Serviço Público Federal
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul

Curso de FÍSICA – BACHARELADO

Trabalho de Conclusão de Curso

Classificação de Argilas com Espectroscopia no Infravermelho Médio e Aprendizagem de Máquina

Júnia Maria da Silva Viana Mendes

Trabalho de Conclusão de Curso apresentado ao curso de Física Bacharelado do Instituto de Física (INFI), da Universidade Federal de Mato Grosso do Sul (UFMS).

Orientador: Prof. Dr. Samuel Leite de Oliveira

Campo Grande – MS

Julho/2025

“42”

(O Guia do Mochileiro das Galáxias)

AGRADECIMENTOS

Agradeço primeiramente à Universidade Federal de Mato Grosso do Sul (UFMS), pela oportunidade de cursar a graduação. Sou grata também ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro concedido por meio da bolsa de Iniciação Científica. Aos meus familiares, amigos e colegas do curso de Física, minha eterna gratidão. Aos integrantes e ex-integrantes do Grupo de Óptica e Fotônica, obrigada por compartilharem diariamente conhecimento, amizade e inspiração.

Ao meu orientador, Prof. Dr. Samuel Leite de Oliveira, agradeço profundamente pela confiança, orientação generosa e pela dedicação incansável. Estendo também meu agradecimento ao Dayvison Felismindo Lima e ao Prof. Dr. Roberto Weider de Assis Franco, cujas contribuições foram essenciais para a concretização deste projeto.

Aos meus amigos Gustavo Sgurscow, Nicolly Yonamine, Angelyck Justen e Sophia Brück, que estiveram comigo nos momentos bons e difíceis, obrigada pela presença constante, pelas escutas sinceras e pela amizade que me fortaleceu ao longo desses anos.

À minha família, meu alicerce, agradeço pelo amor incondicional e apoio inabalável. Em especial, aos meus pais, José Carlos Viana Mendes e Judite Francisca da Silva Viana Mendes, e aos meus irmãos Júlio César da Silva Viana Mendes e Joel Carlos da Silva Viana Mendes que com muito esforço, amor e exemplo me ensinaram o valor do conhecimento e da perseverança.

À minha namorada, Inglid Macêdo, minha parceira de vida, obrigada pela paciência, cumplicidade e por acreditar em mim mesmo nos dias em que eu duvidei.

A todas essas pessoas, meu sincero e emocionado muito obrigada.

RESUMO

A caracterização e a classificação de materiais cerâmicos naturais, como a argila, são fundamentais para aplicações arqueológicas, ambientais e industriais, demandando métodos rápidos e precisos. Este trabalho teve como objetivo classificar amostras de argila coletadas no entorno do sítio arqueológico Alto do Bonfim (GO), utilizando espectroscopia no infravermelho por transformada de Fourier (FTIR) associada à análise multivariada e a algoritmos de aprendizagem de máquina. Três amostras (A1, A2 e A3) foram analisadas por FTIR, e os espectros obtidos passaram por pré-processamento por variância normal padrão (*Standard Normal Variate,* SNV), redução de dimensionalidade via Análise de Componentes Principais (PCA) e classificação utilizando os algoritmos k-vizinhos mais próximos (*K-Nearest Neighbors*, KNN), máquina de vetores de suporte (*Support Vector Machine,* SVM) e Análise Discriminante Linear (*Linear Discriminant Analysis*, LDA). Apesar da similaridade entre os espectros das amostras, os modelos conseguiram evidenciar uma separação progressiva, especialmente da amostra A3, em função de sua composição química distinta. A análise mostrou que a região espectral entre 600 e 1500 cm^{-1} apresenta maior variabilidade relevante para a discriminação. O algoritmo LDA apresentou 100% de acurácia em ambos os intervalos espectrais, superando o KNN e o SVM, demonstrando ser o método mais eficaz. Conclui-se que a combinação de FTIR, PCA e LDA constitui uma abordagem robusta e eficiente para a classificação automatizada de argilas, com potencial de aplicação em diversas áreas do conhecimento.

Palavras-chave: Espectroscopia FTIR, Argila, Aprendizado de Máquina, Análise das componentes principais (PCA).

ABSTRACT

The characterization and classification of natural ceramic materials, such as clay, are essential for archaeological, environmental, and industrial applications, requiring fast and accurate methods. This work aimed to classify clay samples collected near the Alto do Bonfim archaeological site (GO) using Fourier Transform Infrared (FTIR) spectroscopy combined with multivariate analysis and machine learning algorithms. Three samples (A1, A2, and A3) were analyzed by FTIR, and the resulting spectra were preprocessed using Standard Normal Variate (SNV), followed by dimensionality reduction through Principal Component Analysis (PCA) and classification using *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), and *Linear Discriminant Analysis* (LDA) algorithms. Despite the visual similarity of the spectral data, the models gradually revealed a separation - especially for sample A3, which has a distinct chemical composition. The analysis indicated that the 600–1500 cm^{-1} spectral region contains the most relevant variability for discrimination. The LDA algorithm achieved 100% accuracy in both spectral ranges, outperforming KNN and SVM, and proving to be the most effective method. It is concluded that the combination of FTIR, PCA, and LDA provides a robust and efficient approach for automated clay classification, with potential applications in various scientific fields.

Keywords: FTIR Spectroscopy, Clay, Machine Learning, Principal Component Analysis (PCA).

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ACRÔNIMOS E SIGLAS

INFI	Instituto de Física
UFMS	Universidade Federal do Mato Grosso do Sul
PCA	Análise das componentes principais (*Principal Component Analysis*)
PC	Componente principal (Principal componente)
IR	Infravermelho
FTIR	Espectroscopia no Infravermelho por Transformada de Fourier
LDA	Análise Discriminante Linear (*Linear discriminant analysis*)
KNN	K- vizinhos mais próximos (*K – Nearest Neighbors*)
SVM	Máquina de Vetores de Suporte (*Support Vector Machine*)
SNV	Variância Normal Padrão (*Standard Normal Variate*)
CV	Validação Cruzada (*Cross-validation*)

SUMÁRIO

1. INTRODUÇÃO E JUSTIFICATIVAS

O entendimento a respeito do comportamento humano por trás da produção e utilização de cerâmicas antigas se inicia com a compreensão dos recursos utilizados e de suas propriedades físico-químicas.[1,2,3,4,5] O principal desses recursos é a argila, um material de origem mineral composto predominantemente por silicatos de alumínio hidratados, como caulinita, ilita e esmectita, entre outros minerais que conferem propriedades distintas às cerâmicas produzidas a partir delas.[1,2,6,7]

A classificação e a diferenciação entre diferentes tipos de argilas são de grande relevância para áreas como a arqueometria, geologia, engenharia e ciência dos materiais. Essas classificações auxiliam não apenas na compreensão dos aspectos tecnológicos das produções cerâmicas antigas, mas também para a identificação da origem geográfica e das rotas comerciais associadas aos artefatos.[8] Para isso, diversas técnicas analíticas têm sido empregadas, entre as quais se destacam a difração de raios X (XRD), a fluorescência de raios X (XRF), a espectroscopia de emissão óptica com plasma indutivamente acoplado (ICP-OES), espectroscopia Raman e, especialmente, a espectroscopia na região do infravermelho.[1,9,3,10,11,6,7,12]

A espectroscopia no infravermelho por transformada de Fourier (FTIR), em particular na modalidade de reflectância total atenuada (ATR), tem se mostrado uma ferramenta extremamente útil e eficaz para a caracterização de materiais cerâmicos. A técnica permite a identificação de compostos minerais com base na absorção de radiação infravermelha por diferentes grupos funcionais, sendo capaz de distinguir tanto fases cristalinas como pseudo-amorfas. Essa capacidade torna a FTIR especialmente valiosa no estudo de artefatos arqueológicos submetidos a processos térmicos.[7,8,13]

Além de ser uma técnica rápida, de baixo custo e minimamente destrutiva — características ideais para aplicações em patrimônio cultural —, a FTIR produz dados espectrais ricos e complexos.[9,14] Entretanto, para que tais dados se tornem informações úteis sobre a amostra, é necessário aplicar métodos estatísticos e computacionais que permitam a extração de padrões e relações relevantes.[13,14]

Nesse sentido, métodos quimiométricos, como a Análise de Componentes Principais (PCA) e a Análise Discriminante Linear (LDA) têm sido amplamente utilizados na interpretação de espectros infravermelhos. Mais recentemente, o uso de algoritmos de aprendizagem de máquina, como k-vizinhos mais próximos (KNN) e máquina de vetores de suporte (SVM), tem se mostrado promissor para tarefas de classificação, predição e quantificação de fases minerais em argilas.[4,14,12,15]

Modelos supervisionados de aprendizado de máquina já foram aplicados com sucesso na classificação de fragmentos cerâmicos com base em dados elementares e espectrais, alcançando índices de acurácia superiores a 90%, mesmo com amostras de difícil diferenciação visual ou química.[16,17] Além disso, estudos como o de Du Plessis *et al*. demonstraram que algoritmos de aprendizado de máquina treinados com dados de FTIR e XRF poderiam substituir métodos mais caros e demorados, como a difração de raios X, na quantificação de minerais como caulinita e haloisita.[12]

Dessa forma, a combinação entre espectroscopia FTIR e algoritmos de aprendizado de máquina representa uma alternativa eficiente e inovadora para a classificação de argilas. Esta abordagem não apenas permite uma análise mais rápida e acessível, como também aumenta significativamente a capacidade de diferenciação entre amostras com composição mineralógica semelhante.

Neste trabalho, propõe-se a utilização da espectroscopia no infravermelho por transformada de Fourier (FTIR), aliada a métodos quimiométricos e algoritmos supervisionados de aprendizagem de máquina, para a classificação de amostras de argila. Busca-se avaliar o potencial dessa combinação para diferenciar amostras com base em suas composições minerais e características espectrais, contribuindo com ferramentas modernas para a análise de materiais cerâmicos e para os estudos de arqueometria.

1.1 Espectroscopia de infravermelho médio (FTIR)

A espectroscopia no infravermelho é uma técnica amplamente utilizada para caracterização química de diferentes moléculas, baseada na absorção de radiação

eletromagnética na região do infravermelho médio (aproximadamente 400 a 4000 cm^{-1}). Nessa região, as ligações covalentes das moléculas absorvem energia, promovendo transições vibracionais para níveis de energia mais elevados.[18]

Esse processo de absorção é quantizado, isto é, uma molécula só pode absorver fótons com energias específicas, correspondentes às suas frequências naturais de vibração. A energia absorvida aumenta a amplitude dos movimentos vibracionais das ligações químicas.[18]

Cada tipo de ligação molecular apresenta modos vibracionais específicos, que são ativados por diferentes quantidades de energia. Dessa forma, cada molécula tem um padrão de absorção único na faixa do infravermelho. Mesmo que substâncias diferentes tenham o mesmo tipo de ligação química, os espectros ainda serão diferentes por estarem em ambientes químicos diferentes.[19]

O processo de absorção da radiação no infravermelho pode ser entendido como interação entre a molécula e o campo elétrico da radiação eletromagnética. Para que ocorra a transferência de energia, é necessário que a ligação química apresente um dipolo elétrico que mude na mesma frequência da radiação incidente. Nessa condição, a radiação pode interagir com a matéria alterando seus estados vibracionais ou rotacionais.[19,20]

A absorção de radiação infravermelha, portanto, depende da variação do momento de dipolo da molécula durante a vibração. No caso de moléculas homonucleares, essa variação não ocorre, o que impede a absorção da radiação infravermelha. Por isso, ligações simétricas com grupos idênticos ou quase idênticos em ambas as extremidades não absorvem no infravermelho.[19]

1.1.1 Instrumentação

Os espectrômetros de infravermelho por transformada de Fourier (FTIR) são instrumentos que permitem a obtenção dos espectros de absorção no infravermelho de um composto. O princípio de funcionamento baseia-se na diferença de caminho ótico entre dois feixes originários de uma fonte IR. Esses feixes, ao se recombinarem,

geram padrões de interferência construtiva e destrutiva. O sinal resultante, denominado interferograma, é então convertido para o domínio da frequência por meio de uma Transformada de Fourier, permitindo a identificação das bandas espectrais características.[19,20,21]

O espectrômetro FTIR utiliza um interferômetro de Michelson, conforme ilustrado na Figura 1, para gerar variações nos caminhos ópticos da radiação infravermelha incidente. Nesse sistema, a radiação colimada proveniente da fonte atinge um divisor de feixes, que a separa em dois feixes perpendiculares: um deles é refletido em 90° em direção a um espelho fixo, enquanto o outro segue em linha reta até um espelho móvel. Após serem refletidos, ambos os feixes retornam ao divisor, onde se recombinam, gerando um feixe com padrões de interferência. A diferença de caminho óptico entre os dois feixes — causada pelo movimento do espelho móvel — provoca uma diferença de fase relativa entre eles. Isso resulta em interferência construtiva (sinal máximo no detector) quando a diferença de caminho é igual a múltiplo inteiro do comprimento de onda da luz, e interferência destrutiva (sinal mínimo) quando a diferença de caminho é igual a múltiplo ímpar de metade do comprimento de onda. O feixe emergente, agora com a informação interferométrica, atravessa a amostra e, posteriormente, é detectado possibilitando a obtenção do espectro de absorção no espectrômetro.[19,21,22]

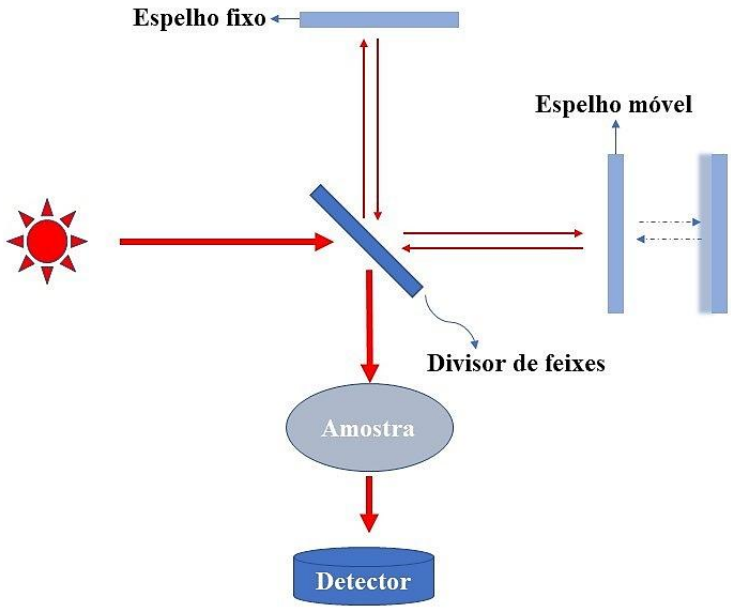

Figura 1. Esquema de interferômetro de Michelson de espectrômetro no infravermelho.
Fonte: elaboração própria.

Quando o feixe recombinado, contendo toda a energia emitida pela fonte no intervalo de 4000 a 400 cm^{-1}, atravessa a amostra, as moléculas presentes absorvem simultaneamente, de forma seletiva, as frequências correspondentes às suas vibrações características. Em seguida, esse feixe é então direcionado a um fotodetector, que registra o sinal resultante (o interferograma), um registro no domínio do tempo que contém todas as informações espectrais do infravermelho. Esse sinal é então comparado a um interferograma de referência, obtido a partir do material do divisor de feixes, cuja contribuição é subtraída por meio de processamento computacional.[19,21,22]

Por fim, uma etapa fundamental na operação dos espectrômetros FTIR é a aplicação da Transformada de Fourier ao interferograma, que é o sinal analógico gerado pelo detector ao registrar a intensidade da radiação com e sem a presença da amostra. Esse tratamento matemático converte o sinal temporal, complexo, em suas

componentes de frequência, revelando as diferentes bandas de absorção contidas no interferograma. O resultado é um espectro digital de absorção óptica no infravermelho, que relaciona a intensidade de absorção da amostra em função do número de onda, permitindo a identificação das propriedades moleculares.[18,21,23]

1.1.2 Reflectância total atenuada (ATR)

A técnica de Reflectância Total Atenuada (ATR), é amplamente utilizada na região do infravermelho para obtenção de espectros de absorção de materiais opacos e filmes finos, proporcionando resultados comparáveis aos obtidos por espectros de transmissão.[18,21]

Seu princípio baseia-se no fenômeno da reflexão interna total, que ocorre quando a radiação infravermelha incide sobre a interface entre um cristal de alto índice de refração e uma amostra de índice de refração inferior, desde que o ângulo de incidência exceda o ângulo crítico, determinado pelos índices das duas superfícies.[18,21]

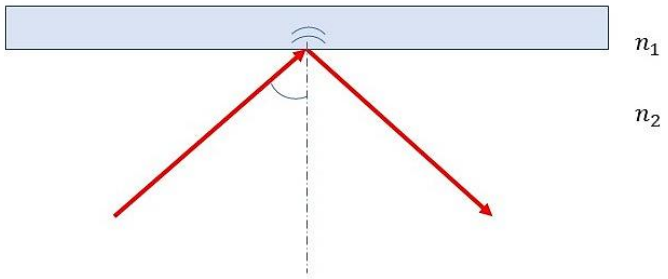

Figura 2. Esquema de Reflectância Total Atenuada (ATR).
Fonte: elaboração própria.

Nessa condição, o feixe de luz é totalmente refletido dentro do cristal, mas uma parte da sua energia do feixe incidente penetra superficialmente na amostra por meio de uma onda evanescente. Essa onda pode ser parcialmente absorvida pela amostra,

dependendo do comprimento de onda e das propriedades ópticas do material, em especial das variações no índice de refração. Como resultado, a intensidade da luz refletida é atenuada em regiões específicas do espectro, o que permite a obtenção do espectro de absorção em função do comprimento de onda correspondente.[18,21]

1.2 Aprendizagem de máquina

A aprendizagem de máquina é um campo de estudo que se dedica ao desenvolvimento de sistemas capazes de melhorar seu desempenho com base na experiência. Para isso, utiliza algoritmos computacionais que simulam aspectos da inteligência humana, com o objetivo de auxiliar na resolução de problemas complexos. Esse processo envolve a modificação do comportamento do sistema a partir da análise de dados e da identificação de padrões, sendo guiado por princípios estatísticos e teóricos que descrevem e explicam o funcionamento dos algoritmos de aprendizagem.[24,25]

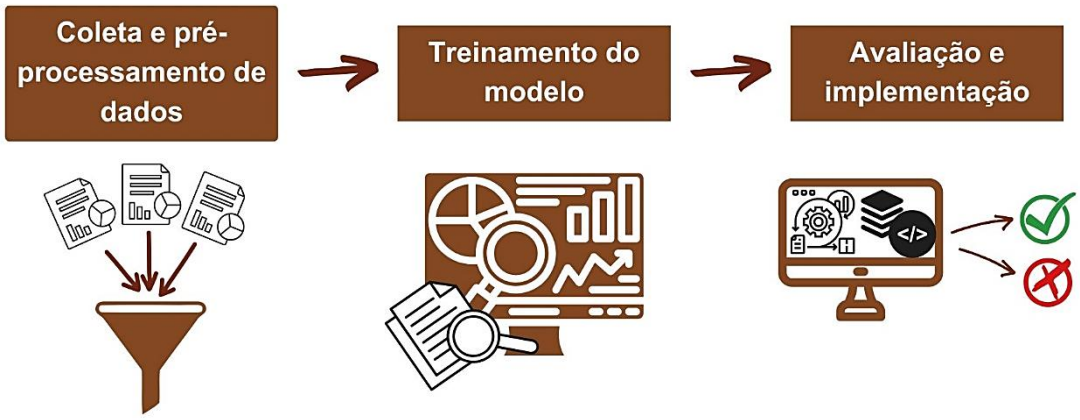

Figura 3. Esquema ilustrativo do processo de aprendizado de máquina. Fonte: elaboração própria.

1.2.1 Análise Multivariada

A aprendizagem de máquina tem grande relevância, especialmente na interpretação de dados experimentais, ao empregar métodos estatísticos e

matemáticos para analisar conjuntos de dados altamente complexos.[14] Esse recurso é particularmente importante na espectroscopia de infravermelho, onde há uma enorme quantidade de variáveis — como as intensidades de radiação associadas aos números de onda — exigindo uma abordagem multivariada para análise eficiente. Nesse contexto, esses algoritmos permitem extrair informações relevantes a partir de grandes volumes de dados, contribuindo para a compreensão e interpretação de fenômenos físico-químicos.

Os algoritmos de aprendizagem de máquina utilizados para identificar semelhanças e diferenças entre diferentes tipos de amostras, com o objetivo de agrupá-las e classificá-las, são divididos em duas categorias principais: métodos supervisionados e não supervisionados de reconhecimento de padrões.[14]

Nos métodos supervisionados, cada amostra é previamente associada a uma classe predefinida, e essa informação é utilizada durante o treinamento para construir modelos de classificação. Esses modelos combinam dados espectrais quantitativos com informações qualitativas provenientes da rotulagem das amostras. Por meio de métodos matemáticos e estatísticos, os modelos aprendem a reconhecer padrões e a realizar a classificação de novos dados com base em comportamentos previamente observados.[4,14]

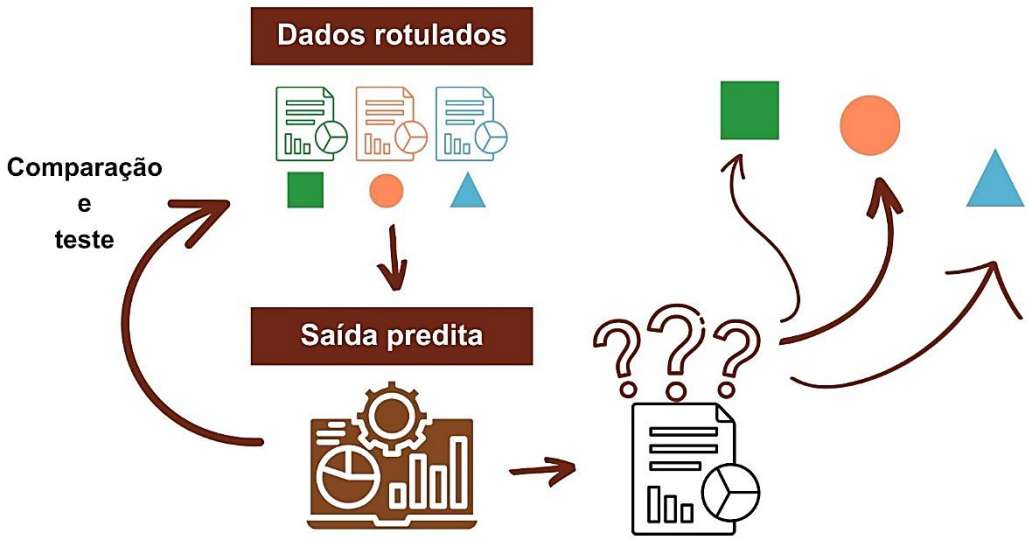

Figura 4. Esquema de aprendizado Supervisionado.

Fonte: elaboração própria.

Já os métodos não supervisionados não utilizam informações prévias sobre as classes das amostras. Tratam-se de técnicas exploratórias que analisam apenas os dados de entrada, agrupando as amostras com base nas similaridades observadas nos dados experimentais, como os espectros, revelando estruturas e padrões naturais sem conhecimento inicial sobre sua categorização.[4,14]

Figura 5. Esquema representativo do aprendizado não supervisionado.

Fonte: elaboração própria.

Antes da escolha do classificador, é necessário realizar um pré-tratamento dos dados previamente organizados, com o objetivo de minimizar variações indesejáveis que não seriam naturalmente corrigidas durante a análise. Nesse contexto, a técnica variância normal padrão (SNV) consiste em uma transformação matemática que corrige efeitos aditivos e multiplicativos, sendo especialmente eficaz na atenuação de interferências causadas por espalhamento de luz e pelo tamanho das partículas sólidas. Esse método consiste em autoescalar cada linha da matriz de dados original, normalizando-a individualmente.[14,26]

1.2.2 Análise das Componentes Principais

A análise das componentes principais (PCA, do inglês *Principal Component Analysis*) é um método não supervisionado de reconhecimento de padrões, ou seja, não requer nenhum conhecimento prévio sobre as classes das amostras. Os resultados são obtidos exclusivamente a partir dos dados experimentais das amostras.[4,14]

O principal objetivo da PCA é reduzir a dimensionalidade dos dados preservando, ao máximo, a variância original, ou seja, a informação relevante presente no conjunto de dados. Isso é feito por meio da transformação dos dados originais em um novo sistema de coordenadas, cujos eixos (componentes principais) correspondem às direções de maior variabilidade dos dados. Desse modo, o método é capaz de realçar as informações relevantes de modo a torná-las mais evidentes a inspeção visual.[14]

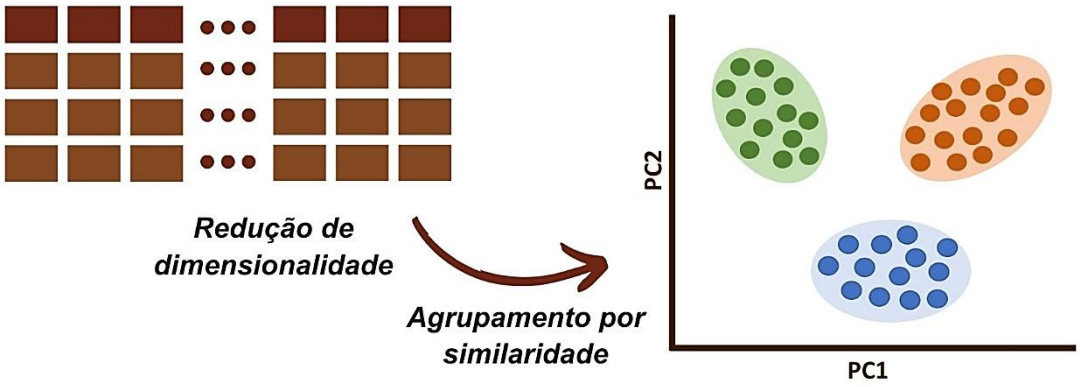

Figura 6. Esquema de Análise das Componentes principais (PCA).

Fonte: elaboração própria.

Para atingir esse objetivo, o PCA estabelece um novo conjunto de variáveis (componentes principais) formado por combinações lineares das variáveis originais. Essas novas variáveis são calculadas de maneira que as primeiras componentes concentrem a maior parte da variância total do conjunto dos dados originais, agrupando, implicitamente, as variáveis que apresentam comportamentos semelhantes.[14,27,28]

1.2.3 Análise supervisionada

Na análise supervisionada, utiliza-se um conjunto de treinamento formado por amostras previamente classificadas. Os dados experimentais dessas amostras são usados para desenvolver um modelo empírico ou uma regra de classificação. Esses métodos são chamados supervisionados porque o conhecimento prévio das classes orienta a construção dos critérios de discriminação. Antes de sua aplicação, o modelo deve ser validado com um conjunto de teste, composto por amostras externas, a fim de avaliar sua capacidade preditiva. Se os resultados forem satisfatórios, o modelo pode ser utilizado para classificar novas amostras com base na propriedade de interesse.

A Análise Discriminante Linear (*Linear Discriminant Analysis,* LDA), é um classificador que utiliza funções discriminantes para criar regras de diferenciação entre classes. O método parte do pressuposto de que cada classe segue uma distribuição normal e compartilha uma mesma matriz de covariância. O objetivo do LDA é encontrar uma ou mais direções (retas ou hiperplanos, dependendo da dimensionalidade dos dados) que maximizem a separação entre os centros das classes, ao mesmo tempo que minimizem a dispersão dentro de cada uma delas. Dessa forma, busca-se definir uma superfície de decisão que separe linearmente as amostras de diferentes classes: amostras de uma classe tendem a se posicionar de um lado, enquanto as da outra classe ficam do lado oposto.[27]

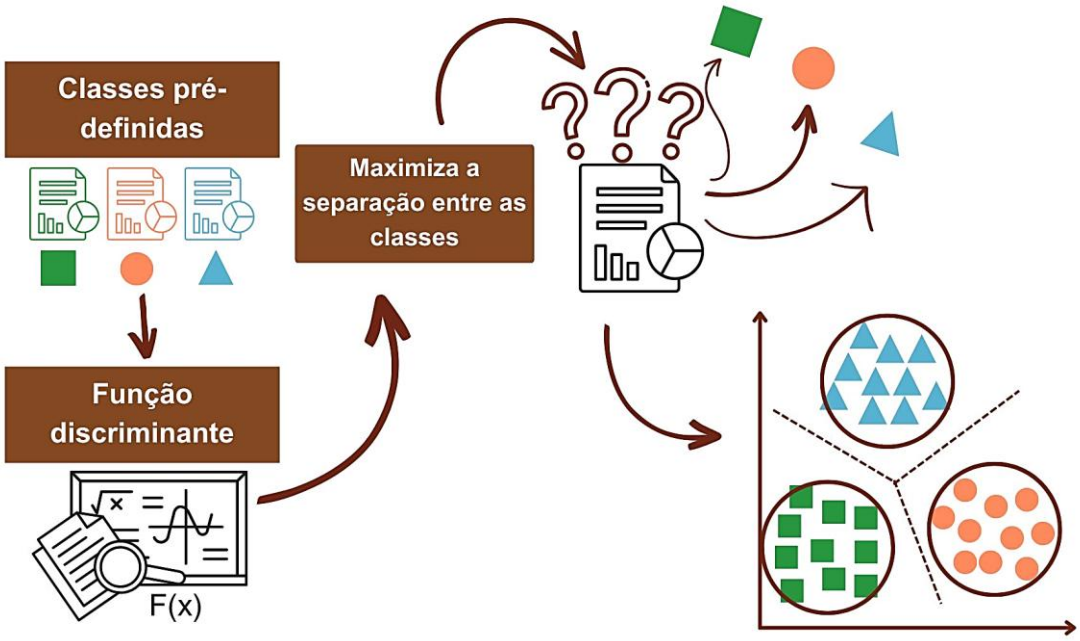

Figura 7. Esquema de Análise de Discriminante Linear (LDA).
Fonte: elaboração própria.

O k-vizinhos mais próximos (KNN) é um classificador baseado na observação de padrões em um conjunto de dados.[29] Ele utiliza a distância euclidiana multivariada para medir a proximidade entre os pontos no espaço amostral.[27] Na fase de classificação, calcula-se a distância entre cada par de amostras, permitindo classificar uma nova amostra desconhecida com base em sua vizinhança em relação às amostras do conjunto de treinamento — ou seja, nas amostras mais semelhantes segundo essa métrica de distância.[14]

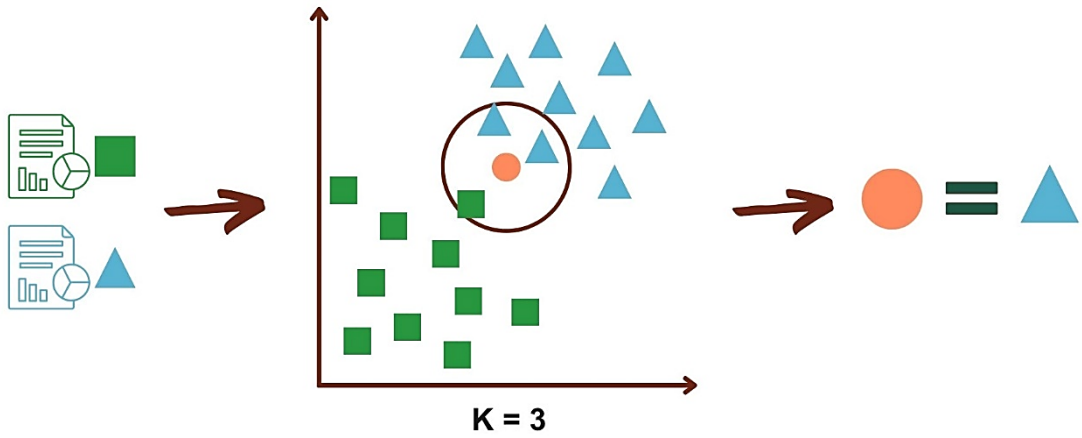

Figura 8. Esquema do método de classificação *K-Nearest Neighbors* (KNN).

Fonte: elaboração própria.

Para isso, cada amostra do conjunto de treinamento é, em determinado momento, excluída e classificada com base nas demais. Calculam-se as distâncias euclidianas entre a amostra excluída e todas as outras, organizando-se essas distâncias em ordem crescente. Em seguida, os k-vizinhos mais próximos são identificados, e a amostra é então atribuída à classe que tiver mais vizinhos. Em caso de empate, a decisão é tomada considerando a menor distância acumulada entre a amostra e seus vizinhos mais próximos.[14,27]

A Máquina de Vetores de Suporte (SVM - *Support Vector Machine*) é um algoritmo de aprendizado de máquina supervisionado utilizado para classificação. Sua principal função é encontrar um hiperplano ótimo que separe as classes no espaço de características, maximizando a margem entre os grupos mais próximos. Essa margem é definida pela distância entre o hiperplano e os vetores de suporte, as amostras mais próximas da fronteira de separação, que são as únicas consideradas pelo SVM na definição da fronteira decisória. A decisão é tomada por meio de uma função que indica de que lado do hiperplano a nova amostra se encontra. O desempenho do SVM está ligado à obtenção de uma margem ampla, o que indica maior confiança na separação entre as classes.[27,30]

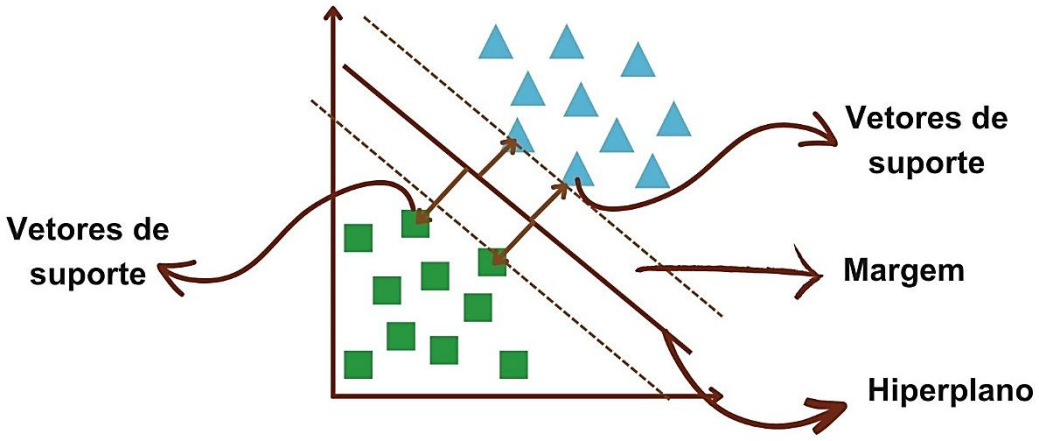

Figura 9. Esquema ilustrativo da Máquina de Vetores de Suporte (SVM).

Fonte: elaboração própria.

2. OBJETIVOS

2.1 Objetivo geral

O presente trabalho tem como objetivo aplicar a espectroscopia no infravermelho médio por transformada de Fourier, em conjunto com técnicas de análise multivariada, para a classificação de diferentes amostras de argila.

2.2 Objetivos específicos

- Realizar medições espectroscópicas utilizando FTIR para caracterizar diferentes amostras de argila
- Usar a Análise das Componentes Principais (PCA) para identificar as variáveis que mais contribuem para a separação e/ou agrupamento das amostras;
- Empregar algoritmos de Aprendizado de Máquina para analisar os dados derivados da PCA e automatizar o processo de classificação das amostras.

3. METODOLOGIA

O experimento foi realizado em três etapas. Na primeira, foram realizadas as medições FTIR para obtenção dos espectros de cada amostra. Em seguida, os dados foram submetidos a tratamento e à análise das componentes principais. Por fim, aplicou-se a etapa de classificação por meio de algoritmos de aprendizagem de máquina, KNN, SVM e LDA.

3.1 Extração de preparação das amostras

Foram analisadas três amostras de argila, denominadas A1, A2 e A3, coletadas em um raio de 17 km ao redor do sítio arqueológico Alto do Bonfim, registrado sob a sigla GO-JA-007 pelo Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN). Esse sítio está localizado no município de Perolândia em Goiás (Figura 10). Após a coleta, as amostras de argila foram destorroadas, homogeneizadas e peneiradas por meio de uma malha com abertura de 106 µm.[1]

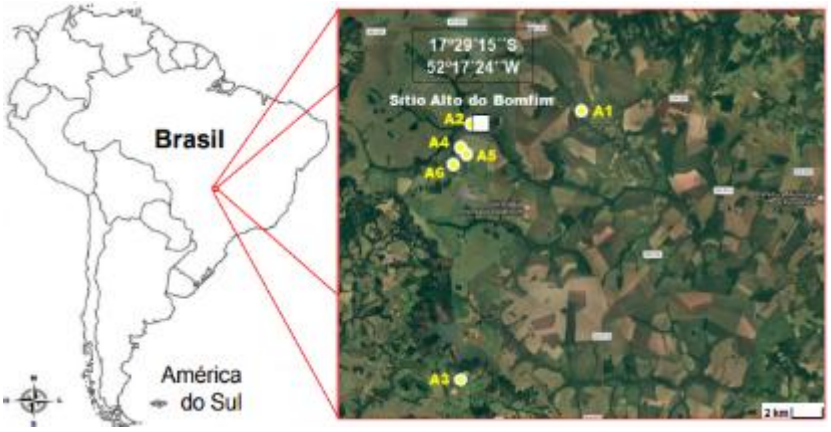

Figura 10. Localização geográfica do Sítio Alto do Bonfim e dos pontos de coleta das amostras de argilas (A1, A2, A3), no município de Perolândia, GO. Fonte: Referência.[1]

3.2 Espectroscopia no infravermelho por transformada de Fourier - FTIR

As análises de espectroscopia de infravermelho na região do infravmelho foram realizadas utilizando um espectrômetro FTIR (Spectrum 100, Perkin Elmer), localizado no Laboratório de Óptica e Fotônica do Instituto de Física da Universidade Federal do Mato Grosso do Sul. As medições foram feitas no modo de transmissão, com o acessório de reflectância total atenuada (ATR). Os espectros foram adquiridos no intervalo espectral de 4000 a 600 cm^{-1}, com resolução de 4 cm^{-1}, velocidade de 0,1 cm/s, passo espectral de 0,5 cm^{-1} e um total de 10 varreduras. As amostras foram posicionadas sobre o cristal do acessório ATR e cuidadosamente pressionadas para garantir o máximo contato com a superfície do cristal durante as análises. As medições foram realizadas de forma alternada entre as amostras A1, A2 e A3, totalizando 36 medidas para cada amostra, e em seguida, após cada medição, elas foram descartadas, de forma que nenhuma fosse reutilizada nas medidas.

3.3 Pré-processamento de dados

Os espectros foram submetidos ao pré-tratamento por meio da normalização pela variância normal padrão (SNV), com o objetivo de eliminar deslocamentos verticais e centralizar os dados em torno da média. Esse procedimento é essencial para minimizar variações experimentais que possam comprometer a comparação entre os espectros.[14,31,32]

O método SNV consiste em subtrair, de cada espectro, sua média, e dividir o resultado pelo respectivo desvio padrão. Assim, cada espectro é reescalado para apresentar média zero e desvio padrão igual a 1. Isso é realizado conforme a Equação 1, onde x é o espectro original, $\bar{x}$ é a média dos valores de x, e s é o desvio padrão:

$$x' = \frac{x - \bar{x}}{s} \quad (1)$$

3.4 Análise das componentes principais

A dimensão do conjunto espectral foi reduzida utilizando o método da Análise de Componentes Principais (PCA), resultando em um conjunto reduzido de variáveis ordenadas e não correlacionadas. Com isso, a matriz original dos dados foi convertida em uma matriz de escores (*scores*), que representa as projeções das amostras nos novos eixos principais (componentes), e em uma matriz de carga (*loadings*), que indica a contribuição de cada variável original para os componentes principais.

A aplicação do PCA foi realizada utilizando um algoritmo implementado em Python, com objetivo de avaliar o potencial de separação entre as amostras. Foram analisados os três primeiros componentes principais (PCs) que concentram a maior variância explicada do conjunto de dados originais.

3.5 Algoritmos de aprendizagem de máquina

Foram utilizados três métodos de classificação: k-vizinhos mais próximos (KNN), Máquina de Vetores de Suporte (SVM) com *kernel* linear e Análise Discriminante Linear (LDA).

A acurácia e a validação dos modelos foram avaliados por meio do método de validação cruzada, que permite estimar o desempenho dos classificadores com maior confiabilidade. Essa abordagem é particularmente indicada para conjuntos de dados pequenos, pois maximiza o aproveitamento das amostras nos processos de treinamento e o teste.[33]

O KNN é um algoritmo baseado na distribuição espacial das amostras, classificando cada ponto de acordo com a classe predominante entre seus k vizinhos mais próximos; neste trabalho, foi adotado $k = 5$. O SVM com *kernel* linear, por sua vez, também se apoia na representação espacial dos dados, mas opera construindo um hiperplano ótimo que separa as classes com a maior margem possível.

Por fim, o LDA parte do pressuposto de que as classes possuem a mesma matriz de covariância e, com base nessa premissa, define limites lineares para realizar a separação entre elas. Os algoritmos foram implementados na linguagem Python, enquanto o software Origin foi utilizado para a elaboração dos gráficos de resultados.

4. RESULTADOS E DISCUSSÃO

4.1 Resultados das medições FTIR

A Tabela 1, adaptada da referência [1], mostra a composição química das amostras de argila analisadas. Observa-se que a amostra A3 não possui valores registrados para SiO_2 e Al_2O_3, o que indica uma diferença composicional relevante em relação às demais. Essa particularidade sugere que a argila A3 pode apresentar um comportamento distinto nos métodos analíticos aplicados, facilitando sua diferenciação em relação às amostras A1 e A2.

Tabela 1. Composições elementares das amostras de argila (% em massa). Dados adaptados de [1].

Argila	A1	A2	A3
Si (%)	48,17	45,57	54,6
Al (%)	34,64	36,79	27,96
Fe (%)	7,63	5,98	8,32
Ti (%)	7,33	9,8	3,13
S (%)	1,81	1,48	1,41
K (%)	-	-	2,18
Ba (%)	-	-	-
SiO_2 (%)	49,05	45,85	-

Al_2O_3 (%)	40,28	42,67	-
Si/Al	1,39	1,24	1,95
Ki	2,1	1,8	2,8 - 3,0

As amostras de argila foram analisadas por espectroscopia no infravermelho por transformada de Fourier (FTIR). Os espectros de transmitância foram obtidos na faixa abrangendo os números de onda de 4000 a 600 cm^{-1}.

A Figura 11 mostra a média dos espectros das diferentes amostras de argila. É possível identificar uma banda discreta entre 3695 e 3620 cm^{-1}, atribuída ao estiramento O-H de grupos hidroxila estruturais típicos de minerais de argila.[34]

Além disso, na região próxima a 1034 cm^{-1}, uma banda de absorção característica das vibrações de estiramento da ligação Si-O. Embora a amostra A3 não apresente SiO_2 em sua composição segundo a Tabela 1, essa banda ainda está presente em seu espectro. Isso ocorre porque essa vibração não está exclusivamente associada ao dióxido de silício (SiO_2), mas sim a uma rede estrutural rica em silício e oxigênio, comum aos minerais de argila.[34,35]

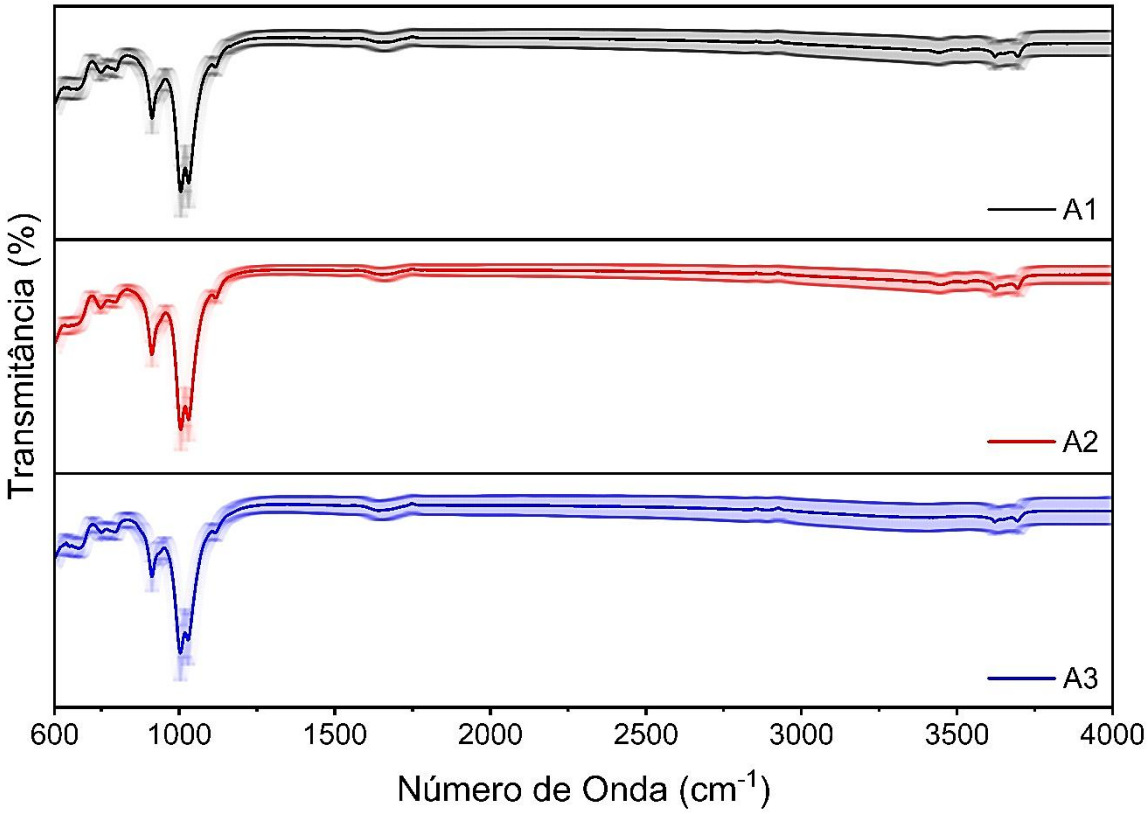

Figura 11. Espectros de infravermelho médio e seus respectivos desvios padrões para um total de 36 medidas das amostras de argila: A1 (preto), A2 (vermelho), e A3 (azul).

Na Figura 11 é possível perceber que os espectros das amostras de argila possuem comportamentos parecidos, diferenciando-se principalmente nos valores de intensidade óptica. Essa similaridade espectral torna difícil a tarefa de diferenciar as amostras de argila, tornando necessária a utilização de ferramentas matemáticas para uma análise mais precisa.

Os espectros foram pré-processados utilizando a normalização SNV para remover ruídos e corrigir deslocamentos verticais. Esse procedimento é necessário para evitar distinção dos dados devido a variações experimentais, assegurando que as diferenças observadas estejam mais relacionadas às características químicas ou estruturais das amostras do que a fatores externos ou instrumentais.

4.2 Análise das componentes principais

A análise PCA foi utilizada para avaliar a separabilidade entre as amostras, através da redução da dimensionalidade dos dados. Isso foi feito a partir de combinações lineares das variáveis espectrais no infravermelho, projetadas nas direções de maior variabilidade dos dados.

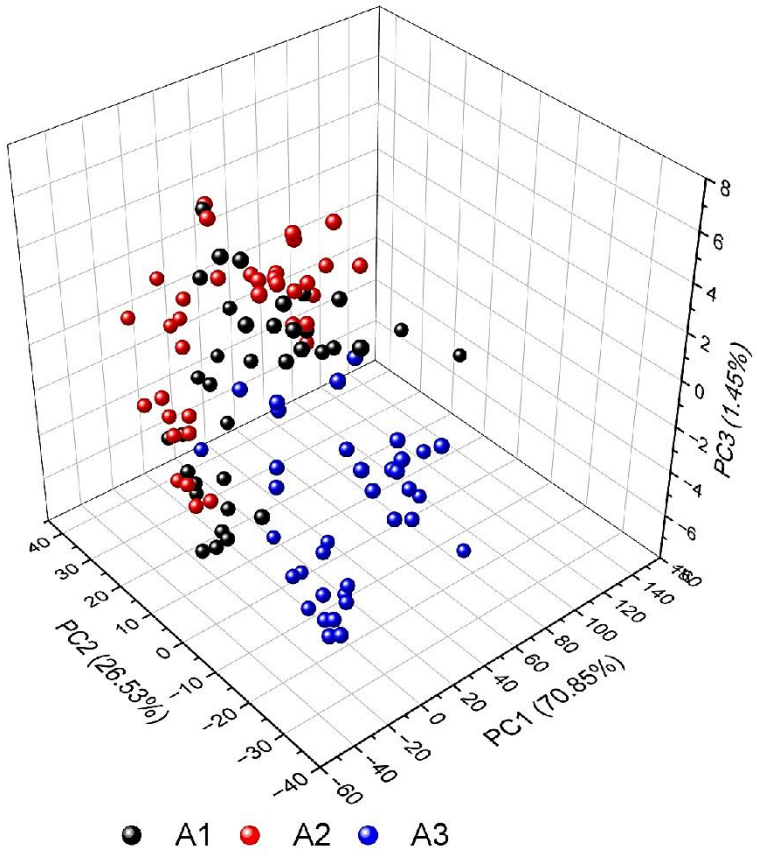

Figura 12. Scores correspondentes às três primeiras componentes principais (PC1, PC2 e PC3) das amostras de argila A1 (preto), A2 (vermelho) e A3 (azul).

A Figura 12 apresenta o resultado da análise PCA a partir das 3 primeiras componentes principais, na região espectral de 4000-600cm^{-1}, para as amostras de argila A1, A2 e A3. A primeira componente (PC1) explica 70,85% da variância total dos dados, a segunda (PC2) com 26,53% e a terceira (PC3) com 1,45%, totalizando 98,83% da variância explicada. O que é suficiente para a análise tridimensional representativa dos dados espectrais.

Entretanto, observa-se na Figura 12 que não é possível obter uma separação clara entre as amostras de argila. Os pontos estão dispostos de modo que as classes A1, A2 e A3 estão muito próximos entre si. Assim, mesmo com a alta variância acumulada pelas três primeiras componentes, o PCA apresenta limitações para diferenciar as amostras nas classes A1, A2 e A3.

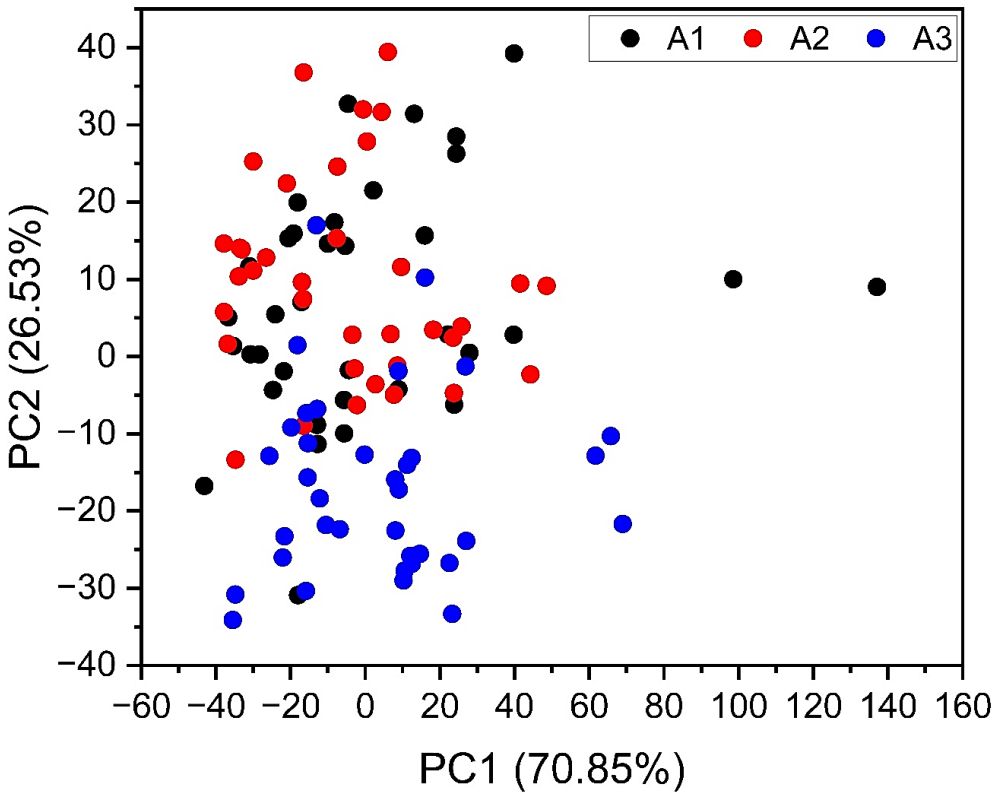

Figura 13. Scores das amostras de argila A1 (preto), A2 (vermelho) e A3 (azul) nas duas primeiras componentes principais (PC1 e PC2).

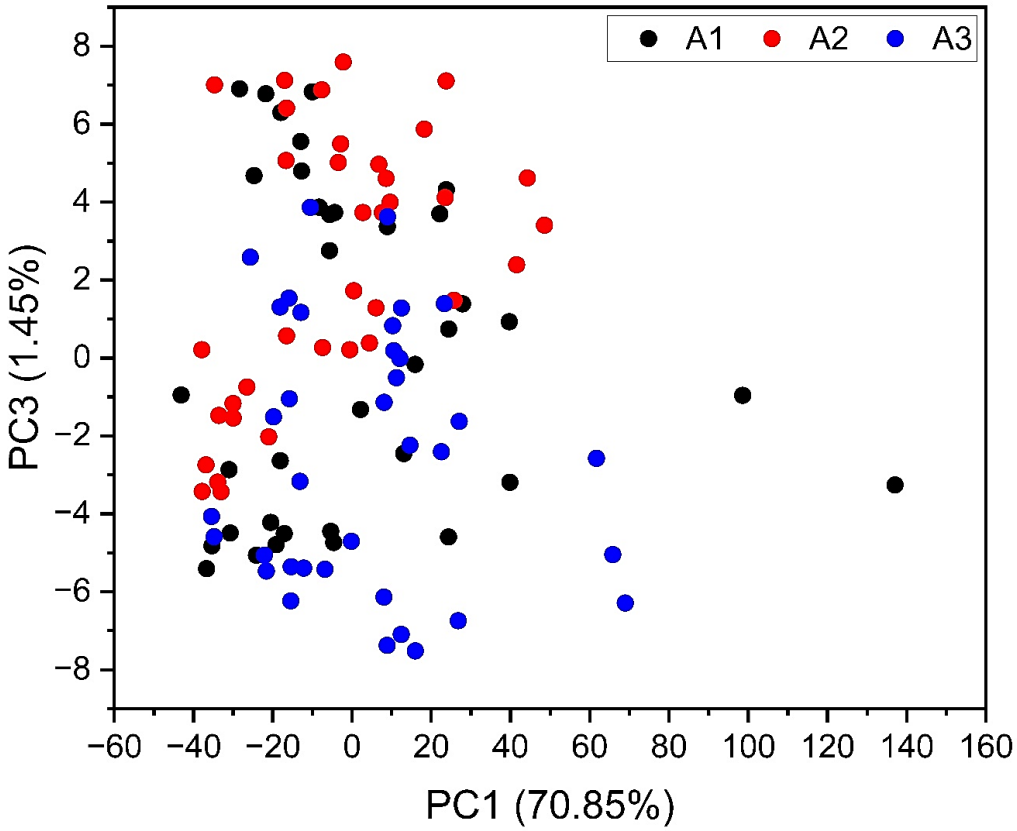

Figura 14. Scores das amostras de argila A1 (preto), A2 (vermelho) e A3 (azul) nas duas primeiras componentes principais (PC1 e PC3).

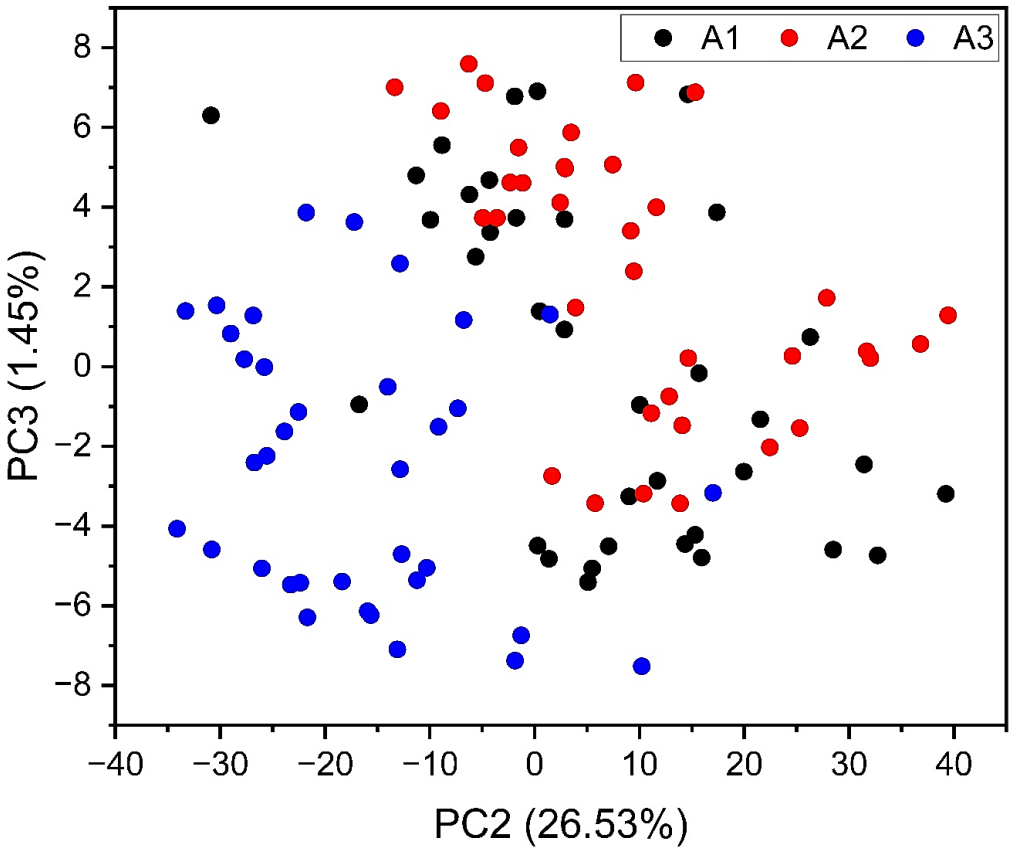

Figura 15. Scores das amostras de argila A1 (preto), A2 (vermelho) e A3 (azul) nas duas primeiras componentes principais (PC2 e PC3).

É possível observar nas Figuras 13, 14 e 15 que, apesar de não obter uma separação satisfatória das amostras, a argila A3 apresenta tendencia de se diferenciar das demais amostras. Isso pode ser explicado pelas diferenças na sua composição química, conforme evidenciado na Tabela 1.

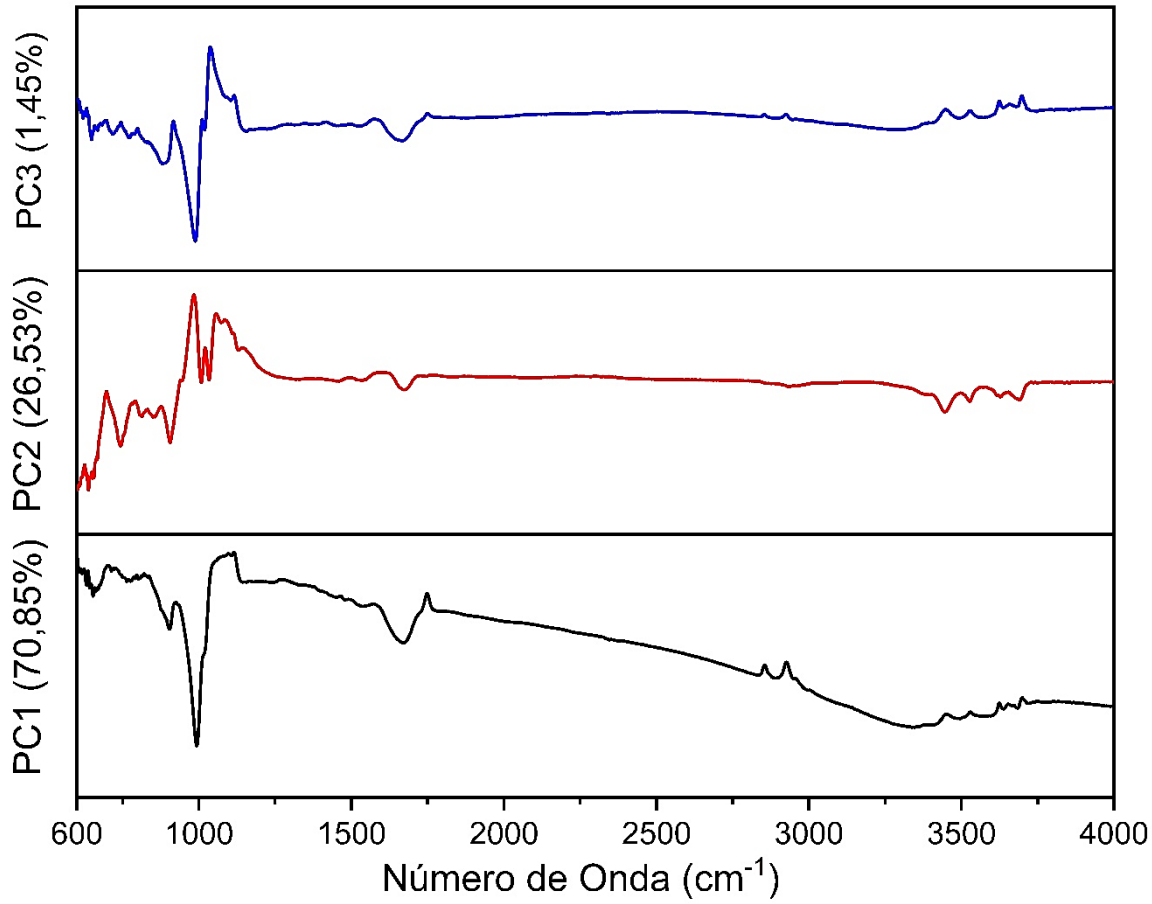

Figura 16. *Loadings* dos dados obtidos pela PCA, no intervalo espectral de 600 a 4000 cm^{-1}.

Além disso, os *loadings* foram analisados com o intuito de avaliar qual foi a região espectral que possa ter maior influência na diferenciação de uma amostra para outra. Verificou-se que, na região de 600 a 1500 cm^{-1}, ocorre uma variação mais expressiva nos dados. Com base nessa observação, a PCA foi também aplicada especificamente a esse intervalo espectral.

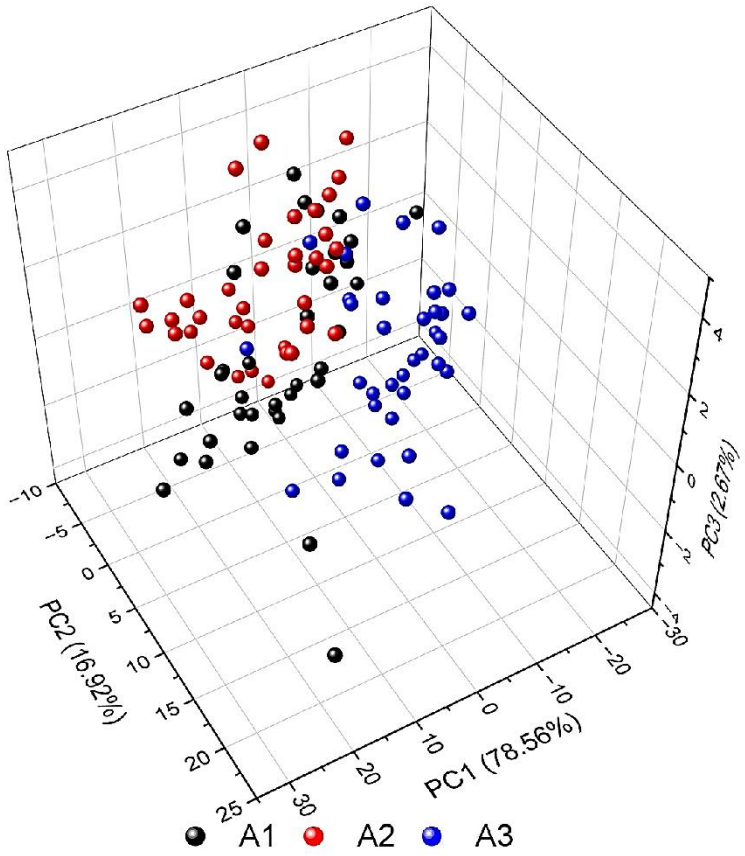

Figura 17. Scores correspondentes às três primeiras componentes principais (PC1, PC2 e PC3) das amostras de argila A1 (preto), A2 (vermelho) e A3 (azul) para a região espectral de 600 à 1500cm^{-1}.

A Figura 10 apresenta os resultados da análise PCA considerando as 3 primeiras componentes principais, na região espectral de 1500-600 cm^{-1}, para as argilas A1, A2 e A3. A primeira componente principal (PC1) tem variância de 78,56%, a segunda (PC2) de 16,92% e a terceira (PC3) de 2,67%, totalizando 98,15% da variância explicada dos dados espectrais.

Na Figura 17 é possível observar uma discreta tendencia de separação entre as amostras, com a amostra A3 separada das demais, enquanto as amostras A1 e A2 já não se apresentam sobrepostas, como observado anteriormente.

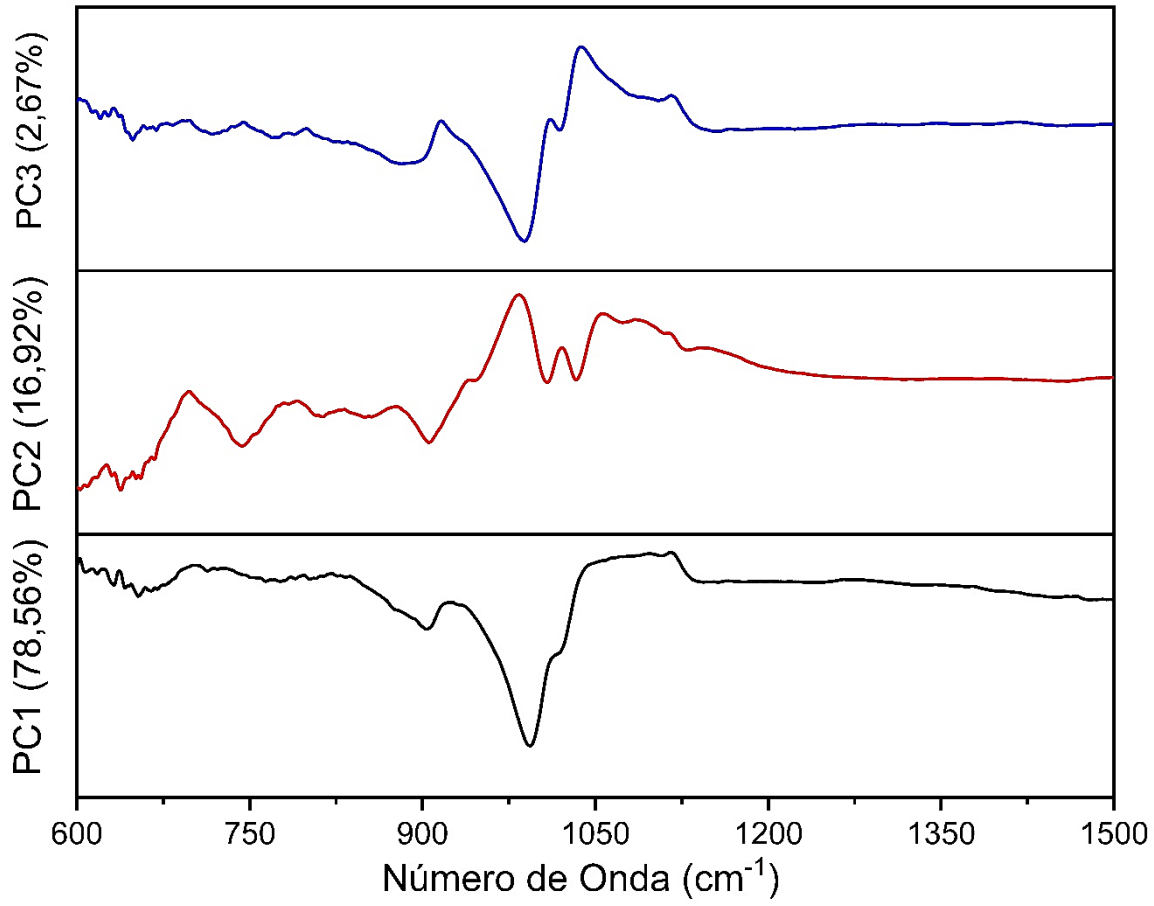

Figura 18. *Loadings* dos dados obtidos pela PCA no intervalo espectral de 600 a 1500 cm^{-1}.

4.3 Aprendizagem de máquina

Os dados obtidos por meio da PCA foram submetidos a etapa de treinamento e validação dos algoritmos de aprendizagem de máquina. Foram aplicadas as técnicas de k – vizinhos mais próximos, Máquina de vetores de suporte, e Análise Discriminante Linear. A matriz de confusão correspondendo a cada modelo de classificação pode ser observada na Figura 19.

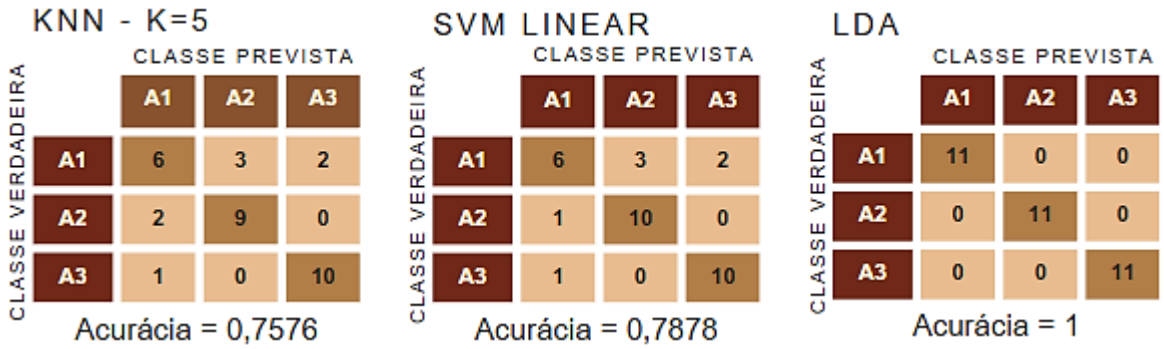

Figura 19. Matrizes de confusão representando o desempenho dos algoritmos de classificação KNN ($k = 5$), SVM com *kernel* linear e LDA aplicados aos dados obtidos por PCA no intervalo espectral de 4000 a 600 cm^{-1}.

A matriz de confusão é uma ferramenta utilizada na avaliação de modelos validados por validação cruzada, permitindo verificar o desempenho do classificador por meio das taxas de acertos e erros. Ela apresenta os valores de verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) e falsos negativos (FN). Com base nesses dados, é possível calcular métricas relevantes como acurácia, sensibilidade e especificidade.

Os valores localizados na diagonal principal da matriz correspondem às amostras classificados corretamente. No caso do algoritmo KNN foi obtida uma acurácia de 75,76%, o menor desempenho entre os métodos testados de classificação no intervalo espectral (4000–600 cm^{-1}). O SVM alcançou uma acurácia de 78,78%, enquanto o LDA obteve o melhor desempenho, com uma acurácia de 100%, classificando corretamente todas as amostras.

Os algoritmos também foram aplicados ao conjunto de dados de PCA correspondente ao intervalo espectral de 1500 a 600 cm^{-1}, cujos resultados estão apresentados na Figura 13. Nessa faixa, o KNN obteve uma acurácia de 72,72%, enquanto o SVM registrou 69,69%, sendo novamente o desempenho mais baixo. O LDA, por sua vez, manteve o desempenho máximo, com 100% de acerto.

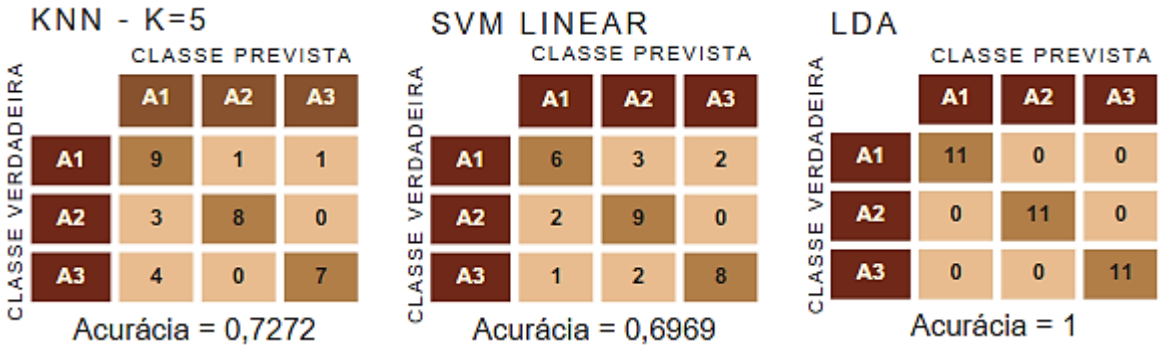

Figura 20. Matrizes de confusão representando o desempenho dos algoritmos de classificação KNN ($k = 5$), SVM com *kernel* linear e LDA aplicados aos dados obtidos por PCA no intervalo espectral de 1500 a 600 cm^{-1}.

Os resultados obtidos revelaram um desempenho expressivo na classificação das amostras de argila, especialmente com o algoritmo LDA, que alcançou acurácia de 100% em ambas as faixas espectrais analisadas (4000–600 cm^{-1} e 1500–600 cm^{-1}). Esse resultado se destaca quando comparado aos dados da revisão sistemática realizada por Ling et al.[36], a qual reuniu estudos que utilizaram aprendizado de máquina para análise de cerâmicas arqueológicas. Segundo os autores, a maioria dos trabalhos alcança acurácias entre 70% e 95%, sendo raros os casos que ultrapassam a marca dos 98%.

Em contrapartida, os algoritmos KNN e SVM apresentaram acurácias inferiores (entre 69% e 78%), o que está em consonância com o observado na literatura. Como apontado por Ling et al.[36], a performance de modelos como o SVM tende a ser sensível à escolha de hiperparâmetros e ao tamanho do conjunto de treinamento, sendo comum a redução da acurácia quando se trabalha com bases de dados reduzidas, como neste estudo.

Dessa forma, os resultados demonstram elevada robustez e confiabilidade da abordagem empregada, especialmente com o uso do LDA. A estratégia adotada — espectroscopia FTIR associada à PCA e algoritmos de aprendizado supervisionado — está alinhada com as tendências da literatura e se mostrou eficaz na classificação de argilas com diferentes origens. Além de validar seu uso como uma metodologia rápida, precisa e de baixo custo, os resultados aqui apresentados superam as acurácias relatadas em diversos estudos, evidenciando a robustez e aplicabilidade da abordagem proposta.

5. CONCLUSÃO

A utilização da espectroscopia FTIR associada à análise multivariada e técnicas de aprendizagem de máquina demonstrou ser uma abordagem eficiente para a discriminação de amostras de argila. Apesar da semelhança visual entre os espectros das amostras A1, A2 e A3, especialmente na região de 4000 a 600 cm^{-1}, a aplicação de pré-processamento (normalização SNV) e da Análise de Componentes Principais (PCA) permitiu explorar variações sutis nos dados, essenciais para a separação das classes. A análise dos *loadings* destacou que a região espectral entre 600 e 1500 cm^{-1} possui maior variabilidade, justificando a aplicação de PCA também nesse intervalo.

Em ambos os intervalos analisados (4000–600 cm^{-1} e 1500–600 cm^{-1}), observou-se uma tendência de separação da amostra A3, condizente com a sua composição química distinta das demais, como indicado na Tabela 1. A projeção tridimensional das três primeiras componentes principais reteve mais de 98% da variância total dos dados em ambos os casos, o que valida o uso dessas componentes para análise discriminante.

Com base nas projeções do PCA, os modelos de classificação KNN, SVM e LDA foram treinados e validados. Para a região de 4000–600 cm^{-1}, as acurácias obtidas foram: 75,76% para o KNN, 78,78% para o SVM e 100% para o LDA, sendo este o modelo com maior acurácia. Para a região de 600–1500 cm^{-1}, os resultados foram similares: 72,72% para o KNN, 69,69% para o SVM e novamente 100% para o

LDA. Esses valores foram confirmados pelas matrizes de confusão, evidenciando que o LDA foi capaz de classificar corretamente todas as amostras nos dois intervalos espectrais.

Conclui-se, portanto, que a combinação entre espectroscopia FTIR, PCA e algoritmos de aprendizagem de máquina, especialmente o LDA, constitui uma metodologia rápida, robusta e precisa para a classificação de argilas com diferentes composições e origens geográficas. Esses resultados têm potencial aplicação em áreas como ciência dos materiais, geologia, arqueometria e ciência ambiental.

6. REFERÊNCIAS BIBLIOGRÁFICAS

[01] Lima, D. F. Determinação dos parâmetros de queima de cerâmicas de tradição tupiguarani de Goiás. Universidade Estadual do Norte Fluminense – UENF. Campos dos Goytacazes, RJ, 2019.

[02] RICE, P. M. Pottery Analysis: A sourcebook. [s.l.] Univ. of Chicago, 2005.

[03] MANGUEIRA, G. M. et al. Evaluation of archeothermometric methods in pottery using electron paramagnetic resonance spectra of iron. Applied Clay Science, v. 86, p. 70–75, dez. 2013.

[04] Anglisano, A.; Casas, L.; Queralt, I.; Di Febo, R. Supervised Machine Learning Algorithms to Predict Provenance of Archaeological Pottery Fragments. Sustainability 2022, 14, 11214.

[05] JORDANOVA, N. et. al. Imprints of paleo-environmental conditions and human activities in mineral magnetic Properties of fired clay remains from neolithic houses. Journal of Archaeological Science: Reports 33, 2020.

[06] Allegretta, I.; Ciasca, B.; Pizzigallo, M. D. R.; Lattanzio, V. M. T.; Terzano, R. A fast method for the Chemical analysis of clays by total-reflection x-ray fluorescence spectroscopy (TXRF). Applied Clay Science 2019, 180, 105201. https://doi.org/10.1016/j.clay.2019.105201.

[07] JOZANIKOHAN, Golnaz; NOSRATI ABARGHOOEI, Mohsen. The Fourier transform infrared spectroscopy (FTIR) analysis for the clay mineralogy studies in a clastic reservoir. *Journal of Petroleum Exploration and Production Technology*, [S. l.], v. 12, p. 2093–2106, 2022. DOI: https://doi.org/10.1007/s13202-021-01449-y.

[08] SHOVAL, Shlomo. Fourier Transform Infrared Spectroscopy (FT-IR) in Archaeological Ceramic Analysis. In: HUNT, Alice (ed.). *The Oxford Handbook of Archaeological Ceramic Analysis*. Oxford: Oxford University Press, 2016. p. 507–527. DOI: https://doi.org/10.1093/oxfordhb/9780199681532.013.28.

[09] LIMA, D.F. et al. Fe^{III} in pottery: Identifying firing temperature and ambiguity. Applied Clay Science, v 190, 2020.

[10] SHOVAL, S. The firing temperature of a persian-period pottery kiln at Tel Michal, Israel, estimated from the composition of its pottery. Journal of Thermal Analysis, v. 42, n. 1, p. 175–185, 1994.

[11] Ikeoka, R.A.; Appoloni, C.R.; Scorzelli, R.B.; dos Santos, E.; Rizzutto, M.d.A.; Bandeira, A.M. Study of Ancient Pottery from the Brazilian Amazon Coast by EDXRF, PIXE, XRD, Mössbauer Spectroscopy and Computed Radiography. Minerals 2022, 12, 1302. https:// doi.org/10.3390/min12101302.

[12] DU PLESSIS, Pieter I. et al. Quantification of Kaolinite and Halloysite Using Machine Learning from FTIR, XRF, and Brightness Data. *Minerals*, Basel, v. 11, n. 12, p. 1350, 2021. DOI: https://doi.org/10.3390/min11121350.

[13] VAHUR, Signe et al. Quantitative mineralogical analysis of clay-containing materials using ATR-FT-IR spectroscopy with PLS method. *Analytical and Bioanalytical Chemistry*, [S. l.], v. 413, p. 5951–5965, 2021. DOI: https://doi.org/10.1007/s00216-021-03617-9.

[14] FERREIRA, M. Quimiometria - Conceitos, Métodos e Aplicações. Editora da Unicamp. Campinas, SP, 2015.

[15] BARONE, Germana et al. *Artificial Neural Network for the provenance study of archaeological ceramics using clay sediment database*. 2019. Disponível em: https://www.researchgate.net/publication/330769006.

[16] WANG, Qian; XIAO, Xuan; LIU, Zi. Using microscopic imaging and ensemble deep learning to classify the provenance of archaeological ceramics. *Scientific Reports*, [S. l.], v. 14, n. 32024, 2024. DOI: https://doi.org/10.1038/s41598-024-83533-x.

[17] RUSCHIONI, G. et al. Supervised learning algorithms as a tool for archaeology: Classification of ceramic samples described by chemical element concentrations. *Journal of Archaeological Science: Reports*, [S. l.], v. 49, p. 103995, 2023. DOI: https://doi.org/10.1016/j.jasrep.2023.103995.

[18] BUIJS, H. Infrared Spectroscopy. Springer Handbooks, p. 607–613, 2006. ISSN 25228706.

[19] PAVIA, D. L. et al. Introdução à espectroscopia. São Paulo: Cengage Learning, 2010.

[20] KHAN, S. A. et al. Fourier transform infrared spectroscopy: Fundamentals and application in functional groups and nanomaterials characterization. Handbook of Materials Characterization, p. 317–344, 2018.

[21] HOLLAS, J.M. Modern Spectroscopy. 2ª ed. Wiley, 1993.

[22] ONES, D. A. E. J. Fourier Transform Infrared Spectra. Fourier Transform Infrared Spectra, 1978.

[23] Kiyoshi Yamamoto, Hatsuo Ishida, Optical theory applied to infrared spectroscopy, Vibrational Spectroscopy. 2031(94)00022-9. https://doi.org/10.1016/0924.

[24] M. I. Jordan T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, Science . 349, 255-260(2015). DOI:10.1126/science.aaa8415

[25] Schmidt, J., Marques, M.R.G., Botti, S. et al. Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater 5, 83 (2019). https://doi.org/10.1038/s41524-019-0221-0.

[26] BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy, v. 43, n. 5, p. 772–777, 1989. ISSN 00037028.

[27] Allegretta, I.; Marangoni, B.; Manzari, P.; Porfido, C.; Terzano, R.; Pascale, O.; Senesi, G, S. Macro-classficication of meteorites by portable energy dispersive X-ray fluorescence spectroscopy (pED-XRF), principal componente analysis (PCA)

and machine learning algorithms. Talanta 2019, 212, 120785. https://doi.org/10.1016/j.talanta.2020.120785.

[28] Jolliffe IT, Cadima J., Principal component analysis: a review and recent developments, Philos Trans A Math Phys Eng Sci. 2016 Apr 13;374(2065):20150202. doi:10.1098/rsta.2015.0202.

[29] Zhang Z., Introduction to machine learning: k-nearest neighbors, Ann Transl Med. 2016 Jun;4(11):218. doi: 10.21037/atm.2016.03.37.

[30] Fichou, Dimitri & Ristivojevic, Petar & Morlock, Gertrud. (2016), Proof-of-Principle of rTLC, an Open-Source Software Developed for Image Evaluation and Multivariate Analysis of Planar Chromatograms, Analytical Chemistry. 88. 10.1021/acs.analchem.6b04017.

[31] ENGEL, J. et al. Breaking with trends in pre-processing? TrAC Trends in Analytical Chemistry, v. 50, p. 96–106, 2013. doi: 10.1016/j.trac.2013.04.015.

[32] Rinnan, Åsmund & Berg, Frans & Engelsen, Søren. (2009), Review of the Most Common preProcessing Techniques for Near-Infrared Spectra, TrAC Trends in Analytical Chemistry. 28. 1201-1222. 10.1016/j.trac.2009.07.007.

[33] PÉREZ-GUAITA, David et al. Assessment of the statistical significance of classifications in infrared spectroscopy based diagnostic models. Analyst, [S. l.], v. 140, n. 21, p. 7468–7477, 2015. DOI: https://doi.org/10.1039/C4AN01783H.

[34] KRIVOSHEIN, P. K. et al. *FTIR Photoacoustic and ATR Spectroscopies of Soils with Aggregate Size Fractionation by Dry Sieving*. ACS Omega, v. 7, p. 2177–2197, 2022. DOI: https://doi.org/10.1021/acsomega.1c05702.

[35] ALVAREZ ACEVEDO, N. I. et al. Caracterização mineralógica de argilas naturais provenientes do Sudeste brasileiro visando aplicações industriais. *Cerâmica*, São Paulo, v. 63, n. 366, p. 253–262, 2017. DOI: https://doi.org/10.1590/0366-69132017633662045.

[36] LING, Ziyao; DELNEVO, Giovanni; SALOMONI, Paola; MIRRI, Silvia. Findings on Machine Learning for Identification of Archaeological Ceramics: A Systematic Literature Review. *IEEE Access*, v. 12, p. 100167–100185, 2024. DOI: https://doi.org/10.1109/ACCESS.2024.3429623.