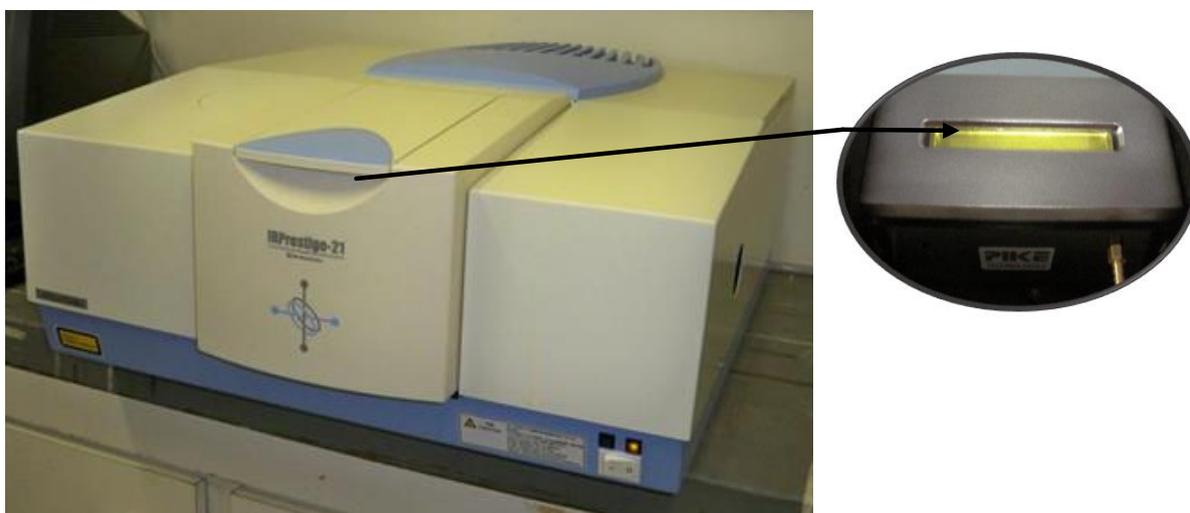
4.2.1 Espectroscopia MIR aliada a Refletância Total Atenuada (ATR)

A maioria dos estudos sobre predição de propriedades de combustíveis opta pela utilização da Espectroscopia no Infravermelho (BAIRD; OJA, 2016). A Espectroscopia envolve a interação da radiação eletromagnética com a matéria tendo como principal objetivo o estudo dos níveis de energia de átomos ou moléculas. A região no Infravermelho Médio (MIR) refere-se a transições vibracionais e rotacionais das ligações das moléculas e se estende de cerca de 4000 a 200 cm^{-1} sendo considerada a mais importante para a caracterização de compostos orgânicos (BARBOSA, 2008). Constitui a região MIR do espectro, também uma faixa chamada região de “impressão digital” (1.200 a 700 cm^{-1}), onde pequenas diferenças na estrutura e na composição de uma molécula resultam em mudanças significativas no perfil e na distribuição das bandas de absorção. Assim, uma simples semelhança entre dois espectros nessa amplitude, bem como em outras, sinaliza fortes evidências da identidade dos compostos que produzem os espectros (HOLLER; SKOOG; CROUCH, 2009).

A primeira vantagem da Espectroscopia MIR é o fato de se tratar de uma técnica quase universal: é possível analisar moléculas em diversos estados físicos (sólidos, semi-sólidos, líquidos, gases), pós, polímeros, substâncias orgânicas, inorgânicas, materiais biológicos, entre outras. A outra vantagem se refere ao fato que seus espectros são ricos em informação, pois as bandas nessa região fornecem as estruturas químicas das moléculas em uma amostra, enquanto as intensidades estão relacionadas as concentrações (PAVIA *et al.*, 2010, SMITH, 2011).

A partir do uso da técnica de Refletância Total Atenuada, conhecida pela sigla ATR (do inglês, *Attenuated Total Reflectance*) os espectros de absorção são obtidos de forma rápida, em menos de um minuto cada, para amostras sólidas ou líquidas, com um mínimo de preparação prévia das amostras, sendo aplicável tanto em aspecto qualitativo quanto quantitativo (FILGUEIRAS, 2014). Além disso, as análises podem ser realizadas *in situ* com aparelhos portáteis. Dessa maneira, as amostras líquidas podem ser adicionadas diretamente no acessório ATR (Figura 4.1), a exemplo de soluções de biodiesel em diesel. Verifica-se ainda, para esse tipo de acessório, que a técnica é considerada não destrutiva, tendo a vantagem de manter a amostra preservada (SANTANA *et al.*, 2015).

Figura 4.1 – Espectrofotômetro com Transformada de Fourier, equipado com um acessório de Refletância Total Atenuada (ATR) de Cristal de ZnSe.

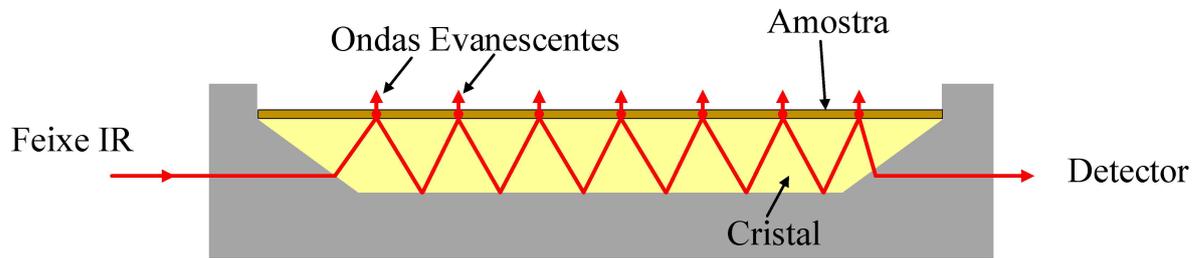


Fonte: A autora, 2021.

O uso da ATR baseia-se no fato da ocorrência da reflexão da radiação infravermelha quando ela passa de um meio mais denso (cristal de ATR), para um menos denso (a amostra). Esse meio mais denso é denominado prisma ou elemento de reflexão interna (PAVIA *et al.*, 2010). Durante o processo de reflexão, o feixe se comporta como se penetrasse há uma pequena distância no meio menos denso antes de ser refletido, como demonstrado na Figura 3.2. A profundidade de penetração, que varia de uma fração até vários comprimentos de onda, depende do comprimento de onda da radiação incidente, dos índices de refração dos dois materiais e do ângulo do feixe incidente em relação à interface. Essa radiação penetrante é denominada de onda evanescente (SMITH, 2011).

Os espectros obtidos na região do MIR fornecem informações químicas características de cada composto, mas a grande sobreposição de sinais dificulta a interpretação visual dos mesmos (SMITH, 2011). Para resolver este problema, é necessário a aplicação de métodos quimiométricos de análise multivariada, sem o qual, devido à complexidade dos dados, se tornaria difícil extrair informações relevantes, sobre a qualidade das amostras (VALDERRAMA *et al.*, 2009).

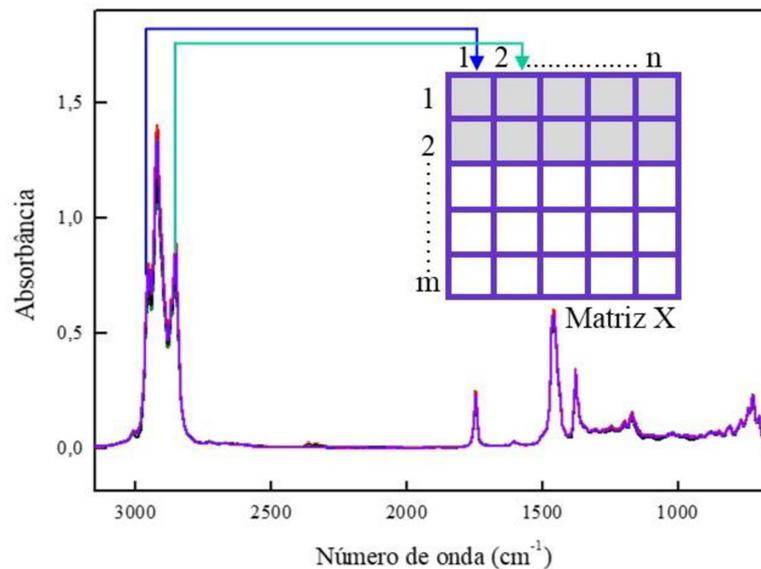
Figura 4.2 – Processo de reflexão no ATR horizontal



Fonte: A autora, 2021.

4.2.2 Organização e pré-processamento dos dados espectrais

Uma vez coletados os dados, estes devem ser organizados em um arranjo ordenado de linhas e colunas, constituindo uma matriz X de dimensão $n \times m$, que pode ser construída a partir de um vetor de resposta instrumental, onde n (linhas) se refere, por exemplo, ao conjunto de espectros das amostras e m (colunas) representam as variáveis, que são valores das absorvâncias em determinados números de ondas (FERREIRA, 2015), conforme demonstrado na Figura 4.3.

Figura 4.3 – Organização dos dados para a construção da matriz X .

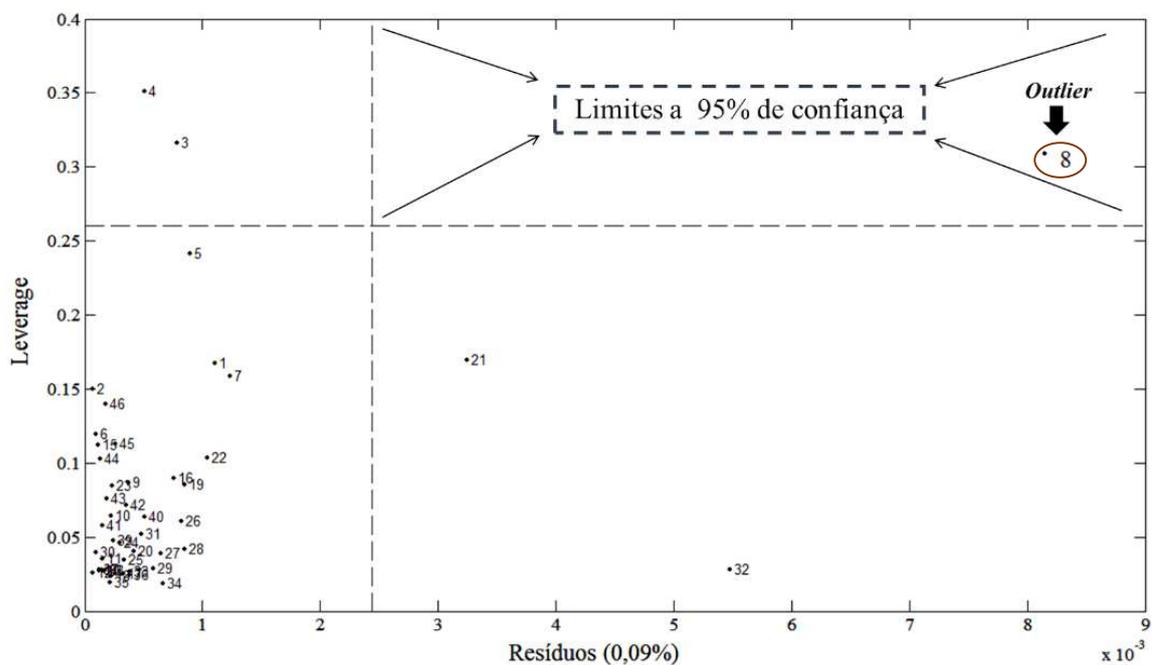
Fonte: A autora, 2021.

O pré-processamento dos dados espectrais, normalmente é realizado para remover da informação espectral tendências que não estão relacionadas a composição química da amostra e preparar os dados para o processamento seguinte, independente do modo de aquisição (WU; SUN, 2013).

Durante a etapa experimental, o erro na realização das medidas instrumentais, de manipulação das amostras, entre outros casos, pode gerar a presença de amostras anômalas (do inglês, *outliers*), definido como a existência de amostras que não se comportam conforme a maioria dos dados (MASSART *et al.*, 1998).

Na Figura 4.4 é possível observar amostra (*outlier*) que se encontra fora dos limites a 95% de confiança (linhas tracejadas), isto é, apresenta maior número de *leverage* e maior número de resíduos. A grandeza *leverage*, pode ser interpretada como a distância de uma amostra da média do conjunto de dados. Os resíduos de *Student* correspondem a diferença entre os valores reais e os valores preditos pelo modelo (SILVA *et al.*, 2012).

Figura 4.4 – Gráfico de *leverage* versus Q Residual



Fonte: A autora, 2021.

Nesta representação (Figura 4.4) as amostras anômalas não afetam a construção do modelo, mas originam erros nas estimativas de previsão \hat{y} para essas amostras que, por sua vez, propagam-se nos erros de previsão do modelo (FERREIRA, 1999, NAES *et al.*, 2004). Portanto, convém retirá-las do modelo, quando a permanência delas implicar no aumento de informação errônea, vindo a diminuir a eficiência de previsão caso tal informação seja correlacionada com as amostras do conjunto de previsão.

Das técnicas de pré-processamento espectrais, geralmente é possível observar ruído aleatório e variação da linha de base. O método mais simples para suavizar tais variações não informativas é realizada por meio do algoritmo *baseline*. Este algoritmo cria uma função polinomial que descreve o comportamento em regiões espectrais escolhidas onde os valores de absorvâncias são próximas de zero (linha de base do espectro). Em seguida toda a região espectral pode ser subtraída ponto a ponto, os quais os valores representam o nível das variações das linhas de base (WISE *et al.*, 2006; WU; SUN, 2013).

4.3 Métodos analíticos

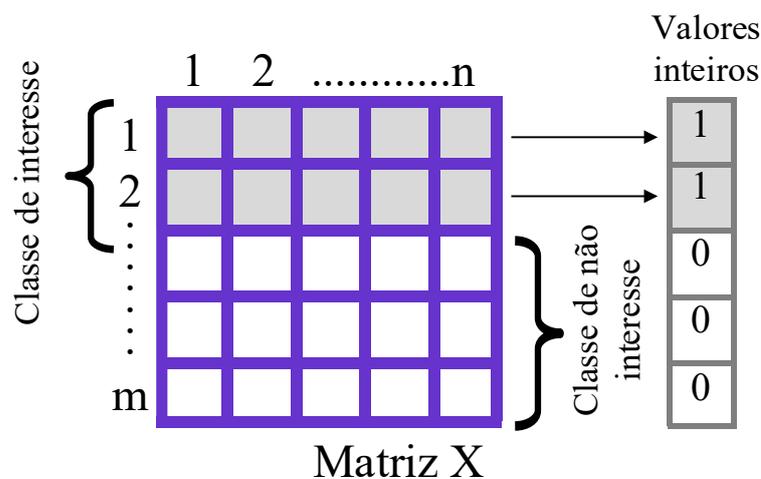
4.3.1 Análise Discriminante por Mínimos Quadrados Parciais

Como um método de reconhecimento de padrões, desenvolvido a partir dos algoritmos de mínimos quadrados parciais (*Partial least Squares*-PLS), a modelagem por PLS-DA busca relacionar as variáveis da matriz \mathbf{X} (dados espectrais), com as propriedades de interesse categórica (classes), discriminadas em um vetor \mathbf{y} . O vetor de respostas (\mathbf{y}) assume valores de $\mathbf{0}$ ou $\mathbf{1}$, onde $\mathbf{0}$ é usado para a classe de não interesse e $\mathbf{1}$ para a classe de interesse, conforme a Figura 4.5 (SILVA *et al.*, 2012; VALDERRAMA; VALDERRAMA, 2016).

Tipicamente, os valores previstos pelo modelo PLS-DA seriam 0 ou 1, mas na prática os valores previstos são próximos de 0 e 1. Faz-se necessário calcular um valor limite para separar as classes (interesse e não-interesse), é baseado no teorema de Bayes (BARKER; RAYENS, 2003, PATACA, 2006). Este leva em consideração que a distribuição das amostras no conjunto de treinamento será semelhante a observada no conjunto de previsão (WISE *et al.*, 2006).

Com essa abordagem, a PLS-DA pode ser utilizada na quimiometria para a classificação de amostras quando é necessária a redução de variáveis e não está claro se as diferenças entre grupos irão dominar a variabilidade total destas. E ainda, consiste em encontrar a similaridade entre amostras a partir de uma grande quantidade de informações, o qual não seria possível distinguir apenas pela simples observação dos dados (BARKER; RAYENS, 2003; VALDERRAMA; VALDERRAMA, 2016).

Figura 4.5 – Representação da matriz de dados X e o vetor y usados no PLS-DA.



Fonte: A autora, 2021.

Uma etapa crucial no processo estrutural do modelo PLS-DA é a definição do número de Variáveis Latentes (VLs). A escolha do número ideal de VLs é realizada com base no método de validação cruzada das amostras de calibração (ALMEIDA *et al.*, 2013).

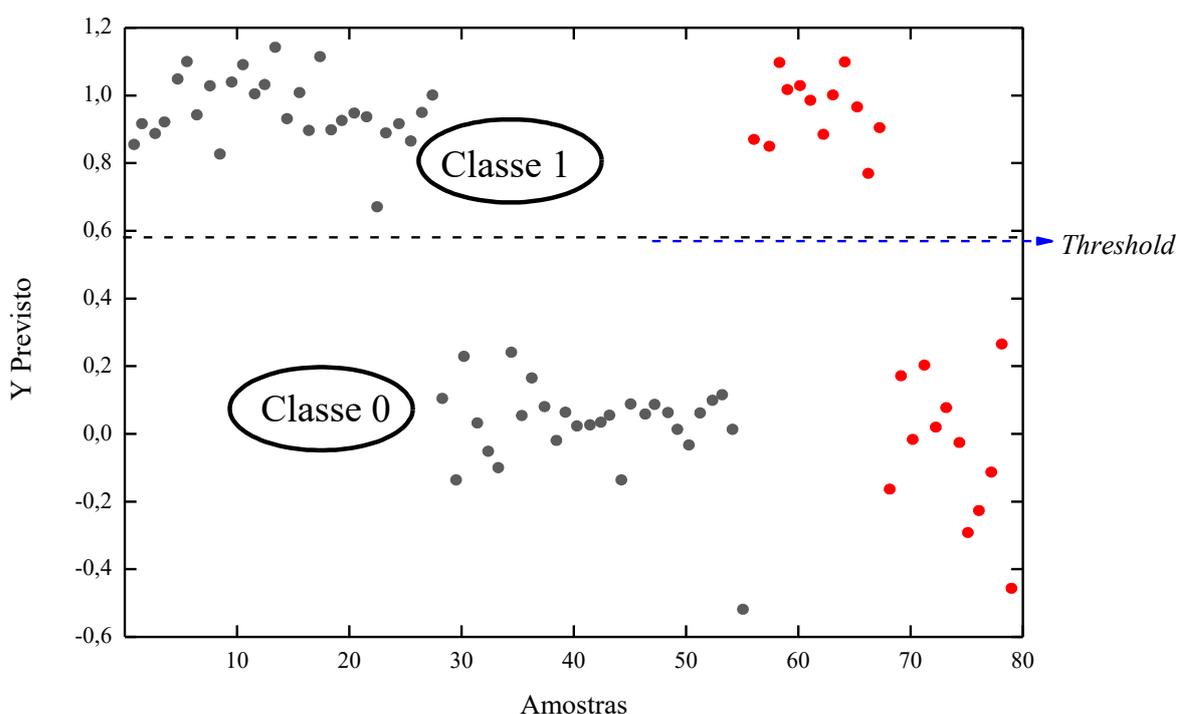
A validação cruzada ou do inglês *Cross Validation* (CV), é um processo de reamostragem, que define o número de VLs no qual o modelo deverá apresentar menor erro de predição (RMSECV, raiz do erro quadrático médio de validação cruzada), isto é, que a diferença entre os valores de referência e os valores previstos pelo modelo seja o menor possível (ANDERSEN; BRO, 2010; ERKINBAEV; HENDERSON, PALIWAL, 2017). Desse modo, o erro quadrático médio da validação cruzada (RMSECV- do inglês, *Root Mean Square Error of Cross Validation*) é calculado conforme a Equação 3.1.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3.1)$$

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

Os tipos de validação cruzada empregam diferentes critérios de reamostragem, de acordo com a natureza dos dados. Para conjuntos de dados pequenos, em torno de 20 amostras, usa-se o *leave one-out*, no qual uma amostra é retirada de cada vez, e em seguida essa amostra é prevista pelo modelo construído (BRERETON, 2000, ALMEIDA *et al.*, 2013). Já para conjuntos de dados maiores são recomendadas a validação por blocos contínuos, nos quais as amostras podem ser organizadas em ordem aleatória, por subconjuntos aleatórios, ou por venezianas (*venetian blinds*). Essa última, aplicada neste trabalho, seleciona as amostras retiradas em intervalos constantes, o que é recomendado para conjuntos de amostras organizados em ordem sistemática de agrupamentos/classes (BOTELHO, 2014; SUN, 2009). Em ambos os casos, o processo é repetido até que todas as amostras tenham sido previstas.

A Figura 4.6 mostra um exemplo do gráfico de estimativas para um modelo PLS-DA, o qual 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 de não interesse que estão abaixo do *threshold*. Figura 4.6 – Representação gráfica do resultado do modelo PLS-DA.



Fonte: A autora, 2021.

4.3.2 Avaliação de desempenho dos modelos PLS-DA

Sempre que um *procedimento* analítico é proposto, existe a necessidade de se averiguar se o método apresenta a performance adequada para as condições nas quais será aplicado. Por regra, nesse processo, deve-se haver uma utilização de amostras cujas variáveis sejam conhecidas (BEEBE; PELL; SEASHOLTZ,1998; VALDERRAMA *et al.*, 2009).

Para avaliar o desempenho dos modelos é utilizada a Tabela de Confusão, a qual permite a análise de parâmetros que melhor podem fornecer base para descrever os resultados do modelo de classificação e caracterizar os erros, vindo a refinar a classificação (SANTANA, 2015; SOUZA, 2018). Dentre estes aqui utilizados, podem ser citados: as taxas de Falsos Positivos (FP), Falsos Negativos (FN), Verdadeiros Positivos (VP) e Verdadeiros Negativos (VN), sensibilidade, especificidade, precisão, eficiência e o Coeficiente de Correlação de Matthew's (CCM).

De acordo com a avaliação das Taxas, descreve-se a possibilidade de:

- (i) FP- amostra negativa (não pertencente à classe) ser classificada como uma amostra positiva (pertencente à classe);
- (ii) FN- amostra positiva (pertencente à classe) ser classificada como uma amostra negativa (não pertencente a classe);
- (iii) VP- amostra positiva (pertencente à classe) ser classificada como uma amostra positiva;
- (iv) VN- amostra negativa (não pertencente à classe) ser classificada como uma amostra negativa;

A **sensibilidade** é a capacidade que o modelo tem de classificar corretamente as amostras verdadeiras positivas e, a **especificidade** é a capacidade que o modelo tem de classificar corretamente as amostras verdadeiras negativas. Os parâmetros sensibilidade e especificidade são calculados de acordo com as Equações 3.2 e 3.3, respectivamente (XU *et al.*, 2012; PEREIRA *et al.*, 2018).

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (3.2)$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \quad (3.3)$$

A precisão é a proporção de classificação correta, independente da classe da amostra (XU et al., 2012), de acordo com a Equação 3.4.

$$\text{Precisão} = \frac{VN + VP}{VN + VP + FN + FP} \quad (3.4)$$

Já os parâmetros **eficiência** e o **Coefficiente de Correlação de Matthew (CCM)**, expressam através de seus valores o grau de eficiência dos modelos construídos. A eficiência é apurada pela média aritmética dos valores de sensibilidade e especificidade, em que o valor 1 indica uma eficiência de 100%, sugerindo, pois, que todas as amostras foram corretamente classificadas, conforme a sua classe de origem (WORLEY; HALOUSKA; POWERS, 2013).

O CCM, representado na equação 3.5, assume resultados com valores entre -1 e $+1$, correspondendo o valor igual a $+1$ à predição perfeita (100%); 0 uma classificação completamente aleatória e -1 , uma classificação inversa (ALMEIDA *et al.*, 2013).

$$CCM = \frac{(VP \times VN - FP \times FN)}{\sqrt{(VP + FN)(VP + FP)(VN + FN)(VN + FP)}} \quad (3.5)$$

4.3.3 Conceitos sobre aprendizado de máquina e fundamentação em Máquina de Vetores de Suporte

|Aprendizado de Máquina

Com o avanço dos processadores computacionais, algoritmos de máquina foram desenvolvidos a fim de reproduzir a habilidade do cérebro humano, como a capacidade de aprender. Este processo envolve o campo da Inteligência Computacional denominada aprendizado de máquina (AM), ou do inglês, *machine learning* (CARVALHO, 2019). As técnicas de AM visam induzir um algoritmo a fazer previsões ou decisões, baseado no conhecimento a partir de um conjunto de dados (BEZERRA, 2016; REZENDE, 2019). Os algoritmos de máquina utilizam-se de técnicas estatísticas e de otimização matemática que permitem a extração de informações, o reconhecimento de padrões, e a predição para exemplos novos e desconhecidos, que é a capacidade de generalizar. A aprendizagem ocorre quando a rede atinge uma solução generalizada para uma classe de problemas. Os exemplos de treinamento vêm de algumas, geralmente, desconhecidas distribuições de probabilidade (consideradas representativas do espaço de ocorrência) e o aprendiz tem de construir um modelo geral sobre o espaço que permita prever os dados não conhecidos quando ele estava aprendendo (MOHRI; ROSTAMIZADEH, TALWALKAR, 2012).

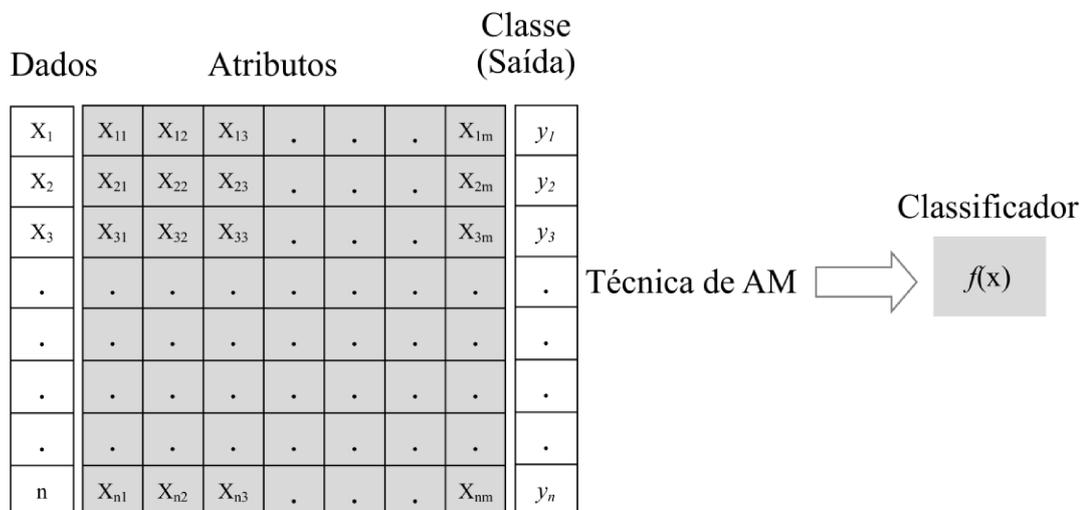
No contexto geral, os resultados teóricos do aprendizado de máquina tendem principalmente a lidar com um tipo de inferência indutiva, e no caso deste trabalho, é utilizada a chamada aprendizagem supervisionada (LORENA; CARVALHO, 2007).

|Aprendizagem Supervisionada

Ocorre quando o algoritmo de aprendizado extrai a representação do conhecimento com base em exemplos rotulados, referenciado como um conjunto de treinamento (LORENA; CARVALHO, 2007). São apresentados ao computador exemplos de entradas e saídas desejadas, fornecidas por um agente externo “professor” (PELLUCCI *et al.*, 2011; KORBUT, 2017).

No exemplo de aprendizagem supervisionada da Figura 4.7, cada exemplo é descrito como uma representação numérica, que consiste em uma ou mais medidas que são denominadas atributos, além de um valor (rótulo) indicando a qual classe que pertence (SANTOS, 2016). O objetivo do classificador é mapear as características dos vetores (\mathbf{x}) em rótulos de classes (\mathbf{y}), a fim de construir uma hipótese que permita prever valores de \mathbf{y} para exemplos ainda não rotulados (BATISTA, 2003).

Figura 4.7 – Esquema de indução de um classificador em aprendizado supervisionado.



Fonte: A autora, 2021.

A figura 4.7 mostra um conjunto treinamento com n dados $\{(x_1, x_2, \dots, x_n)\}$, onde cada dado \mathbf{x} indica os valores das entradas, descrito por m atributos, enquanto as variáveis \mathbf{y} representam as classes (saída/resposta) para cada um dos exemplos fornecidos, definido pelo classificador (LORENA; CARVALHO, 2007). Essa tarefa, nada mais é do que gerar uma função de mapeamento $y=f(\mathbf{x})$ da relação entre essas variáveis (CASTRO; BRAGA, 2011). Portanto, a classificação de um determinado conjunto de padrões pode ocorrer se sua capacidade de separação permitir ou, também, se um mapeamento for feito de forma que esta capacidade aumente (TAKAHASHI, 2015).

Existem diversas técnicas de AM que utilizam a estratégia de aprendizado supervisionado e são capazes de gerar um ótimo desempenho. Entre essas técnicas, Máquina de Vetores de Suporte (*Support Vector Machine* - SVM), têm sido aplicadas com sucesso na classificação de padrões lineares e não-lineares (bi-classes e multiclases), apresentando bom desempenho de generalização em diversas áreas de estudo. Pesquisas científicas utilizando algoritmos de máquina foram reportadas nos campos da física, medicina, engenharia de

materiais, e outras investigações (MAHMOD; MOSTAFAEI; MIRZAAEE-GHALEH, 2019). Notam-se também estudos crescentes aplicados na modelagem de dados quimiométricos, com excelentes resultados (ALVES; POPPI, 2016; LOPES; 2017; FAKHOURI, HUDAIB; FAKHOURI, 2019; ROCHA; PRADO; BLONDER, 2020; WANG *et al.*, 2020). Entretanto, poucos estudos estão relacionados quanto à predição do conteúdo de biodiesel, podendo ser uma alternativa na análise de dados envolvendo matrizes complexas para os problemas de classificação, proposta neste trabalho.

4.3.4 Fundamentação em Máquina de Vetores de Suporte

SVM é uma técnica discriminante do aprendizado de máquina, originalmente desenvolvida para a classificação binária (VAPNIK, 1995). O classificador SVM é fundamentado nos princípios indutivos da Minimização do Risco Estrutural (SRM, *Structural Risk Minimization*). A partir deste princípio, buscam minimizar o erro médio com relação ao conjunto de teste (risco de generalização), que é limitado pela soma do erro de treinamento (erro empírico) e mais um índice (a dimensão VC) (*Vapnik-Chervonenkis*, 1974), considerada a medida da capacidade de um conjunto de funções classificadoras (ZHOU; FAKHOURY, 2019). O objetivo da SVM consiste em obter um equilíbrio entre esses erros, minimizando o excesso de ajustes (*overfitting*), que podem reduzir a capacidade de generalização do classificador (BURGERS, 1998). Diante disso, oferecem uma maior robustez para dados de grandes dimensões, visto que minimizam um limite na dimensão VC e o número de erros de treinamento ao mesmo tempo (BONESSO, 2013).

Tendo em vista a aplicação prática da SVM, existem também outros conceitos e propriedades inerentes, que devem ser consideradas para a obtenção de um classificador (CALDATO, 2017; RUFINO, 2011; SANTOS, 2019), como descrito:

- As SVMs possuem um modelo matemático que consiste em um *Problema de Otimização Quadrático*, o que permite garantir uma solução ótima única (global) para diversos problemas.

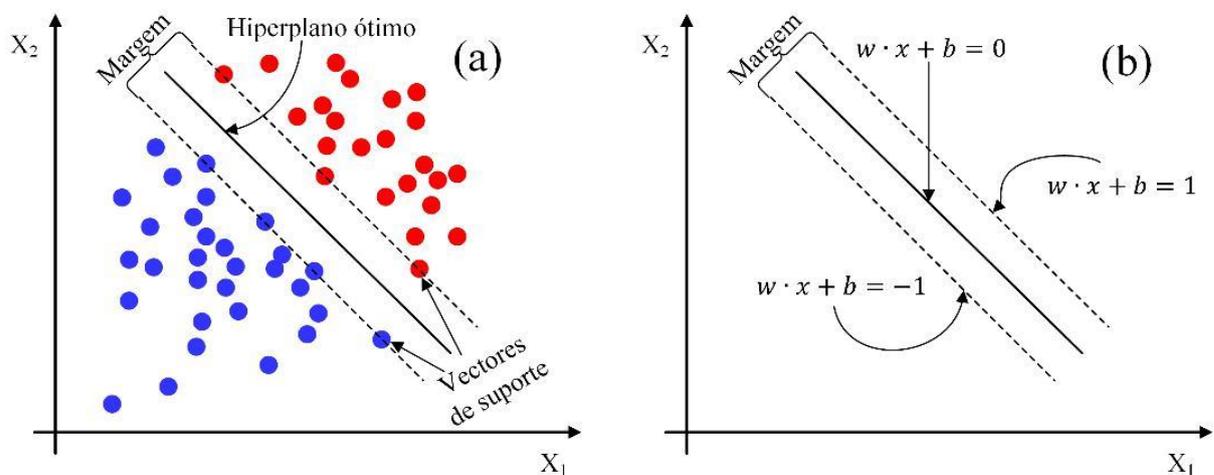
- Na fase de formulação do modelo são definidos os tipos de variáveis e o nível de relação entre elas, as representações das restrições e das funções de desempenho, denomina-se função objetivo. A meta em um problema de otimização é encontrar a solução que maximiza ou minimiza o valor dessa função.
- Como resultado do mapeamento de dados é obtida uma função classificadora (ou aproximadora) com pesos esparsos. A partir do momento que todos os dados de treinamento se encontram carregados, o algoritmo aprende a classificação e armazena na forma de vetores de suporte, responsáveis por definir a superfície de separação.
- O classificador SVM é um separador de margem máxima e para a determinação do hiperplano canônico faz a separação em termos que a distância entre os conjuntos das amostras de treinamento é maximizada, chamada margem funcional.
- O uso das chamadas funções *Kernel* da SVM permite que um separador hábil discrimine amostras não linearmente separáveis a partir do mapeamento dos dados em um novo espaço dimensional (espaço de características).

|Classificação Linear

Classificadores SVM buscam mapear os dados em um espaço em que seja possível encontrar fronteiras lineares permitindo a separação entre as classes. Esta separação é obtida a partir de um hiperplano ótimo (OSH, do inglês *Optimal Separating Hyperplane*), do qual separa-se um conjunto de vetores sem erros e a distância entre os vetores (das classes opostas) mais próximos ao hiperplano é maximizada (VAPNIK, 1999). Os vetores de suporte, representam as únicas amostras que demarcam a fronteira entre as classes, com as quais o modelo final é construído. Uma vez que o modelo é treinado, um número significativo de amostras pode ser descartado e apenas estes pesos permanecem (ZHOU; FAKHOURY, 2019) para resolver o problema de classificação. Na Figura 4.8 é representado um exemplo que mostra a separação em duas classes utilizando o conceito de vetor suporte no qual pode-se observar o hiperplano ótimo que separa as duas classes cujos pontos estão representados em cores diferentes para destacar a classificação.

A *priori*, ao se tratar de um conjunto de dados de aprendizado linearmente separável, podem ser obtidos diversos hiperplanos separadores possíveis. Seja um conjunto treinamento $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, onde y possui valor 1 ou -1, indicando a que classe x pertence, e x é um vetor de p dimensões normalizado (com valores 0, 1 ou -1), pode ser descrita a seguinte função de decisão (FLETCHER, 2009), conforme a Equação 3.6.

Figura 4.8 – Representação gráfica de um conjunto treinamento com separação ótima, hiperplano e vetores de suporte (a). Margem de separação entre fronteiras das classes positivas e negativas (b).



Fonte: Adaptado de FILGUEIRAS, 2014.

Essa função linear (Equação 3.6) divide o espaço em duas regiões $w \cdot x + b = 1$ (classe positiva) e $w \cdot x + b = -1$ (classe negativa), como demonstrado na Figura 4.8-b. Sendo w o vetor de pesos, x o vetor de amostras e b o *bias*, uma variável que permite que a margem do hiperplano seja maximizada, e a função $f(x)$ o hiperplano ótimo de separação.

$$f(x) = w \cdot x + b \quad (3.6)$$

Como forma de maximizar a margem de separação, ou seja, obter a distância máxima entre os pontos de dados das duas classes, assume-se que todos os dados de treinamento satisfazem as seguintes restrições (Equação 3.7 e 3.8).

$$x_i + w + b \geq +1, \quad \text{se } y_i = +1 \quad (3.7)$$

$$x_i + w + b \leq -1, \quad \text{se } y_i = -1 \quad (3.8)$$

A geometria de vetores simples mostra que a distância entre estes dois hiperplanos é $2/\|w\|$, portanto deseja-se minimizar o valor de $\|w\|$. Minimizar $\|w\|$ é o equivalente a minimizar o termo $\frac{1}{2}\|w\|^2$, tornando a otimização mais factível por Programação Quadrática (*Quadratic Programming* – QP), nas condições:

$$\text{Minimize } \frac{1}{2}\|w\|^2, \quad \text{tal que: } y_i(x_i + w + b) - 1 \leq 0 \forall_i \quad (3.9)$$

Uma forma de resolver um problema de otimização quadrática é utilizando multiplicadores de Lagrange. A técnica consiste em introduzir uma nova variável (por exemplo, λ), chamada de multiplicador de Lagrange (VAPNYARSKII, 2001), descrita na Equação 3.10.

$$L(w, b, \lambda) = \frac{1}{2}[w \cdot w] - \sum_{i=1}^N \lambda_i [y_i(w + x_i + b) - 1] \quad (3.10)$$

Onde λ é o coeficiente de Lagrange e w , a função objetiva.

A ideia principal é procurar por pontos nos quais as linhas de contorno sejam tangentes entre si. Isso se resume a igualar o gradiente de uma certa função Lagrangiana, ao vetor nulo (Equação 3.11 e 3.12).

$$W = \sum_{i=1}^N y_i \lambda_i x_i \quad (3.11)$$

$$\sum_{i=1}^N y_i \lambda_i = 0 \quad (3.12)$$

Uma vez determinados os multiplicadores de Lagrange pode se determinar o vetor w e o coeficiente b (também chamado de limiar). Os resultados apresentados nestas equações então sugerem:

$$b = -w \cdot x_i + y_i \quad \text{para algum } \lambda_i > 0 \quad (3.13)$$

Os dados que possuem $\lambda_i > 0$ são considerados os vetores de suporte, responsáveis pela determinação da função do hiperplano separador.

Classificação não Linear e Margem suave

Embora desejável estabelecer condições cujos dados sejam linearmente separáveis, nem sempre é possível obter um hiperplano que separa os dados com margem máxima, em grande parte dos problemas de classificação. Existem diversos fatores que em um conjunto de teste real limitam esse processo, podendo ser a presença de ruídos e *outliers* no conjunto dos dados ou a própria natureza do problema, que é não linear (SMOLA; SCHÖLKOPF, 2004). Porém, é possível obter um hiperplano que minimiza a probabilidade do erro de classificação junto às amostras de treinamento. O ajuste pode ocorrer a partir da inserção de uma variável (ξ_i) não negativa no modelo, denominada variável de folga (CORTES; VAPNIK, 1995), a qual mede os desvios dos pontos (x_i, d_i) para a condição ideal de separação das classes (Equação 3.14).

$$d_i(w \cdot x + b) \geq +1 - \xi_i \quad (3.14)$$

Em que w é o vetor dos parâmetros, b é a constante e ξ a variável folga do erro fora da margem de aceite, calculada pelos vetores de suporte. O índice i rotula os n casos de treinamentos.

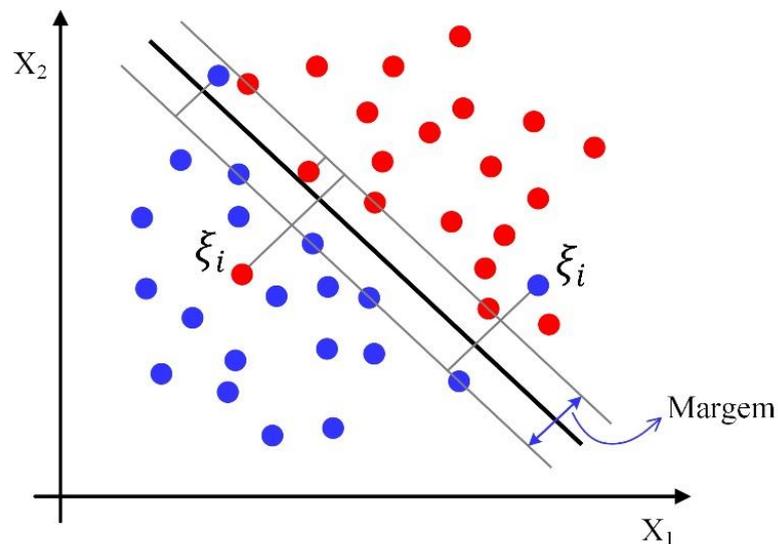
As variáveis ξ no modelo, permitem suavizar as margens de separação do classificador, consentindo que algumas amostras possam ter um erro associado à sua classificação. Esse erro é proporcional à distância da amostra ao hiperplano definido pelos vetores de suporte da classe à qual pertencem (RUSCHEL, 2017). Essa condição é representada na Figura 4.9.

Para que ocorra um erro é necessário que o respectivo ξ_i seja maior que 1, assim:

$\sum_{i=1}^N \xi_i$ é o limite superior para o número de erros de treinamento.

No caso onde $0 > \xi_i < 1$, a amostra correspondente não terá margem máxima, mas será rotulada corretamente. Quando for $\xi_i \geq 1$, a amostra poderá ser rotulada com erros.

Figura 4.9 – Figura exemplificativa de inserção da variável ξ_i em um modelo de classificação binária não linearmente separável.



Fonte: Adaptado de FILGUEIRAS, 2014.

Tipicamente, a margem para essa nova condição, tende a ser muito grande, tornando-se necessário que os erros de classificação sejam ponderados por uma constante C adicionada ao problema de otimização (RUSCHEL, 2017). Portanto, tem-se a função da Equação 3.9 reformulada, tornando-a:

$$\text{Minimize:} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (3.15)$$

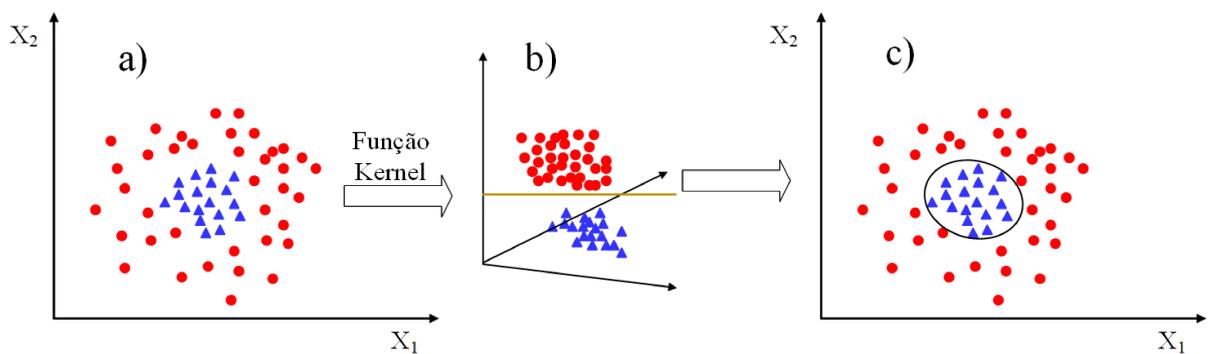
$$\text{Sujeito a:} \quad \begin{cases} w \cdot x + b \geq +1 - \xi_i \\ w \cdot x + b \leq -1 + \xi_i \\ \xi_i \geq 0 \end{cases} \quad (3.16)$$

A constante C é o termo de regularização que impõe um peso à minimização dos erros no conjunto de treinamento, contrabalanceando a complexidade da estrutura das margens e os valores de desvios ξ (PASSERINI, 2004). Portanto, a robustez do modelo depende da escolha do valor de C . De acordo com Wamser (2020), quanto maior o valor de C , maior a penalização por amostras classificadas erroneamente e maior precisão. Um valor baixo de C pode fazer com que o algoritmo faça previsões pobres devido à falta de flexibilidade, o que se conhece por *underfitting* (KUHN; JOHNSON, 2013).

Funções de Kernel

O *Kernel* é uma função que após receber os pontos do espaço de entrada (x_i e x_j) efetua o produto escalar desses argumentos, de seu espaço original para outro de maior dimensão, onde um limite de decisão linear possa ser aplicado (LORENA; CARVALHO, 2008; BONESSO, 2013). Este novo ambiente é chamado espaço de características (*feature space*). O esquema ilustrativo dessa transformação pode ser visualizado na Figura 4.10.

Figura 4.10 – (a) Mapeamento de características pela função de *Kernel* no espaço original. (b) Projeção das amostras de entrada no espaço de características de alta dimensão. (c) Representação ilustrativa de uma função não linear de classificação.



Fonte: Adaptado de ZHOU; FAKHOURY, 2019.

As funções *Kernel* são definidas segundo as condições do teorema de *Mercer*. No caso, os *kernels* devem ser matrizes positivamente definidas, ou seja, $K_{ij} = K(x_i \cdot x_j)$, para todo $i, j = 1 \dots N$, devem ter auto-vetores maiores que 0.

Dentre as diferentes funções *Kernel* mais utilizadas em SVM temos, a polinomial, sigmoide e a função de base radial (do inglês, *Radial Basis Function- RBF*). Cada uma delas apresentam parâmetros a serem definidos pelo usuário, e que podem influenciar no desempenho da máquina (ÂNGELO, 2014; KOWALCZYK, 2017), como é descrito abaixo.

- a) *kernels* polinomiais: são kernels lineares acrescentados da constante 1 e elevados ao expoente d , onde d é um número natural qualquer e a representação deste *kernel* é dada por:

$$K_{i,j} = (x_i \cdot x_j + 1)^d \quad (3.17)$$

Para as demais situações, há outros *kernels* polinomiais, com características não lineares.

- b) *kernels* sigmoidais: Satisfaz as condições de Mercer apenas para alguns valores de δ e κ e permitem que a SVM tenha o comportamento semelhante ao de uma rede neural.

$$K_{i,j} = \tanh(\delta(x_i \cdot x_j) + K) \quad (3.18)$$

- c) *kernels* RBF: uma função é de base radial quando seu valor funcional depende apenas da distância de seu argumento à origem, dada por:

$$K_{i,j} = \exp(-\gamma \|x_i - x_j\|^2) \quad (3.19)$$

Onde γ (*Gama*) é um parâmetro de ajuste usado para controlar o raio de ação do *Kernel*.

O valor padrão de γ é correspondente à $\gamma=1/k$ (k representa o número de variáveis usadas no conjunto de dados de treinamento) (ALVES; POPPI, 2013). Geralmente, quanto maior o valor de γ , maior a largura da curva e menor a precisão, de modo que reduz o *kernel* para uma função constante, tornando difícil o processo de aprendizagem. Por outro lado, valores pequenos de γ permitem ao classificador o ajuste dos dados de entrada, minimizando o risco de sobreajustamento. A tendência é um limite de decisão linear (LIMA, 2014; WAMSEN, 2020).

Na rede RBF, este parâmetro deve ser ajustado junto com o índice C (custo), que é a constante de regularização usada para equilibrar exemplos mal classificados de treinamento e a superfície de decisão (CRISTIANINI; SHAWE-TAYLOR, 2000; BONESSO, 2013). Entretanto, no processo de modelagem, esse tipo de *kernel* apresenta certa complexidade, no sentido de que é difícil de se chegar a um valor ótimo para γ , e o valor correspondente de C que funcione adequadamente para um determinado problema (NOVAKOVIC; VELJOVIC, 2011), porém é o que tem apresentado melhores resultados em aprendizado de máquina e depende do processo de busca (*Grid Search*).

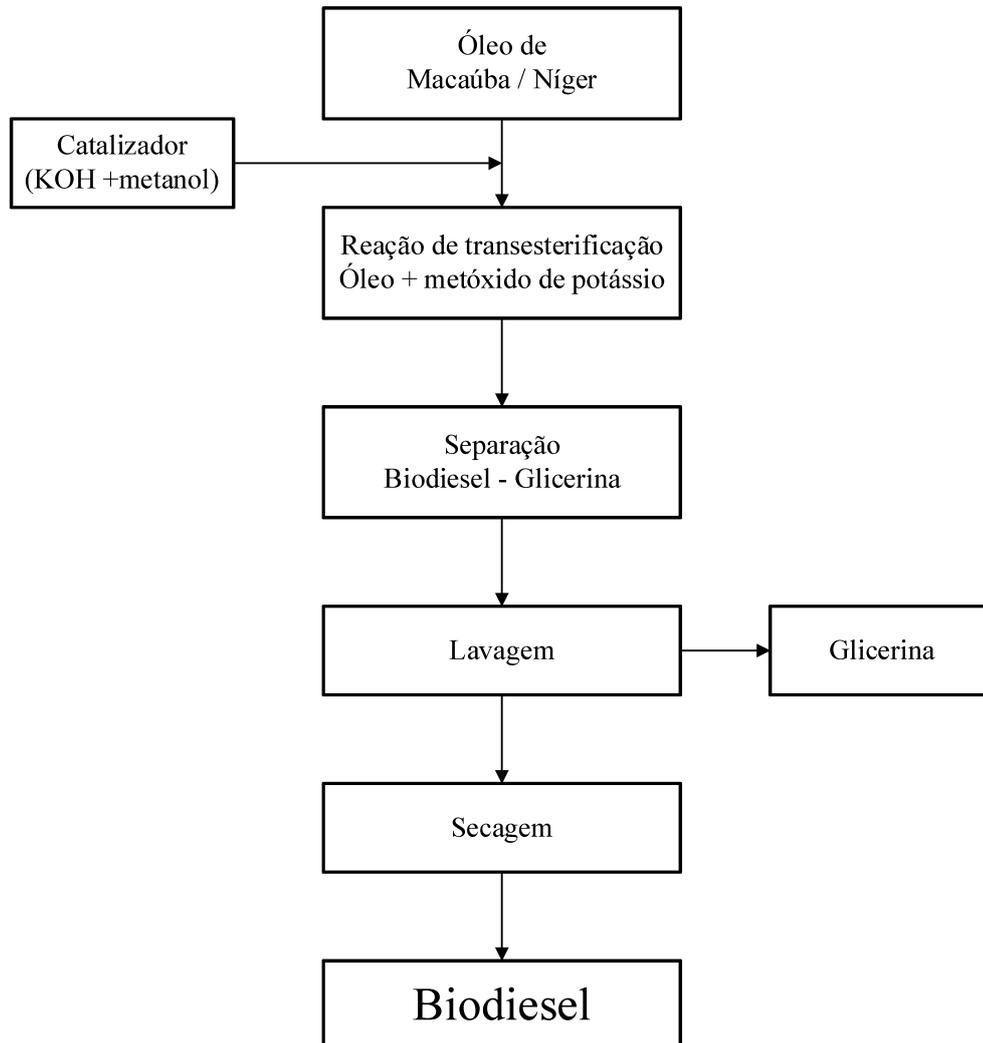
5 PROCEDIMENTO EXPERIMENTAL

5.1 Produção do Biodiesel

Os biodieseis metílico de macaúba (mesocarpo) e níger (semente), utilizados nos modelos PLS-DA e SVM desenvolvidos, foram produzidos no Laboratório de Biocombustível e Tecnologia Ambiental- LABTA da Universidade Federal de Uberlândia, Minas Gerais, Brasil. Cada tipo de biodiesel foi obtido a partir da reação de transesterificação com catálise básica, dada a baixa acidez dos óleos, seguindo metodologia padrão (ROSSI *et al.*, 2018).

Inicialmente pesou-se 100g de óleo (macaúba/níger) e 20,0 g de metanol em uma razão molar de 1:5 óleo/álcool na presença de 1,0 g de hidróxido de potássio (KOH) como catalisador. Para uma melhor homogeneização da mistura a mesma foi colocada em banho ultrassônico por aproximadamente 10 minutos, até que todo o KOH fosse dissolvido. Misturou-se o metóxido de potássio a 100 g de óleo colocando-o sobre uma placa de aquecimento com agitação por barra magnética por 40 minutos. Ao término da reação de transesterificação, essa mistura foi transferida para um funil de decantação para separar as diferentes fases, inferior (glicerina), superior (biodiesel). Após 24 horas de repouso, a glicerina foi recolhida e o biodiesel obtido passou pelo processo de lavagem com água destilada a 90°C, até que sua fase aquosa apresentasse uma aparência límpida. Ao final, o biodiesel foi seco utilizando um rotoevaporador durante 1 hora, à 88 rpm e aproximadamente 80° C. O fluxograma com as etapas do processo de produção dos biodieseis pode ser verificado na Figura 5.1.

Figura 5.1 – Fluxograma da produção de biodiesel.



Fonte: A autora, 2021.

5.2 Preparo das amostras

Para o desenvolvimento dos modelos de classificação, produziu-se um total de 74 combinações de diesel BX, que foram efetuadas para cada tipo de biodiesel metílico (níger e macaúba) na mistura com o diesel puro, em um intervalo de concentração de 0,25% a 30,00% (v/v). O diesel mineral utilizado foi fornecido pela empresa Transpetro S/A (Uberlândia, Minas Gerais, Brasil).

A Tabela 5.1 apresenta o número de amostras usadas na construção do modelo PLS-DA. O primeiro grupo de amostras foi preparado utilizando a proporção de 10% de biodiesel para 90% de diesel puro (B10), e o segundo grupo com diferentes percentagens BX (B3 à B9 e B11 à B30 contendo, respectivamente, de 3 a 9% e 11 a 30% de biodiesel).

Tabela 5.1 – Número de amostras aplicadas na construção do modelo PLS-DA.

Modelo	Número de amostras	Conjunto treinamento	Conjunto Teste
PLS-DA	37 de B10	25	12
	37 de BX (B3 a B9 e B11 a B30)	25	12

Fonte: A autora, 2021.

5.3 Obtenção dos Espectros MIR

Os dados espectrais MIR de todas as amostras foram obtidos em quintuplicata, com 16 varreduras, na região de 4.000 a 600 cm^{-1} , com resolução de 4 cm^{-1} . Utilizou-se um espectrômetro FT-MIR de marca Perkin Elmer, modelo Spectrum Two (Perkin Elmer, Waltham, MA, EUA), equipado com um acessório de Refletância Total Atenuada Horizontal (HATR) com cristal seleneto de zinco (ZnSe).

5.4 Análise Quimiométrica

5.4.1 Modelos PLS-DA

Para realizar os cálculos e construir os modelos PLS-DA, foi usado o ambiente computacional MATLAB versão 7.5 (*Mathworks Inc.*) e o pacote de programas do PLS_Toolbox, versão 8.0.2 (*Eigenvetor Research*), efetuando-se a organização dos dados e o desenvolvimento da análise conforme descrito a seguir.

Os modelos PLS-DA foram desenvolvidos com base nos algoritmos PLS (WOLD *et al.*, 2001), onde as variáveis no bloco \mathbf{X} (dados espectrais) foram relacionadas às classes discriminadas no vetor \mathbf{y} . O vetor de respostas \mathbf{y} assumiu valores de $\mathbf{0}$ e $\mathbf{1}$, em que o $\mathbf{0}$ correspondeu a classe de não interesse (BX), e o $\mathbf{1}$ a classe de interesse (B10). O conjunto de dados da matriz \mathbf{X} foi dividido em dois subconjuntos, um para treinamento e outro para teste.

O número de variáveis latentes (VL) foi definido de acordo com a porcentagem de variância observada nos blocos \mathbf{X} (absorbância por comprimento de onda) e \mathbf{y} (classe) na comparação conjunta contendo a parcela do Erro Quadrático Médio de Validação Cruzada (RMSECV). Essa escolha foi realizada conforme o método padrão ASTM E1655-05, o qual estabelece que o número mínimo de amostras para um conjunto de treinamento deve ser igual a $6(k + 1)$, onde k é o número de variáveis latentes), e o conjunto teste deve ser igual a $4k$ (NORMA ASTM, 2012).

O conjunto de dados foi centrado na média, em que foi calculada a média das intensidades para cada comprimento de onda e subtraiu-se cada intensidade do respectivo valor médio; desta forma o eixo cartesiano teve sua origem centrada no valor médio das medidas, conforme a Equação 4.1.

$$x_{ij(cm)} = x_{ij} - x_j \quad (4.1)$$

Onde $x_{ij(cm)}$ é o valor centrado na média para a variável j na amostra i , x_{ij} é o valor da variável j na amostra i e x_j representa a média dos valores das amostras na coluna j .

A validação cruzada foi realizada utilizando o critério das venezianas (*venetian blinds*), com 12 divisões e duas amostras por bloco. Como o conjunto de treinamento é formado por um grande número de amostras e foi organizado agrupando-as sequencialmente em função da classe, esse critério é o mais apropriado para garantir uma seleção representativa e homogênea das amostras de todas as classes (BOTELHO, 2014).

A exatidão dos modelos foi determinada pelo cálculo dos valores de erro quadrático médio de calibração (RMSEC), de validação cruzada (RMSECV) e predição (RMSEP), os quais foram calculados de acordo com a Equação 4.2 (MASSART, 1998).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4.2)$$

Em que, n é o número de amostras, y_i o valor de referência e \hat{y}_i é o valor previsto pelo modelo de calibração. Quando aplicado nas amostras do conjunto de calibração, o RMSE é denominado de RMSEC e quando é calculado para as amostras do conjunto de previsão, é denominado de RMSEP.

A identificação de amostras anômalas (*outliers*) foi realizada através da análise do gráfico de *leverage versus Q residuals*, com 95% de confiança para o modelo PLS-DA desenvolvido. Este gráfico é dividido em 4 regiões definidas pelo limite de resíduos e pela influência crítica do valor de *leverage* (igual a $3k/n$, onde k é o número de variáveis latentes e n o número de amostras) (SHIROMA; RODRIGUEZ-SAONA, 2009). Nesta representação, *outliers* encontram-se à região superior direita.

Para avaliar o desempenho dos modelos PLS-DA foi utilizada a Matriz de Confusão, sendo verificados os seguintes parâmetros: Taxas de Falsos Positivos (FP), Falsos Negativos (FN), Verdadeiros Positivos (VP) e Verdadeiros Negativos (VN), sensibilidade, especificidade, precisão, eficiência e o Coeficiente de Correlação de Matthew's (CCM).

5.4.2 Modelos SVM

|Desenvolvimento dos modelos

Para a construção e otimização dos modelos SVM foi utilizada a biblioteca *e1071* em linguagem R (Versão 3.6.2). O R é uma linguagem de programação que permite ao pesquisador propor novas sub-rotinas, refinar os dados e implementar novos métodos de análise (RAMOS *et al.*, 2019). Desenvolvida por Ross Ihaka e Robert Gentleman, oferece uma interface flexível e expansível, que em geral, vêm acompanhada de pacotes com funções implementadas em R, possibilitando que o programador tenha acesso a várias técnicas e conjuntos de códigos/comandos dos quais necessita (R CORE TEAM, 2018) e a biblioteca *e1071* é uma versão clássica de códigos implementado em Fortran que pode ser utilizada em linguagem R e

que oferece uma interface para *libsvm* (*Library for Support Vector Machines*), biblioteca aplicada à máquinas de vetores de suporte. Essa biblioteca fornece métodos de visualização e ajuste de parâmetros, podendo-se testar a grade de parâmetros fornecidos, e a função *svm*, para treinar os modelos (KARATZOGLOU; MEYER; HORNIK, 2006).

|*Divisão dos dados*

Buscando-se a confiabilidade estatística quanto ao desempenho dos modelos classificadores SVM, a base de dados foi dividida em três subconjuntos: treinamento, teste e validação, compreendendo 50/12/12 amostras, respectivamente. *À priori*, a fase de teste e uma etapa extra de validação, pode ser utilizada para estimar quão bem os modelos foram treinados (que depende do tamanho dos dados e da estimativa das propriedades do modelo, como de precisão e dos erros de classificação), usando uma base de dados totalmente desconhecida (HASTIE; TIBSHIRANI; FRIEDMAN, 2017). Alguns algoritmos de aprendizado de máquina operam usando as três fases.

|*Otimização e Predição*

Diferentes funções do *Kernel*, como a linear, polinomial, sigmoidal e a função de base radial (RBF) foram testadas, visando explorar a capacidade de generalização do classificador SVM. Para isto foram ajustados alguns parâmetros essenciais às configurações dos modelos, como é demonstrado na Tabela 5.2.

Tabela 5.2 – Parâmetros de ajuste das funções *Kernel*.

<i>Kernel</i>	Parâmetros ajustáveis
Linear	C
Polinomial	C, γ , <i>coef0</i> e grau
RBF	C, γ , <i>coef0</i>
Sigmoidal	C, γ , <i>coef0</i>

Fonte: CHANG; LIN (2011).

O parâmetro $coef_0$ é utilizado para normalizar os dados de entrada para um intervalo de -1 até 1 ; seu valor padrão é $coef_0 = 0$. O parâmetro *Degree* define o grau do polinômio no *kernel* polinomial; seu valor padrão é $Degree = 3$ (CHANG; LIN, 2011).

No caso dos valores para C e γ , foi feita uma busca em grade (*Grid-search*), no qual diferentes combinações desses índices em valores discretos são explorados em testes sucessivos, gerando um par com o melhor resultado entre todas as diferentes combinações possíveis entre os valores discretos (HUANG; WANG, 2006; BONESSO, 2013). A constante C é o termo de regularização do algoritmo de aprendizagem (Custo/penalização), e γ (Gama) controla a flexibilidade das curvas de *kernel* (AZEVEDO, 2018; WAMSEN, 2020). Todos os valores testados, foram usados com base em resultados positivos referenciados na literatura (PAULA, 2017).

Métricas de avaliação

A *matriz de confusão* foi empregada como um dos índices de avaliação dos modelos de classificação, sendo analisados os parâmetros de sensibilidade, especificidade, precisão e eficiência, conforme a Tabela 5.3.

Tabela 5.3 – Parâmetros da matriz de confusão para análise SVM.

Parâmetros	Fórmula
Sensibilidade	$VP/(VP+FN)$
Especificidade	$VN/(FP+VN)$
Precisão	$VP/VP+FP$
Eficiência	$VN+VP/(VN+VP+FN+FP)$

Fonte: A autora, 2021.

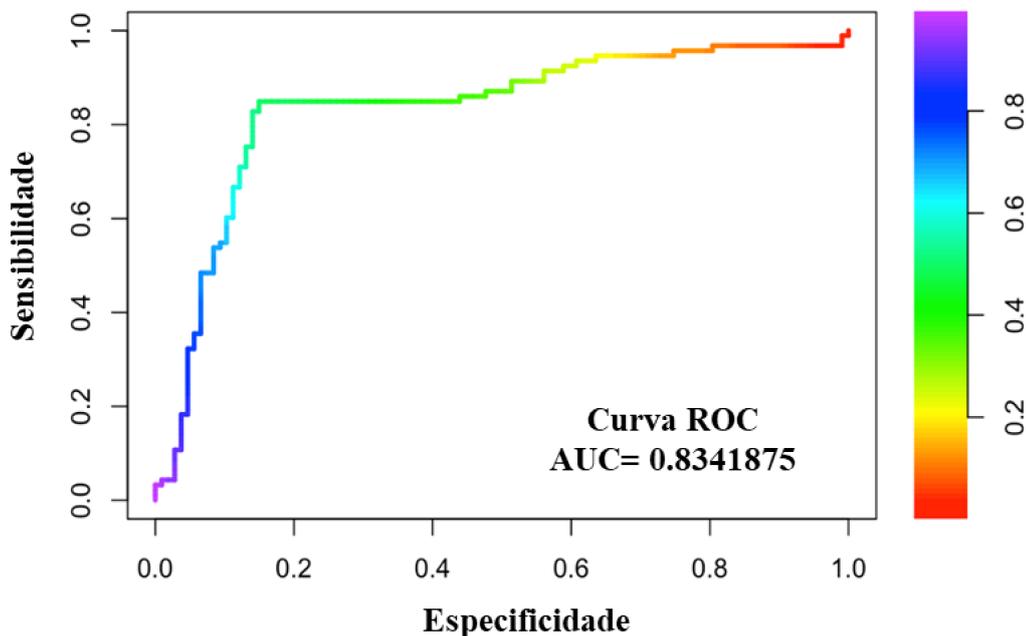
Um outro índice verificado foi o *F-Score*. O *F-score* ou *F1* é definido como a média harmônica da precisão e revocação (*recall*), que mostra o quanto o classificador aprendeu sobre uma determinada classe. Isso por que considera os erros falsos positivos (precisão) e falsos negativos (revocação), aplicando-se bem para os sistemas de classificação binários. Pode-se assumir valores entre 0 e 1, quanto mais próximo de 1 melhor o desempenho (SANTOS, 2013).

A precisão é a proporção de classificação correta, independente da classe da amostra (já antes definida), enquanto o *recall* é o número de verdadeiros positivos / pelo número total de verdadeiros positivos e falsos negativos. Como o *F1* é uma forma de combinar as duas medidas, temos a equação representada.

$$F1 = \frac{2 \times \text{precisão} \times \text{recall}}{\text{precisão} + \text{recall}} \quad (5.3)$$

Outra maneira de se quantificar a eficiência de um algoritmo de classificação é pela área sob a curva (AUC, *área Under the ROC Curve*). Aqui foi desenvolvida para fornecer uma representação gráfica e resumida de desempenho dos modelos de classificação. Essa métrica se baseia nos conceitos de especificidade e sensibilidade, representando o grau ou medida de separabilidade, explicitando o quanto um determinado modelo é capaz de distinguir entre classes (BRADLEY, 1997). A Figura 5.2 mostra um exemplo gráfico de curva ROC.

Figura 5.2 – Exemplo gráfico de Curva ROC.



Fonte: A autora, 2021.

Os gráficos ROC, ou simplesmente curva ROC (*Receiver Operating Characteristic*) são gráficos bidimensionais nos quais a taxa de verdadeiros positivos (TVP) é plotada no eixo vertical (ordenada) e a taxa de falsos positivos (TFP) no eixo da horizontal (abcissa) ao considerar-se diferentes valores de limiares ou de parâmetros (BRAGA, 2000; REZENDE, 2019). O valor da AUC varia de 0,0 até 1,0 e o limiar entre as classes é 0,5. Quanto mais o

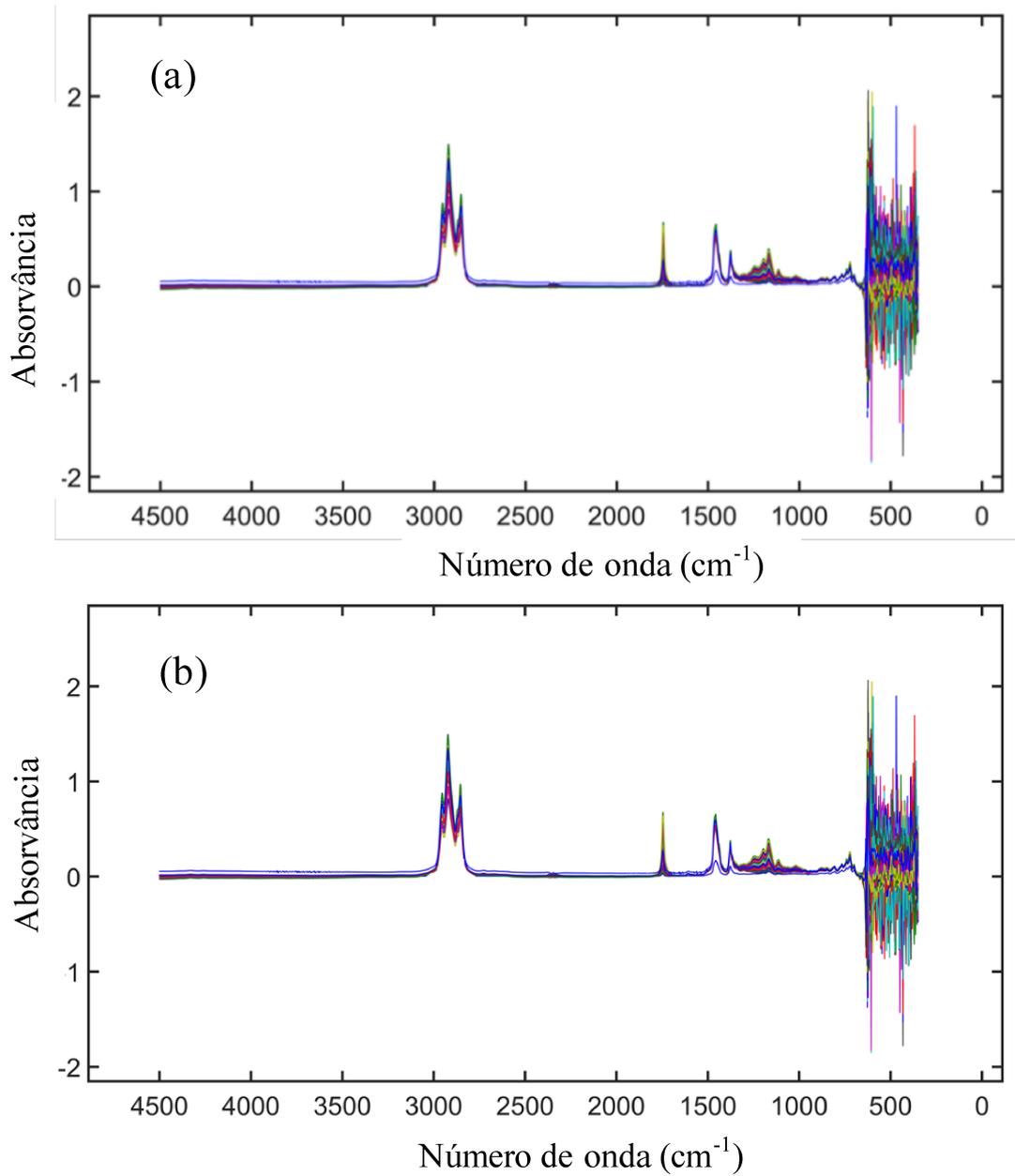
valor da AUC se aproxima de 1,0, melhor é a capacidade discriminativa dos modelos. Um modelo cujas previsões são 100% incorretas tem uma AUC de 0, enquanto 100% corretas possui uma AUC igual a 1,0 (SÁ, 2018; REZENDE, 2019).

6 RESULTADOS E DISCUSSÃO

6.1 Caracterização dos Espectros MIR

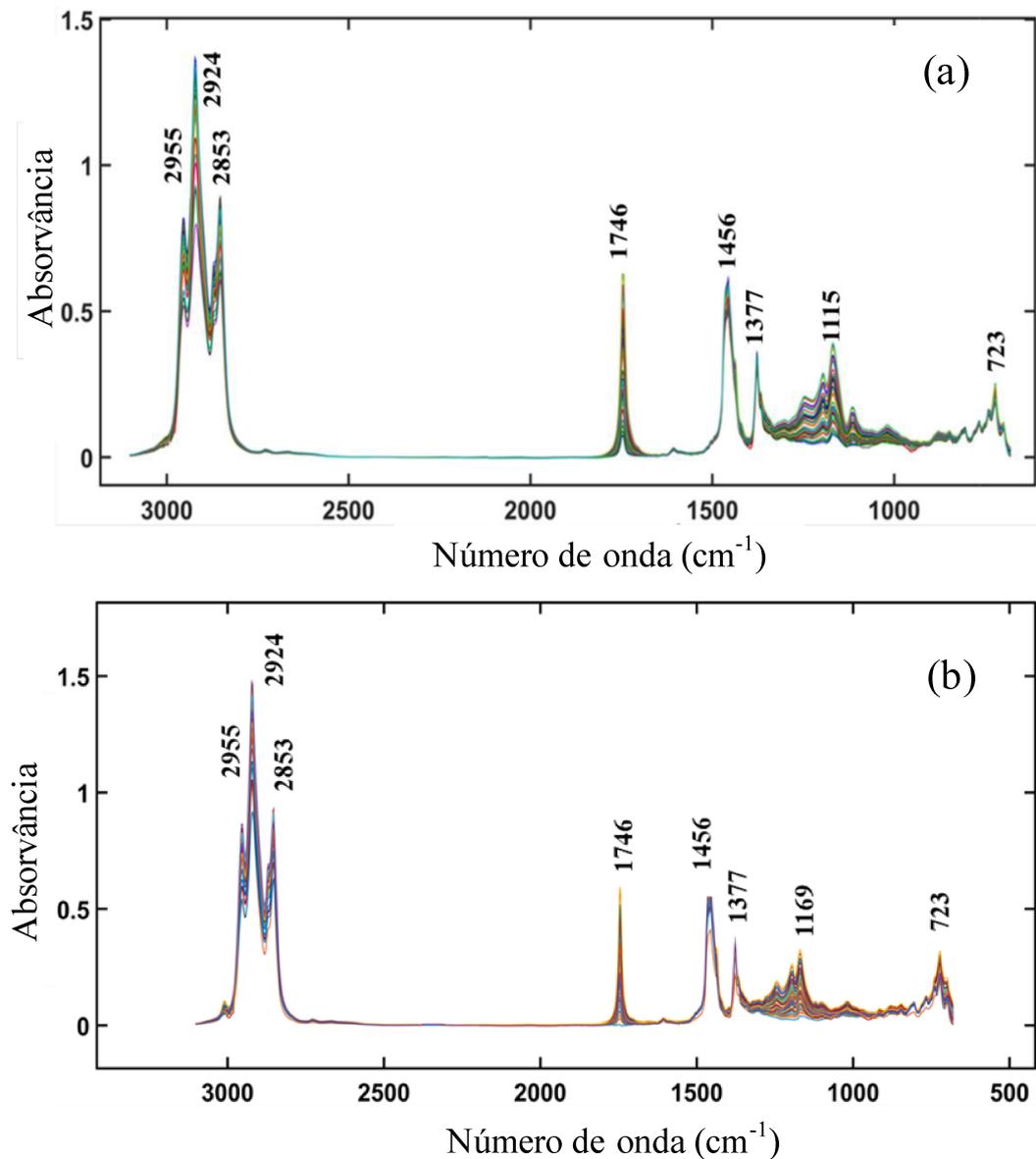
Os espectros visualizados dos biodieseis metílicos de Macaúba e de Níger (Figura 6.1), se referem à região compreendida entre 4.500 e 450 cm^{-1} , antes do pré-tratamento. É possível observar variações na linha de base, extremidade ruidosas e sem bandas de absorção, o que evidenciou a necessidade de se aplicar métodos para a correção da linha de base, removendo as regiões sem informação. Logo, os espectros foram corrigidos pelo método *baseline*, aplicado nas regiões de 2570 a 1850 cm^{-1} e 4000 a 3200 cm^{-1} (regiões sem bandas de absorção). Após esta correção, as regiões espectrais abaixo de 680 cm^{-1} e acima de 3100 cm^{-1} foram descartadas, conforme apresentado na Figura 6.2. Essa remoção foi devida a presença de ruídos instrumentais e ausência de bandas de absorção, respectivamente.

Figura 6.1 – Espectros MIR originais dos biodieseis metílicos em mistura com diesel: (a) Macaúba; (b) Níger, sem correção da linha de base e sem cortes.



Fonte: A autora, 2021.

Figura 6.2 – Espectros MIR corrigidos dos biodieseis metílicos em mistura com diesel:
(a) Macaúba; (b) Níger.



Fonte: A autora, 2021.

A Figura 6.2 exibe os espectros MIR para as 74 combinações de biodiesel/diesel utilizadas no presente trabalho. Na avaliação desses espectros, não foi possível, a partir da inspeção visual dos comprimentos de onda, identificar ou correlacionar simultaneamente cada tipo de mistura de biodiesel/diesel. Nesse caso, uma vez que esta determinação é de interesse, torna-se imprescindível a aplicação de um método quimiométrico, como da modelagem por PLS-DA.

Na Tabela 6.1, os espectros MIR correspondentes às bandas de absorção da mistura biodiesel/diesel estão associadas às vibrações, estiramento e deformação angular (BARBOSA, 2008; SILVERSTEIN *et al.*, 2004), que compreendem as faixas:

Tabela 6.1 – Bandas de absorções correspondentes aos modos vibracionais comuns aos biodieseis metílicos de macaúba e níger em mistura com o diesel.

Número de onda (cm ⁻¹)	Atribuição
2957 a 2.853	Apresentam bandas de absorção de maior intensidade nessa região espectral, provenientes das vibrações de deformação axial das ligações C-H da ligação dupla.
1.746	Banda referente às vibrações de deformação axial da ligação carbonila (C=O), presentes nos ésteres.
1.500 a 1300	Vibrações de deformação angular das ligações C-H dos grupamentos metila e metileno.
1.300 a 900	Bandas referentes às vibrações de deformação axial da ligação C-O dos ésteres constituintes do biodiesel.
845 a 700	Representam os grupos metilênicos ligados– (CH ₂) _n –, sendo a banda mais intensa em 723 cm ⁻¹ .
723	Banda de baixa intensidade, correspondente a vibração de deformação angular tipo balanço de C-H do grupo metileno ρ[CH ₂] _n , n≥4.

Fonte: A autora, 2021.

6.2 Modelos PLS-DA

A qualidade dos modelos PLS-DA construídos, foi avaliada dentre outros parâmetros, comparando os valores de RMSECV, RMSEC, RMSEP e da variância capturada em relação ao número de variáveis latentes (VL), conforme Tabela 6.2.

Na avaliação dos erros quadráticos médios observa-se valores relativamente baixos, com percentagens dentro da reprodutibilidade do método normatizado. O valor de RMSEP foi inferior a 1% (v/v), em concordância com o método padrão ABNT NBR 15568 (ABNT, 2008). Além disso, o RMSEP apresentou uma pequena diferença em relação ao RMSEC, demonstrando que não houve sobreajuste do modelo.

Tabela 6.2 – Parâmetros de Exatidão dos modelos PLS-DA.

Parâmetros	Valores dos modelos PLS-DA	
	Macaúba	Níger
RMSECV (% v/v)	0,21	0,22
RMSEC (% v/v)	0,16	0,18
RMSEP (% v/v)	0,22	0,19
Variáveis Latentes	6	6
Variância em X (%)	99,84	99,88
Variância em y (%)	89,49	87,11

Nota: % v/v (% volume/volume)

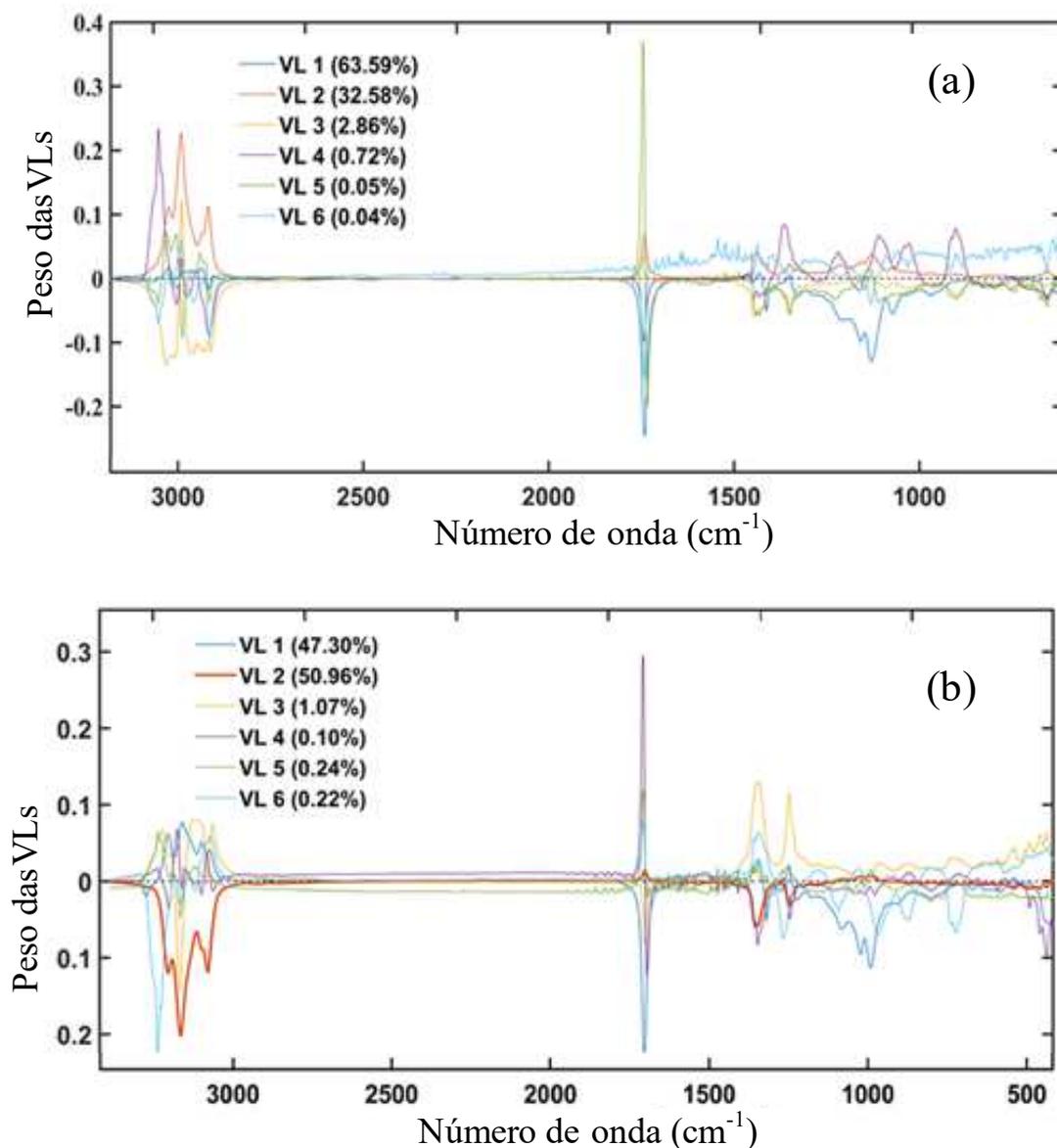
Fonte: A autora, 2021.

Para a classificação do teor de macaúba e níger na mistura B10/BX com o diesel, foram utilizadas 6 VL em cada caso (Tabela 6.2), que proporcionaram (99,84 e 99,88%) da variância explicada na matriz **X** e (89,49 e 87,11%) no vetor **y**, respectivamente.

Na Figura 6.3 é observado pela da inspeção dos pesos das VL, as que mais contribuíram para a diferenciação entre as classes. Essa avaliação foi realizada a partir do perfil dos espectros MIR da mistura B10/BX do biodiesel metílico de macaúba.

As variáveis de VL1 e VL2 apresentaram os maiores pesos influenciando, respectivamente, com 63,59% e 32,58% para a obtenção do resultado de classificação. Estas variáveis podem ser localizadas nas seguintes regiões espectrais: de 2979 a 2850 cm^{-1} , correspondente às vibrações de estiramento assimétrico da ligação C-H do grupo metila (CH_3) e às vibrações de estiramento simétrico e assimétrico da ligação C-H do grupo metileno (CH_2); de 1750 a 1730 cm^{-1} , relacionada à deformação axial ou estiramento de carbonila ($\text{C}=\text{C}$) de éster; de 1500 a 1300 cm^{-1} , referentes às vibrações de deformação angular das ligações C-H dos grupamentos metila e metileno; de 900- 700 cm^{-1} , correspondente à deformação angular (tipo balanço) assimétrica C-H de grupo metileno (CH_2), conforme Barbosa, 2008; Ruschel, 2014.

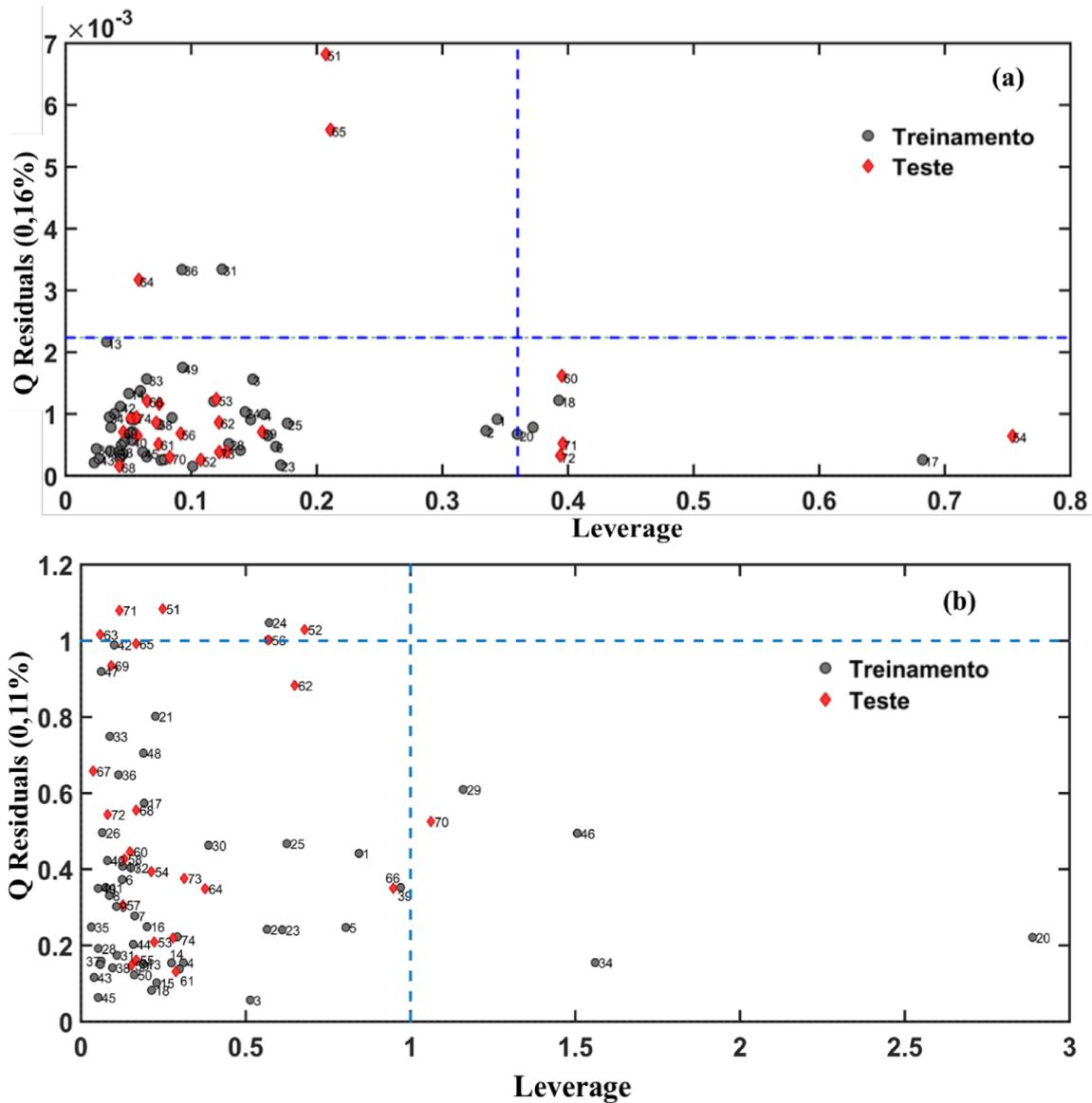
Figura 6.3 – Gráfico dos pesos de VL1, VL2, VL3, VL4, VL5 e VL6 versus número de onda do modelo PLS-DA de: (a) Macaúba e (b) Níger.



Fonte: A autora, 2021.

A Figura 6.4 exibe o gráfico de *leverage* versus *Q residuals* para o modelo PLS-DA desenvolvido. Nota-se que nenhuma das amostras foi considerada como *outlier*, uma vez que a região superior direita definida pelas linhas tracejadas não contém nenhum ponto. Entretanto, houve a ocorrência de amostras que apresentaram altos valores de resíduos, ou seja, que perpassaram o limite de resíduos. Contudo, estas não são consideradas anômalas, pois os valores de *leverage* apresentados para estes pontos estão à esquerda da influência crítica do valor de *leverage*.

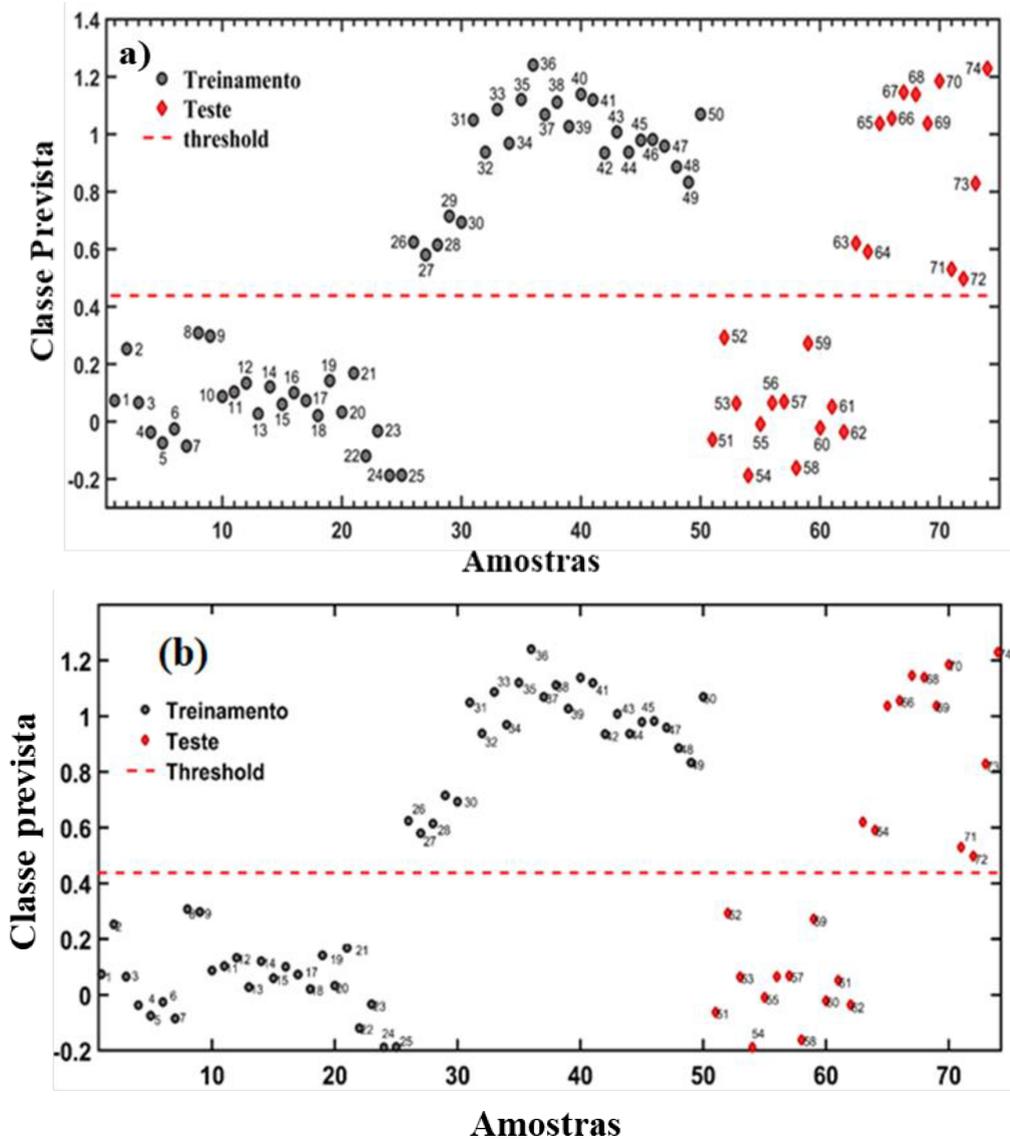
Figura 6.4 – Gráfico de *leverage* versus *Q residuals* com 95% de confiança para os modelos PLS-DA dos biodieseis metílicos: (a) Macaúba e (b) Níger.



Fonte: A autora, 2021.

A Figura 6.5 mostra clara separação das amostras de treinamento e teste para as classes discriminadas de B10 (classe de interesse) e BX (classe de não interesse) do biodiesel metílico de macaúba e níger na mistura com diesel. As amostras pertencentes à blenda B10, se apresentaram acima do limiar *Threshold*, e de BX abaixo desse limite, demonstrando a capacidade do modelo em classificar corretamente as amostras em suas respectivas classes, com uma previsão de 100%.

Figura 6.5 – Valores de classe estimados para conjuntos de treinamento (●) e de teste (◆) pelo modelo PLS-DA na discriminação de amostras de B10 e BX do biodiesel metílico: (a) Macaúba e (b) Níger.



Fonte: A autora, 2021.

O desempenho dos modelos construídos também foi avaliado com base na Tabela de Confusão, como expresso a partir dos parâmetros de classificação para o conjunto treinamento e de teste (Tabela 6.3).

De acordo com os resultados obtidos, observa-se tanto para o conjunto teste quanto para o conjunto treinamento, valores iguais a 100% na avaliação de sensibilidade, especificidade, precisão e eficiência. Assim, evidenciou-se a excelente capacidade do modelo PLS-DA desenvolvido, em classificar corretamente as amostras, o que também pode ser percebido pelo

Coefficiente de Correlação de Matthews (igual a + 1). Resultados similares também foram obtidos por outros autores. Mazivila e colaboradores (2015) verificaram precisão de 100% na discriminação da mistura biodiesel/ diesel (B5) usando espectroscopia no infravermelho médio e PLS-DA. Nespeca (2018) obteve coeficientes de correlação próximos a 99% e sensibilidade e especificidade de 100% ao detectar adulterantes na gasolina usando cromatografia em fase ultrarrápida e análise por PLS-DA.

Tabela 6.3 – Parâmetros de classificação dos modelos PLS-DA obtidos dos biodieseis metílico de Macaúba e Níger.

Parâmetros	Valores dos modelos PLS-DA	
	Conjunto treinamento	Conjunto Teste
Taxa de Falso Positivo e Falso Negativo (%)	0	0
Taxa de Verdadeiro Positivo e Negativo (%)	100	100
Sensibilidade (%)	100	100
Especificidade (%)	100	100
Precisão (%)	100	100
Eficiência (%)	100	100
Coefficiente de Matthew's	1	1

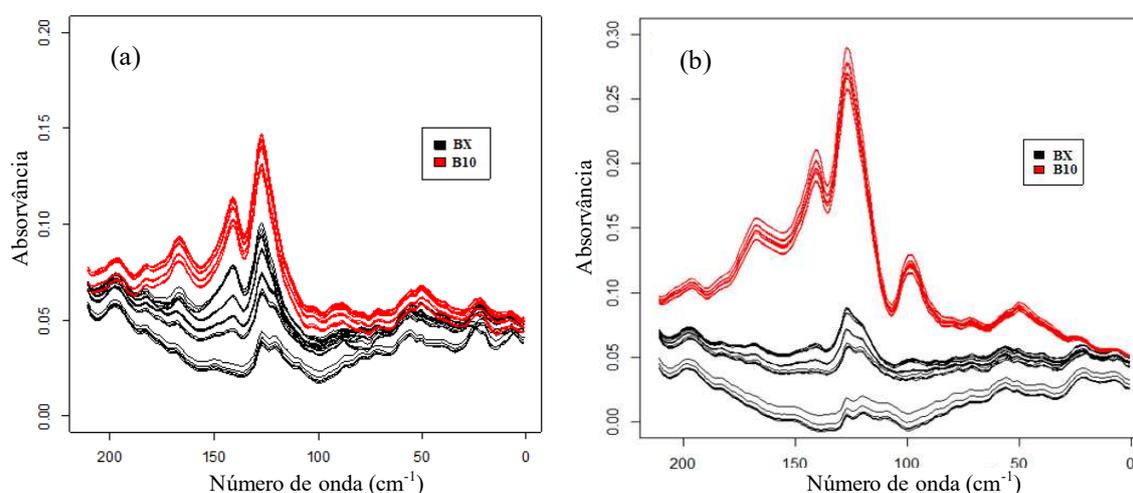
Fonte: A autora, 2021.

6.3 Modelos SVM

Na Figura 6.6, podemos observar o comportamento das amostras quanto ao número de ondas (cm^{-1}) em relação a absorvância entre as classes de B10 e BX do biodiesel metílico de níger e de macaúba para região de número de onda entre 0 e 200 cm^{-1} . Observa-se, nesta região, que para as amostras da classe B10, a taxa de absorvância é superior à taxa de absorvância para as amostras da classe BX. Essa diferença é bastante nítida principalmente quando se trata da macaúba, em que é possível observar uma “lacuna” entre as taxas de absorvância de B10 e BX.

No caso do biodiesel metílico de níger, esta discriminação entre a taxa de absorvância de B10 e BX é muito pequena, sendo que em alguns intervalos de número de ondas, existe uma intersecção entre as taxas para os dois casos em estudo.

Figura 6.6 – Relação entre número de onda (cm^{-1}) e absorvância entre as classes B10 e BX para o biodiesel metílico de (a) Níger e (b) Macaúba.



Fonte: A autora, 2021.

Para ajustar os modelos SVM para cada tipo de biodiesel, foi realizado um ajuste de parâmetros, o qual permitiu definir os hiperparâmetros C e γ através de diversas simulações, conforme nas Tabela 6.4 e Tabela 6.5.

Como o intuito é estimar um modelo que apresente a menor dispersão e erro possível com o menor custo e flexibilidade das curvas de *kernel*, para o modelo SVM do biodiesel de níger, adotamos como hiperparâmetro de custo (C) o valor 10 e para a flexibilidade de curva γ o valor 10^{-4} . Tais valores foram selecionados pois os hiperparâmetros proporcionaram uma performance com erro e dispersão iguais a zero, como podemos observar na Tabela 6.4.

Tabela 6.4 – Resultados de performance de acordo com os hiperparâmetros γ e C para os dados do biodiesel metílico de níger.

γ	C	Erro	Dispersão
10^{-6}	10	0,64	0,0843
10^{-5}	10	0,02	0,0632
10^{-4}	10	0,00	0,0000
10^{-3}	10	0,00	0,0000
10^{-6}	100	0,02	0,0632
10^{-5}	100	0,00	0,0000
10^{-4}	100	0,00	0,0000
10^{-3}	100	0,00	0,0000

Fonte: A autora, 2021.

Já para o modelo SVM do biodiesel metílico de macaúba acabamos por adotar como hiperparâmetro de custo (C) o valor 10 e para a flexibilidade de curva γ o valor 10^{-5} . Assim como no modelo SVM do biodiesel metílico de níger, tais valores foram selecionados pois os hiperparâmetros proporcionaram uma performance com erro e dispersão iguais a zero, como podemos observar na Tabela 6.5.

Tabela 6.5 – Resultados de performance de acordo com os hiperparâmetros γ e C para os dados do biodiesel metílico de macaúba.

γ	C	Erro	Dispersão
10^{-6}	10	0,66	0,0966
10^{-5}	10	0,00	0,0000
10^{-4}	10	0,00	0,0000
10^{-3}	10	0,00	0,0000
10^{-6}	100	0,00	0,0000
10^{-5}	100	0,00	0,0000
10^{-4}	100	0,00	0,0000
10^{-3}	100	0,00	0,0000

Fonte: A autora, 2021.

Definido os hiperparâmetros para cada um dos modelos de biodiesel, foi feita a busca do *kernel* que melhor se ajustou aos dados de cada um dos biodieseis e que apresentaram as melhores medidas de validação de modelo. As Tabela 6.6 e Tabela 6.7 apresentam os resultados obtidos do treino dos modelos SVM referente a cada *kernel*, para o biodiesel metílico de níger e macaúba respectivamente.

De acordo com os resultados apresentados pela Tabela 5.6, em relação aos modelos SVM do biodiesel metílico de níger, podemos verificar que o modelo com o *kernel* RBF foi o que apresentou melhor acurácia (0,92) e *F-Score* (0,91) em relação aos demais. Desta maneira, estas medidas, assim como os hiperparâmetros $C = 10$ e $\gamma = 10^{-4}$ citados é o que melhor discriminaram as classes B10 e BX para os dados do biodiesel metílico de níger. Já para os modelos obtidos do biodiesel metílico de macaúba, conforme os resultados apresentados pela Tabela 5.7, foram geradas excelentes medidas de acurácia e *F-Score* para todos os *Kernel* avaliados. No entanto, levando-se em consideração que as classes podem ser separadas por um hiperplano, nos auxilia na escolha do *kernel* Linear para estimar o modelo em questão. Esta escolha se deve, neste caso, além de o problema ser linearmente separável, a simplicidade para o modelo.

Tabela 6.6 – Treino dos modelos segundo diferentes *kernel* para os dados do biodiesel metílico de níger.

<i>Kernel</i>	Acurácia	<i>F- Score</i>
RBF	0,92	0,91
Sigmoidal	0,52	0,08
Polinomial (<i>grau 3</i>)	0,42	0,06
Linear	0,52	0,08

Fonte: A autora, 2021.

Tabela 6.7 – Treino dos modelos segundo diferentes *kernel* para os dados do biodiesel metílico de macaúba.

<i>Kernel</i>	Acurácia	<i>F- Score</i>
RBF	0,83	0,85
Sigmoidal	1,00	1,00
Polinomial (<i>grau 3</i>)	1,00	1,00
Linear	1,00	1,00

Fonte: A autora, 2021.

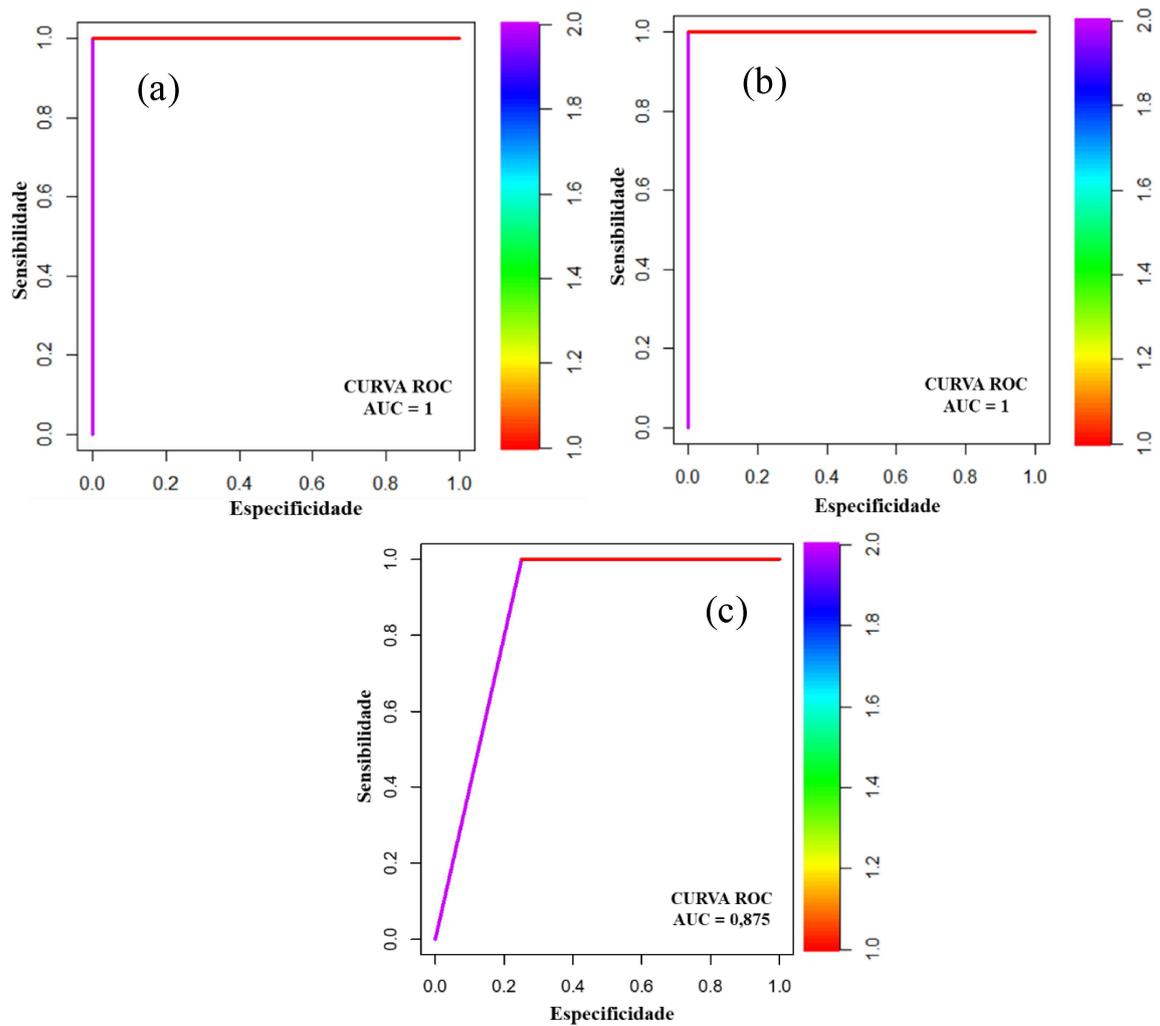
Para uma validação mais detalhada de cada um dos modelos ajustados conforme hiperparâmetros citados acima, as Figura 6.7 e Figura 6.8 apresentam as curvas ROC das amostras de treino, teste e validação para o biodiesel metílico níger e macaúba, respectivamente.

Conforme se observa na Figura 6.7 (a), em relação aos dados do biodiesel metílico de níger, a AUC para a amostra de treino foi 1, o que é um alto índice discriminativo e de certa maneira é esperado em dados amostrais de treino. Já na Figura 6.7 (b) também constatamos uma AUC de 1, o que nos indica um alto índice discriminativo para novos dados, visto que esses são os dados amostrais de teste. Por fim, em relação aos dados amostrais de validação do modelo, observados pela Figura 6.7 (c), identificamos uma AUC de 0,875 que também é um alto índice discriminativo e que reforça a qualidade do modelo ajustado para a classificação entre B10 e BX para o biodiesel metílico de níger.

Quanto aos dados da macaúba, verifica-se na Figura 6.8 (a), em relação à amostra de treino, a AUC igual a 1 que, semelhantemente como nos dados da níger, é um alto índice discriminativo e que é esperado por se tratar de dados de treino. Em relação aos dados de teste observamos uma AUC de 0,8333, apresentado pela Figura 6.8 (b), que também é um alto índice discriminativo, evidenciando que o modelo selecionado consegue classificar bem os dados entre as classes B10 e BX.

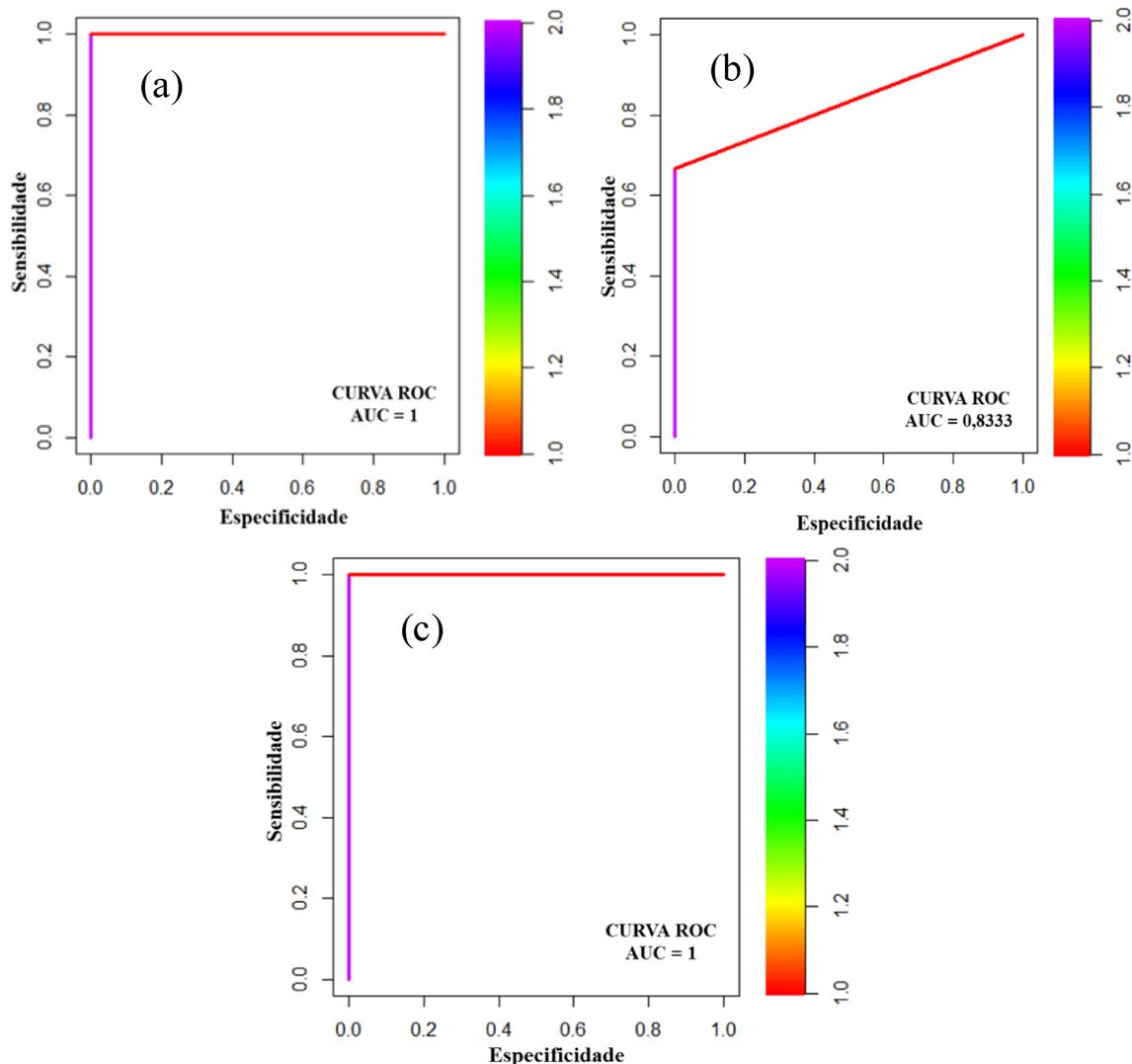
Quanto aos dados amostrais de validação observamos uma AUC de 1, como mostra a Figura 6.8 (c), o que evidencia um bom ajuste do modelo SVM com os hiperparâmetros estabelecidos anteriormente para os dados do biodiesel metílico de macaúba.

Figura 6.7 – Curvas ROC e AUC referentes ao modelo SVM de *kernel* RBF aplicado às amostras de (a) treino, (b) teste e (c) validação para o biodiesel metílico de níger.



Fonte: A autora, 2021.

Figura 6.8 – Curvas ROC e AUC referentes ao modelo SVM de *kernel* Linear aplicado às amostras de (a) treino, (b) teste e (c) validação para o biodiesel metílico de macaúba.



Fonte: A autora, 2021.

Por fim, temos a Tabela 6.8, a qual apresenta um resumo da matriz de confusão e as principais medidas de acurácia para cada um dos modelos ajustados, os quais foram aplicados às amostras utilizadas tanto no treino quanto na validação dos mesmos.

Tanto para os dados do biodiesel metílico de macaúba quanto para os do níger, podemos observar pela Tabela 6.8 que o modelo selecionado apresentou bons resultados quanto às medidas de qualidade discriminativa. A precisão e eficiência de classificação da SVM é superior a 70,0% nas análises de teste e validação, em ambos os modelos desenvolvidos. Os resultados também revelaram valor igual a 100% para os dados de treino, o que reforça a

capacidade de discriminação dos modelos SVM, garantindo bom desempenho de generalização.

Tabela 6.8 – Resumo da matriz de confusão e principais medidas de acurácia do modelo.

<i>Macaúba</i>	<i>Nº de amostras</i>	Parâmetros (%)		<i>Sensib.</i>	<i>Espec.</i>	<i>Precisão</i>	<i>Eficiência</i>
		<i>Falso Positivas</i>	<i>Falso Negativas</i>				
Treino	25	0	0	100	100	100	100
Teste	12	4	0	100	66,66	75,00	83,33
Validação	12	0	0	100	100	100	100
<i>Níger</i>							
Treino	25	0	0	100	100	100	100
Teste	12	0	0	100	100	100	100
Validação	12	0	3	75,00	100	100	87,50

Fonte: A autora, 2021.

Estes resultados podem ser comparados com as de outros trabalhos, também demonstrando a qualidade dos modelos desenvolvidos com a aplicação da SVM. Ruschel (2017) analisou conjuntos de amostras de gasolina e etanol, adulteradas e não-adulteradas, que foram caracterizados através da espectroscopia no infravermelho com transformada de Fourier com acessório de reflectância total atenuada (FTIR-ATR). Os dados foram aplicados para os modelos discriminativos PLS-DA e SVM. No geral, ambos os modelos de classificação apresentaram sensibilidade e especificidade máximas, demonstrando que as metodologias aplicadas podem ser ótimas alternativas para distinguir as amostras adulteradas das não-adulteradas para os combustíveis estudados (biodiesel metílico e etílico de óleo de soja, de óleo residual de fritura e de gordura vegetal hidrogenada).

Mahmodi, Mostafaei, Mirzaee-Ghaleh (2019) abordaram a detecção e classificação de biodiesel usando um nariz eletrônico (sensores). A análise de dados foi realizada por meio da análise discriminante linear e quadrática (LDA e DDA), bem como da SVM. Com base nos resultados, SVM, QDA e LDA tiveram precisões de classificação de 94,8%, 94,1% e 87,1%, respectivamente. Além disso, a precisão de discriminação e classificação da SVM foi maior (cerca de 95,4%) para as várias misturas de diesel e biodiesel. Os combustíveis analisados foram obtidos dos óleos de canola e de milho, que foram misturados com diferentes porcentagens de volume (2,5,10,20,80,100) do diesel de petróleo.

Mancini e colaboradores (2019) desenvolveram três métodos, incluindo PLS-DA, SVM e análise discriminante linear de análise de componente principal (PCA-LDA). Para tanto, três conjuntos de dados espectrais no infravermelho próximo foram investigados para avaliar a possibilidade de se obter informações sobre as características qualitativas dos biocombustíveis: (a) discriminação entre madeira tratada / virgem; (b) discriminação entre casca / madeira; e (c) discriminação entre biomassa herbácea / lenhosa. Com base nos resultados, todos os métodos puderam lidar com sucesso na tarefa de classificação, mas o SVM obteve o melhor desempenho de classificação. Os valores médios da proporção das previsões corretas do conjunto de teste independente dos três métodos avaliados foram PLS-DA (73,0%), PCA-LDA (78,0%) e SVM (83,0%).

7 CONCLUSÕES

7.1 Conclusões gerais

A combinação das técnicas analíticas/quimiométricas de espectroscopia MIR e dos métodos preditivos PLS-DA e SVM permitiu discriminar, de maneira rápida e eficiente, as amostras de B10 e BX dos biodieseis metílicos de macaúba e níger na mistura biodiesel/diesel, oferecendo a possibilidade de minimizar o uso de solventes e reagentes.

Quanto à análise dos modelos de PLS-DA, na avaliação dos erros quadráticos médios observou-se valores baixos (de até 0.22 %), com percentagens dentro da reprodutibilidade do método normatizado (Norma ABNT NBR 15568). Além disso, o RMSEP apresentou uma pequena diferença em relação ao RMSEC, demonstrando que não houve sobreajuste dos modelos.

Os critérios de desempenho obtidos para os modelos desenvolvidos, tais como sensibilidade, especificidade, precisão e eficiência, foram iguais a 100%. Conferiu-se também para o Coeficiente de Correlação de Matthew o grau de eficiência dos modelos construídos (igual a + 1). Em geral, teve-se demonstrada a excelente capacidade dos modelos PLS-DA em

classificar corretamente os teores de biodiesel metílico de macaúba e níger em suas respectivas blends.

O algoritmo de SVM também mostrou ser uma boa opção para discriminação entre as classes B10 e BX do biodiesel metílico de níger e macaúba. Para os modelos de níger, o *kernel* RBF foi o que proporcionou melhor acurácia, com os hiperparâmetros $C = 10$ e $\gamma = 10^{-4}$. Já para os modelos de macaúba, foram obtidas excelentes medidas de acurácia para todos os *Kernel* avaliados, com os hiperparâmetros definidos $C = 10$ e $\gamma = 10^{-5}$.

A precisão e eficiência de classificação da SVM foi superior a 70,0% nas análises de teste e validação, em ambos os modelos desenvolvidos. Os resultados também revelaram valor igual a 100% para os dados de treino, o que reforça a capacidade discriminativa dos modelos SVM, garantindo bom desempenho de generalização.

Como obtido neste estudo, as metodologias analíticas empregadas são viáveis, e poderiam ser utilizadas pelos órgãos de fiscalização para o controle da qualidade destes combustíveis.

7.2 Recomendações e perspectivas futuras

Avaliar a utilização dos algoritmos de classificação SVM e PLS-DA para misturas biodiesel/diesel, de diferentes matérias-primas.

Efetuar novos estudos com outras combinações (biodiesel metílico e etílico).

Fazer um estudo comparativo das metodologias utilizadas com outros modelos classificadores como Naive Bayes, Árvore de Decisão, Random Forest, Gradient Boosting.

REFERÊNCIAS

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS (Brasil). **Resolução ANP 15/2006**. Estabelece as especificações do óleo diesel de uso rodoviário, para comercialização pelos diversos agentes econômicos em todo o território nacional. Rio de Janeiro: ANP, 2006. Disponível em: www.anp.gov.br/petro/legis_qualidade.asp. Acesso em: 9 out. 2018.

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS (Brasil). **Boletim mensal do monitoramento da qualidade dos combustíveis líquidos automotivos**. Rio de Janeiro: ANP, 2016. Disponível em: <http://www.anp.gov.br/?pg=824801473367496451>. Acesso em: 8 mar. 2018.

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS (Brasil). **Anuário Estatístico Brasileiro do Petróleo e do Gás Natural e Biocombustíveis**. Rio de Janeiro: ANP, 2018. Disponível em: <http://www.anp.gov.br/publicacoes/anuario-estatistico/anuario-estatistico-2018>. Rio de Janeiro: ANP, 2018. Acesso em: 26 dez. 2018.

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS (Brasil). **Boletim mensal do monitoramento da qualidade dos combustíveis líquidos automotivos**. Rio de Janeiro: ANP, ano 19, set. 2020. Disponível em: <http://www.anp.gov.br/publicacoes/boletins-anp/boletim-de-monitoramento-da-qualidade-dos-combustiveis>. Acesso em: 14 nov. 2020.

ALMEIDA, M. R. *et al.* Discrimination between authentic and counterfeit bank notes using Raman spectroscopy and PLS-DA with uncertainty estimation. **Microchemical Journal**, New York, v. 109, p.170-177, 2013. <https://dx.doi.org/10.1016/j.microc.2012.03.006>.

ALVES, J. C. L.; POPPI, R. J. Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM). **Talanta**, Amsterdam, v. 104, n. 30, p. 155-161, 2013. <https://doi.org/10.1016/j.talanta.2012.11.033>

ALVES, J. C. L.; POPPI, R. J. Quantification of conventional and advanced biofuels contents in diesel fuel blends using near-infrared spectroscopy and multivariate calibration. **Fuel**, Amsterdam, v. 165, p. 379-388, 2016. Disponível em: https://www.cheric.org/research/tech/periodicals/doi.php?art_seq=1371208. Acesso em: 14 nov. 2020.

AMBROSANO, L. **Avaliação de plantas oleaginosas potenciais para cultivo na safrinha**. 2012. 81 f. Dissertação (Mestrado em Agronomia) – Universidade Federal de Lavras, Lavras, 2012.

AMORIN, L. H. C.; MARTINS, L. da S.; URBANO, A. Commitment between roughness and crystal lite size in the vanadium oxide thin film opto-electrochemical properties. **Materials Research**, São Carlos, v. 22, n. 1, p. 1-6, 2018. <https://doi.org/10.1590/1980-5373-mr-2018-0245>

AMERICAN SOCIETY FOR TESTING AND MATERIALS. **ASTM E1655-05 International**. Standard Practices for Infrared Multivariate Quantitative Analysis. West Conshohocken: ASTM, 2012. <https://doi.org/10.1520/E1655-05R12>

ANDERSEN, C. M.; BRO, R. Variable selection in regression - a tutorial. **Journal of Chemometrics**, New York, v. 24, p. 728-737, 2010. <https://doi.org/10.1002/cem.1360>. <https://doi.org/10.1002/cem.1360>

ANGHEBEN, A. A. **Estudo sobre a combustão de blendas de hidrogênio e diesel em motores ciclo diesel**. 2013. 40 f. Trabalho de Conclusão de Curso (Graduação em Tecnologia de Biocombustíveis) – Universidade Federal do Paraná, Rio Grande do Sul, 2013.

ANTONIASSI, R. *et al.* Impact of genotype on fatty acid profile, oil content and nutritional value of the sweet fruits of *Acrocomia aculeata*. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 42, n. 6, p. 1-10, 2020. <https://doi.org/10.1590/0100-29452020796>

ARANSIOLA, E. F. *et al.* A review of current technology for biodiesel production: State of the art. **Biomass and Bioenergy**, Oxford, v. 61, p. 276-297, 2014. <https://doi.org/10.1016/j.biombioe.2013.11.014>

ARINE, B. B. R. **Melhoramento do índice de detecções na espectrometria gama em amostras ambientais usando inteligência artificial**. 2017. 151 f. Dissertação (Mestrado em Ciências Ambientais) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Sorocaba, 2017.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 15568**: Biodiesel: determination of biodiesel content in diesel fuel oil by mid infrared spectroscopy. Rio de Janeiro: ABNT, 2008. Disponível em: <https://www.target.com.br/produtos/normas-tecnicas/40641/nbr15568-biodieseldeterminacao-do-teor-de-biodiesel-em-oleo-diesel-por-espectroscopia-na-regiao-do-infravermelhomedio>. Acesso em: 14 jan. 2018.

ATABANI, A. E. *et al.* Non-edible vegetable oils: A critical evaluation of oil extraction, fatty acid compositions, biodiesel production, characteristics, engine performance and emissions production. **Renewable and Sustainable Energy Reviews**, [s. l.], v. 18, p. 211-245, 2013. <https://doi.org/10.1016/j.rser.2012.10.013>

AZEVEDO, A. N. G.; LIMA, B. G. A. Biocombustíveis: desenvolvimento e inserção internacional. **Revista Direito Ambiental e Sociedade**, Caxias do Sul, v. 6, n. 1, p. 77-100, 2016. Disponível em: <http://www.ucs.br/etc/revistas/index.php/direitoambiental/article/view/2693/2513>. Acesso em: 18 fev. 2018.

AZEVÊDO, L. A. **Métodos estatísticos em aprendizado de máquinas para problemas de classificação**. 2018. 131 f. Dissertação (Mestrado em Estatística) – Universidade de Brasília, Brasília, DF, 2018.

BAIRD, Z. S.; OJA, V. Predicting fuel properties using chemometrics: a review and an extension to temperature dependent physical properties by using infrared spectroscopy to predict density. **Chemometrics and Intelligent Laboratory Systems**, [s. l.], v. 158, p. 41-47, 2016. <https://doi.org/10.1016/j.aca.2013.07.047>

BALLABIO, D.; CONSONNI, V. Classification Tools in Chemistry. Part 1: Linear Models. PLS-DA. **Analytical Methods**, [s.l.], v.5, p. 3790-3798, 2013. <https://doi.org/10.1039/c3ay40582f>

BARBOSA, L. C. D. A. **Espectrometria no infravermelho**: na caracterização de compostos orgânicos. Viçosa: Editora UFV, 2008.

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

BATISTA, G. E. A. P. **Pré-processamento de dados em aprendizado de máquina supervisionado**. 2003. 232 f. Tese (Doutorado em Ciências da Computação e Matemática Computacional) – Instituto de Ciências Matemática e Computacional, Universidade de São Paulo, São Paulo, 2003.

BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. **Journal of Chemometrics**: a practical guide. New York: J. Wiley, 1998. <https://doi.org/10.1021/ci980426l>

BERTON, L. H. C. **Avaliação de populações naturais, estimativas de parâmetros genéticos e seleção de genótipos elite de macaúba (*Acrocomia aculeata*)**. 2013. 163 f. Tese (Doutorado em Agricultura Tropical e Subtropical) – Instituto Agrônômico, Campinas, 2013.

BEZERRA, E. Introdução à aprendizagem profunda. In: OGASAWARA, V. (Ed.). **Tópicos em gerenciamento de dados e informações**. 1rd. ed. [S. l.: s. n.]: 2016. v. 3, p. 57-86. Disponível em: <https://www.researchgate.net/publication/309321510>. Acesso em: 30 ago. 2019.

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

BORSATO, D.; GALÃO, O, F.; MOREIRA, I. **Combustíveis fósseis**: carvão e petróleo. Londrina: Editora Eduel, 2009.

BONESSO, D. **Estimação dos parâmetros do kernel em um classificador svm na classificação de imagens hiperespectrais em uma abordagem multiclasse**. 2013.108 f.

Dissertação (Mestrado em Sensoriamento Remoto) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.

BOTELHO, B. G. **Desenvolvimento e validação de métodos para análise direta de alimentos usando ferramentas quimiométricas, espectroscopia no infravermelho e imagens digitais**. 2014. 136 f. Tese (Doutorado em Ciências- Química) – Universidade Federal de Minas Gerais, Belo Horizonte, 2014.

BOTTEGA, S. P. **Desempenho agrônômico do Niger em função de época de semeadura, adubação e arranjo de plantas**. 2012. 72 f. Tese (Doutorado em Agronomia) – Universidade Federal da Grande Dourados, Mato Grosso do Sul, 2012.

BURGES, C. J. A tutorial on support vector machines for pattern recognition. **Data Mining and Knowledge Discovery**, [s. l.], v. 2, n. 121, p. 121-167, 1998. Disponível em: <http://staff.icar.cnr.it/manco/Teaching/2005/datamining/articoli/Burges98.pdf>. Acesso em: 24 maio 2018.

CALDATO, E. B. **Avaliação da support vector machine na classificação de fácies no campo de Peregrino, Bacia de Campos-RJ, Brasil**. 2017. 59 f. Dissertação (Mestrado em Geociências) – Universidade Estadual de Campinas, Campinas, 2017.

CAMPOS RODRÍGUEZ, C. E. *et al.* Exergetic and economic comparison of ORC and Kalina cycle for low temperature enhanced geothermal system in Brazil. **Applied Thermal Engineering**, [s. l.], v. 52, n. 1, p. 109-119, abr. 2013. <https://doi.org/10.1016/j.applthermaleng.2012.11.012>.

CANCELA, B. R.; CESCÓN, L. S. Titulometric Methodology for Determining Biodiesel Content in Solutions with Petroleum Diesel. **Perspectivas da Ciência e Tecnologia**, [s. l.], v. 12, p.138-153, 2020. <http://dx.doi.org/10.22407/1984-5693>.

CARDOSO, A. *et al.* Opportunities and challenges for sustainable production of *Acrocomia aculeata* through agroforestry systems. **Industrial Crops and Products**, [s. l.], v. 107, p. 573-580, 2017. <https://doi.org/10.1016/j.indcrop.2017.04.023>

CARDOSO, A. *et al.* Macauba extractivism at the Cariri Cearense region: commercialization and opportunities. **Brazilian Journal of Development**, Curitiba, v. 6, n. 5, p. 25261-25279, 2020. <https://doi.org/10.34117/bjdv6n5-108>

CARNEIRO, M.A.C. *et al.* Produção de fitomassa de diferentes espécies de cobertura e suas alterações na atividade microbiana de solo de cerrado. **Revista Bragantia**, Campinas, v. 67, n. 2, p. 455-462, 2008. Disponível em: <https://www.scielo.br/pdf/brag/v67n2/a21v67n2.pdf>. Acesso em: 15 maio 2018.

CARVALHO, V. P. **Aprendizado de máquina e estatístico na discriminação de populações na presença de matrizes de covariâncias heterogêneas e vetores aleatórios não normais multivariados**. 2019. 47 f. Tese (Doutorado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa, 2019.

CASTRO, L. C.; BRAGA P. B. Aprendizado supervisionado com conjuntos de dados desbalanceados. **Sba: Controle & Automação Sociedade Brasileira de Automática**, Campinas, v. 22, n. 5, p. 441-466, set./out. 2011. <http://doi.org/10.1590/S0103-17592011000500002>.

CÉSAR, A. D. S. *et al.* The prospects of using *Acrocomia aculeata* (macaúba) a non-edible biodiesel feedstock in Brazil. **Renewable and Sustainable Energy**, [*s. l.*], v. 49, p. 1213–1220, 2015. <https://doi.org/10.1016/j.rser.2015.04.125>.

CICONINI, G. **Caracterização de frutos e óleo de polpa de macaúba dos biomas Cerrado e Pantanal do estado de Mato Grosso do Sul, Brasil**. 2012. 150f. Dissertação (Mestrado em Biotecnologia) – Universidade Católica Dom Bosco, Campo Grande, 2012.

CICONINI, G. *et al.* Biometry and oil contents of *Acrocomia aculeata* fruits from the Cerrados and Pantanal biomes in Mato Grosso do Sul, Brazil. **Industrial Crops and Products**, Oxford, v. 45, p. 208-214, 2013. <https://doi.org/10.1016/j.indcrop.2012.12.008>

COLOMBO, A. C. *et al.* Macauba: a promising tropical palm for the production of vegetable oil. **Oilseeds & fats Crops and Lipids**, Montrouge, v. 25, n. 108, p. 1-9, 2018. <https://doi.org/10.1051/ocl/2017038>

CONCEIÇÃO, L. D. H. C. S. *et al.* Macaúba. *In*: LOPES, R.; OLIVEIRA, M. S. P.; CAVALLARI, M. M.; BARBIERI, R. L.; CONCEIÇÃO, L. H. C. S. C. (Ed.). **Palmeiras Nativas do Brasil**. Brasília, DF: Embrapa, 2015. p. 269-305.

CONSELHO NACIONAL DE POLÍTICA ENERGÉTICA (Brasil). Resolução CNPE nº 11, de 14 de dezembro de 2016. Dispõe sobre adição obrigatória de biodiesel ao óleo diesel vendido ao consumidor final, em qualquer parte do território nacional. **Diário Oficial da**

União, Brasília, DF, 1 mar. 2017. Disponível em: http://www.mme.gov.br/documents/10584/3201726/Resolu%C3%A7%C3%A3o_CNPE_11_Biodiesel.pdf/db551997-10a9-4ce6-96957479b41ead2f. Acesso em: 18 mar. 2019.

CONSELHO NACIONAL DE POLÍTICA ENERGÉTICA (Brasil). Resolução CNPE nº 16, de 29 de outubro de 2018. Dispõe sobre a evolução da adição obrigatória de biodiesel ao óleo diesel vendido ao consumidor final, em qualquer parte do território nacional. **Diário Oficial da União**, Brasília, DF, 8 nov. 2018. Disponível em: www.mme.gov.br/documents/10584/71068545/Resolucao_16_CNPE_29-10-18.pdf/. Acesso em: 18 mar. 2019.

COPPEL, I. A. F. *et al.* Life cycle analysis of macauba palm cultivation: A promising crop for biofuel production. **Industrial Crops and Products**, [s. l.], v. 125, p. 556-566, 2018. <https://doi.org/10.1016/j.indcrop.2018.09.036>.

CORDEIRO, C. S. *et al.* Catalisadores heterogêneos para a produção de monoésteres graxos (biodiesel). **Química Nova**, São Paulo, v. 34, n. 3, p. 477-486, 2011. <https://doi.org/10.1590/S0100-40422011000300021>

CORTES, C.; VAPNIK, V. Support-Vector Networks. **Machine Learning**, Boston, v. 20, n. 3, p. 273-297, 1995.

COSTA, E. *et al.* Seedlings of *Acrocomia aculeata* in different substrates and protected environments. **Engenharia Agrícola**, Jaboticabal, v. 34, n. 3, p. 395-404, 2014.

CRISTIANINI, N.; SHAW-TAYLOR, J. **An Introduction to Support Vector Machines and Other Kernel-based Learning Methods**. [S. l.]: Cambridge University Press, 2000. <https://doi.org/10.1017/CBO9780511801389>.

ELUM, Z. A.; MOMODU, A. S. Climate change mitigation and renewable energy for sustainable development in Nigeria: A discourse approach. **Renewable and Sustainable Energy Reviews**, [s. l.], v. 76, p. 72-80, Sept. 2017. <https://doi.org/10.1016/j.rser.2017.03.040>.

EMPRESA DE PESQUISA ENERGÉTICA. **Plano decenal de expansão de energia 2029**. Rio de Janeiro: EPE, 2020. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/Documents/PDE%202029.pdf>. Acesso em: 16 nov. 2020.

EMPRESA DE PESQUISA ENERGÉTICA. **Análise de conjuntura dos biocombustíveis 2019**. Rio de Janeiro: EPE, 2019. Disponível em: https://www.epe.gov.br/sites-pt/publicacoesdadosabertos/publicacoes/PublicacoesArquivos/publicacao489/Analise_de_Conjuntura_Ano_2019.pdf. Acesso em: 12 de nov. 2020.

EMPRESA DE PESQUISA ENERGÉTICA. **Balço energético nacional 2017**. Rio de Janeiro: EPE, 2017. Disponível em: https://ben.epe.gov.br/downloads/Relatorio_Final_BEN_2017.pdf. Acesso em: 14 set. 2018.

ERKINBAEV, C.; HENDERSON, K.; PALIWAL, J. Discrimination of gluten free oats from contaminants using near infrared hyperspectral imaging technique. **Food Control**, Guildford, v. 80, p. 197-203, 2017. <https://doi.org/10.1016/j.foodcont.2017.04.036>

EVARISTO, A. B. *et al.* Actual and putative potentials of macauba palm as feedstock for solid biofuel production from residues. **Biomass and Bioenergy**, Oxford, v. 85, p. 18-24, 2016. <https://doi.org/10.1016/j.biombioe.2015.11.024>

FAKHOURI, S. N.; HUDAIB, A.; FAKHOURI, H. N. Enhanced Optimizer Algorithm and its Application to Software Testing. **Journal of Experimental & Theoretical Artificial Intelligence**, v. 32, n. 6, p. 1-45, 2019. <https://doi.org/10.1080/0952813X.2019.1694591>

FAVARO, S. *et al.* **Armazenamento e processamento da macaúba**: contribuições para manutenção da qualidade e aumento do rendimento de óleo da polpa. Brasília, DF: Embrapa Agropecuária, 2018. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/186831/1/BPD-16-CNPAE.pdf>. Acesso em: 9 jun. 2019.

FATIMA, A. *et al.* Quality and chemistry of niger seeds (*Guizotia abyssinica*) grown in the United States. **Industrial Crops and Products**, Oxford, v. 75, p. 40-42, 2015. <https://doi.org/10.1016/j.indcrop.2015.07.031>

FERNÁNDEZ, C. M. *et al.* Supercritical extraction and fractionation of *Jatropha curcas* L. oil for biodiesel production. **The Journal of Supercritical Fluids**, Amsterdam, v. 97, p. 100-106, 2015.

FERREIRA, M. M. C. *et al.* Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, São Paulo, v. 22, p. 1678-7064, 1999. <http://doi.org/10.1590/S0100-40421999000500016>

FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Campinas: Universidade Estadual de Campinas, 2015. <https://doi.org/10.7476/9788526814714>

FLETCHER, T. **Support vector machines explained: Introductory course**. London: University College London, 2009. (Technical internal report). Disponível em: <http://www.tristanfletcher.co.uk/SVM%20Explained.pdf>. Acesso em: 23 set. 2019.

FILGUEIRAS, P. R. *et al.* Determination of Saturates, Aromatics, and Polars in Crude Oil by ¹³C NMR and Support Vector Regression with Variable Selection by Genetic Algorithm. **Energy Fuels**, [s. l.], v. 30, n. 3, p. 1972-1978, 2016. <https://doi.org/10.1021/acs.energyfuels.5b02377>

FILGUEIRAS, P. R. **Regressão por vetores de suporte aplicado na determinação de propriedades físico-químicas de petróleo e biocombustíveis**. 2014. 115 f. Tese (Doutorado em Química) – Universidade Estadual de Campinas, Campinas, 2014.

FRANCHINI, R. G. **Estudos de sistemas de produção, envolvendo rotação de culturas, com ênfase em culturas para produção de grãos, de óleo e de torta em plantio direto**. 2014. 77 f. Tese (Doutorado em Agronomia) – Universidade Federal da Grande Dourados, Dourados, 2014.

GEISSDOERFER, M.; VLADIMIROVA, D.; EVANS, S. Sustainable business model innovation: A review. **Journal of Cleaner Production**, Amsterdam, v. 198, p. 401–416, 2018. <https://doi.org/10.1016/j.jclepro.2018.06.240>

GELETA, M; STYMNE, S; BRYNGELSSON, T. Phylogenetics and taxonomic delimitation of the genus *Guizotia* (Asteraceae) based on sequences derived from various chloroplast DNA regions. **Plant Systematics and Evolution**, New York, v. 1, n. 289, p. 77-89, 2010. <https://doi.org/10.1007/s00606-010-0334-x>

GELETA, M.; STYMNE, S.; BRYNGELSSON, T. Variation and inheritance of oil content and fatty acid composition in niger (*Guizotia abyssinica*). **Journal of Food Composition and Analysis**, Amsterdam, v. 24, p. 995-1003, 2011. <http://dx.doi.org/10.1016/j.jfca.2010.12.010>

GELINSKI, E, K. *et al.* Biodiesel Synthesis Monitoring using Near Infrared Spectroscopy. **The Open Chemical Engineering Journal**, [S. l.], v. 12, p. 95-110, 2018. <https://doi.org/10.2174/1874123101812010095>.

GETINET, A.; SHARMA, S. M. **Níger**. *Guizotia abyssinica* (L. f.) Cass Roma: International Plant Genetic Resources Institute: International Usina Genetic Resources Institute, 1996. (Promoting the conservation and use of underutilized and neglected crops, 5). Disponível em: <https://cgspace.cgiar.org/bitstream/handle/10568/104987/136.pdf?sequence=3&isAllowed=y>. Acesso em: 19 out. 2019.

GORDIN, C. R. B.; SCALON, S. P. Q.; MASETTO, T. E. Disponibilidade hídrica do substrato e teor de água da semente na germinação de níger. **Pesquisa Agropecuária Tropical**, Goiânia, v. 45, n. 3, p. 312-318, 2015. Disponível em: <https://www.scielo.br/pdf/pat/v45n3/1517-6398-pat-45-03-0312.pdf>. Acesso em: 18 set. 2018.

GUARIEIRO, L, L, N. *et al.* Determination of biodiesel percentage in biodiesel: diesel mixtures using mid-infrared spectroscopy. **Química Nova**, São Paulo, v.31, n.2, p. 421-426, 2008. <https://doi.org/10.1590/S0100-40422008000200041>

HASSAN, N. M. S.; RASUL, M. G.; HARCH, C. A. Modelling and Experimental Investigation of Engine Performance and Emissions Fuelled with Biodiesel Produced from Australian Beauty Leaf Tree. **Fuel**, Amsterdam, n. 150, p. 625-635, 2015.

HOCHSCHEIDT, B. D. *et al.* Neutralização ou Esterificação como Técnicas de Melhoria dos Insumos para Produção de Biodiesel. **Revista Virtual Química**, São Paulo, v. 12, n. 2, p. 325-334, mar./abr. 2020. <https://doi.org/10.21577/1984-6835.20200025>

HUANG, D.; ZHOU, H.; LIN, L. Biodiesel: an alternative to conventional fuel. **Energy Procedia**, [s. l.], v. 16, p. 1874-1885, 2012. <https://doi.org/10.1016/j.egypro.2012.01.287>

IHA, O. K. *et al.* . Physicochemical properties of *Syagrus coronata* and *Acrocomia aculeata* oils for biofuel production. **Industrial Crops and Products**, [s. l.], v. 62, p. 318-322, Dec. 2014. <https://doi.org/10.1016/j.indcrop.2014.09.003>

INTERNATIONAL ENERGY AGENCY. **Task 42: Biobased Chemicals: a 2020 update**. [S.l.]: IEA, 2020.

JOTHIRAMALINGAM, R.; WANG, M. K. Review of Recent Developments in Solid Acid, Base, and Enzyme Catalysts (Heterogeneous) for Biodiesel Production via Transesterification. **Industrial & Engineering Chemistry Research**, Washington,, v. 48, p. 6162-6172, 2009. <https://doi.org/10.1021/ie801872t>

JUNQUEIRA, N. T. V. *et al.* **Caracterização de populações naturais de macaúba e avaliação do potencial produtivo**. Brasília, DF: Embrapa Agropecuária, 2019. Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1122305/1/Doc-354-biblioteca.pdf> . Acesso em: 18 nov. 2020.

KOWALCZYK, A. **Support Vector Machines Succinctly**. Morrisville: Syncfusion, 2017.. Disponível em: https://s3.amazonaws.com/ebooks.syncfusion.com/downloads/support_vector_machines_succinctly/support_vector_machines_succinctly.pdf?AWSAccessKeyId=AKIAWH6GYCX3T2N7V5N5&Expires=1608944863&Signature=6EmeBZg2CEiZ1yd4ciKURjdiTOY%3. Acesso em: 30 set. 2019.

KUHN, M.; JOHNSON, K. **Applied predictive Modeling**. New York: Springer, 2013. <http://doi.org/10.1007/978-1-4614-6849-3>.

LEVERMANN, R. A.; SOUZA, J. P. M. Óleo de palma: o crescimento da indústria global. **Agroanalysis**, Rio de Janeiro, v. 8, p. 13-15, 2014. Disponível em: http://www.agroanalysis.com.br/materia_detalhe.php?idMateria=1659. Acesso em: 17 jun. 2018.

LI, D. *et al.* Prediction of soil organic matter content in a litchi orchard of South China using spectral indices. **Soil & Tillage Research**, Amsterdam, v. 123, p. 78-86, 2012. <https://doi.org/10.1016/j.still.2012.03.013>

- LIMA, R. A. G. **Utilização de sistemas inteligentes para classificação de tráfego malicioso**. 80 f. 2014. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal Viçosa, Viçosa, 2014.
- LIMA, N. E. *et al.* A review of the palm genus *Acrocomia*: Neotropical green gold. **Organisms Diversity & Evolution**, Jena, v. 18, n. 2, p. 151-161, 2018. <https://doi.org/10.1007/s13127-018-0362-x>
- LOBATO, B. **Macaúba é matéria-prima promissora para biodiesel**. [Brasília, DF]: Embrapa, 2014. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/2329636/macauba-e-materia-prima-promissora-para-biodiesel>. Acesso em: 16 dez. 2019.
- LÔBO, I. P.; FERREIRA, S. L. C.; CRUZ, R. S. D. Biodiesel: parâmetros de qualidade e métodos analíticos. **Química Nova**, São Paulo, v. 32, n. 6, p. 1596-1608, 2009. <https://doi.org/10.1590/S0100-40422009000600044>
- LOPES, M. V. S. **Aplicação de classificadores para determinação de conformidade de biodiesel**. 82 f. 2017. Tese (Doutorado em Engenharia de Eletricidade) – Universidade Federal do Maranhão, Maranhão, 2017.
- LORENA, A. C.; CARVALHO, A. C. D. Evolutionary tuning of svm parameter values in multiclass problems. **Neurocomputing**, [s. l.], v. 71, n. 16-18, p. 3326-3334, 2008. <https://doi.org/10.1016/j.neucom.2008.01.031>.
- LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, Porto Alegre, v. 14, n. 2, p. 43-67, 2007. <https://doi.org/10.22456/2175-2745.5690>.
- LU, D.; ZHAO, J.; LENG, Y.; JIANG, P.; ZHANG, C. Novel porous and hydrophobic POSSionic liquid polymeric hybrid as highly efficient solid acid catalyst for synthesis of oleate. **Catalysis Communications**, Amsterdam, v. 83, p. 27-30, 2016. <http://dx.doi.org/10.1016/j.catcom.2016.05.004>
- MAHMODI, M.; MOSTAFAEI, M.; MIRZAEI-GHALEH, E. Detection and classification of diesel-biodiesel blends by LDA, QDA and SVM approaches using an electronic nose, **Fuel**, Amsterdam, v. 258, p.1-10, 2019. <https://doi.org/10.1016/j.fuel.2019.116114>

MALIK, A.; TAULER, R. Extension and application of multivariate curve resolution-alternating least squares to four-way quadrilinear data-obtained in the investigation of pollution patterns on Yamuna River, India-A case study. **Analytica Chimica Acta**, Amsterdam, v. 794, p. 20-28, 2013. <https://doi.org/10.1016/j.aca.2013.07.047>

MANCINI, M. *et al.*, Characterization of cardoon accessions as feedstock for biodiesel production. **Fuel**, Amsterdam, V. 235, p. 1287-1293, 2019. <https://doi.org/10.1016/j.fuel.2018.08.123>

MARCHETTI, J. M.; MIGUEL, V. U.; ERRAZU, A. F. Possible methods for biodiesel production. **Renewable and Sustainable Energy Reviews**, [*s. l.*], v. 11, p. 1300-1311, 2007. <https://doi.org/10.1016/j.rser.2005.08.006>

MARÇO, P. H. *et al.* R. Resolução multivariada de curvas com mínimos quadrados alternantes: descrição, funcionamento e aplicações. **Química Nova**, São Paulo, v. 37, p. 1525-1532, 2014.

MARTINS, E. R. F.; CRUZ, N. D. Pesquisas em desenvolvimento com Pinhão-manso paraguaio no Instituto Agronômico. **O Agrônomo**, Campinas, v. 37, n. 2, p. 109-113, 1985.

MARX, S. Glycerol-free biodiesel production through transesterification: a review. **Fuel Processing Technology**, Amsterdam, v. 151, p. 139-147, Oct. 2016. <https://doi.org/10.1016/j.fuproc.2016.05.033>

MASSART, D. L. *et al.* (ed.). Process Modelling and Sampling. *In*: MASSART, D. L. *et al.* (ed.). **Data Handling in Science and Technology**: Part A, Amsterdam: Ed. Elsevier, 1998. cap. 20. Disponível em: <https://www.sciencedirect.com/bookseries/data-handling-in-science-and-technology/vol/20/part/PA>. Acesso em: 22 jun. 2019.

MAZIVILA, S. J. *et al.* Fast Detection of Adulterants/Contaminants in Biodiesel/Diesel Blend (B5) Employing Mid-Infrared Spectroscopy and PLS-DA. **Energy Fuels**, New York, v. 29, p. 227-232, 2015.

MENEGHETTI, S. M. P.; MENEGHETTI, M. R.; BRITO, Y. C. A. Reação de transesterificação, algumas aplicações e obtenção de biodiesel. **Revista Virtual de Química**, São Paulo, v. 5, n. 1, p. 63-73, 2013. Disponível em: <http://www.uff.br/rvq>. Acesso em: 16 out. 2019.

MUELLER, D. F, *et al.* Fourier Transform Infrared Spectroscopy (FTIR) and multivariate analysis for identification of different vegetable oils used in biodiesel production. **Sensors**, [s. l.], v. 13, n. 4, p. 4258-4271. <http://dx.doi.org/10.3390/s130404258>.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of Machine Learning**. 2nd ed. [S. l.]: MIT Press, 2012.

NABI, M. N.; AKHTER, M. S.; SHAHADAT, M. M. Z. Improvement of engine emissions with conventional diesel fuel and diesel-biodiesel blends. **Bioresource Technology**, Amsterdam, v. 97, p. 372-378, 2006.

NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. **A user-friendly guide to Multivariate Calibration and Classification**. Chichester: NIR Publications, 2004. <https://doi.org/10.1255/nirn.658>

NAVARRO-DÍAZ, H. J. *et al.* Macauba oil as an alternative feedstock for biodiesel: Characterization and ester conversion by the supercritical method. **Journal Supercritical Fluids**, [s. l.], v. 93, p.130-137, 2014. <https://doi.org/10.1016/j.supflu.2013.11.008>

NESPECA, M. G. *et al.* Rapid and Simultaneous Prediction of Eight Diesel Quality Parameters through ATRFTIR Analysis. **Journal of Analytical Methods in Chemistry**, New York, v. 215, n. 211, p. 1-10, 2018. <https://doi.org/10.1155/2018/1795624>

NOBRE, D. A. C. *et al.* Macaúba: palmeira de extração sustentável para biocombustível. **Colloquium Agrariae**, [s. l.], v. 10, p. 92-105, 2014. <https://doi.org/10.5747/ca.2014.v10.n2.a112>

NUCCI, S. M. **Desenvolvimento, caracterização e análise da utilidade de marcadores microssatélites em genética de população de macaúba**. 2007. Dissertação (Mestrado em agricultura tropical e subtropical) – Instituto Agronômico, Campinas, 2007.

NUNES, A. A. *et al.* Good practices of harvest and processing provide high quality Macauba pulp oil. **European Journal of Lipid Science and Technology**, Weinheim,, v. 117, n. 12, p. 2036-2043, 2015. <https://doi.org/10.1002/ejlt.201400577>

PATACA, L. C. M. **Análises de mel e própolis utilizando métodos quimiométricos de classificação e calibração**. 2006. 97 f. Tese (Doutorado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2006.

PAVIA, D. L. *et al.* **Introdução à espectroscopia**. Tradução da 4. ed. norte-americana. São Paulo: [s. n.], 2010.

PELLUCCI, P. R. S.; PAULA, R. R. de; SILVA, W. B. de O.; LADEIRA, A. P. Utilização de técnicas de aprendizado de máquina no reconhecimento de entidades nomeadas no português. **E-xacta**, Belo Horizonte, v. 4, n. 1, p. 73-81, 2011. Disponível em: www.unibh.br/revistas/exata/. Acesso em: 4 maio 2019.

PEREIRA, J. M. *et al.* Fast Discrimination of Milk Contaminated with Salmonella sp. Via Near-Infrared Spectroscopy. **Food Analytical Methods**, [s. l.], v. 11, p. 1878-885, 2018. <https://doi.org/10.1007/s12161-017-1090-0>

PIMENTA, T. V. **Metodologias de obtenção e caracterização dos óleos do fruto da macaúba com qualidade alimentícia: da coleta á utilização**. 2010. 114 f. Dissertação (Mestrado em Engenharia Química) – Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

RAMADAN, M. F.; MÖRSEL, J. T. Proximate neutral lipid composition of Niger (*Guizotia abyssinica* Cass.) seed. **Czech Journal Food Science**, [s. l.], v. 20, p. 98-104, 2002. <https://doi.org/10.17221/3518-CJFS>

RAMADAN, M. F.; MÖRSEL, J. T. Analysis of glycolipids from black cumin (*Nigella sativa* L.), coriander (*Coriandrum sativum* L.) and Niger (*Guizotia abyssinica* Cass.) oilseeds. **Food Chemistry**, Amsterdam, v. 80, p. 197-204, 2003. [https://doi.org/10.1016/s0308-8146\(02\)00254-6](https://doi.org/10.1016/s0308-8146(02)00254-6)

RAMADAN, M. F.; MÖRSEL, J. T. Oxidative stability of black cumin (*Nigella sativa* L.), coriander (*Coriandrum sativum* L.) and niger (*Guizotia abyssinica* Cass.) crude seed oils

upon stripping. **European Journal of Lipid Science and Technology**, Weinheim, v. 106, p. 35-43, 2004. <https://doi.org/10.1002/ejlt.200300895>

RENEWABLE FUELS ASSOCIATION. **Statistics**. Washington, DC: RFA. Disponível em: <https://ethanolrfa.org/statistics/annual-ethanol-production>. Acesso em: 19 jun. 2019.

REZENDE, V. C. **Metodologia para a classificação automática de doenças em plantas utilizando redes neurais convolucionais**. 2019. 81 f. Dissertação (Mestrado em Engenharia Elétrica) – Instituto de Tecnologia, Universidade Federal do Pará, Belém, 2019.

RIBEIRO, R. A. Estado e biocombustíveis no Brasil: uma parceria para o desenvolvimento? **Ponto de Vista**: perspectivas sobre o desenvolvimento, Rio de Janeiro, n. 8, p. 1-31, 2013. Disponível em: <http://neic.iesp.uerj.br/pontodevista/pdf/Ponto%20de%20Vista%20N.8,%20agosto%202013.pdf>. Acesso em: 5 maio 2018.

RICHARD, R.; DUBREUIL, B.; THIEBAUD-ROUX, D.; PRAT, L. On-line monitoring of the transesterification reaction carried out in microreactors using near infrared spectroscopy. **Fuel**, Amsterdam, v. 4, p. 318-325, 2013. <http://doi.org/10.1016/j.fuel.2012.07.054>

ROCHA, W. F. C.; PRADO, C. B. P.; BLONDER, N. Comparison of chemometric problems in food analysis using non-linear methods. **Molecules**, [s. l.], v. 25, p.1-55, 2020. <http://10.3390/molecules25133025>

ROSSI, G. Z. *et al.* Análise técnica da produção do biodiesel a partir do óleo de fritura residual. **Journal of Engineering and Exact Science**, [s. l.], v. 4, n. 1, p. 1075-2527, 2018. Disponível em: <https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.18540%2Fjcecv14i1ss1pp0101-0108>. Acesso em: 14 jun. 2020.

RUFINO, H. L. P. **Algoritmo de aprendizado supervisionado**: baseado em máquinas de vetores de suporte: uma contribuição para o reconhecimento de dados desbalanceados. 2011. 107 f. Tese (Doutorado em Engenharias) – Universidade Federal de Uberlândia, Uberlândia, 2011.

RUSCHEL, C. F. C. **Aplicação de ferramentas quimiométricas e técnicas espectroscópicas na análise de combustíveis**. 2017. 85 f. Tese (Doutorado em Química) – Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2017.

SÁ, C. C. **Métodos de validação tradicional e temporal aplicados à avaliação de classificadores de RNAs codificantes e não codificantes**. 2018. 102 f. Dissertação (Mestrado em Bioinformática) – Universidade de São Paulo, São Paulo, 2018. <http://doi.org/10.11606/D.95.2018.tde-19052018-122805>

SAMADI, S. H. *et al.* Potential saving in energy using combined heat and power technology for drying agricultural products (*banana slices*). **Journal of the Saudi Society of Agricultural Sciences**, [s. l.], v. 13, n. 2, p. 174-182, 2014. <http://doi.org/10.1016/j.jssas.2013.09.001>

SANTANA, F. B. D. **Uso de espectroscopia no infravermelho médio e análise discriminante por quadrados mínimos parciais na determinação de adulterações em óleos de andiroba, primula e rosa mosqueta**. 2015. 79 f. Dissertação (Mestrado em Química) – Universidade Federal de Uberlândia, Uberlândia, 2015.

SANTANA, F. B. D. *et al.* Experimento didático de quimiometria para classificação de óleos vegetais comestíveis por espectroscopia no infravermelho médio combinado com análise discriminante por mínimos quadrados parciais: um tutorial, parte v. **Química Nova**, São Paulo, v. 43, n. 3, p. 371-381, 2020. <http://doi.org/10.21577/0100-4042.20170480>

SANTOS, M. C.; TOGNOLLI, J. O.; OLIVEIRA, O. M. M. F. Quimiometria como ferramenta analítica para definição das condições de ensaio da enzima peroxidase de soja. **Eclética Química Journal**, [s. l.], v. 35, n. 4, p. 121-131, 2010. <http://doi.org/10.1590/S0100-46702010000400016>

SANTOS, A. B. **Combinação de múltiplas abordagens de classificação para interpretação de imagens hiperespectrais de sensoriamento remoto**. 2013. 83 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

SANTOS, C. M. *et al.* Assessment of ANN and SVM models for estimating normal direct irradiation (H_b). **Energy Conversion and Management**, Oxford, v. 126, p. 826-836, 2016. <http://doi.org/10.1016/j.enconman.2016.08.020>

SANTOS, R. *et al.* Empirical study on rotation and information exchange in particle swarm optimization. **Swarm and Evolutionary Computation**, [s. l.], v. 48, p. 312–328, 2019. <https://doi.org/10.1016/j.swevo.2019.04.011>

- SARIN, R.; SHARMA, M.; KHAN, A. A. Studies on *Guizotia abyssinica* L. oil: Biodiesel synthesis and process optimization. **Bioresource Technology**, Amsterdam, v. 100, n. 18, p. 4187-4192, 2009. <https://doi.org/10.1016/j.biortech.2009.03.072>
- SHAHIR, V. K.; JAWAHAR, C. P.; SURESH, P. R. Comparative study of diesel and biodiesel on CI engine with emphasis to emissions. A review. **Renewable and Sustainable Energy Reviews**, [s. l.], v. 45, p. 686-697, 2015. <https://doi.org/10.1016/j.rser.2015.02.042>
- SHIROMA, C.; RODRIGUEZ-SAONA, L. Application of NIR and MIR spectroscopy in quality control of potato chips. **Journal of Food Composition Analysis**, [s. l.], v. 22, p. 596-605, 2009. Disponível em: <https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1016%2Fj.jfca.2008.09.003>. Acesso em: 14 jul. 2019.
- SILVA, G. W. B. *et al.* Biodiesel/Diesel Blends Classification with Respect to Base Oil Using NIR Spectrometry and Chemometrics Tools. **Journal of the American Oil Chemists' Society**, Chicago, v. 89, n. 7, p. 1165-1171, 2012. <https://doi.org/10.1007/s11746-012-2028-8>
- SILVA, G. N. *et al.* Air drying of macauba fruits: maintaining oil quality for biodiesel production. **Acta Scientiarum**, Maringá, v. 42, p. 1-10, 2020. <https://doi.org/10.4025/actasciagron.v42i1.43451>
- SINGH, I.; TAGGAR, M. S. Recent Trends in Biodiesel Production-An Overview. **International Journal of Applied Engineering**, Berlin, v. 9, n. 10, p. 1151-1158, 2014. <https://doi.org/10.13031/2013.21213>
- SITOE, B. V. *et al.* Quantification of *Jatropha* methyl biodiesel in mixtures with diesel using mid-infrared spectrometry and interval variable selection methods. **Analytical Letters**, New York, v. 53, n. 4, p. 589-605, 2020. <https://doi.org/10.1080/00032719.2019.1659805>
- SCHÖLKOPF, B. *et al.* Comparing Support Vector Machines with gaussian kernels to radial basis function classifiers. **IEEE Transactions on Neural Networks**, New York, v. 45, n. 11, p. 2758- 2765, 1997.
- SMITH, B. C. **Fundamentals of Fourier transform infrared spectroscopy**. 2nd ed. Boca Raton: CRC Press, 2011.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on Support Vector Regression.

Statistics and Computing, London, v. 14, p. 199-222, 2004. Disponível em: <https://alex.smola.org/papers/2004/SmoSch04.pdf>. Acesso em: 15 ago. 2019.

STOYCHEVA, M. *et al.* **Biodiesel: feedstocks and processing technologies**. [S. l.]: Intech, 2011.

SOUZA, C. L. S. **Novas possibilidades para determinação do teor de biodiesel em óleo diesel para aplicação em distribuidoras de combustíveis**. 2015. 61 f. Dissertação (Mestrado em Química) – Universidade Federal de Goiás, Goiânia, 2015.

SOUZA, L. M. **Controle de qualidade de óleos de linhaça e Ginkgo biloba usando espectrometria no infravermelho médio e ferramentas quimiométricas**. 2018. 95 f. Tese (Doutorado em Ciências Exatas e da Terra) – Universidade Federal de Uberlândia, Uberlândia, 2018.

SUAREZ, C. A. G. **Otimização da produção de etanol 2G a partir de hexoses e pentoses**. 2014. 165 f. Tese (Doutorado em Engenharia Química e Área de Concentração em Pesquisa e Desenvolvimento de Processos Químicos) – Universidade Federal de São Carlos, São Carlos, 2014.

SUN, D. W. **Infrared spectroscopy for food quality analysis and control**. 1st ed. New York: Academic Press, 2009. Disponível em: <https://www.elsevier.com/books/infrared-spectroscopy-for-food-quality-analysis-and-control/sun/978-0-12-374136-3>. Acesso em: 15 jul. 2019.

TABATABAEI, M. *et al.* Recent trends in biodiesel production. **Biofuel Research Journal**, [s. l.], v. 7, p. 258-267, 2015. <http://dx.doi.org/10.18331/BRJ2015.2.3.4>

TAKAHASHI, C. C. **Mapeamento explícito como Kernel em aprendizado de máquinas de vetores de suporte**. 2015. 57 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Minas Gerais, Belo Horizonte, 2015. Disponível em: https://repositorio.ufmg.br/bitstream/1843/RAOA-BBSNWX/1/dissertacao_carla_c_takahashi.pdf. Acesso em: 30 maio 2019.

TOSATO, F. *et al.* Análise de amostras de combustíveis por fotometria, NIR portátil e RMN de ^1H - uma comparação com os resultados encontrados por técnicas normatizadas. **Química Nova**, São Paulo, v. 43, n. 2, p. 155-167, 2020. <http://dx.doi.org/10.21577/0100-4042.20170473>

TUBINO, M.; ROCHA JUNIOR, J. G.; BAUERFELDT, G. F. Biodiesel synthesis: A study of the triglyceride methanolysis reaction with alkaline catalysts. **Catalysis Communications**, Amsterdam, v. 75, p. 6-12, 2016. <https://doi.org/10.1016/j.fuel.2019.116114>

UNITED STATES DEPARTMENT OF AGRICULTURE. **Biofuels Annual**: Indonesia. Washington, DC: USDA, 2019. Disponível em: https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=Biofuels%20Annual_Jakarta_Indonesia_8-9-2019. Acesso: 18 dez. 2020.

VALDERRAMA, P.; BRAGA, J. W.; POPPI, R. J. Estado da arte de figuras de mérito em calibração multivariada. **Química Nova**, São Paulo, v. 32, n. 5, p. 1278-1287, 2009. <http://dx.doi.org/10.1590/S0100-40422009000500034>

VALDERRAMA, L, VALDERRAMA, P. Nondestructive identification of blue pen inks for documentoscopy purpose using iPhone and digital image analysis including an approach for interval confidence estimation in PLS-DA models validation. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 156, p. 188-195, 2016. <https://doi.org/10.1016/j.chemolab.2016.06.009>

VAPNIK, V. N. An Overview of Statistical Learning Theory. **IEEE Transactions on Neural Networks**, New York, v. 10, n. 5, p. 988-999, 1999.

VAPNIK, V. N. **The Nature of Statistical Learning Theory**. New York: Springer-Verlag, 1995.

VIANNA, S.A. *et al.* Biometric characterization of fruits and morphoanatomy of the mesocarp of *Acrocomia* species (Arecaceae). **International Journal of Biology**, Ontario, v. 9, p. 78-92, 2017. <https://doi.org/10.5539/ijb.v9n3p78>

WAMSER, G. M. **Identificação de promotores em sequências de DNA de bactérias *Escherichia coli* através de Máquinas de Vetores de Suporte**. 2020. 55 f. Trabalho de Conclusão de Curso (Graduação em Ciências da Computação) – Universidade de Caxias do Sul, Rio Grande do Sul, 2020.

WANG, S. *et al.* Simultaneous detection of different properties of diesel fuel by near infrared spectroscopy and chemometrics. **Infrared Physics and Technology**, Exeter, v. 104, p. 1-9, Jan. 2020. <https://doi.org/10.1016/j.infrared.2019.103111>

WEISS, E. A. **Oilseed crops**. 2nd ed. Malden: Blackwell Science, 2000.

WISE, B. M. *et al.* PLS_Toolbox Version 4.0 for use with MATLAB™. Wenatchee: Eigenvector Research, 2006. Disponível em: http://mitr.p.lodz.pl/raman/jsurmacki/pliki/zajecia/LMDiT/cw3/LMDiT_PLS_Manual_4.pdf. Acesso em: 31 jun. 2019.

WOLD, S.; SJOSTROM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 58, p. 109-130, 2001. <https://doi.org/10.1016%2FS0169-7439%2801%2900155-1>

WORLEY, B.; HALOUSKA, S.; POWERS, R. Utilities for quantifying separation in PCA/PLS-DA scores plots. **Analytical Biochemistry**, [s. l.], v. 433, n. 2, p. 102-104, 2013. <https://doi.org/10.1016/j.ab.2012.10.011>

WU, D.; SUN, D-W. Application of visible and near infrared hyperspectral imaging for non-invasively measuring distribution of water-holding capacity insalmon flesh. **Talanta**, Amsterdam, v. 116, p. 266-276, 2013. <https://doi.org/10.1016/j.talanta.2013.05.030>

XU, L. *et al.* Rapid discrimination of pork in Halal and non-Halal Chinese ham sausages by Fourier transform infrared (FTIR) spectroscopy and chemometrics. **Meat Science**, Amsterdam, v. 92, p. 506-510, 2012. <https://doi.org/10.1016/j.meatsci.2012.05.019>

YADAV, S. *et al.* Genetic divergence studies in niger (*Guizotia abyssinica*) germplasm. **Biomass and Bioenergy**, Oxford, v. 44, p. 64-69, 2012. <http://dx.doi.org/10.1016/j.biombioe.2012.04.011>

ZHOU, X.; FACKOURY, V. **User Equipment Characterization using Machine Learning**. 2017. 42 f. Thesis (Master) –Faculty of Engineering, Lund University, 2017. Disponível em: <https://www.eit.lth.se/srapport.php?uid=1223>. Acesso em: 23 ago. 2019.