PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DOS MATERIAIS

“Análise e Diferenciação de espécies de madeiras comerciais utilizando espectroscopia óptica no infravermelho por transformada de Fourier e análise multivariada”

Everton Chaves Prates de Jesus

Campo Grande – MS
2021

EVERTON CHAVES PRATES DE JESUS

“Análise e Diferenciação de espécies de madeira comerciais utilizando espectroscopia óptica no infravermelho por transformada de Fourier e análise multivariada”

Dissertação apresentada à Universidade Federal de Mato Grosso do Sul, como parte dos requisitos do Programa de Pós-Graduação em Ciência dos Materiais, para obtenção do título de Mestre.

Prof. Dr. Cícero Cena
Orientador

Campo Grande – MS
2021

DEDICATÓRIA

A Deus, que em todo o momento guia meus passos. Graças a seu amor, até mesmo os obstáculos podem ser dádivas, porque nos ensinam as maiores lições da vida.

A minha mãe, Irma, meu pai, Otacílio, meus irmãos Everson e Ângela, minhas cachorras, Biluzinha e Todinha, pelo seu carinho ao me ver chegar todo dia em casa, e toda minha família pelo apoio, incentivo e carinho que sempre foi essencial para minha caminhada e que não mediram esforços para me ajudar a chegar até aqui.

A minha noiva Karina por todo apoio psicológico e emocional que me proporcionou, me incentivando, me dando forças e permanecendo ao meu lado em todos os momentos, fáceis ou difíceis, para que fosse possível a realização desse trabalho.

AGRADECIMENTOS

A Deus, que em todo momento guia os meus passos.

A minha mãe, Irma, meu pai, Otacílio, meus irmãos, Ângela e Everson, minhas cachorras Biluzinha e Todinha, e toda minha família pelo apoio, incentivo e carinho.

A minha noiva, Karina, por todo apoio psicológico e emocional que me proporcionou, permanecendo ao meu lado sempre que necessário.

Ao meu orientador, Prof. Dr. Cícero Cena, por ter aceitado fazer parte do desenvolvimento desse trabalho e pela confiança depositada. Agradeço a sua simplicidade e respeito nos seus ensinamentos, pelas discussões e por compartilhar até seus itens pessoais como seu netbook durante parte da pesquisa, não medindo esforços para tentar a cada dia espalhar a ciência na sociedade como um todo.

Agradeço ao grupo de pesquisa GOF/UFMS pelo apoio nos debates e discussões, como Matheus Cícero, Gustavo Larios, por dedicar seu tempo a troca de informações. Em especial ao colega, Thiago França por ter compartilhado seu conhecimento e ajudado com o aprendizado de máquina, e por todo tempo reservado para troca de informações.

Ao Instituto de Física da Universidade de Mato Grosso do Sul, e a todo seu corpo docente, pela união e troca de conhecimento, orientações e conselhos, pela disponibilização dos laboratórios e de toda infraestrutura fornecida.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Número do Processo 88887.501966/2020-00.

Faça o teu melhor,
na condição que você tem,
enquanto você não tem condições melhores,
para fazer melhor ainda.
(Mario Sergio Cortella)

Resumo

A identificação de espécies de madeira é feita usualmente com base em análise sensorial humana, observando sua cor, cheiro, textura, quantidade e a distribuição dos poros como alguns dos fatores que formam a identidade de cada espécie de madeira. Nesta vertente, pesquisas voltadas a associação das técnicas de Aprendizagem de Máquina (ML), juntamente a técnica de Espectroscopia de Infravermelho por Transformadas de Fourier (FTIR) para discriminar diferentes tipos de materiais, que resolvam as dificuldades dos métodos tradicionais e apresentem vantagens para implementação, como rapidez, custo, vem crescendo nos últimos anos. Neste trabalho associamos a técnica de FTIR e ML para solucionar os problemas encontrados na identificação incorreta da madeira comercial, otimizando a rapidez na análise e preparação da amostra em relação ao método tradicional. Os espectros de absorção no infravermelho de cinco espécies de madeira: *Hymenolobium petraeum Ducke,* Angelim-pedra (ANG); *Gochnatia polymorpha*, Cambara (CAM); *Erisma uncinatum,* Cedrinho (CED)*Dipteryx odorata,* Champagne (CHA); *Goupia glabra Aubl,* Peroba do Norte (PER) foram utilizadas como na análise multivariada usando ML. Os resultados demonstraram que a técnica de FTIR juntamente com análise multivariada, foram capazes de diferenciar as cinco espécies de madeira com sensibilidade e especificidade de 100%. O método desenvolvido possibilita verificar o potencial da técnica associado ao ML para que indústrias, laboratórios, empresas e/ou órgãos de controle possam identificar a natureza do produto após extraídos e semi-manufaturado.

Palavras-chave: FTIR, Análise Multivariada, Análise de Componentes Principal, Validação Cruzada, Madeiras Comerciais, Aprendizagem de Máquina.

Abstract

The identification of wood species is usually done based on human sensory analysis, observing its color, smell, texture, quantity and pore distribution as some of the factors that form the identity of each wood species. In this aspect, researches aimed at the association of Machine Learning (ML) techniques, together with the Fourier Transform Infrared Spectroscopy (FTIR) technique to discriminate different types of materials, which solve the difficulties of traditional methods and present advantages for implementation, such as speed, cost, has been growing in recent years. In this work we associate the FTIR and ML technique to solve the problems found in the incorrect identification of commercial wood, optimizing the speed of analysis and sample preparation in relation to the traditional method. Infrared absorption spectra of five wood species: Hymenolobium petraeum Ducke, Angelim-stone (ANG); Gochnatia polymorpha, Cambara (CAM); Erisma uncinatum, Cedrinho (CED)Dipteryx odorata, Champagne (CHA); Goupia glabra Aubl, Northern Peroba (PER) were used as in multivariate analysis using ML. The results showed that the FTIR technique together with multivariate analysis were able to differentiate the five wood species with 100% sensitivity and specificity. The developed method makes it possible to verify the potential of the technique associated with ML so that industries, laboratories, companies and/or control bodies can identify the nature of the product after extracted and semi-manufactured.

Keywords: FTIR, Multivariate Analysis, Principal Component Analysis, Cross Validation, Commercial Timber, Machine Learning.

LISTA DE FIGURAS

LISTA DE ABREVIATURAS E SIGLAS

AM	Análise Multivariada
ANG	Angelim-pedra
ATR	Refletância Total Atenuada
CAM	Cambará
CED	Cedrinho
CGT	Lixo Gin de Algodão
CHA	Champagne
FTIR	Espectroscopia de Infravermelho por Transformada de Fourier
IR	Infravermelho
k-NN	*k*-vizinhos mais próximos
LDA	Análise de Discriminante Linear
LOOCV	Validação Cruzada deixando um de fora
ML	Aprendizado de Máquina
PC	Componentes Principais
PCA	Análise de Componentes Principais
PER	Peroba do Norte
PLS	Mínimos Quadrados Parciais
SNV	Variável Normal Padrão
SVM	Máquina de Suporte de Vetores
VC	Validação Cruzada

SUMÁRIO

1 INTRODUÇÃO

A madeira é um composto abundante, renovável e biodegradável com muitas aplicações úteis, constituído basicamente 50% de celulose, de 20 a 30% de lignina, 20 a 25% de hemicelulose (WANGAARD, 1979; SHEBANI, 2008; POPESCU, 2009) e 2 a 5% de outros constituintes das madeiras que pode incluir lipídios, compostos fenólicos, terpenóides, ácidos graxos, ácidos de resina e ceras (SHEBANI, 2008; POPESCU, 2009). A celulose é um polímero linear de unidades de glicose que pode formar ligações intracadeias e intercadeias, produzindo uma macromolécula cristalina com maior peso molecular que outros componentes da madeira (JOHN, 2008). As hemiceluloses compreendem um grupo de polissacarídeos compostos por uma combinação de açúcares em anel de 5 e 6 carbonos (JOHN, 2008). A lignina é um polímero condensado aleatório com muitos grupos aromáticos mais hidrofóbico do que a celulose ou a hemicelulose (POPESCU, 2011). Os demais componentes podem ser: lipídios, compostos fenólicos, terpenóides, ácidos graxos, ácidos de resina e ceras (SHEBANI, 2008; POPESCU, 2009).

A identificação de espécies de madeira é feita usualmente com base em análise sensorial humana, observando sua cor, cheiro, textura, quantidade e a distribuição dos poros como alguns dos fatores que formam a identidade de cada espécie de madeira. A análise pode ser feita a olho nu ou com auxílio de uma lente portátil de 10 vezes de aumento. A rastreabilidade da madeira é ainda mais difícil de ser realizada sendo empregado apenas aspectos relativos à fiscalização de papéis de compra, venda, plantio etc. A ausência de um método de análise mais robusto para classificação e rastreabilidade, independente do treinamento ou análise sensorial humana acarreta a classificação errônea da mesma, prejudicando controle de qualidade e fiscalização da madeira comercializada. A identificação da composição química da madeira é uma ferramenta importante pois nos setores produtivos, é difícil e depende de treinamento, conhecimento e habilidade do responsável.

A caracterização tradicional das amostras de madeira pode ser mais complexa envolvendo várias etapas em que os componentes da madeira são isolados ou degradados em fragmentos monoméricos (CHEN, 2010). O método Van Soest consiste em quantificar o teor total de fibra das plantas utilizando um surfactante específico capaz de solubilizar a bicamada lipídica dos alimentos, tornando o conteúdo celular solúvel, separando por filtração o resíduo insolúvel

constituído pelos constituintes da parede celular como celulose, hemicelulose e lignina (VAN SOEST, 1963; VAN SOEST, 1967). Esses procedimentos destroem a matriz da madeira e requerem grandes tamanhos de amostras ao longo tempo de análise (FERRAZ, 2000). Dessa maneira, a técnica de Espectroscopia Óptica por Transformada de Fourier (FTIR) possui a vantagem em relação ao método tradicional referente ao tempo de análise da amostra.

A relação entre os parâmetros estruturais, como a correlação entre a composição química e as propriedades físicas da madeira ainda não foram totalmente explorados (POLETO, 2012) assim, esse trabalho tem como objetivo principal aplicação do FTIR e análise multivariada para diferenciação de espécies de madeira. Além disso, o método desenvolvido possibilita que indústrias, laboratórios, empresas e/ou órgãos de controle possam identificar a natureza do produto após extraídos e semi-manufaturado.

2 OBJETIVOS

2.1 Objetivo Geral

Aplicação da espectroscopia óptica no infravermelho por transformada de Fourier e análise multivariada para diferenciação de espécies de madeira.

2.2 Objetivos Específicos

- Caracterizar as diferentes espécies de madeira utilizando espectroscopia óptica no infravermelho por transformada de Fourier;

- Identificar os modos vibracionais associados as moléculas que compõem a madeira;

- Usar Análise das Componentes Principais (PCA) para avaliar as principais contribuições para separação e/ou classificação dos grupos de espécies de madeira;

- Usar Aprendizado de máquina (do inglês, *Machine Learning – ML*) para analisar dados do PCA e automatizar a classificação;

3 REVISÃO BIBLIOGRÁFICA

3.1 Métodos de identificação da madeira

A correta identificação da madeira necessita empregar técnicas das quais pode-se destacar o método de Van Soest que consiste em quantificar o teor total de fibra das plantas utilizando um surfactante específico capaz de solubilizar a bicamada lipídica da membrana celular dos alimentos, tornando o conteúdo celular solúvel, separando por filtração o resíduo insolúvel constituído pelos constituintes da parede celular como celulose, hemicelulose e lignina (VAN SOEST, 1963; VAN SOEST, 1967) para analisar a composição química da madeira, sua desvantagem é o tempo requerido para análise, que é de 3 a 4 dias, se comparado com a técnica FTIR que o tempo requerido de análise por amostra é de 10 minutos (CHEN, 2010), sendo assim é uma opção realização da técnica de espectroscopia óptica pela vantagem de possuir o tempo de preparação da amostra rápido (CHEN, 2010).

Na literatura científica existem trabalhos que reportam a classificação de madeira utilizando técnica de Espectroscopia com emissão Ótica por Plasma Induzido por Laser (LIBS) combinada com rede neural artificial (RNA), foi investigada para classificar quatro espécies de madeira (pau-rosa-africano, bubinga do Brasil, padauk de Myanmar e Pterocarpus erinaceus, os dados espectrais do recurso foram selecionados com base nos carregamentos do PCA e normalizada usando a soma de todos os dados de espectros de recursos, os resultados experimentais mostraram que LIBS integrado com RNA pode ser aplicado para analisar e reconhecer espécies de madeira (CUI et al, 2019;2021).

DOS SANTOS et al. 2021, emprega espectroscopia no infravermelho próximo (NIR) para análise e classificação de espécies de louro branco, louro pimenta, louro preto, louro rosa, itaúba amarela, embora utilize assinatura espectral referente a modos vibracionais moleculares, esse trabalho utilizou três aprendizado de máquina: SVM, PLS-DA e k-NN, obtendo com algoritmo PLS-DA 97% de precisão dos resultados, esta analisa uma região diferente do espectro infravermelho em comparação com o FTIR e restringe-se à análise de espécies da mesma família. Por fim, uma patente sobre uso de espectroscopia Raman para análise de qualidade de madeira foi encontrada, além da técnica também diferir o foco da análise, também é diferente.

Lazzari et al (2018) utilizou o FTIR em combinação com PCA pela primeira vez para classificar a biomassa por meio de sua composição de bio-óleo, esses foram produzidos por pirólise para as quinze fontes de biomassa disponíveis no Brasil e sua caracterização química foi observada usando análises cromatográficas, o uso do PCA permitiu discriminar as biomassas em três grupos principais, que apresentaram composições distintas de bio-óleo, por meio de seus dados espectrais no infravermelho. Usando metodologia não destrutiva usando ATR-FTIR com PCA para diferenciar entre tecidos de casca históricos do Pacífico, esse estudo preliminar de tecido de casca de árvore contemporâneo e histórico, a análise multivariada de espectros de FTIR na região de 1200–1600 cm^{-1} foi mostrado como sendo usada para agrupar tecidos de casca históricos originários de diferentes espécies, a análise de PCA identificou três grupos para o pano histórico com os gráficos de carregamento destacando onde as diferenças entre os espectros FTIR são predominantes para cada PC (SMITH, 2019).

Mais recentemente Larios et. al (2020) utilizou da espectroscopia FTIR associada a ferramentas quimiométricas para discriminar a qualidade fisiológica (baixo e alto vigor) dos lotes de sementes de soja. A melhor classificação foi obtida pelo método de análise discriminante linear (LDA) com análise de componentes principais (LARIOS, 2020). Esse estudo, mostrou o fácil preparo da amostra e rápida análise, mostrando o potencial da espectroscopia FTIR aliada aos métodos quimiométricos para serem utilizados em testes de vigor de sementes. Dessa maneira, a espectroscopia de FTIR mostrou ser uma ferramenta de alto potencial que forneceu detalhes sobre as características estruturais de diferentes amostras de materiais (PLÁCIDO, 2014; POPESCU,2009; RUDAKIYA, 2019; LAZZARI, 2018; SMITH, 2019; LARIOS, 2020) sem a demora na sua preparação, associados com métodos de classificação o que demonstra a versatilidade que a técnica FTIR possui em desenvolver a rápida classificação em diferentes tipos de materiais, com alta sensibilidade e especificidade.

3.2 Composição química da madeira

A madeira é constituída principalmente por substâncias orgânicas. Os principais elementos constituintes apresentam-se nas seguintes porcentagens aproximadas, independentemente da espécie vegetal considerada (WANGAARD, 1979): Carbono 50%,

Oxigênio 44%, Hidrogênio 6%. O composto orgânico predominante é a celulose, que constitui cerca de 50% da madeira, formando os filamentos que reforçam as paredes das fibras longitudinais (PFEIL, 2003). Outros dois componentes importantes são as hemiceluloses (constituindo 20 a 25% da madeira) e a lignina (20 a 30%) que envolvem as macromoléculas de celulose ligando-as (YOUNG, 1998). A lignina provê rigidez e resistência à compressão às paredes das fibras (WANGAARD, 1979).

A estrutura da madeira apresenta ainda pequenas quantidades (0,2 a 1%) de sais minerais, que constituem os alimentos dos tecidos vivos (PFEIL,2003). Esses minerais produzem as cinzas quando a madeira é queimada. As espécies vegetais apresentam ainda mais materiais, como resinas, óleos, ceras, que são depositados nas cavidades das células, produzindo coloração e cheiro característicos da espécie (PFEIL,2003).

A madeira também contém uma pequena quantidade de extrativos que pode incluir lipídios, compostos fenólicos, terpenóides, ácidos graxos, ácidos de resina e ceras (SHEBANI, 2008; POPESCU, 2009). Geralmente, o conteúdo de extrativos varia entre 2% e 5%, mas pode chegar a 15% (DESHAVATH, 2019).

Dos três componentes principais da madeira, a celulose é o recurso natural mais abundantemente disponível e a demanda por ela é cada vez maior por sua natureza ecologicamente correta e biocompatível (POPESCU, 2009; WATIKINS, 2015). Nas plantas o principal componente da parede celular é o polissacarídeo conhecido como celulose, o qual determina em grande parte a estrutura da planta. A madeira é composta por cerca de 50% de celulose, sua fórmula genérica é $(C_6H_{12}O_5)n$. A celulose é um polímero composto de monômeros de glicose, que é um componente estrutural da parede celular vegetal. A figura 1 representa a estrutura química da celulose (WATIKINS, 2015) as longas cadeias se unem lateralmente por pontes de hidrogênio entre grupos de hidroxila e originam as micelas, que unidas formam as fibrilas e irão constituir as paredes do tecido do xilema.

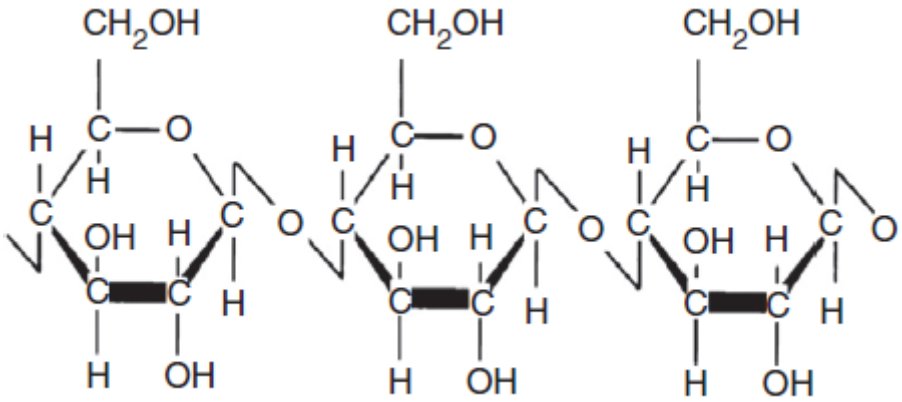


Figura 1 – Representação da estrutura química da celulose (POPESCU, 2010).

A hemicelulose é o segundo carboidrato polimérico mais abundante na natureza e um dos principais componentes da biomassa lignocelulósica (DESHAVATH, 2019). Ela possui uma estrutura mais irregular com grupos laterais, grupos substituintes e açúcares presentes ao longo do comprimento da cadeia (JOHN, 2008). Por conta da presença de agrupamentos hidroxilas conectadas à sua cadeia principal e expostos devido à sua condição amorfa, a hemicelulose torna-se mais suscetível às reações químicas de degradação e é menos tolerante à ação do calor.

Os grupamentos -OH e as ramificações de açúcar da hemicelulose fazem com que ela desempenhe um papel importante de ligação e de estabilização entre as microfibrilas, que são altamente polares e unidas via ligações de hidrogênio (DESHAVATH, 2019). Ao contrário da celulose homóloga, as hemiceluloses são heteropoliméricas por natureza, a figura 2 representa diferentes tipos de compostos, como pentoses (xilose, ramnose e arabinose) unidades de açúcares que possuem apenas cinco átomos de carbono, hexoses (glicose, manose e galactose) possuem seis átomos de carbono e ácidos orgânicos (acético, 4-O ácidos -metilglucurônico, D-glucurônico e D-galacturônico) (GÍRIO, 2000; DESHAVATH, 2019).

Os xiloglucanos são constituintes hemicelulósicos das paredes celulares primárias da biomassa da madeira. Uma estrutura típica de xiloglucano é composta por um esqueleto de D-glicose ligado a β-1,4, e três das quatro unidades de glicose são substituídas na posição O-6 com D-xilose (DESHAVATH, 2019). Além disso, os Galactoglucomananos ou O-acetil-galactoglucomananos são os principais constituintes hemicelulósicos de madeiras macias, respondendo por 20% a 25% de sua massa seca (PEREIRA, 2003). Assim, as hemiceluloses

isoladas das madeiras são misturas complexas de polissacarídeos, sendo os mais importantes, glucouranoxilanas, arabinoglucouranoxilanas, glucomananas, arabinogalactanas e galactoglucomananas (PHILIPP, 1988). Vale ressaltar que o termo hemiceluloses não designa um composto químico definido, mas sim uma classe de componentes poliméricos presentes em vegetais fibrosos, possuindo cada componente propriedades peculiares (PHILIPP, 1988).

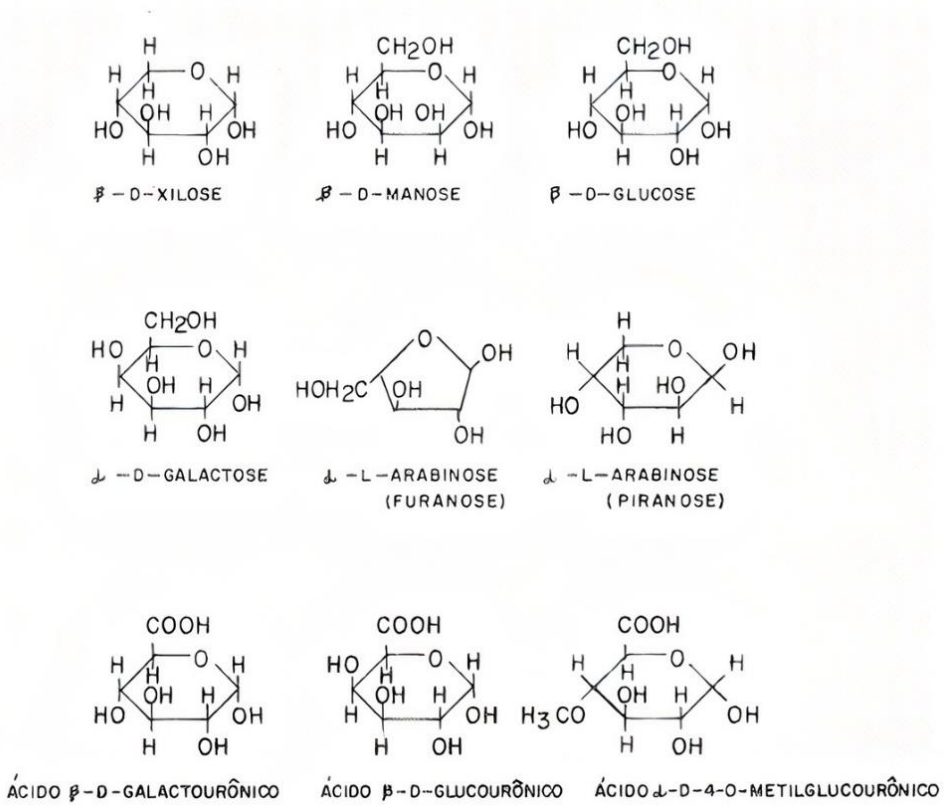


Figura 2 – Açucares que compõem as hemiceluloses (PHILIPP, 1988).

A lignina é altamente aromática por natureza e o terceiro polímero mais abundante do mundo. Ao contrário da celulose e da hemicelulose, a lignina não contém carboidratos (açúcares) em sua estrutura polimérica. A lignina é um composto fenilpropano tridimensional metoxilado que é o único responsável pela rigidez estrutural da biomassa lignocelulósica, que geralmente cobre a hemicelulose e a celulose (ZHOU, 2011).

Na Figura 3 é uma representação da estrutura química da lignina, as substâncias fenólicas englobam uma grande variedade de compostos, dentre eles a lignina, todos eles apresentando um grupo de hidroxila (-OH) ligado a um anel aromático (WATIKINS, 2015). Na figura 4 está

indicado a estrutura química dos componentes da lignina formada por três tipos de monômeros: álcoois p-cumarílico, coniferílico e sinápilico (WATIKINS, 2015).

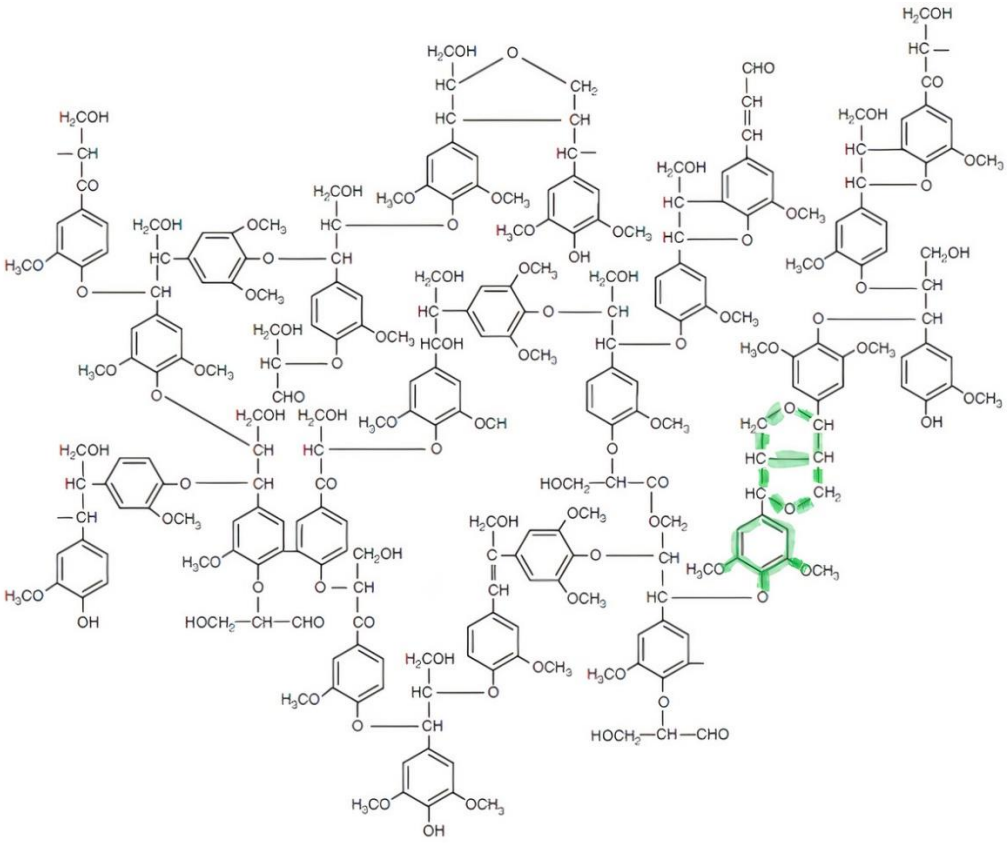


Figura 3 – Estrutura química da lignina (WATIKINS, 2015).

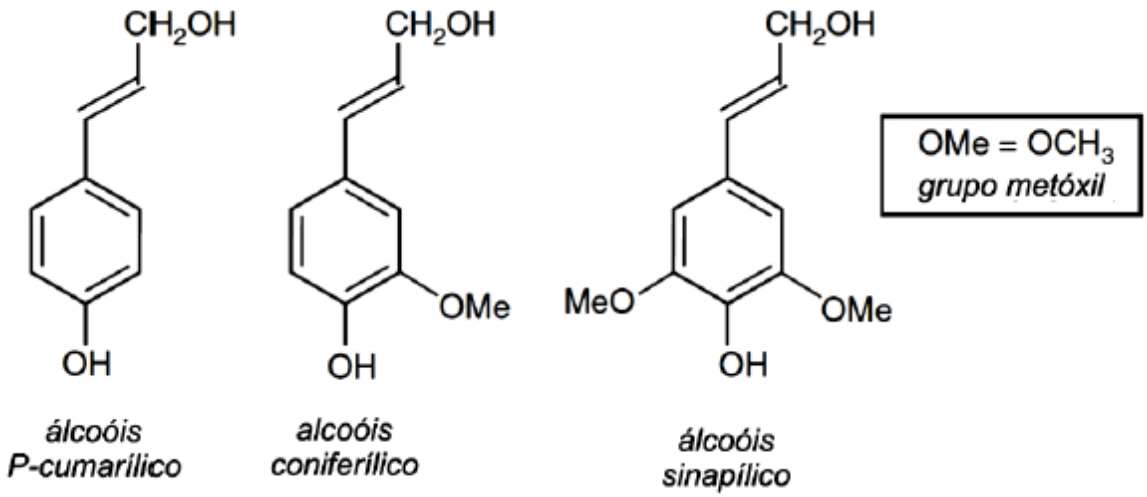


Figura 4 – Representação da estrutura química dos componentes da lignina (SHEBANI, 2008).

3.3 Espectroscopia FTIR para classificação de madeiras

Estudos recentes mostram o uso FTIR para determinar mudanças na estrutura química das madeiras tratadas termicamente, o pinheiro silvestre (*Pinus sylvestris L.*), faia oriental (*Fagus orientalis L.*) e abeto oriental *(Picea orientalis L.*) espécies de madeira foram tratadas termicamente em diferentes temperaturas (OZLEM, 2016). A discriminação de amostras de madeira usando análise multivariada, incluindo análise de agrupamento hierárquico (HCA) e PCA com algoritmos LDA também foi usada para classificar as amostras de madeira desconhecidas na respectiva classe (SHARMA, 2020).

O FTIR um método útil para estudar madeiras, tem sido aplicada a fim de medir um eventual gradiente de água absorvida (mais precisamente, grupos hidroxila, OH) entre a superfície e a parte interna de uma antiga e moderna escultura de madeira (VARTANIAN, 2014). O uso do FTIR, juntamente PLS para a discriminação entre dois espécies de madeira de nogueira mostrou que a discriminação entre espécies tratadas e não tratadas com vapor de Juglans nigra é muito clara (HOBRO, 2010).

A espectroscopia foi aplicada a 120 amostras de anéis de cerne de oito indivíduos pinheiros de diferentes locais da Espanha, resultados mostram que FTIR em combinação com análises multivariadas pode ser uma ferramenta útil para a identificação de espécies e comprovação de amostras de madeira de pinho de origem desconhecida (TRAORE, 2018). Três espécies diferentes de madeira dura, nomeadamente choupo (*Populus spp*), cal (*Tilia spp*) e bétula (*Betula spp*), foram investigados por meio de espectroscopia de FTIR, diferenças claras foram encontradas nos espectros das três amostras, confirmando que FTIR é uma ferramenta poderosa para a discriminação do tipo de madeira (BUOSO, 2016).

.

3.4.1 Análise de Componentes Principais (*Principal Component Analysis -PCA*)

A literatura mais antiga sobre PCA data de Pearson (1901) e Hotelling (1933), mas só depois que os computadores eletrônicos se tornaram amplamente disponíveis foi possível usar em um conjunto de dados que não fosse trivial (JOLLIFFE e CADIMA, 2016). Resumidamente, a PCA é uma ferramenta de estatística aplicada para reduzir a dimensionalidade das variáveis originais por transformação ortogonal linear, convertendo um conjunto de variáveis correlacionadas em variáveis não correlacionadas (ERIKSSON, 1999; JOLLIFFE, 2016). Assim, um gráfico de componentes principais pode destacar variações e tendências do sistema, fornecendo uma abordagem simples para visualizar o conjunto de dados (JOLLIFFE, 1973; KANNO, 2019; BELLOU, 2020).

Na figura 5, duas classes de amostras uma na cor verde e outra na amarela, em um plano tridimensional, com suas variáveis (x_1, y_1 , z_1). Usando o PCA suas coordenadas são reescritas por ordem decrescente de variância em um novo plano tridimensional (PC1, PC2 e PC3). Como as duas primeiras PC's são suficientes para descrever a variância, é possível reduzir o número de variáveis (v1e v2) sem prejuízo de informação.

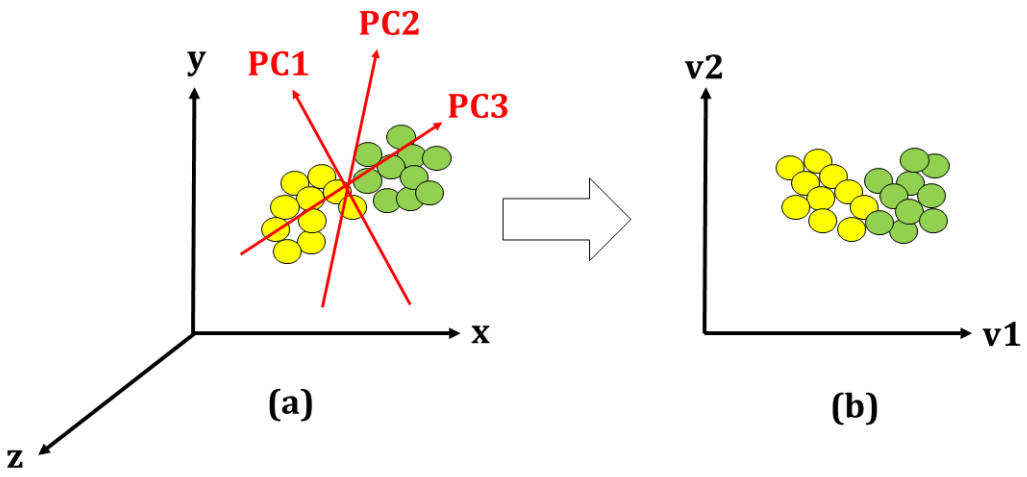


Figura 5 – (a) Representação de duas classes no espaço tridimensional, (b) duas componentes principais. (SILVA, 2021).

Ela consiste na extração da matriz de covariância do sistema e a partir dela obtém-se os autovalores e autovetores, que definem a direção e a parcela de maior variabilidade dos dados, referente a direção da PC. O contexto padrão para PCA como uma ferramenta de análise exploratória de dados envolve um conjunto de dados com observações sobre *p* variáveis

numéricas, para cada uma de *n* entidades ou indivíduos (JOLLIFFE, 2016). Dessa forma, esse método se torna uma ferramenta potente para auxiliar na visualização do comportamento do sistema, quando ele apresenta um grande volume de dados ou amostras, preservando a qualidade dos dados.

3.3.2 Análise de Discriminante Linear (*Linear Discriminant Analysis - LDA*)

A LDA proposta por Fisher (1936) é um dos métodos mais utilizadas para classificação de dados. Funciona maximizando a proporção da variância entre as classes para a variação dentro da classe, sendo ótima sob as premissas de probabilidade Gaussiana e matrizes de covariância igual entre grupos. Muitas modificações foram propostas para este método como LDA Ortonormal (OKADA, 1985) e LDA não paramétrico (FUKUNAGA, 1990) embora o clássico seja satisfatório (FERNÁNDEZ, 2012). O LDA foi aplicado com sucesso para classificação e reconhecimento de padrões em diversos campos como engenharia, economia, ciência da computação, biologia, entre outros (FERNÁNDEZ, 2012). Os algoritmos LDA criam limites (figura 6) que delimitam as classes usando o centro e a distribuição de cada cluster (LARIOS, 2020).

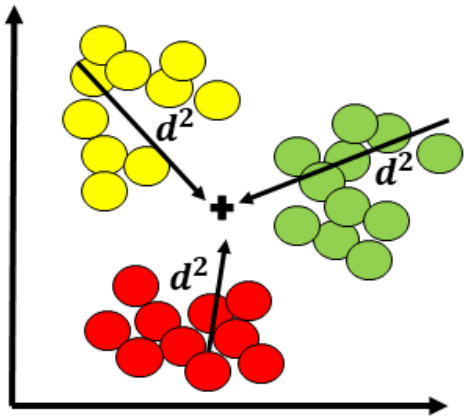


Figura 6 – Representação esquemática do funcionamento do LDA. (SILVA, 2021).

3.3.3 Máquina de Suporte de Vetores (*Suport Vector Machine - SVM*)

O SVM é uma técnica de AM desenvolvida por Vapnik (1998). O SVM é um classificador não probabilístico (FERNÁNDEZ, 2012). Ele funciona construindo um hiperplano ou conjunto de hiperplanos maximizando sua distância dos dados mais próximos de um ponto de cada lado, conseguindo assim a maior separação (Figura 7) (FERNÁNDEZ, 2012). Esse hiperplano é denominado hiperplano de margem máxima, para casos não lineares, o hiperplano é construído por uma função kernel não linear no lugar de produtos escalares (FERNÁNDEZ, 2012). O SVM organiza os dados em classes com base em sua distribuição espacial, onde o processo de diferenciação é obtido ajustando hiperplanos entre eles (LARIOS, 2020). Polinômios homogêneos e funções de base radial gaussiana são os núcleos mais usados (FERNÁNDEZ, 2012). Na forma original, SVM são classificadores binários, ou seja, eles discriminam entre duas classes (FERNÁNDEZ, 2012).

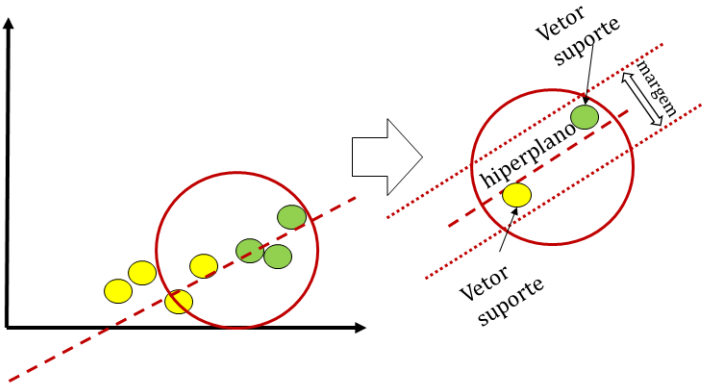


Figura 7 – Representação esquemática do funcionamento do SVM (SILVA, 2021).

3.3.4 *k*-NN (*k*-Nearest Neighboor)

O *k*-NN, *k* vizinhos mais próximos é um método multivariado não paramétrico de classificação supervisionada introduzido por Fix e Hodges (1989). A ideia básica é a seguinte: a classe de uma dada amostra será a classe mais repetida correspondendo aos vizinhos circundantes (Figura 8), também pode ser aplicado em populações onde a suposição de normalidade não é necessária (FERNÁNDEZ, 2012). O procedimento começa escolhendo uma distância apropriada (principalmente as distâncias Euclidiana ou Mahalanobis) entre as amostras, representadas por vetores de feições, em seguida, as distâncias entre a amostra de teste, x_0, e as outras amostras são

calculadas, as *k* amostras mais próximas daquelas que queremos classificar são selecionadas a seguir, é calculada a proporção dessas *k* amostras pertencentes a cada uma das populações estudadas (FERNÁNDEZ, 2012). Finalmente, a amostra x_0 é classificada dentro da população correspondente à maior frequência existente.

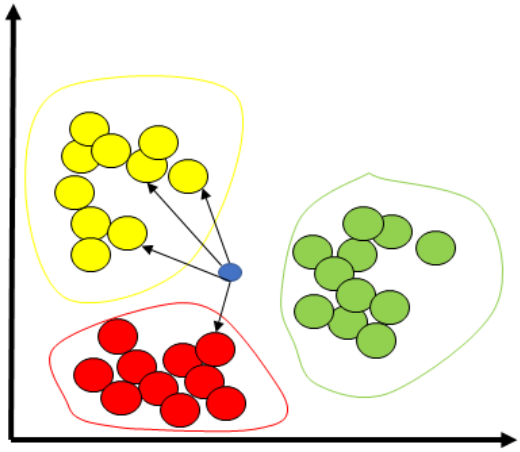


Figura 8 – Representação esquemática do funcionamento do k-NN (SILVA, 2021).

3.4 Métodos de Validação dos Dados

Para avaliar a precisão do método supervisionado de reconhecimento de padrões, métodos de avaliação são empregados para avaliar a robustez do protocolo. A técnica de validação cruzada (VC ou CV, do inglês *Cross-Validation*) foi proposta inicialmente por Geisser (1975), como um modelo analítico de previsão e validação de conjunto de dados. De modo que permita uma análise que não conflite ou interfira em um processo de aprendizado, evitando a tendenciosidade em sistemas computacionais de análise de dados (RODRIGUEZ, 2010). Sua metodologia é baseada na divisão do conjunto de dados em k pastas ou dobras (do inglês *k-fold*) aleatoriamente, onde as informações de cada pasta possuam a mesma ou quase a mesma quantidade de informações (GEISSER, 1975). Com isso, a VC treina o sistema desenvolvido em *k* vezes, utilizando *k-1* em cada um dos treinamentos, e através de sucessíveis repetições treina todos os dados e os utiliza para validar o sistema, sem que estes interfiram no processo (Figura 9) (GEISSER, 1975).

Outro método utilizado para validação dos dados é chamado de *Holdout Validation*, na qual uma porcentagem dos dados é selecionada para ser usada como dados de teste. O modelo é

treinado com o restante dos dados não utilizados para teste e a performance do modelo é avaliada com os dados de teste. Então, esse tipo de avaliação consiste em um modelo que utiliza somente uma porção dos dados para determinar a sua acurácia.

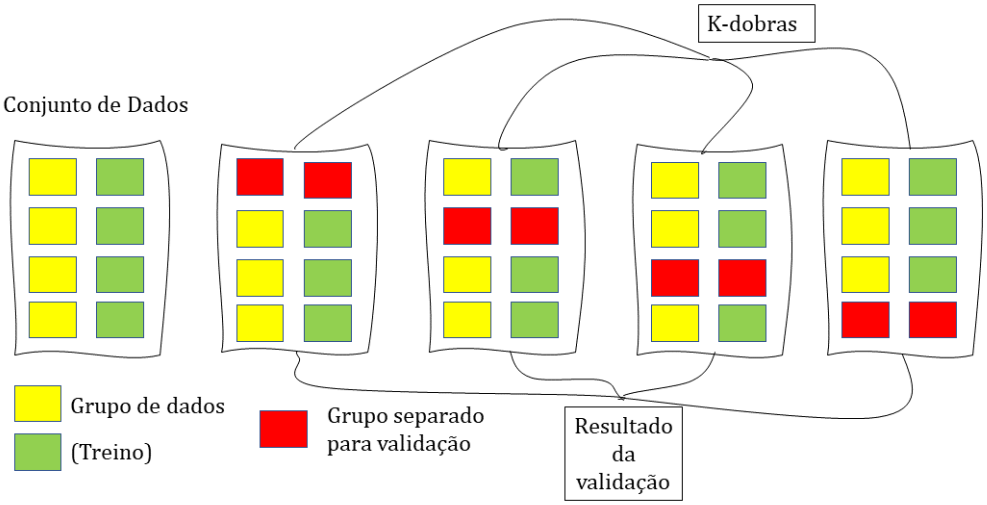


Figura 9 – Metodologia para validação cruzada (LARIOS, 2020).

4 MATERIAIS E MÉTODOS

4.1 Coleta e preparação das amostras

As amostras analisadas foram madeiras das espécies: Angelim-pedra (ANG) (*Hymenolobium petraeum Ducke*), Cambará (CAM) (*Gochnatia polymorpha*), Cedrinho (CED) (*Erisma uncinatum*), Champagne (CHA) (*Dipteryx odorata*) e Peroba do Norte (PER) (*Goupia glabra Aubl*), obtidas comercialmente na cidade de Campo Grande/MS. Foram coletadas duas alíquotas de pó de madeira de 26 lotes diferentes para cada espécie de madeira, totalizando 52 amostras por espécie.

O pó da madeira foi obtido através do processo de abrasão utilizando uma serra de fita. A granulometria do pó foi uniformizada utilizando uma peneira analítica de mesh 45 (355 µm), garantindo uma melhor uniformidade e acondicionamento da amostra na porta amostra para medida.

4.2 Caracterização das amostras e análise dos dados

Os espectros de absorção das amostras foram obtidos através da técnica FTIR em um espectrômetro Perkin-Elmer, modelo Spectrum 100N FT-NIR, utilizando acessório de refletância total atenuada (ATR). Foram feitas medidas no modo absorbância na faixa espectral de 4000 a 600 cm^{-1} utilizando 10 varreduras e resolução de 4 cm^{-1}. Para cada amostra analisada foi obtido um espectro médio das duplicatas medidas para cada lote.

Os espectros médios para cada lote foram pré-tratados utilizando o método de variável normal padrão (do inglês "*Standard Normal Variable – SNV"*), conforme descrito pela equação 01. O método de SNV foi aplicado com objetivo de remover os deslocamentos verticais, centrando os espectros na média – este procedimento é requerido para evitar distinção dos dados devido a variações experimentais. Em que x_{ij} é o valor da intensidade espectral, u_i é a média da observação a ser tratada e o σ_i é o desvio padrão (ENGEL, 2013).

$$x_{ij\,(a)} = \frac{x_{ij} - \mu_i}{\sigma_i} \qquad \text{Equação 01}$$

Em seguida foi realizada a PCA uma ferramenta estatística multivariada não supervisionada que permite reduzir as dimensões dos dados, a matriz de dados é convertida em uma matriz de pontuação e uma matriz de carregamento. A transformação converte um conjunto de variáveis correlacionadas em um conjunto de variáveis não correlacionadas, onde o primeiro PC tem a maior variância, o segundo PC a segunda maior variância e assim por diante. Dessa maneira as PC's podem destacar variações e tendências do sistema, fornecendo uma maneira simples de visualizar o conjunto de dados. Dentre seus elementos temos os autovetores chamados escores (*scores*) e os autovalores que são denominados seus pesos (*loadings*) (ENGEL, 2013).

Nesse trabalho toda análise foi desenvolvida utilizando o software MATLAB R2018a (*Mathworks INC, Natick, USA).* Inicialmente, foi analisado a faixa espectral de 4000 a 600 cm^{-1}, em seguida o intervalo espectral de 3000 a 2700 cm^{-1} e por fim o intervalo de 2000 a 700 cm^{-1} das espécies de madeiras.

4.3 Classificação das amostras

Foram aplicados três tipos de algoritmos de ML na análise multivariada supervisionada como: LDA, SVM, k-NN. O algoritmo LDA usa os valores treinados de amostras para determinar limites entre diferentes classes, calculando o centro de distribuição linear e o contorno. O SVM organiza os dados para realizar uma representação espacial das amostras agrupando os pontos em categorias, a diferenciação é obtida por uma linha virtual que separa os grupos e divide as classes em hiperplanos.

A principal diferença entre LDA e SVM é que o LDA assume uma distribuição de probabilidade normal entre as amostras, mesma matriz covariância para todas as classes, enquanto nenhuma suposição é necessária para o SVM (ENGEL, 2013). Nesse trabalho o SVM foi aplicado em seis configurações diferentes, ou seja, SVM linear, quadrático, cúbico, fino, médio, grosseiro. O k-NN é baseado na distribuição espacial de pontos e classifica as amostras com base nos vizinhos mais próximos. Para a classificação, o k-NN processa os dados e classifica uma nova amostra de acordo com a classe dos vizinhos mais próximos, ou seja, usa os "k" vizinhos mais próximos para classificação (fino = 1, médio = 10 e grosso = 100) ou usando pesos para as distâncias de modo que o mais próximo apresente maior peso. Nesse trabalho o k-NN foi aplicado em cinco configurações diferentes, ou seja, k-NN fino, médio, cosseno, cúbico e ponderado.

O número ideal de (PC's) com maior acurácia foi obtido de forma otimizada através de uma sequência de testes, primeiramente com incremento de 5 PC's e posteriormente de uma em uma, objetivando a maior acurácia possível com o menor número de PC's. A acurácia foi medida por LOOCV. Nesse caso, uma amostra é retirada do conjunto de treinamento e usada para testar o poder preditivo da modelagem. O processo é repetido, alternando a amostra retirada para testar todos os conjuntos de amostras. A partir do processo de validação é realizada a classificação da amostra, adotando-se os próprios algoritmos de aprendizado de máquina, em que cada amostra foi testada e atribuída ao seu grupo correspondente. As amostras são classificadas corretamente quando a classe prevista da classificação do ML, corresponde a classe verdadeira, identificada a partir do método padrão.

Faz LOOCV e varia número de PCs para evitar a ocorrência de "*overfiting"* e "*underfiting",* foi feito tal procedimento pois o "*overtfiting"* se adapta muito bem com os dados que está sendo treinado, porém, não generaliza bem para novos dados, ou seja, o modelo

"decorou" o conjunto de dados e não aprendeu de fato o que diferencia aqueles dados. *O "underfiting"* ocorre quando o modelo não se adapta bem sequer aos dados com que foi treinado (LARIOS, 2020).

5 RESULTADOS E DISCUSSÕES

Os espectros médios de FTIR obtidos com desvio padrão médio para os diferentes grupos de espécies de madeira são representados na Figura 10.

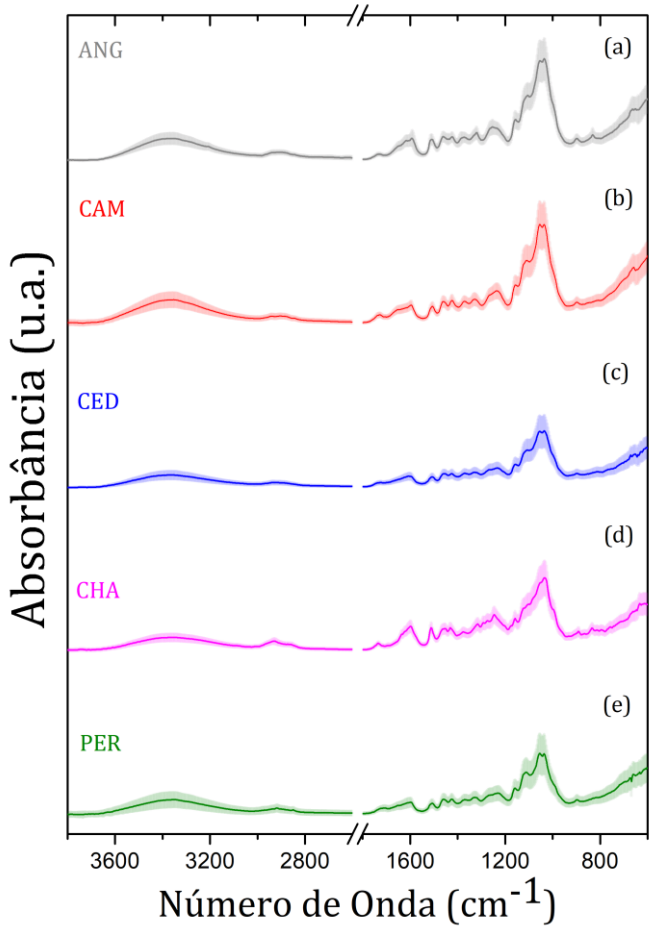


Figura 10: Espectro FTIR médio dos grupos com desvio padrão médio ilustrado pela parte sombreada para diferentes espécies de madeira. (a) *Hymenolobium petraeum Ducke,* Angelim-pedra (ANG); *(b) Gochnatia polymorpha*, Cambara (CAM); (c) *Erisma uncinatum,* Cedrinho (CED);(*d) Dipteryx odorata*, Champagne (CHA); (e) *Goupia glabra Aubl*, Peroba do Norte (PER). Fonte: Autoria própria.

Os espectros FTIR apresentam uma banda larga em torno de 3360 cm^{-1}, que pode ser associado a diferentes modos de alongamento de O-H, usualmente associado a água (PANDEY, 2003; MÜLLER, 2009; POPESCU, 2009; RUDAKIYA, 2019) e as bandas em torno de 2930 cm^{-1}, associadas a grupos de alongamento metoxil C-H presentes nos espectros de todos os componentes da madeira, mas principalmente nos espectros de celulose (PANDEY, 2003; MÜLLER, 2009; POPESCU, 2009; POLETO 2012).

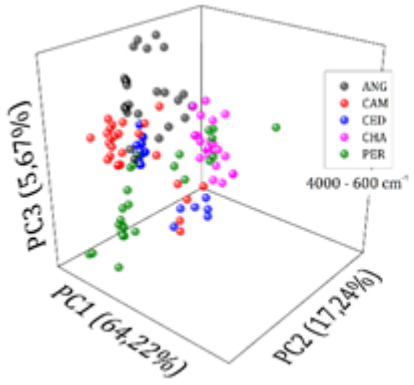
A banda a 1731 cm^{-1} é atribuída às vibrações de alongamento de C=O dos grupos carboxila e acetila na hemicelulose (PANDEY, 2003; SCHWANNINGER, 2004; POPESCU, 2009; POLETO, 2012). As bandas de 1597, 1510 cm^{-1} são atribuídas a vibrações de alongamento ou flexão de C=C, C-O, presentes na lignina (PANDEY, 2003; POPESCU 2009; CHEN, 2010; POLETO, 2012).

As bandas 1461, 1427, 1371, e 1103 cm^{-1} são características das vibrações de C-H, deformação de C-O, flexão ou alongamento de grupos de lignina e carboidratos (POLETO, 2012). A banda de 1427 cm^{-1} está relacionada a vibrações aromáticas associadas à C-H na deformação plana da celulose. As bandas de 1731, 1371, 1245, 1158, 1103, 1035 cm^{-1} são atribuídas à deformação de C=O, C-H, C-O-C, C-O ou vibrações de alongamento de diferentes grupos em carboidratos (SCHWANNINGER, 2004; POPESCU, 2009). A banda em torno de 892 e 833 cm^{-1} é atribuída a material amorfo na região da celulose (PANDEY, 2003; POPESCU, 2009).

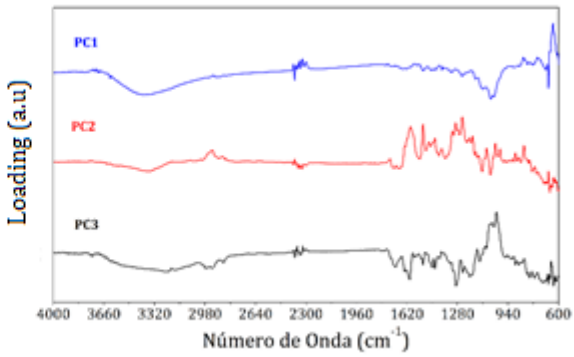
Os espectros médios para cada lote foram pré-tratados utilizando o SVN com objetivo de remover deslocamentos verticais *"Offset"*, centrando os espectros na média, este procedimento é requerido para evitar distinção dos dados devido a variações experimentais (ENGEL, 2013).

Para avaliar a potencial separação das espécies usando FTIR, o conjunto de dados foi analisado pelo método de PCA. O resultado da análise PCA na faixa espectral de 4000-600 cm^{-1} das diferentes espécies de madeiras: angelim-pedra (ANG), cambará (CAM), cedrinho (CED), Champagne (CHA) e peroba (PER) mostrado na figura 11.

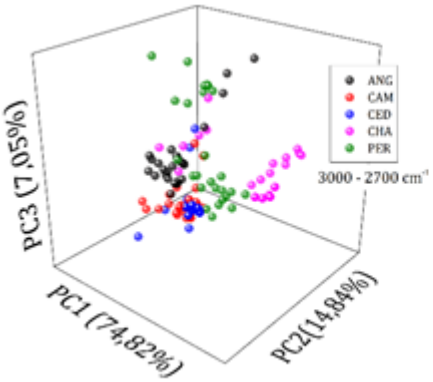
(A)


ANG
CAM
CED
CHA
PER
4000 - 600 cm-1
PC3 (5,67%)
PC1 (64,22%)
PC2 (17,24%)

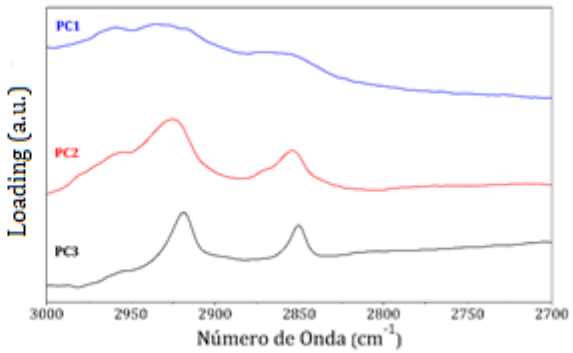
(B)


PC1
PC2
PC3
Loading (a.u)
4000 3660 3320 2980 2640 2300 1960 1620 1280 940 600
Número de Onda (cm-1)

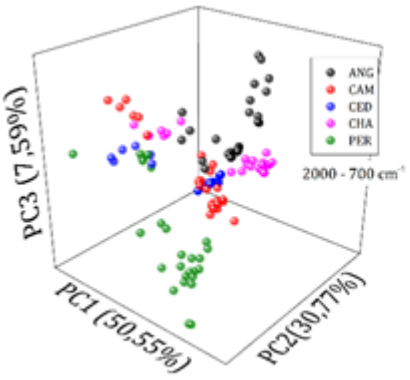
(C)


ANG
CAM
CED
CHA
PER
3000 - 2700 cm-1
PC3 (7,05%)
PC1 (74,82%)
PC2(14,84%)

(D)


PC1
PC2
PC3
Loading (a.u.)
3000 2950 2900 2850 2800 2750 2700
Número de Onda (cm-1)

(E)


ANG
CAM
CED
CHA
PER
2000 - 700 cm-1
PC3 (7,59%)
PC1 (50,55%)
PC2(30,77%)

(F)

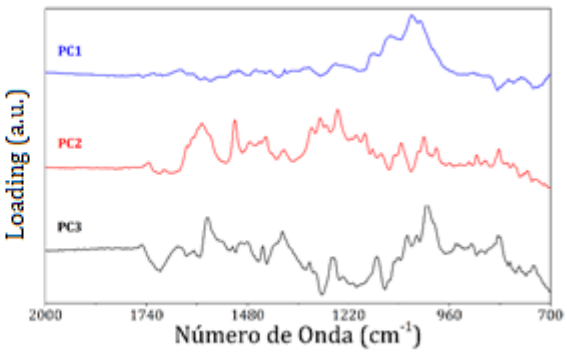

PC1
PC2
PC3
Loading (a.u.)
2000 1740 1480 1220 960 700
Número de Onda (cm-1)

Figura 11: Gráfico de Scores (esquerda) e Loadings (direita) obtidos da análise de PCA do FTIR após SNV. (a,b) 4000-600 cm^{-1}; (c,d) 3000-2700 cm^{-1}; (e,f) 2000-700 cm^{-1} para as espécies: *Hymenolobium petraeum Ducke,* Angelim-pedra (ANG); *Gochnatia polymorpha*, Cambara (CAM); *Erisma uncinatum,* Cedrinho (CED); *Dipteryx odorata,* Champagne (CHA); *Goupia glabra Aubl,* Peroba do Norte (PER). Fonte: Autoria própria.

Foram analisados os intervalos espectrais que incluem os principais picos observados no gráfico de loadings, com intuito de verificar as bandas vibracionais que melhor contribuem para separação das classes. A Figura 11 (a,b) representa gráfico de scores e loadings das três primeiras componentes principais (PCs) obtidas por PCA, nele é exposto os coeficientes da 1ª, 2ª e 3ª PCs, com os respectivos índices de variância 64,22%, 17,24% e 5,67% totalizando 87,13%.

A análise de PCA aplicado ao espectro FTIR completo, figura 11-a, não apresenta uma clusterização das espécies que inclui todas as variáveis das 5 espécies de madeiras, o que não possibilita uma separação clara entre os grupos. Dessa forma, foi analisado os gráficos de scores e loadings entre o intervalo de 3000-2700 cm^{-1} (figura 11 c,d) e o intervalo 2000-700 cm^{-1} (figura 11 e,f) com o objetivo de explorar qual intervalo melhor contribui para a maior variabilidade de cada espécie, focando nas bandas atribuídas aos principais componentes da madeira.

O PCA aplicado ao espectro FTIR 3000-2700 cm^{-1} figura 11 (c) não produz uma clusterização das espécies que inclui todas as variáveis das 5 espécies de madeiras, não possibilita uma separação clara entre os grupos. A figura 11 (d) representa gráfico de loadings das três primeiras componentes principais (PCs) obtidas por PCA, nele é exposto os coeficientes da 1ª, 2ª e 3ª PCs, com os respectivos índices de variância 74,82%, 14,84% e 7,05% totalizando 96,71%.

Dessa forma, avaliando o terceiro intervalo proposto de 2000-700 cm^{-1}foi observado no gráfico de scores, figura 11 (e), uma melhor tendencia de separação das espécies de madeiras. A Figura 11 (f) mostra o gráfico de loadings das três primeiras componentes principais (PCs) obtido por PCA no intervalo de 2000-700 cm^{-1}, o elemento predominante nessa faixa espectral são os componentes principais da madeira: celulose, hemicelulose e lignina, além de possuir

quantidades de extrativos, que podem incluir, lipídios, compostos fenólicos, terpenóides, ácidos graxos, ácidos de resina, carboidratos e ceras (SCHWANNINGER, 2004; SHEBANI, 2008; POPESCU, 2009; POLETO, 2012).

Analisando os três intervalos, foi demonstrado através da figura 11 (e,f) que aponta a observação tridimensional dos scores e gráfico de loadings , como sendo a região para melhor discriminar as cinco espécies de madeira. Demonstrando que a melhor diferenciação entre as espécies de madeira está no intervalo entre 2000-700 cm^{-1} . Essas bandas vibracionais que estão contribuindo para melhor diferenciação são as moléculas de celulose, lignina, hemicelulose e pequenas quantidades de extrativos de madeira que podem provavelmente estar relacionadas a diferenças intra e interespecíficas entre os lotes, ajudando na melhor separação entre os grupos, pois cada espécie de madeira possui uma quantidade de extrativos, dependendo da espécie da madeira, idade da madeira e a localização da madeira na árvore (SHEBANI, 2008).

É verificado no gráfico de loadings (Fig. 11, f) um salto na banda de 1597 cm^{-1} onde se encontra vibração aromática C=C do anel benzeno característico da lignina, também é verificado uma tendência negativa do gráfico em torno de 1035 cm^{-1} onde se encontra vibração aromáticas associadas à C-H na deformação plana da celulose.

Embora se observa uma tendencia de separação para melhor classificação das espécies empregamos os algoritmos de ML nos dados de PCA. Na figura 12, é possível verificar a acurácia para cada um tipo de classificador diferente. A classificação multivariada foi aplicada com o objetivo agrupar amostras com características similares em classes pré-determinadas, ou seja, classificá-las a partir de modelos (matemáticos ou estatísticos) LOOCV e teste de validação empregadas na LDA, QDA, k-NN: fino, médio, cúbico, cosseno, ponderado, subespaço, SVM: fino, médio, grosseiro, linear, quadrático, utilizando o gráfico de scores obtidos pelo PCA no intervalo supracitado.

A validação dos modelos foi testada com LOOCV. O número ideal de PCs com maior acurácia foi obtido através de uma sequência de testes, primeiramente com incremento de 5 PCs e posteriormente de uma e uma, objetivando um compromisso da maior acurácia possível com menor número de PCs. Todos os classificadores apresentaram uma acurácia maior que 80%, entretanto, o classificador SVM-Linear apresentou 100% de acurácia com o menor número de PCs, ou seja, foi observado que com 7 PCs o método SMV-Linear alcançou 100% de acurácia.

O melhor desempenho de classificação é alcançado para os dados na faixa espectral 2000-700 cm^{-1}, como também sugerido pela figura 11 (e,f) , as madeiras de diferentes espécies são discriminadas pelo método SVM-Linear com a utilização de 7 PCs obtendo uma acurácia de 100%, isso pode ser explicado pelo fato dessa região possuir uma maior quantidade de modos vibracionais fundamentais. O SVM-Linear classificou corretamente 100% das espécies de madeira, independente do espectro de alcance, mostrando alta sensibilidade e especificidade.

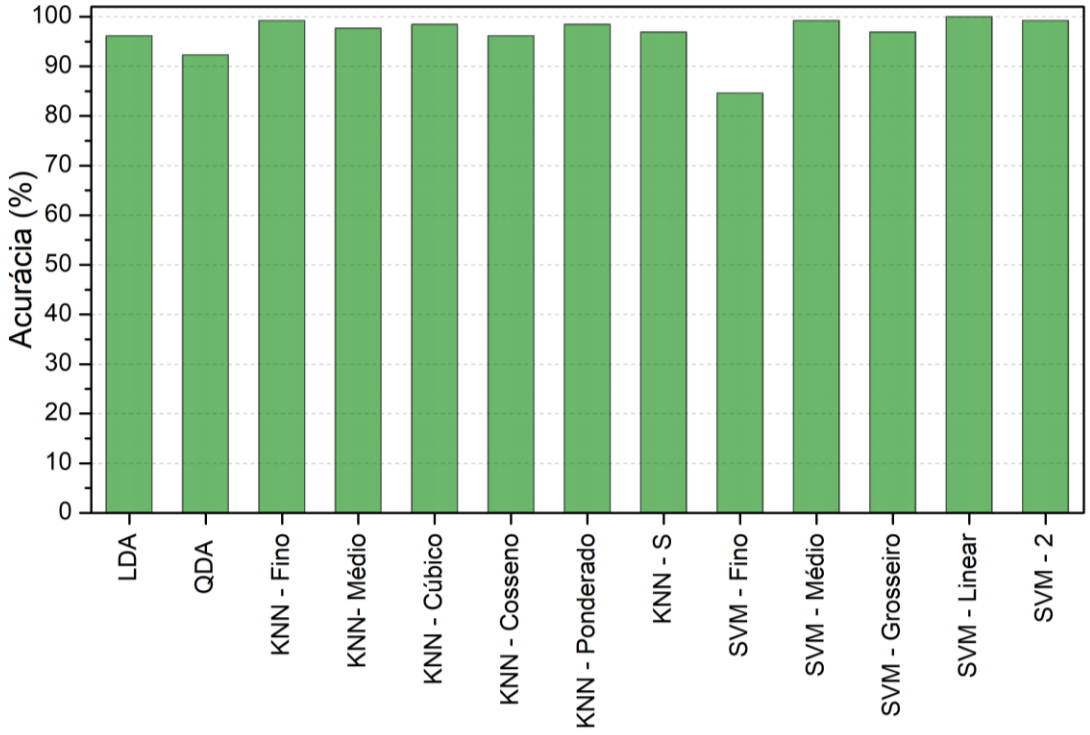


Figura 12: Acurácia para os métodos de classificação obtidos por LOOCV. Fonte: Autoria própria.

A Figura 13 mostra a matriz de confusão pelo método ML e SVM-Linear com 7 PCs no intervalo 2000-700 cm^{-1}. Dentre as abordagens mais frequentes para análise de resultados da validação cruzada está a matriz de confusão, com a indicação da taxa de sucesso e a porcentagem de verdadeiros negativos (VN), verdadeiros positivos (VP), falsos positivos (FP), falsos negativos (FN) previstos, que são utilizados para qualificar as predições e determinar os valores de exatidão, sobre a sensibilidade (taxa de classe verdade) e a especificidade (taxa de classe

prevista) parâmetros fundamentais para avaliar a robustez do protocolo (LARIOS, 2020). As amostragens da diagonal principal são aquelas que foram corretamente classificadas (DE LIMA; BARBORA, 2019) resultando em 100%.

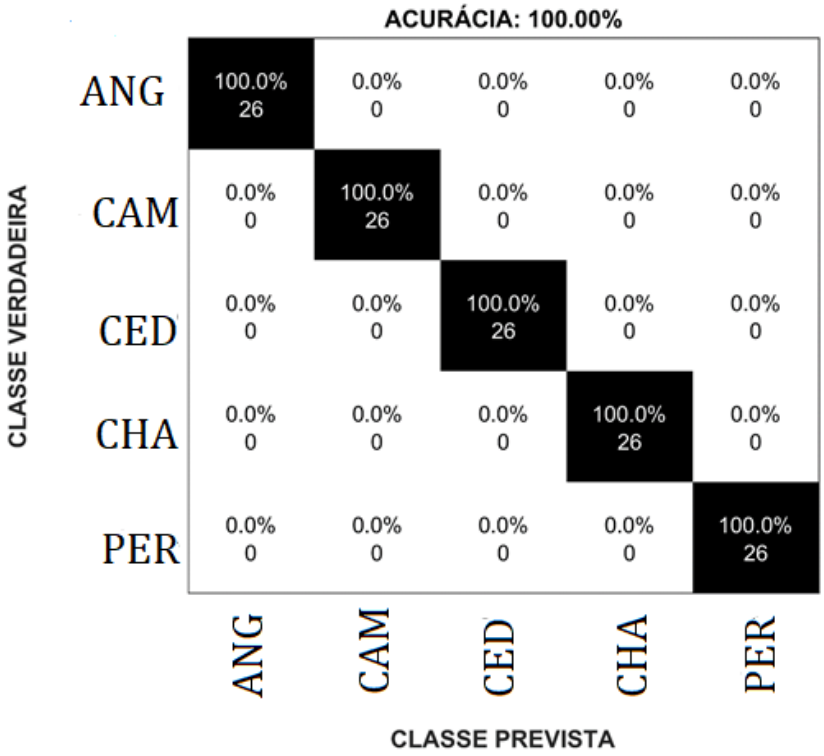


Figura 13: Matriz de confusão pelo método de aprendizagem de máquina pelo método SVM-Linear com 7 PCs no intervalo 2000-700 cm^{-1}. Fonte: Autoria própria.

Por fim, ressaltamos que a espectroscopia de FTIR foi escolhida por ser uma técnica simples para obter a estrutura dos constituintes das cinco diferentes espécies de madeira, em contraste com análise química convencional, esse método requer amostras pequenas, tempo de análise curto, depois de realizado, aplicou-se o SNV aos dados do FTIR, para evitar erros de deslocamentos verticais para não prejudicar a análise dados.

O uso da análise estatística multivariada se justifica pelo motivo dessa técnica conseguir simultaneamente estudar três ou mais variáveis (características) como as espécies de madeiras, simplificando e tornando mais eficiente e completo o que seria feito por inúmeras análises univariada e bivariadas, bem como analisar inter-relações em muitas variáveis e explicá-las em termos de suas componentes. Isso é de suma importância pois conseguimos analisar diferentes

classes com um conjunto grande de variáveis, e verificar quais são essas variáveis que estão contribuindo para tal ou não.

Foi feito LOOCV e variou o número de PCs para evitar a ocorrência de "*overfiting"* e "*underfiting*". Vale ressaltar que a opção pelo classificador SVM-Linear foi não somente pelo alcance de 100% de acurácia, mas que é excelente classificador para problemas relacionados com multiclasse e não-lineares, ou seja, construindo um novo espaço de dados onde a classificação se torne mais simples (separação linear).

Os resultados e discussões dessa pesquisa demonstram que através de uma técnica simples e rápida (FTIR) em combinação com análise multivariada e algoritmos de aprendizado de máquina, tem um grande potencial em classificar diferentes espécies de madeira com um grande conjunto de dados, com maior rapidez se comparado a técnicas convencionais, pois requer amostras pequenas, tempo de preparo de amostra, obtenção de espectros e tempo de análise menor. Além disso, essa pesquisa pode contribuir para grandes empresas no ramo de madeireiras, como por exemplo, a indústria de celulose, pois é possível verificar os principais componentes de madeira (de forma rápida e confiável), e na indústria de celulose a maior dificuldade está ligado a extração da lignina.

6 CONCLUSÃO

O uso do FTIR combinado com PCA e métodos de classificação de aprendizado de máquina mostra ser uma ferramenta de alto potencial para diferenciar espécies de madeira. O melhor resultado é obtido analisando o intervalo na faixa entre 2000-700 cm^{-1} no qual obteve melhor separação no grupo de espécies de madeira estudadas utilizando o classificador SVM-Linear.

O teste de LOOCV indica que pode ser diferenciada com 100% de acurácia as espécies de madeira. O resultado obtido e a simples preparação da amostra mostram alto potencial que o FTIR tem combinado com análise multivariada para ser utilizado na diferenciação entre as espécies de madeira, se comparado com métodos tradicionais em relação ao tempo de preparo da amostra.

Essa pesquisa pode contribuir para grandes empresas no setor de madeireiras, como por exemplo, no controle e fiscalização para comercialização.

7 REFERÊNCIAS

BARATIERI, M.; BAGGIO, P.; FIORI, L.; & GRIGIANTE, M. **Biomass as an energy source: Thermodynamic constraints on the performance of the conversion process**. Bioresource Technology. Vol. 99, p. 7063–7073, 2008.

BELLOU, E.; GYFTOKOSTAS, N.; STEFAS, D.; GAZELI, O. e COURIS S. **Laser- induced breakdown spectroscopy assisted by Machine learning for olive oils classification: the effect of the experimental parameters**. Spectrochim. Acta. 2020.

BUOSO, C.M.; POLI, M.; MATTHAES, P.; SILVESTRIN, L.; ZAFIROPOULOS, D. **Nondestructive wood discrimination: FTIR – Fourier Transform Infrared Spectroscopy in the characterization of diferente wood species used for artistic objects**. International Journal of Modern Physics: Vo. 44, 2016.

CASARIL, E.A; SANTOS, G.C.; MARANGONI, B.S.; LIMA, M.S.; ANDRADE, L.H.C.; FERNANDES, W.S.; INFRAN, O.M.J.; ALVES, N.O.; BORGES, D.G.L.M.; CENA, C.; OLIVEIRA, A.G. **Intraspecific differentiation of sandflies specimens by optical spectroscopy and multivariate Analysis. Journal of biophotonics**. 2020.

CHEN, H.; FERRARI, C.; ANGIULI, M.; YAO, J.; RASPI, C.; BRAMANTI, E. **Qualitative and quantitative Analysis of wood samples by Fourier transform infrared spectroscopy and multivariate Analysis**. Carbohydrate Polymers. Vol. 82, p. 772-778, 2010.

CUI, X.; WANG, Q.; WEI, K.; TENG, G.; XU, X. **Laser-induced breakdown spectroscopy for the classification of wood materials using machine learning methods combined with feature selection.** Plasma Science and Technology, v. 23, n. 5, 2021.

CUI, X.; WANG, Q.; ZHAO, Y.; QIAO, X.; TENG, G. **Laser-induced breakdown spectroscopy (LIBS) for classification of wood species integrated with artificial neural network (ANN)**. Applied Physics B, v. 125, n. 56, 2019.

DESHAVATH, N. N.; VEERANKI, V. D.; GOUD, V. V. **Lignocellulosic feedstocks for the production of bioethanol availability, structure, and composition**. Sustainable Bioenergy. p. 1-19, 2019.

DOS SANTOS, J.X.; VIEIRA, H.C.; SOUZA, D.V. et al. **Discrimination of "Louros" wood from the Brazilian Amazon by near-infrared spectroscopy and machine learning techniques**. Eur. J. Wood Prod. 79, 989–998, 2021.

ENGEL, JASPER ET AL. **Breaking with trends in pre-processing**. TrAC Trends in Analytical Chemistry. V.50, p.96-106, 2013.

ERIKSSON, L.; JOHANSSON, E.; KETTANEH-WOLD, N.; & WOLD, S. **Multi- and megavariate data Analysis: Principles and applications**. Umea, Sweden: Umetrics Academy. Vol. 16, p. 261-262, 1999.

FERNÁNDEZ, M. F.; SAAVEDRA, J. T.; MALLIK, A.; NAYA, S. **A comprehensive classification of wood from thermogravimetric curves**. Chemometrics and Intelligent Laboratory Systems. Vol. 118, p. 159-172, 2012

FERRAZ, A.; BAEZA, J.; RODRIGUEZ, J.; & FREER, J. **Estimating the chemical composition of biodegraded pine and Eucalyptus wood by DRIFT spectroscopy and multivariate Analysis**. Biosource tecnology. Vol. 74, p. 201-212, 2000.

FISHER, R.A. **The use of multiple measurements in taxonomic problems**. Annual Eugenics. Vol,7, p.179-188, 1936.

FIX, E.; HODGES, J. L. **Discriminatory Analysis nonparametric discrimination**: **Consistency properties**. International Statistical Review, vol. 57, n.3, p. 238-247, 1989.

FUKUNAGA, K. **Introduction to Statistical Pattern Recognition**. Academic Press Professional. Inc., San Diego, CA, 1990.

GÍRIO, F.M.; FONSECA, C.; CARVALHEIRO, F.; DUARTE, L. C.; MARQUES, S.; BOGEL-LUKASIK, R. **Hemicelluloses for fuel ethanol: a review**. Bioresour Technol. Vol. 101, p. 4775-4800, 2010.

HOBRO, J.A.; KULIGOWSKI, J.; DOLL, M. **Differentiation of walnut wood species and steam treatmente using ATR-FTIR and partial Least squares discriminant Analysis (PLS-DA)**. Anal Bioanal Chem. Vol. 398, p. 2713-2722, 2010.

HOTELLING, H. **Analysis of a complexo f statistical variables into principal components**. J. Educ. Psychol. Vol.24, p.417-441, 498-520, 1933.

J. D. RODRIGUEZ.; A. PEREZ.; J. A. LOZANO. **Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation.** IEEE Trans. Pattern Anal. Mach.Intell. Vol. 32, p. 569–575, 2010.

JOHN, M. J.; THOMAS, S. **Biofibres and biocomposites**. Carbohydrate Polymers. Vol. 71, p. 343-364, 2008.

JOLLIFFE, I. T. **Discarding variables in a principal component Analysis**. II: Real Data. J. R. Stat. Soc. Vol. 22, p.21-31, 1973.

JOLLIFFE, I.T.; CADIMA, J. **Principal Component Analysis: a review and recent developments**. Phil. Trans. R.Soc. A 374: 20150202, 2016.

KANNO, Y.; KANEKO, H**. Improvement of predictive accuracy in semi-supervised regression Analysis by selecting unlabeled chemical structures**, Chemom. Intell. Syst. Vol. 191, p. 82-87, 2019.

LARIOS, G.S.; NICOLODELLI, G.; SENESI G.S.; RIBEIRO M.C.S.; XAVIER, A.A.P.; MILORI, D.M.B.P.; ALVES, C.Z.; MARANGONI, B.; CENA C**. Laser-induced breakdown spectroscopy as a powerful tool for distinguishing high and low-vigor soybean seed lots**. Food Analytical Methods. Vol. 13, p.1691-1698, 2020.

LARIOS, G.; NICOLODELLI, G.; RIBEIRO, M.; CANASSA, T.; REIS, A. R.; OLIVEIRA, S. L.; ALVES, C. Z.; MARANGONI, B. S.; CENA, C. **Soybean seed vigor discriminaiton by using infrared spectroscopy and machine learning algorithms**. Analytical Methods. Vol. 12, p. 4247-4396, 2020.

LAZZARI, E.; SCHENA, T.; MARCELO, M.C.A.; PRIMAZ, C.T.,;SILVA, A.N.; FERRÃO, M.F.; BJERK, T.; CARAMÃO, E.B**. Classification of biomass through their pyrolytic bio-oil composition using FTIR and PCA analysis**. Industrail Crops & Products. Vol. 111, p. 856–864, 2018.

M. SCHWANNINGER.; J.C RODRIGUES, H. PEREIRA.; B. HINTERSTOISSER. **Effects of short-time vibratory ball milling on the shape of FT-IR spectra of wood and celulose**. Vibrational Spectroscopy. Vol 36, p. 23-40, 2004.

MORRIS, D. R., STEWARD, F. R., & GILMORE, C. A. **Comparative analysis of the consumption of energy of two wood pulping processes.** Energy Conversion and Management. Vol. 41, p. 1557–1568, 2000.

MÜLLER, G., SCHÖPPER, C., VOS, H., KHARAZIPOUR, A., & POLLE, A. **FTIR-ATR spectroscopicanalysis of chages in wood properties during particle- and fibreboard production of hard- and softwood trees**. Bioresources. Vol. 4, 49–71, 2009.

OKADA, T.; TOMITA, S. **An optimal orthonormal system for discriminant analysis.** Pattern Recognition. Vol. 18, p. 139-144, 1985.

OLVEIRA, C. I.; FRANCA, T.; NICOLODELLI, G.; MORAIS, P.C.;MARANGONI, B.; BRACCHETTA, G.; MILORI, D.M.B.P.; ALVES, Z.C; CENA, C. **Fast and Accurate discrimination of Brachiaria brizantha (A.Rich.) stapf seeds by molecular spectroscopy and machine learning.** Agricultural Science & technology. 2021.

OZLEM, O¸C; SEFA, D; ISMAIL, H.B; HASLET, E. **Determination of chemical changes in heat-treated wood using ATR-FTIR and FT Raman spectrometry**, 2016.

PANDEY, K.K.; PITMAN, A.J. **FTIR studies of the changes in wood chemistry following decay by brown-rot and white-rot fungi**. International Biodeterioration & Biodegradation. Vol. 52, p. 151–160, 2003.

PANDEY, K.K. **A note on the influence of extractives on the photo-discoloration and photo-degradation of wood**. Polymer Degradation and Stability. Vol. 87, p. 375-379, 2005.

PFEIL, W.; PFEIL, M. **Estruturas de Madeira**. 6ed. Rio de Janeiro. LTC. 2003.

PLÁCIDO, J.; CAPAREDA, S. **Analysis of alkali ultrasonication pretreatment in bioethanolproduction from cotton gin trash using FT-IR spectroscopy and principal component analysis**. Bioresources Bioprocess. Vol. 1, p. 23, 2014.

PEARSON, K. **On lines and planes of closest fit to systems of points in space**. Phil. Mag. Vol. 2, p. 559-572, 1901

PEREIRA, H.; GRACA, J.; RODRIGUES, J. C. **Wood chemistry in relation to quality.** 2003.

PHILIPP, P.; D'ALMEIDA, M.L.O. **Celulose e Papel. Volume I. Tecnologia de Fabricação da Pasta Celulósica**. Instituto de Pesquisas Tecnológicas do Estado de São Paulo – Centro Técnico em celulose e papel. São Paulo,2ª edição, 1988.

POLETTO, M.; ZATTERA, A. J.; SANTANA, R.M.C.**Structural Differences Between Wood Species: Evidence from Chemical Composition, FTIR Spectroscopy, and Thermogravimetric Analysis**. Journal of Applied Polymer Science (Online), v. 126, p. E336-E343, 2012.

POPESCU, C.M; SINGUREL, G.; POPESCU, M-C; VASILE, C; ARGYROPOULOS, D.S.; WILLFOR, S. **Vibrational spectroscopy and X-ray diffraction methods to establish the differences between hardwood and softwood. Carbohydrate Polymers.** Vol 77, p. 851-857, 2009.

POPESCU, C.M.; POPESCU, M.C.; VASILE, C. **Structural changes in biodegraded lime wood**. Carbohydate Polymers. Vol. 79, p. 362-372, 2010.

POPESCU, M-C; POPESCU, C.M; LISA, G; SAKATA, Y. **Evaluation of morphological and chemical aspects of diferente wood species by spectroscopy and termal methods**. Journal of molecular structure. Vol. 988, p. 65-72, 2011.

RUDAKIYA, D.M.; GUPTE, A. **Assessment of White rot fungus mediated hardwood degradation by FTIR spectroscopy and multivariate Analysis.** Journal of Microbiological Methods. Vol. 157, p. 123-130, 2019.

S. GEISSER. **The Predictive Sample Reuse Method with Applications.** Journal of the American Statistical Association. Vol. 70, p. 320–328, 1975.

SHARMA, V; YADAV, J; KUMAR, R; TESAROVA, D; EKIELSKI, A; MISHRA, P.K. **On the rapid and non-destructive approach for wood identification using ATR-FTIR spectroscopy and Chemometric methods**. Vibrational Spectroscopy. 2020.

SHEBANI, A. N.; VAN REENEN, A.J.; MEINCKEN, M. **The effect wood extractives on the termal stability of different wood-LLDPE composites**.Thermochimica acta. Vol. 481, p. 52-56, 2008.

SILVA, T.F. **Diferenciação de Blendas de PVA/PVP utilizando Análise Multivariada: Limites de Aplicação do Método**. Dissertação de Mestrado, 2021.

SMITH, M. J.; SMITH, A. S. H.; LENNARD, F. **Development of non- destructive methodology using ATR-FTIR with PCA to differentiate between historical Pacific barkcloth**. Jounral of Cultural Heritage. Vol. 39, p. 32-41, 2019.

TRAORÉ, M.; KAAL, J.; CORTIZAS, A.M. **Application of FTIR spectroscopy to the characterization of archeological wood.** Spectrochimica Acta. Vol. 153, p. 63–70, 2016.

TRAORÉ, M. KALL, J. CORTIZAS. A.M. **Differentiation between pine woods according to species and growing location using FTIR-ATR. Wood** Sci Technol. Vol. 52, p. 487-504, 2018.

VAN SOEST, P. J. **Use of detergents in the Analysis of fibrous feeds. II. A rapid method for the determination composition of fiber and lignin**. Journal of the Association of Official Analytical Chemists. Vol.46, p. 829-835, 1963.

VAN SOEST, P. J.; WINE, R. H. **Use of detergents in the Analysis of fibrous feeds. IV. Determination of plant cell-wall constituents**. Journal of the Association of Official Analytical Chemists. Vol. 50, p. 50-55, 1967

VAPNIK, V. **Statistical Learning Theory**. Willey, 1998.

VARTANIAN, E; BARRES, O; ROQUE, C. **FTIR spectroscopy of woods: a new approach to study the weathering of the carving face of a sculpture**. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2014.

WATIKINS, D.; NURUDDIN, MD.; MAHESH, H.; TCHERBI-NARTEH, A.; JEELANI, S. **Extraction and characterization of lignina from diferent biomass resources.** Journal of Materials Research and Tecnology. Vol. 4, p.26-32, 2015.

WANGAARD, F.F. **Wood: its structure and properties**. The Pennsylvania State University, USA, 1979.

YOUNG, F.; MINDNESS, S.; GRA, Y, R.; BENTUR, A. **The Science and Technology of Civil Engineering Materials**. Frentice Hall, USA, 1998.

.

ZHOU, C.H.; XIA, X.; LIN, C. X.; TONG, D.S.; BELTRAMINI, J. **Catalytic conversion of lignocellulosic biomass to fine chemicals and fuels**. Chem Soc Rev. Vol. 40, p. 5588-5617, 2011.