



FUNDAÇÃO UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL

Instituto de Física - INFI

Programa de Pós-Graduação em Ciência dos Materiais - PPGCM

MATHEUS CICERO DA SILVA RIBEIRO

**USO DE ESPECTROSCOPIA LIBS ACOPLADOS A ROTINAS DE  
APRENDIZADO DE MÁQUINAS PARA DIFERENCIAÇÃO DE AMOSTRAS  
MINERAIS**

CAMPO GRANDE, MS

2023

MATHEUS CICERO DA SILVA RIBEIRO

**USO DE ESPECTROSCOPIA LIBS ACOPLADOS A ROTINAS DE  
APRENDIZADO DE MÁQUINAS PARA DIFERENCIAÇÃO DE AMOSTRAS  
MINERAIS**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência dos Materiais da Universidade Federal de Mato Grosso do Sul, como requisito á obtenção do título de Doutor em Ciência dos Materiais.

Orientador: Prof. Dr. Bruno Spolon Marangoni

CAMPO GRANDE, MS

2023

## FOLHA DE APROVAÇÃO

Os componentes da banca de avaliação, abaixo listados, consideram esse trabalho aprovado

---

Prof. Dr. Bruno Spolon Marangoni

---

Prof. Dr. Giorgio Saverio Senesi

---

Prof. Dr. Cicero Rafael Cena da Silva

---

Prof. Dr. Gustavo Nicolodelli

---

Prof. Dr. Jader de Souza Cabral

---

Prof. Dr. Carlos Renato Menegatti

---

Prof. Dr. Diogo Duarte dos Reis

## RESUMO

Minerais são definidos como uma substância natural inorgânica, formada através de processos geológicos, que possuem uma estrutura cristalina regular e sólida a temperatura ambiente. São formados através de ligações químicas entre elementos químicos eletricamente carregados, onde a nomenclatura dos minerais é dada pelo seu grupo ânion predominante. A correta classificação de amostras minerais é uma tarefa de grande importância para vários campos de estudos. Amostras minerais podem ser erroneamente classificadas em diferentes grupos, se analisadas somente de modo visual, devido a haver semelhanças e/ou diferenças nas características visuais entre amostras de um mesmo grupo mineral. Os testes padrões mais utilizados, além das inspeções visuais, que necessitam de experiência do observador, são as técnicas analíticas como espectroscopia por raio-x ou de emissão óptica por plasma acoplado indutivamente (ICP). Porém, essas técnicas analíticas não são vantajosas, pois apresentam o uso de radiação ionizante e não possibilidade de análises *in situ*, respectivamente. O uso da técnica LIBS surge como uma alternativa para análises *in situ*, de forma rápida e eficaz, o que é de grande importância para a diferenciação e classificação de amostras minerais. Neste trabalho, o uso da técnica de espectroscopia óptica de emissão atômica por plasma induzido por laser (LIBS), combinados com métodos quimiométricos de análise de componentes principais (PCA), e com algoritmos supervisionados de aprendizado de máquinas, k-vizinhos próximos (KNN), máquinas de vetores de suporte (SVM) e análise discriminante (LDA e QDA), foram avaliados na diferenciação e classificação de amostras minerais das classes silicato, oxido, fosfato, carbonato e fluorita. Os resultados mostraram que o algoritmo *Fine KNN* foi capaz de classificar corretamente todas as amostras de validação externa, para a classe silicato, com acurácia de 100%, para os três tipos de abordagens distintas realizadas. Limitações para a implementação desse procedimento como técnica de análise ainda existem, principalmente para diferenciação das amostras não-silicato.

**Palavras-chave:** LIBS portátil, Aprendizado de Máquinas, Minerais.

## ABSTRACT

Minerals are defined as natural inorganic substances that are formed due geological process with a regular atomic lattice and solid format in room temperature. Their chemical structure are formed between charged particles, where they are grouped into classes based of their predominant anion group. The correct classification of minerals samples plays an important role in several fields of study. Minerals samples can be mistakenly classified into different groups, if analyzed only visually, due to similarities and/or differences in visual characteristics between samples of the same mineral groups. The most commonly used tests, apart from the visual analysis that requires skillful employees, are analytical techniques such as X-ray spectroscopy or Inductively Coupled Plasma Optical Emission Spectroscopy (ICP). But, these analytical techniques are not advantageous because they require the use of ionizing radiation and no possibility to perform in situ analysis, respectively. The LIBS technique is an alternative to be used in situ, that provides fast and efficient elemental analysis, what is of great importance to discriminate and classify mineral samples. In the present study, the use of the spectroscopy technique Laser Induced Breakdown Spectroscopy (LIBS), combined with chemometrics methods as principal component analysis (PCA) and machine learning algorithms, k-nearest neighbour (KNN), support vector machine (SVM) and discriminant analysis (DA), were evaluated in the differentiation and classification of minerals samples from the classes silicate, oxide, phosphate, carbonate and fluoride. The results showed that the Fine KNN algorithm were able to correctly classify all samples in the external validation scheme, for the silicate class, reaching accuracy of 100% in three different approaches implemented. There are still some limitations in the use of this procedure, mainly related to the differentiation of the non-silicate minerals samples.

**Key-word:** *handheld LIBS, Machine Learning, Minerals.*

## Lista de Tabelas

- Tabela 1 - Nomenclatura padrão dos algoritmos KNN com base no número de vizinhos próximos, distância métrica e normalização de distância. Fonte: Mathworks.com ..... 27
- Tabela 2 - Nomenclatura padrão dos algoritmos SVM com base na sua função kernel e normalização. Fonte: Mathworks.com ..... 30
- Tabela 3 - Comparativo da quantidade de amostras iniciais, antes do método de detecção de outliers SAM, a quantidade de amostras excluídas em cada classe, e a quantidade de amostras mantidas após a aplicação do método de detecção de outliers SAM. .... 36
- Tabela 4 - Regiões espectrais LIBS selecionadas, para separar amostras da classe mineral Silicato versus as amostras dos restantes das classes minerais. O comp. de onda inicial e o Comp. de onda final indicam, respectivamente, o comprimento de onda inicial e final, onde o pico central do elemento químico foi selecionado. O elemento químico correspondente a cada pico central é mostrado na última coluna da tabela. .... 40
- Tabela 5 - Regiões espectrais LIBS selecionadas, para separar amostras da classe mineral Oxido versus as amostras dos restantes das classes minerais. O comp. de onda inicial e o Comp. de onda final indicam, respectivamente, o comprimento de onda inicial e final, onde o pico central onde o elemento químico foi selecionado..... 43
- Tabela 6 - Regiões espectrais LIBS selecionadas, para separar amostras da classe mineral Fosfato versus as amostras dos restantes das classes minerais. O comp. de onda inicial e o Comp. de onda final indicam, respectivamente, o comprimento de onda inicial e final, onde o pico central do elemento químico foi selecionado. .... 46
- Tabela 7 - Regiões espectrais LIBS selecionadas, para separar amostras da classe mineral Carbonato versus as amostras dos restantes das classes minerais. O comp. de onda inicial e o Comp. de onda final indicam, respectivamente, o comprimento de onda inicial e final, onde o pico central do elemento químico foi selecionado. .... 49
- Tabela 8 - Regiões espectrais LIBS selecionadas para separar amostras da classe mineral Fluorita versus as amostras dos restantes das classes minerais. O comp. de onda inicial e o Comp. de onda final indicam, respectivamente, o comprimento de onda inicial e final, onde o pico central do elemento químico foi selecionado. .... 51

## Lista de Figuras

- Figura 1 – Amostras minerais, de diferentes classes minerais, a serem analisadas, diferenciadas e classificadas. .... 19
- Figura 2 - Configuração básica da instrumentação LIBS. .... 21
- Figura 3 - Fluxograma da caracterização óptica LIBS das amostras minerais, seguida de análise por PCA e algoritmos de aprendizado de máquinas, para diferenciação das amostras com base em suas composições elementares. .... 23
- Figura 4 - Representação da classificação de uma amostra externa, em um plano 2D, utilizando o algoritmo k-vizinhos próximos. O número de vizinhos próximos,  $k$ , foi definido como 10, e estão destacados pelos círculos em laranja. Em azul, são as amostras da classe A, em verde as amostras da classe B, e em vermelho a amostra externa que será classificada pelo KNN. .... 26
- Figura 5 - a esquerda, representação do classificador LDA, a direita, representação do classificador QDA. Elipse em azul representa a matriz covariância de amostras pertencentes a classe A, elipse em verde representa a matriz covariância de amostras pertencentes a classe B. Em vermelho, linha de decisão que separam as amostras em classes, onde a probabilidade posterior das amostras a pertencerem as duas classes são iguais. Figura adaptada de (SAVORANI et al., 2013). .... 28
- Figura 6 - a esquerda, representação do SVM linear, a direita, representação de um SVM não-linear utilizando função kernel. Em azul, amostras da classe A, com seu respectivo vetor de suporte,  $x_i$ , em verde, amostras da classe B, com seu respectivo vetor de suporte  $x_j$ . Para o SVM linear, em laranja tracejado é a linha de decisão que separa as classes, localizada no meio do hiperplano que separam as duas classes amostrais. Para o SVM não-linear, em laranja tracejado é a curva de decisão que separa as classes amostrais proveniente de uma função não-linear (kernel). .... 30
- Figura 7 - Para classificação binária, a esquerda, matriz de confusão: TP (true positive – verdadeiros positivos) – amostras da classe A que foram corretamente classificadas, FN (false negatives – falsos negativos) – amostras da classe A que foram incorretamente classificadas, TN (true negatives – verdadeiros negativos) – amostras da classe B que foram corretamente classificadas, FP (false positives – falsos positivos) – amostras da classe B que foram incorretamente classificadas. A direita, cálculo da acurácia, sensibilidade e especificidade. .... 32
- Figura 8 - Para classificação multi-classe de três classes, a esquerda, matriz de confusão:  $TP_i$  (true positives – verdadeiros positivos) amostras da classe  $i$  que foram corretamente classificadas,  $FN_{ij}$  (false negatives – falsos negativos) amostras da classe  $i$  que foram incorretamente classificados como da classe  $j$ . A direita, cálculo da acurácia, sensibilidade e especificidade. .... 32

Figura 9 - Princípio de funcionamento do algoritmo seletor de componentes principais, mostrando todas as componentes principais (PC`s) dos dados amostrais (em cinza), as componentes principais (PC`s) momentaneamente excluídas (em laranja) e a componente principal (PC) permanentemente excluída (em roxo) na 1ª iteração do algoritmo.....	34
Figura 10 – Espectro médio SNV-LIBS das amostras das classes Silicato (verde), Fosfato (azul), Carbonato (cian), Fluorita (preto) e óxido (vermelho). Em destaque, alguns elementos químicos observados em cada classe. ....	37
Figura 11 - Dispersão PCA 3D das amostras minerais. Classe mineral: Silicato (esferas em verde), Fosfato (esferas em azul), Carbonato (esferas em cian), Fluorita (esferas em preto) e Oxido (esferas em vermelho). Cada esfera colorida representa uma amostra diferente do conjunto de dados.....	38
Figura 12 - Loadings da PCA dos espectros SNV-LIBS das amostras minerais, para PC1 (azul), PC2 (roxo) e PC3 (verde).....	39
Figura 13 - Dispersão PCA 3D (PC1 x PC2 x PC3), após a seleção das regiões espectrais LIBS da Tabela 4. Em verde, são os scores das amostras pertencentes a classe mineral Silicato. Em cinza, são os scores das amostras pertencentes as classes minerais Fosfato, Carbonato, Fluorita e Oxido. ....	41
Figura 14 – Loadings das componentes principais 1 (em azul), 2 (em roxo) e 3 (em verde), para PCA das regiões LIBS selecionadas da tabela 4. Diferenciação das amostras silicato das amostras não-silicato. Em destaque, principais elementos químicos responsáveis pela variância em cada componente principal. ....	42
Figura 15 - Dispersão PCA 3D (PC1 x PC2 x PC3), após a seleção das regiões espectrais LIBS da Tabela 5. Em vermelho, são os scores das amostras pertencentes a classe mineral Oxido. Em cinza, são os scores das amostras pertencentes as classes minerais Fosfato, Carbonato, Fluorita e Silicato. ....	43
Figura 16 - Loadings das componentes principais 1 (em azul), 2 (em roxo) e 3 (em verde), para PCA das regiões LIBS selecionadas da tabela 5. Diferenciação das amostras oxido das amostras não-oxido. Em destaque, principais elementos químicos responsáveis pela variância em cada componente principal. ....	45
Figura 17 - Dispersão PCA 3D (PC1 x PC2 x PC3), após a seleção das regiões espectrais da Tabela 6. Em azul, são os scores das amostras pertencentes a classe mineral Fosfato. Em cinza, são os scores das amostras pertencentes as classes minerais Oxido, Carbonato, Fluorita e Silicato. ....	47
Figura 18 - Loadings das componentes principais 1 (em azul), 2 (em roxo) e 3 (em verde), para PCA das regiões LIBS selecionadas da tabela 6. Diferenciação das amostras fosfato das amostras não-fosfato. Em destaque, principais elementos químicos responsáveis pela variância em cada componente principal. ....	48



Figura 19 - Dispersão PCA 3D (PC1 x PC2 x PC3), após a seleção das regiões espectrais da Tabela 7. Em cian, são os scores das amostras pertencentes a classe mineral Carbonato. Em cinza, são os scores das amostras pertencentes as classes minerais Oxido, Fosfato, Fluorita e Silicato.....	49
Figura 20 - Loadings das componentes principais 1 (em azul), 2 (em roxo) e 3 (em verde), para PCA das regiões LIBS selecionadas da tabela 7. Diferenciação das amostras carbonato das amostras não-carbonato. Em destaque, principais elementos químicos responsáveis pela variância em cada componente principal.....	50
Figura 21 - Dispersão PCA 3D (PC1 x PC2 x PC3), após a seleção das regiões espectrais da Tabela 8. Em preto, são os scores das amostras pertencentes a classe mineral Fluorita. Em cinza, são os scores das amostras pertencentes as classes minerais Oxido, Fosfato, Carbonato e Silicato. ....	52
Figura 22 - Loadings das componentes principais 1 (em azul), 2 (em roxo) e 3 (em verde), para PCA das regiões LIBS selecionadas da tabela 8. Diferenciação das amostras fluorita das amostras não-fluorita. Em destaque, principais elementos químicos responsáveis pela variância em cada componente principal. ....	53
Figura 23 - Matriz de confusão para classe mineral silicato versus não-silicato, mostrando os valores da acurácia de classificação, sensibilidade e especificidade. Matriz de confusão para os classificadores Weighted KNN e Cubic SVM.....	55
Figura 24 - Matriz de confusão para classe mineral oxido versus não-oxido, mostrando os valores da acurácia de classificação, sensibilidade e especificidade. Matriz de confusão para os classificadores Fine KNN e Cosine KNN. ....	56
Figura 25 - Matriz de confusão para classe mineral fosfato versus não-fosfato, mostrando os valores da acurácia de classificação, sensibilidade e especificidade. Matriz de confusão para os classificadores Fine KNN, Weighted KNN, Cubic SVM, Linear SVM e Quadratic SVM. ....	57
Figura 26 - Matriz de confusão para classe mineral carbonato versus não-carbonato, mostrando os valores da acurácia de classificação, sensibilidade e especificidade. Matriz de confusão para os classificadores Fine KNN, Weighted KNN, Cubic SVM, Linear SVM e Quadratic SVM. ....	58
Figura 27 - Matriz de confusão para classe mineral fluorita versus não-fluorita, mostrando os valores da acurácia de classificação, sensibilidade e especificidade. Matriz de confusão para os classificadores KNN (Fine e Weighted) e SVM (Cubic, Fine Gaussian e Medium Gaussian). ....	59
Figura 28 - Matriz de confusão para o classificador Cubic SVM, após a seleção de componentes principais. Classe mineral Silicato versus não-silicato. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem. ....	60

Figura 29 - Matriz de confusão para o classificador Weighted KNN, após a seleção de componentes principais. Classe mineral Silicato versus não-silicato. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem. ....	61
Figura 30 - Matriz de confusão para o classificador Cosine KNN, após a seleção de componentes principais. Classe mineral Oxido versus não-oxido. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem. ....	62
Figura 31 - Matriz de confusão para o classificador Fine KNN, após a seleção de componentes principais. Classe mineral Oxido versus não-oxido. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem. ....	63
Figura 32 - Matriz de confusão para os classificadores KNN e SVM, após a seleção de componentes principais. Classe mineral Fosfato versus não-fosfato. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem. ....	65
Figura 33 - Dispersão PCA 3D (PC1 x PC2 x PC3), para as amostras de treinamento. Em verde, são os scores das amostras pertencentes a classe mineral Silicato. Em cinza, são os scores das amostras pertencentes a classe não-silicato. ....	68
Figura 34 - Matriz de confusão para o classificador Fine KNN, para as amostras de treinamento, classe mineral silicato versus classe mineral não-silicato. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem. ....	69
Figura 35 - Matriz de confusão para o classificador Fine KNN, após a seleção de componentes principais. Classe mineral Silicato versus não-silicato, para as amostras de treinamento. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem. ....	70
Figura 36 - Dispersão PCA 3D (PC1 x PC2 x PC3), para as amostras de treinamento, após a seleção de regiões espectrais LIBS da tabela 4. Em verde, são os scores das amostras pertencentes a classe mineral Silicato. Em cinza, são os scores das amostras pertencentes a classe mineral não-silicato. ....	71
Figura 37 - Matriz de confusão para os classificadores KNN e SVM, para seleção de regiões espectrais LIBS da tabela 4. Classe mineral Silicato versus não-silicato, para as amostras de treinamento. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem. ....	72
Figura 38 - Matriz de confusão para o classificador Fine KNN, após o processo de otimização de PC`s. Classe mineral Silicato versus não-silicato, para as amostras de validação externa. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem. ....	73
Figura 39 - Matriz de confusão para o classificador Fine KNN, após a seleção de componentes principais. Classe mineral Silicato versus não-silicato, para as amostras de validação externa. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem. ....	74

Figura 40 - Matriz de confusão para os classificadores KNN e SVM, após a seleção de regiões espectrais LIBS. Classe mineral Silicato versus não-silicato, para as amostras de validação externa. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem..... 75

Figura 41 - Fluxograma mostrando como seriam as etapas para a classificação de uma amostra mineral externa dentre as classes silicato, oxido, carbonato, fosfato ou fluorita, na qual os algoritmos supervisionados de máquinas foram treinados..... 78

## SUMÁRIO

1. INTRODUÇÃO.....	13
2. OBJETIVOS.....	17
2.1. Objetivos Gerais .....	17
2.2. Objetivos Específicos .....	17
3. MATERIAIS E MÉTODOS .....	19
3.1. Amostras Minerais .....	19
3.2. Caracterização Óptica das Amostras Minerais por Espectroscopia Óptica por Plasma Induzido por <i>Laser</i> – LIBS.....	20
3.2.1. Espectroscopia Óptica por Plasma Induzido por Laser – LIBS	20
3.2.2. Caracterização Óptica das Amostras Minerais	21
3.3. Análise Multivariada e Algoritmos de Aprendizado de Máquinas.....	22
3.3.1. Análise das Componentes Principais - PCA	23
3.3.2. Algoritmos de Aprendizado de Máquinas	25
3.3.2.1. K-vizinhos próximos (KNN)	26
3.3.2.2. Análise Discriminante	27
3.3.2.3. Máquinas de Vetores de Suporte	28
3.4. Métricas de Avaliação dos Algoritmos de Aprendizado de Máquinas.....	31
3.5. Algoritmo Seletor de Componentes Principais.....	33
4. RESULTADOS E DISCUSSÕES.....	35
4.1. Potencial de Diferenciação de uma Classe Mineral Específica <i>versus</i> Restante das Classes de Minerais .....	36
4.1.1. Seleção de Regiões Espectrais LIBS para Separação de Classes	39
4.1.2. Aplicação de Algoritmos Supervisionados de Aprendizado de Máquinas para Diferenciação e Classificação das Amostras Minerais	53
4.1.3. Algoritmo Seletor de Componentes Principais	59
4.2. Classe Silicato <i>versus</i> não-silicato: Treinamento e Validação Externa.....	66
4.2.1. Treinamento dos Algoritmos Supervisionados de Aprendizado de Máquinas	67
4.2.2. Algoritmo Seletor de Componentes Principais – Amostras de Treinamento	69
4.2.3. Seleção de Regiões Espectrais LIBS - Amostras de Treinamento	70
4.2.4. Validação Externa	72
5. CONCLUSÃO.....	76
6. PERSPECTIVAS FUTURAS .....	78
7. REFERÊNCIAS BIBLIOGRÁFICAS .....	80

## 1. INTRODUÇÃO

Rochas são os principais constituintes geológicos que formam a maior parte da crosta terrestre. As rochas são formadas, principalmente, por aglomerados de um ou mais grupo de minerais. Minerais são definidos como uma substância natural inorgânica, formada através de processos geológicos, que possuem uma estrutura cristalina regular e sólida a temperatura ambiente (SALT LAKE, 2023). Os minerais são provenientes da cristalização do magma terrestre, de precipitações de soluções aquosas ou precipitações biológicas através da ação de organismos naturais (PÉREZ-AGUILAR; CUÉLLAR-CRUZ, 2022). Os minerais são formados através da ligação química entre elementos químicos diferentes, onde essas ligações ocorrem entre um ânion (elemento químico eletricamente carregado que ganhou elétrons) e um cátion (elemento químico eletricamente carregado que perdeu elétrons). A nomenclatura dos minerais é definida pelo seu grupo ânion predominante. O grupo de minerais Silicato é um composto de silício-oxigênio, com a predominância do ânion  $\text{SiO}_4^{4-}$  (KAHLENBERG; KAINDL; KONZETT, 2007), que possui uma estrutura cristalina tetraédrica. Essa estrutura permite que haja a formação de ligações químicas com vários outros íons, formando assim o grupo de minerais silicato. É o mineral que se encontra em maior abundância, com grandes quantidades tanto no manto quanto na crosta terrestre. Exemplos de minerais silicato são as olivinas ( $(\text{Fe,Mg})_2\text{SiO}_4$ ), quartzo ( $\text{SiO}_2$ ) e feldspat (  $\text{NaAlSi}_3\text{O}_8$ ) (WANG et al., 2021). O grupo de minerais Óxidos possuem o grupo ânion de oxigênio  $\text{O}^{2-}$  como predominante (FENG et al., 2022). O grupo Óxido apresenta ligações com os elementos ferro (Fe), manganês (Mn) e alumínio (Al), onde são exemplos os minerais corindo ( $\text{Al}_2\text{O}_3$ ) e hematita ( $\text{Fe}_2\text{O}_3$ ) (ZHU et al., 2020). O grupo de minerais Carbonato apresenta a predominância do grupo ânion  $\text{CO}_3^{2-}$  (DUFRESNE; RUFLEDT; MARSHALL, 2018), que formam ligações químicas com os elementos cálcio (Ca), magnésio (Mg), ferro (Fe) e manganês (Mn), formando os mais conhecidos minerais calcita ( $\text{CaCO}_3$ ), dolomita ( $\text{CaMg}(\text{CO}_3)_2$ ) e aragonita ( $\text{Ca}(\text{CO}_3)$ ) (COIMBRA et al., 2021). O grupo de minerais Fosfato apresenta a predominância do grupo ânion  $\text{PO}_4^{3-}$  (YANG et al., 2022), que formam ligações químicas com elementos como o Alumínio (Al) e Cálcio (Ca). São representados pela variscita ( $\text{Al}(\text{PO}_4)2\text{H}_2\text{O}$ ) (FRITSCH; KARAMPELAS; MEVELLEC, 2017) e pela apatita  $\text{Ca}_5(\text{PO}_4)_3(\text{OH})$ , mineral do grupo fosfato em maior abundância (ZHOU et al., 2020). O grupo de

minerais Fluorita apresenta o ânion  $F^-$ , cuja ligação química com o elemento químico cálcio (Ca), forma o mineral cálcio fluorita ( $CaF_2$ ) (WANG et al., 2022).

A classificação e a diferenciação de amostras minerais são de grande relevância para vários campos de estudo. Para identificação e classificação das amostras minerais, um tipo de teste padrão comum é a inspeção visual (BUSTILLO et al., 2009). No geral, amostras minerais de classes diferentes apresentam características visuais diferentes, como coloração, textura ou opacidade. Porém, pode ocorrer de amostras pertencentes a diferentes classes, apresentarem características visuais similares. Ou, vice-versa, amostras que são pertencentes a uma mesma classe, apresentarem características visuais diferentes. Essa análise macroscópica visual dificulta a diferenciação de amostras minerais, o que exige certas habilidades de um especialista para a correta classificação da mesma. Um outro tipo de teste padrão utilizado é a aplicação de técnicas analíticas espectroscópicas, como Espectroscopia por Raio-X (ROLDÁN et al., 2015) ou Espectroscopia de Emissão Óptica por Plasma Acoplado Indutivamente (*Inductively Coupled Plasma – ICP*) (NAVAZO et al., 2008), para a quantificação de elementos químicos, e assim serem capazes de classificar as amostras em suas respectivas classes, com base nas suas composições químicas. A ICP apresenta como desvantagem a necessidade de rigoroso pré-preparo das amostras, comumente com a necessidade do uso de amostras diluídas para serem pulverizadas e analisadas pelo equipamento (GARCIA PINTO; ESCALFONI; SAINT'PIERRE, 2012). Também, seu uso *in situ* não é possível devido ao aparato experimental, o que a torna não factível para o uso de classificação de amostras minerais de maneira rápida. Por sua vez, a espectroscopia por raio-x apresenta a vantagem de instrumentação portátil (*handheld XRF instruments*), o que permite rápida caracterização de amostras *in situ*, porém essa técnica apresenta dificuldade na detecção de elementos leves, com baixos números atômicos, o que é fundamental para identificar e distinguir amostras minerais (SWEETAPPLE; TASSIOS, 2015).

Recentemente, um outro tipo de técnica analítica rápida, de caráter multi-elementar, que necessita de pouco preparo de amostra, é a Espectroscopia de Emissão Óptica por Plasma Induzido por Laser - LIBS, do acrônimo *Laser Induced Breakdown Spectroscopy* (LARIOS et al., 2020b). A técnica LIBS é uma técnica de espectroscopia óptica de emissão atômica, utilizada para análises quantitativas e qualitativas, dos elementos químicos presentes em amostras (RIBEIRO et al., 2020). Nela, um plasma é

gerado na superfície da amostra, devido a pulsos de laser de alta intensidade. Quando o plasma atinge a superfície da amostra, a sua elevada temperatura (acima de 50.000 K) provoca a ruptura de uma pequena porção da superfície da amostra, a sua vaporização e consequentemente a produção de íons e átomos excitados. Quando o plasma começa a se resfriar (cerca de 15.000 K), os elementos presentes no plasma começam a emitir em determinadas linhas espectrais, como “impressões digitais” específicas de cada elemento, o que permite a identificação, caracterização e quantificação desses elementos presentes na amostra (NICOLODELLI et al., 2019).

Trabalhos utilizando a técnica LIBS se mostraram promissores na caracterização (WIENS et al., 2005), diferenciação (MCMILLAN et al., 2007), quantificação e identificação (EL HADDAD et al., 2019) de amostras minerais. Combinados com métodos quimiométricos e algoritmos supervisionados de aprendizado de máquinas (GOTTFRIED et al., 2009; PALÁSTI et al., 2022) mostraram a aplicação dessa combinação para diferenciação e classificação de amostras minerais. Utilizando um instrumento portátil LIBS (hLIBS – *Handheld LIBS*), (KIM; FABRE; CAUZID, 2022) quantificaram e caracterizaram amostras minerais, enquanto (LAWLEY; SOMERS; KJARSGAARD, 2021) utilizaram análises qualitativas para identificação e diferenciação das amostras minerais. Combinando hLIBS com algoritmos supervisionados de aprendizado de máquinas, (SENESI et al., 2023) provaram ser capazes de classificar e diferenciar amostras rochosas (*Cherts*) *in situ*. Essa combinação apresentou vantagens, como de ser um procedimento não destrutivo, e que serve como um teste rápido para identificação e classificação desse tipo de amostras.

Um problema comum em museus, é a obtenção da correta classificação de amostras minerais de maneira rápida, principalmente utilizando técnicas portáteis e de modo que seja não destrutiva. Neste trabalho, o uso da espectroscopia óptica LIBS, combinada com métodos quimiométricos, como a análise das componentes principais (PCA), e algoritmos supervisionados de aprendizado de máquinas (em inglês *Machine Learning*), como k-vizinhos próximos (KNN – *k-nearest neighbor*), máquinas de vetores de suporte (SVM – *support vector machines*) e análise discriminante (DA – *discriminant analysis*), foram aplicados a amostras minerais provenientes do Museu de Bari - Itália, para diferenciação e classificação entre as classes silicato, óxido, fosfato, carbonato e fluorita. Como os minerais apresentam estruturas cristalinas, para manter tal estrutura cristalina, apenas certos elementos químicos podem formar ligações entre si, o

que faz com que diferentes minerais apresentem composições químicas diferentes. Assim, será avaliado o potencial dessa combinação, como um procedimento rápido e eficaz para a diferenciação e classificação de amostras minerais, baseadas em suas composições elementares.



## **2. OBJETIVOS**

### **2.1. Objetivos Gerais**

Avaliar o potencial de aplicação de técnica de espectroscopia óptica, combinado com métodos quimiométricos de análise das componentes principais e algoritmos supervisionados de aprendizado de máquinas, para a diferenciação e classificação de amostras minerais baseadas em suas composições elementares.

### **2.2. Objetivos Específicos**

Utilizar a espectroscopia de emissão óptica por plasma induzido por laser (LIBS) como técnica de caracterização das amostras minerais, analisando os principais constituintes de sua estrutura atômica;

Aplicar a análise das componentes principais (PCA) nos espectros LIBS das amostras minerais, utilizando seus *scores* para verificar possíveis agrupamentos ou diferenciação de diferentes classes de minerais, e análise das regiões espectrais que possam melhor influenciar na diferenciação ou classificação das amostras;

Utilização de métodos supervisionado de aprendizado de máquinas, combinado com a PCA, para treinamento e validação de classificadores, verificando qual possui um maior potencial de aplicação para diferenciar ou classificar as amostras minerais em suas respectivas classes;

Aplicação de um algoritmo seletor de componentes principais, baseado no tipo de funcionamento *backward interval variable selection* (SENESI et al., 2023), para selecionar as melhores combinações de componentes principais, a fim de melhorar a acurácia de diferenciação e classificação dos algoritmos supervisionados de aprendizado de máquinas;

Análise quantitativa do poder de diferenciação das amostras minerais, entre as suas classes de minerais, antes e depois da seleção de componentes principais, através da acurácia, sensibilidade, especificidade e matriz de confusão, dos classificadores supervisionado de aprendizado de máquinas. Comparar e verificar qual classificador é o

ideal para a aplicação proposta, verificando os resultados da validação cruzada, e também da validação externa.

### 3. MATERIAIS E MÉTODOS

#### 3.1. Amostras Minerais

As amostras minerais foram fornecidas pela Universidade de Bari Aldo Moro (*Università Degli Studi di Bari Aldo Moro*) – Itália.

No total, 135 (cento e trinta e cinco) amostras foram fornecidas e analisadas por espectroscopia óptica por plasma induzido por laser – (LIBS), sendo 95 amostras pertencentes a classe mineral Silicato, 27 amostras pertencentes a classe mineral Óxido, 7 amostras pertencentes a classe mineral Fosfato, 3 amostras pertencentes a classe mineral Carbonato e 3 amostras pertencentes a classe mineral Fluorita.

As amostras não sofreram nenhum tipo de pré-preparo e nenhuma preparação da superfície de análise. A figura 1 mostra algumas das amostras minerais que foram analisadas.

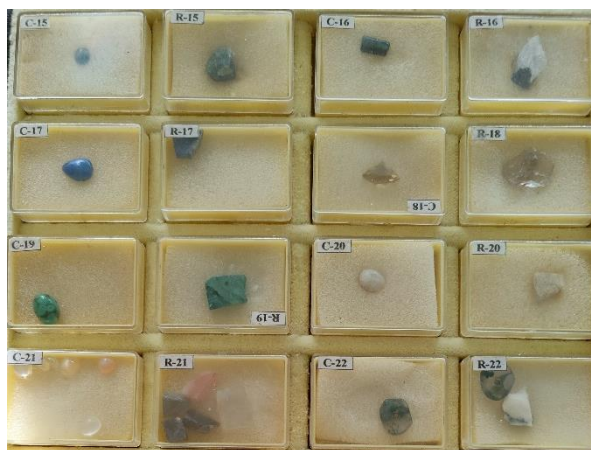


Figura 1 – Amostras minerais, de diferentes classes minerais, a serem analisadas, diferenciadas e classificadas.

### 3.2. Caracterização Óptica das Amostras Minerais por Espectroscopia Óptica por Plasma Induzido por *Laser* – LIBS

#### 3.2.1. Espectroscopia Óptica por Plasma Induzido por Laser – LIBS

A espectroscopia óptica por plasma induzido por laser – LIBS, do acrônimo em inglês *Laser Induced Breakdown Spectroscopy*, é uma técnica óptica de caráter multielementar, que obtém informações a respeito da composição atômica das amostras, ao se coletar a radiação eletromagnética emitida quando os átomos são excitados devido a incidência de um laser de alta potência (RANULFI et al., 2018)

A caracterização elementar se inicia quando um laser de alta potência é focado e incide a superfície da amostra. O laser possui irradiância na ordem de  $1 \frac{\text{GW}}{\text{cm}^2}$ , o que junto com sua curta duração de incidência, remove uma pequena quantidade de massa da superfície da amostra, na ordem de ng, em um processo chamado de ablação. A incidência do laser provoca um grande aumento de temperatura local, o que faz com que a massa se vaporize, fazendo com que ocorra a formação de um plasma. O plasma atinge elevadas temperaturas, cerca de 30.000 K (SENESI, 2014), o que fornece energia para que os átomos dentro do plasma se excitam, formando íons e elétrons livres. Conforme o plasma começa a se resfriar, os íons e elétrons livres começam a se recombinar, e durante essa recombinação, energia é liberada em forma de radiação eletromagnética. A radiação eletromagnética liberada consiste em linhas espectrais atômica e iônicas características de cada átomo presente na amostra (MARKIEWICZ-KESZYCKA et al., 2017).

A técnica LIBS pode ser aplicada em amostras líquidas (BELLOU et al., 2020), sólidas ou gases (SEZER et al., 2018), tanto para análises qualitativas ou quantitativas (AKHTAR et al., 2018; STEFAS et al., 2019). É considerada uma técnica não destrutiva, de caráter multielementar, que não necessita ou necessita de pouco preparo de amostra, e apresenta aparato experimental portátil, o que permite análises *in situ* (PÉREZ-DIEZ et al., 2021).

Em relação a instrumentação LIBS, consistem em um laser pulsado de alta potência Nd:YAG (Neodymium-doped Yttrium/Aluminum Garnet), operando no comprimento de onda de 1064 nm. Ao incidir a superfície da amostra, na qual o laser é focalizado em uma pequena área ao passar através de lentes focalizadoras, se inicia o

processo de ablação, vaporização da massa e geração do plasma. Quando o plasma começa a se resfriar, alguns  $\mu\text{s}$  após sua formação, os átomos começam a se recombinar, e então um conjunto de lentes é utilizado para coletar as radiações eletromagnéticas, em forma de luz, que são emitidas por esses átomos. Através de fibras ópticas, a luz coletada é conduzida a um espectrômetro, que difrata a luz em diferentes comprimentos de onda, e posteriormente é enviada ao um detector. Através de um computador, os sinais são armazenados, processados e visualizados por meio de um software dedicado. A figura 2 abaixo mostra a configuração básica para a instrumentação LIBS.

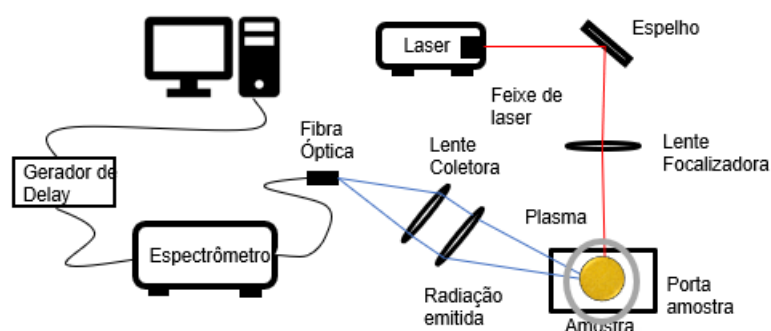


Figura 2 - Configuração básica da instrumentação LIBS.

### 3.2.2. Caracterização Óptica das Amostras Minerais

A caracterização óptica das amostras minerais foram obtidas utilizando um equipamento LIBS portátil Z-300 (SciAps, Inc, Woburn, MA, USA), com dimensões 21 x 29 x 11 cm, com peso de 1,8 kg, alimentado por uma bateria de íons de lítio. O equipamento utiliza um laser Nd:YAG pulsado de estado sólido bombeado por diodo, PULSARTM classe 3, 1064 nm, produzindo um feixe focado de  $100\mu\text{m}$ , que transmite um pulso de 5 – 6 mJ para a amostra, com pulso de duração de 1 ns, para taxa de disparo entre 1 e 50 Hz.

O equipamento Z-300 coleta larga faixa espectral de emissão de luz do plasma, de 190 a 950 nm. O sinal de emissão de luz LIBS foi coletado através de um cabo de fibra óptica, e transmitido a três espectrômetros internos, utilizando detectores CCD com bloqueio de tempo, com faixas espectrais e resoluções de, respectivamente, 190-365 nm com FWHM (*Full Width at Half Maximum* – largura total na metade do

máximo) de 0,18 nm, 365-620 nm com FWHM de 0,24 nm e 620-950 nm com FWHM de 0,35 nm. Os dados espectrais foram coletados com um atraso de 650 ns sobre tempo de integração de 1 ms. O analisador contém uma camera de video para visualização da amostra antes da análise, e um localizador do foco do laser, para mostrar onde o feixe de laser irá atingir a amostra.

A obtenção dos espectros LIBS ocorreu no laboratório do Conselho Nacional de Pesquisa (*Consiglio Nazionale delle Ricerche – CNR*), localizado na cidade de Bari, Itália. Para caracterização elementar, médias de cinco espectros coletados foram calculadas, em uma taxa de disparo de laser de 50 Hz sobre uma grade de 4 x 3. As medições ocorreram ao se colocar o equipamento portátil LIBS contra as amostras, onde a obtenção dos espectros se iniciou através do acionamento de um gatilho.

### **3.3. Análise Multivariada e Algoritmos de Aprendizado de Máquinas**

Após a caracterização óptica por LIBS das amostras minerais, métodos quimiométricos, como a análise das componentes principais (PCA), e posteriormente, algoritmos de aprendizado de máquinas, foram combinados e utilizados para analisar o potencial de separação das amostras, em cinco diferentes classes de minerais. Antes de iniciar essas análises, os espectros LIBS das amostras foram pré-processados pelo procedimento de normalização SNV (*Standard Normal Variate*), que seta a média em zero e o desvio padrão em 1. O procedimento de pré-processamento SNV foi feito para que todos os espectros amostrais apresentem a mesma variância de distribuição, fazendo com que as análises não sejam influenciadas por variações espectrais provenientes de flutuações externas (KUSUMANINGRUM et al., 2018). A partir dos espectros SNV-LIBS em quintuplicata, 1 (um) espectro médio para cada amostra foi obtido, totalizando assim 135 espectros SNV-LIBS. A figura 3 mostra o fluxograma das amostras minerais, desde a caracterização óptica LIBS até a diferenciação das amostras utilizando métodos quimiométricos e de aprendizado de máquinas. As análises quimiométricas e os algoritmos de aprendizado de máquinas foram executados no software Matlab<sup>®</sup> r2015b.

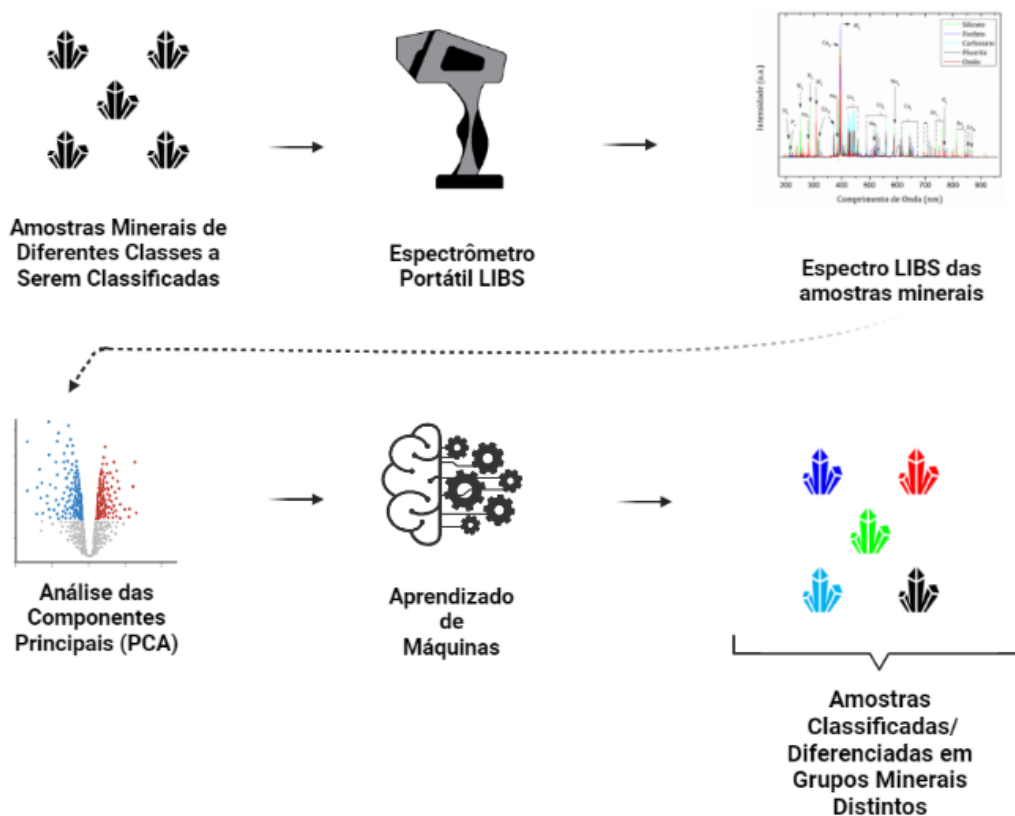


Figura 3 - Fluxograma da caracterização óptica LIBS das amostras minerais, seguida de análise por PCA e algoritmos de aprendizado de máquinas, para diferenciação das amostras com base em suas composições elementares.

### 3.3.1. Análise das Componentes Principais - PCA

A Análise das Componentes Principais, PCA do acrônimo em inglês *Principal Component Analysis*, é um procedimento matemático utilizado, principalmente, para se fazer inferências, ou seja, verificar tendências de similaridades ou diferenças, e para reduzir a dimensionalidade, ou seja, a quantidade de variáveis relacionadas a cada amostra, em um conjunto de dados (RUI ALVES, 2012).

Matematicamente, a PCA é um procedimento de transformação linear, na qual as variáveis iniciais do conjunto de dados amostrais, que são linearmente dependentes, são transformadas em um novo conjunto de variáveis, de quantidade reduzidas e linearmente independentes, chamadas de componentes principais (PC's). De uma forma mais geral, a PCA, através da decomposição de valores singulares (SVD do acrônimo em inglês *singular value decomposition*), centra na média a matriz inicial dos dados

amostrais, e depois a decompõe em três matrizes, chamadas de matriz coluna [**U**], vetores singulares esquerdo, matriz diagonal [**S**], valores singulares e matriz coluna [**V**], vetores singulares direito (AKRITAS; MALASCHONOK, 2004). Ordenando de forma decrescente os autovalores ( $\lambda$ ) da matriz inicial dos dados amostrais, que podem ser calculados através dos valores singulares da matriz diagonal [**S**], e seus respectivos vetores singulares esquerdo [**U**] e direito [**V**], a transformação linear da PCA é obtida ao se projetar a matriz centrada inicial dos dados na matriz ordenada dos vetores singulares direito [**V**]. Os coeficientes das componentes principais, chamados de *loadings*, são os responsáveis pela transformação linear dos dados amostrais, e são representados por cada coluna da matriz ordenada [**V**]. Os *loadings* são ortogonais entre si, ou seja, linearmente independentes, e eles representam a relação das variáveis iniciais com cada componente principal no espaço da PCA. A primeira componente principal, PC1, carrega a maior variância amostral, seguida da PC2, e assim sucessivamente, até a última PC. Durante a PCA, a variância total amostral, dado pela soma de cada autovalor ( $\lambda$ ) ordenado, pode ser obtida por uma quantidade de PC's muito inferior a quantidade de variáveis originais, sem muita perda de informações referente aos dados amostrais. Assim, a redução de dimensionalidade é obtida na PCA. A matriz de *scores* da PCA, representa a projeção da matriz inicial dos dados amostrais no espaço das componentes principais, e relacionam as componentes principais com cada amostra. Dessa maneira, as similaridades e diferenças entre as amostras podem ser vistas através da formação de agrupamentos, permitindo a interpretação e análise de um conjunto de dados.

A análise das componentes principais foi utilizada como um procedimento prévio de organização dos dados amostrais para os algoritmos supervisionados de aprendizado de máquinas (*Machine Learning*). Como a maioria dos classificadores supervisionados de aprendizado de máquinas possuem seus critérios de classificação de amostras desenvolvidos para serem aplicados em dispersões ortogonais, como por exemplo um espaço 2D (plano x-y), 3D (x-y-z) ou com maiores de três dimensões, a PCA garante que a dispersão dos dados amostrais siga tal exigência, de modo que a discriminação entre as amostras possa ser possível.



### 3.3.2. Algoritmos de Aprendizado de Máquinas

Algoritmos de aprendizado de máquinas, ou mais comumente utilizado o termo *Machine Learning*, é uma técnica computacional que tem por objetivo ensinar o computador a realizar determinada tarefa, deixando-o aprender através de informações colhidas diretamente da matriz de dados amostrais.

Os algoritmos de aprendizado de máquinas são subdivididos em duas principais categorias: algoritmos supervisionados e algoritmos não-supervisionados. Os algoritmos supervisionados são utilizados, principalmente, para tarefas de classificação ou de regressão, onde informações a respeito da categoria, ou rotulagem, de cada amostra é informada. Dessa maneira, os algoritmos supervisionados, além de extraírem informações estatísticas da matriz de dados amostrais, possuem também a informação de qual categoria cada amostra pertence, construindo seus parâmetros de classificação para discriminar as amostras nos grupos informados. Exemplos de algoritmos supervisionados são k-vizinhos próximos (KNN), máquinas de vetores de suporte (SVM) e análise discriminante (D.A.) Os algoritmos não-supervisionados, por outro lado, não possuem as informações das categorias de cada amostra, na qual sua principal função é extrair e verificar se há padrões na matriz de dados amostrais que possam levar a agrupamento (clusterização) ou subdivisões das amostras em certos grupos. Assim, algoritmos não-supervisionados são utilizados para clusterização de amostras. A análise das componentes principais (PCA) é um exemplo de algoritmo não supervisionado (HANDELMAN et al., 2019).

A aplicação de algoritmos supervisionado de aprendizado de máquinas, tanto para classificação ou regressão, comumente se dá em duas etapas. A primeira etapa é chamada de treinamento do classificador, onde os dados amostrais juntos com suas respectivas classes são informados a determinado classificador, para que o mesmo possa construir seus parâmetros de classificação, e aprender a diferenciar as amostras. A segunda etapa se dá pela validação interna do classificador, para avaliar sua performance em diferenciar as amostras. Nessa etapa, as mesmas amostras do conjunto de treinamento do classificador são utilizadas para avaliar sua performance, através do método de validação cruzada (*cross-validation*). Recentemente, uma nova etapa está sendo aplicada, a de validação externa, na qual amostras que não fizeram parte do

conjunto de treinamento do classificador são utilizadas para avaliar seu poder de diferenciação em novos dados.

### 3.3.2.1. K-vizinhos próximos (KNN)

O algoritmo supervisionado k-vizinhos próximos (KNN do acrônimo em inglês *k-nearest neighbour*) é um algoritmo que classifica uma nova amostra na classe majoritária de seus k-vizinhos próximos (BEUREN; ANZANELLO, 2019).

Dado um conjunto amostral, onde cada amostra é pertencente a uma determinada categoria (ou classe), após o treinamento do algoritmo KNN e a construção de seus parâmetros de classificação, uma amostra externa ao conjunto de dados amostrais pode ser classificada dentre uma das classes na qual o algoritmo foi treinado. Essa classificação se dá por meio de uma distância métrica, e a quantidade de vizinhos próximos,  $k$ , a amostra externa, atribuindo-a a classe majoritária de seus k-vizinhos próximos. A figura 4 representa a classificação de uma amostra externa, em um plano 2D, para o k-vizinhos próximos com número de vizinhos 10 e distância métrica.

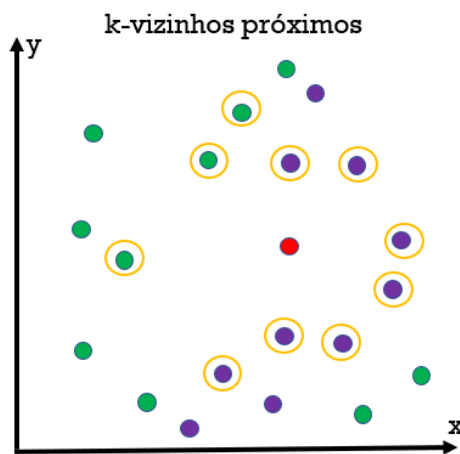


Figura 4 - Representação da classificação de uma amostra externa, em um plano 2D, utilizando o algoritmo k-vizinhos próximos. O número de vizinhos próximos,  $k$ , foi definido como 10, e estão destacados pelos círculos em laranja. Em azul, são as amostras da classe A, em verde as amostras da classe B, e em vermelho a amostra externa que será classificada pelo KNN.

Os algoritmos KNN possuem nomenclatura padrões, com base no número de vizinhos próximos, distância métrica e normalização utilizados durante a construção dos parâmetros de classificação. A tabela 1 mostra essas nomenclaturas.

Tabela 1 - Nomenclatura padrão dos algoritmos KNN com base no número de vizinhos próximos, distância métrica e normalização de distância. Fonte: Mathworks.com

Nomenclatura	Número de Vizinhos	Distância Métrica	Normalização
Fine KNN	1	Euclidiano	-
Medium KNN	10	Euclidiano	-
Coarse KNN	100	Euclidiano	-
Cosine KNN	10	Coseno	-
Cubic KNN	10	Cúbica	-
Weighted KNN	10	Euclidiano	$1/(distância)^2$

### 3.3.2.2. Análise Discriminante

A análise discriminante é um algoritmo supervisionado que tem como critério de classificação regras de decisão baseada no Teorema de Bayes, onde tem por objetivo a construção de parâmetros de classificação baseado na combinação das variáveis, dos dados amostrais, que caracterizam ou separem as amostras em duas classes. A classificação de uma nova amostra é baseada na sua probabilidade posterior de pertencer a determinada classe. A análise discriminante é dividida em duas categorias: análise discriminante linear (*Linear Discriminant Analysis – LDA*) e análise discriminante quadrática (*Quadratic Discriminant Analysis – QDA*).

A análise discriminante linear, LDA, é uma técnica de classificação que visa projetar os dados amostrais em um subespaço de menor dimensão, onde a supressão da dimensão das variáveis faça com que a extração de informações de classificação seja feita. Na LDA, para uma classificação binária, as linhas de decisões que separam as classes amostrais são sempre lineares (linhas ou hiperplanos), e se é assumido que as amostras de cada classe sigam uma distribuição multivariada normal, com iguais matrizes covariâncias, mas com médias diferentes. Baseado nesses critérios, a probabilidade posterior de uma amostra pertencer a uma determinada classe é calculada,

e a amostra é atribuída (classificada) a essa determinada classe (MOULIN et al., 2014). A linha de decisão que separa as classes amostrais está situada onde a probabilidade posterior da amostra pertencer as classes são iguais.

A análise discriminante quadrática, QDA, tem seus critérios de classificação da mesma maneira da que a relatada na descrição da análise discriminante linear. Na QDA, as amostras de cada classe também seguem uma distribuição multivariada normal, porém com diferentes matrizes covariâncias, onde cada grupo amostral possui sua específica matriz. Assim, as linhas de decisões que separam as classes amostrais se tornam não lineares, podendo assumir diferentes formatos (SAVORANI et al., 2013). Para problemas de classificação multi-classe (três classes ou mais), a LDA e a QDA podem ser aplicadas, desde que seja feito certas combinações matemáticas de suas matrizes (YANG et al., 2018). A figura 5 mostra a representação de um espaço 2D, com aplicação da LDA e QDA.

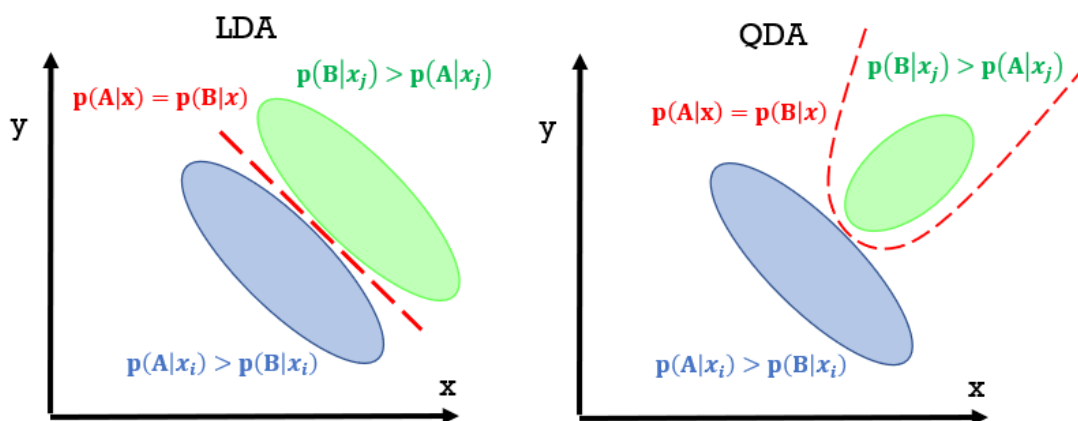


Figura 5 - a esquerda, representação do classificador LDA, a direita, representação do classificador QDA. Elipse em azul representa a matriz covariância de amostras pertencentes a classe A, elipse em verde representa a matriz covariância de amostras pertencentes a classe B. Em vermelho, linha de decisão que separam as amostras em classes, onde a probabilidade posterior das amostras a pertencerem as duas classes são iguais. Figura adaptada de (SAVORANI et al., 2013).

### 3.3.2.3. Máquinas de Vetores de Suporte

O classificador máquinas de vetores de suporte (SVM do acrônimo em inglês *support vector machine*), em classificação binária (somente duas classes), é um tipo de

classificador que tem por objetivo a separação das amostras por classe, através de um hiperplano que maximize a separação entre elas.

Em um espaço multi-dimensional, o classificador SVM separa as classes amostrais, maximizando a margem de separação entre elas. O hiperplano que separa as classes amostrais é definido pelas amostras de cada classe que estão mais separadas uma das outras, formando uma região de separação sem que haja nenhuma outra amostra entre elas. Essas amostras que maximizam a separação entre as classes são chamadas de vetores de suporte. Os vetores de suporte ajudam a definir onde a margem de separação entre as classes irá ser construída, que será o critério de decisão de classificação de uma nova amostra. A margem é construída na metade do hiperplano que maximiza a separação entre as classes, e exige que não haja nenhuma amostra dentro do hiperplano. Como em muitos casos uma separação linear total não é possível de ser feita, um termo de penalidade foi introduzido, tornando os classificadores SVM mais flexíveis (AKBARZADEH et al., 2018).

Para casos em que a separação linear das classes amostrais não possa ser feita, os algoritmos SVM trabalham com a chamada *kernel trick*, onde os dados amostrais são mapeados para um outro hiperplano, onde nesse hiperplano o SVM linear é apto de ser aplicado. Existem vários tipos de funções *kernel* que mapeiam essa transformação, sendo as mais comuns as Gaussianas e as Polinomiais (ZHANG; ZHOU; JIAO, 2004). Quando utilizado com funções *kernel*, os algoritmos SVM passam a ser de nomeados de SVM não-lineares.

Os algoritmos SVM foram construídos para análise e classificação de amostras binárias, ou seja, a diferenciação somente entre duas classes. Quando se deseja aplicar os SVM em problema multi-classe, com três classes ou mais, o método de análise *one-against-one* e *one-against-all* permitem que os classificadores SVM sejam aplicados a determinado tipos de dados. Na análise *one-against-one*, o classificador SVM aprende a discriminar todas as classes amostrais, uma da outra, treinando um classificador para cada par de classes. Na análise *one-against-all*, o classificador SVM aprende a discriminar uma classe amostral das outras, treinando um classificador para cada classe (GÖNEN; TANUGUR; ALPAYDIN, 2008). A figura 6 representa o classificador SVM linear e o SVM não-linear.

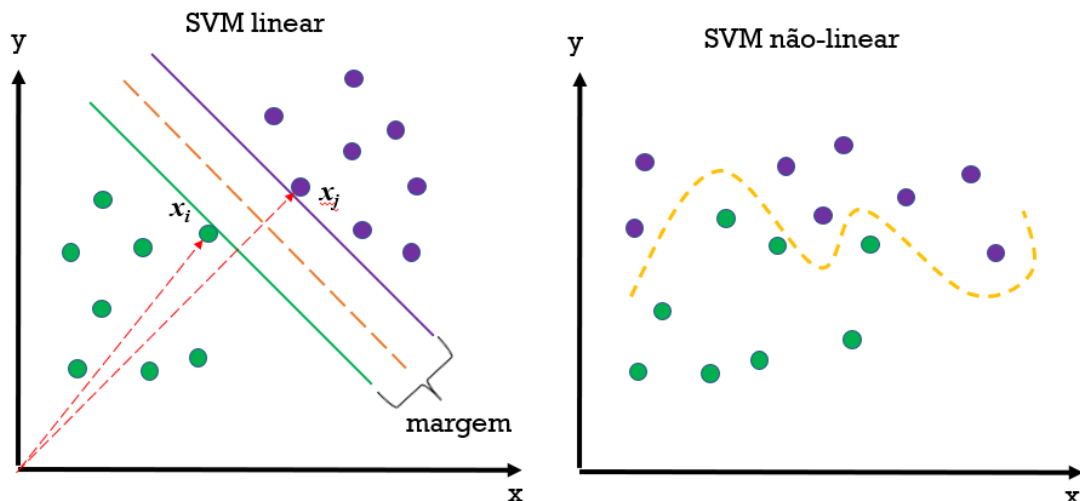


Figura 6 - a esquerda, representação do SVM linear, a direita, representação de um SVM não-linear utilizando função kernel. Em azul, amostras da classe A, com seu respectivo vetor de suporte,  $x_i$ , em verde, amostras da classe B, com seu respectivo vetor de suporte  $x_j$ . Para o SVM linear, em laranja tracejado é a linha de decisão que separa as classes, localizada no meio do hiperplano que separam as duas classes amostrais. Para o SVM não-linear, em laranja tracejado é a curva de decisão que separa as classes amostrais proveniente de uma função não-linear (kernel).

Os algoritmos SVM possuem nomenclatura padrões, de acordo com sua função *kernel* e normalização. A tabela 2 representa essas nomenclaturas.

Tabela 2 - Nomenclatura padrão dos algoritmos SVM com base na sua função kernel e normalização. Fonte: Mathworks.com

Nomenclatura	Função kernel	Normalização
Linear SVM	linear	-
Quadratic SVM	polinomial (grau 2)	-
Cubic SVM	polinomial (grau 3)	-
Fine Gaussian SVM	gaussiana	$\sqrt{(P)} / 4$
Medium Gaussian SVM	gaussiana	$\sqrt{(P)}$
Coarse Gaussian SVM	gaussiana	$\sqrt{(P)} * 4$

### 3.4. Métricas de Avaliação dos Algoritmos de Aprendizado de Máquinas

A avaliação da performance dos algoritmos de aprendizado de máquinas é realizada para avaliar e decidir qual classificador é o mais adequado de ser utilizado para determinada tarefa de diferenciação ou classificação de amostras. A validação cruzada (*cross-validation*), permite a obtenção da acurácia do modelo durante o seu treinamento, e também a construção de sua matriz de confusão.

A validação cruzada é o método mais utilizado para a avaliação da performance de um classificador. É nela que, dado o conjunto de dados amostrais, as amostras são separadas em dois conjuntos, chamados de conjunto de treinamento e conjunto de teste. As amostras do conjunto de treinamento são responsáveis por criar os parâmetros de diferenciação e classificação, na qual o classificador irá posteriormente utilizar para classificar as amostras de teste. Na validação cruzada, os dados amostrais são divididos em  $k$ -partições (*k-fold*), onde a cada  $k$ -iteração, diferentes amostras são atribuídas aos conjuntos de treinamento e de teste, fazendo com que o classificador tenha visto pelo menos uma vez, todas amostras dos dados amostrais. Os métodos mais comuns de validação cruzada são *5-fold cross-validation* (5-fold CV), onde as amostras são divididas em cinco partições (JIA, 2017), e a *leave-one-out cross-validation* (LOO-CV), onde o número de partições é igual ao número de amostras (LARIOS et al., 2021).

A partir do processo de validação cruzada, representações numéricas da performance do classificador são obtidas. A acurácia do classificador, que é a taxa de correta predição de classe das amostras, é a principal métrica avaliada. Seus valores variam de 0 a 100%, e quanto mais alto for esse valor de acurácia, melhor é o poder de discriminação do classificador. Além da acurácia do classificador, a sensibilidade, taxa de correta predição da classe das amostras de uma determinada classe, e especificidade, taxa de correta predição da classe das amostras de outra determinada classe, também são avaliadas para decidir a escolha do melhor classificador a ser utilizado (HANDELMAN et al., 2019).

A quantidade de amostras que foram corretamente classificadas, assim como também a quantidade de amostras que foram incorretamente classificadas, podem ser obtidas através da análise da matriz de confusão de cada classificador. A matriz de confusão representa, de forma visual, a performance individual de cada classificador em

predizer as classes das amostras durante o processo de validação cruzada. É a partir da matriz de confusão que as métricas de performance do classificador, como acurácia, sensibilidade e especificidade, são calculadas (THARWAT, 2018). A matriz de confusão pode ser obtida tanto para problemas de classificação binária (diferenciação entre duas classes), figura 7, ou problemas de classificação multi-classe (diferenciação de três classes ou mais), figura 8.

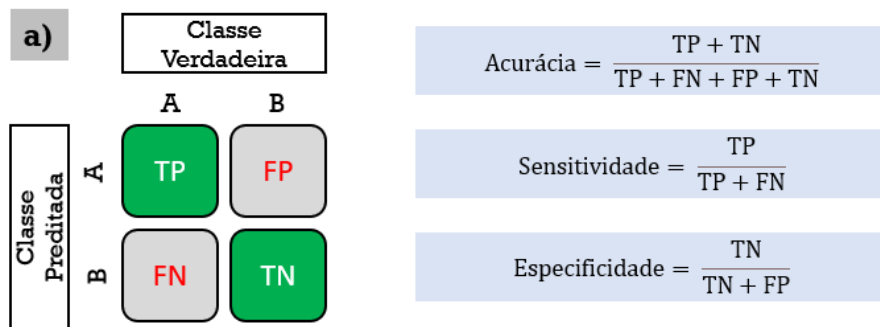


Figura 7 - Para classificação binária, a esquerda, matriz de confusão: TP (true positive – verdadeiros positivos) – amostras da classe A que foram corretamente classificadas, FN (false negatives – falsos negativos) – amostras da classe A que foram incorretamente classificadas, TN (true negatives – verdadeiros negativos) – amostras da classe B que foram corretamente classificadas, FP (false positives – falsos positivos) – amostras da classe B que foram incorretamente classificadas. A direita, cálculo da acurácia, sensibilidade e especificidade.

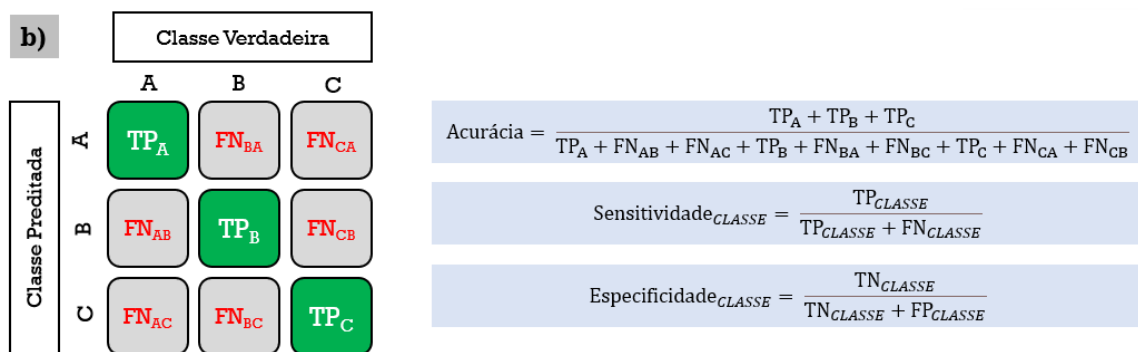


Figura 8 - Para classificação multi-classe de três classes, a esquerda, matriz de confusão: TP<sub>i</sub> (true positives – verdadeiros positivos) amostras da classe i que foram corretamente classificadas, FN<sub>ij</sub> (false negatives – falsos negativos) amostras da classe i que foram incorretamente classificados como da classe j. A direita, cálculo da acurácia, sensibilidade e especificidade.



### 3.5. Algoritmo Seletor de Componentes Principais

A seleção de variáveis é um procedimento utilizado para selecionar variáveis do conjunto de dados amostrais, visando o melhoramento da acurácia do classificador e o seu poder de diferenciação das amostras. Uma estratégia de seleção de variáveis, que possui uma boa interpretação do espectro das amostras como um todo, é a seleção de variáveis por intervalo (*interval variable selection*) (XIONG et al., 2020).

Neste trabalho, um algoritmo seletor de variáveis foi desenvolvido com base no funcionamento do seletor de variáveis *backward interval variable selection* de (YANG et al., 2019). O algoritmo seletor de variáveis desenvolvido, foi aplicado na seleção de componentes principais da análise de PCA. O algoritmo funciona da seguinte maneira: dado o conjunto amostral LIBS, após a análise de redução de dimensionalidade PCA, as componentes principais são tratadas como  $n$  divisões (intervalos). Na primeira iteração do algoritmo, um determinado classificador é treinado e validado internamente, utilizando como variáveis todas as PC`s, e também retirando, momentaneamente, cada uma das  $n$  PC`s por vez. Se a retirada de uma determinada PC produz uma melhora na acurácia do classificador, comparada com a utilização de todas as PC`s, ela é então permanentemente retirada do conjunto de dados amostrais. Na segunda iteração do algoritmo, o classificador é novamente treinado e validado com as  $n-1$  PC`s restantes, fazendo a mesma comparação das acurácias após a remoção momentânea de cada PC. Quando a retirada de alguma PC não melhora a acurácia do classificador, a seleção de variáveis atingiu sua melhor combinação, e então o algoritmo seletor de componentes principais cessa seu processo de iteração. A figura 9 mostra o princípio de funcionamento do algoritmo seletor de variáveis.

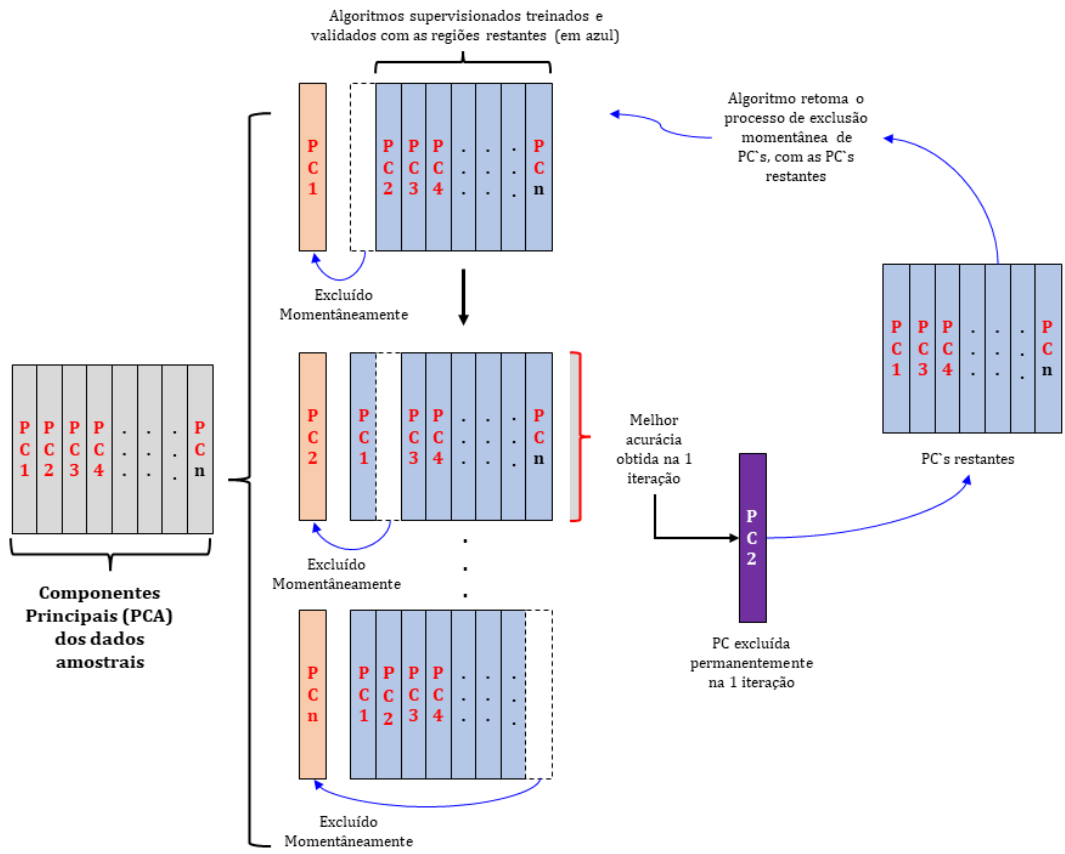


Figura 9 - Princípio de funcionamento do algoritmo seletor de componentes principais, mostrando todas as componentes principais (PC`s) dos dados amostrais (em cinza), as componentes principais (PC`s) momentaneamente excluídas (em laranja) e a componente principal (PC) permanentemente excluída (em roxo) na 1ª iteração do algoritmo.

#### 4. RESULTADOS E DISCUSSÕES

A espectroscopia óptica por plasma induzido por laser (LIBS), combinada com a análise das componentes principais (PCA) e algoritmos supervisionados de aprendizado de máquinas, como KNN, SVM e D.A. (LDA e QDA), foram utilizados para análise de separação e identificação das amostras minerais, em cinco diferentes classes, sendo elas Silicato, Oxido, Fosfato, Carbonato e Fluorita. As amostras minerais foram avaliadas e classificadas, dentre as cinco diferentes classes de minerais, em dois tipos diferentes de análises:

1 – Potencial de diferenciação de uma classe mineral específica *versus* restante das classes de minerais:

- Classe Silicato *versus* não-silicato (Oxido + Fosfato + Carbonato + Fluorita)
- Classe Oxido *versus* não-óxido (Silicato + Fosfato + Carbonato + Fluorita)
- Classe Fosfato *versus* não-fosfato (Silicato + Oxido + Carbonato + Fluorita)
- Classe Carbonato *versus* não-carbonato (Silicato + Oxido + Fosfato + Fluorita)
- Classe Fluorita *versus* não-fluorita (Silicato + Oxido + Fosfato + Carbonato)

2 – Classe Silicato *versus* Restante das Classes Minerais: amostras de treinamento e amostras de validação externa.

Antes do início das análises, os espectros médios SNV-LIBS foram analisados por um método de detecção de *outliers* (amostras discrepantes do conjunto de dados). O método utilizado foi o mapeador de ângulo espectral (SAM – do acrônimo em inglês *Spectral Angle Mapper*). No método SAM, cada espectro amostral é tratado como um vetor, e dois espectros são comparados ao se calcular o cosseno do ângulo entre eles. Quanto mais próximo de 1 for esse valor, mais similares são esses espectros (SENESI et al., 2023). As amostras *outliers* de cada classe foram detectadas e excluídas dos seus respectivos conjunto de dados, onde essas foram as que obtiveram a média dos valores do cosseno do ângulo menor que 0,5. Esse processo de detecção e exclusão de *outliers* foi realizado devido as amostras minerais, pertencentes a uma mesma classe, apresentarem muita discrepância entre si. Assim, a tabela 3 mostra a quantidade inicial de amostras em cada classe, quantas amostras foram excluídas e a quantidade total de amostras que permaneceram nos conjuntos de dados amostrais, após a aplicação do

SAM. O conjunto final de amostras, após o SAM, foi o utilizado na análise de dados que se segue.

Tabela 3 - Comparativo da quantidade de amostras iniciais, antes do método de detecção de outliers SAM, a quantidade de amostras excluídas em cada classe, e a quantidade de amostras mantidas após a aplicação do método de detecção de outliers SAM.

SNV-LIBS: CONJUNTO INICIAL DE AMOSTRAS		AMOSTRAS OUTLIERS	SNV-LIBS: CONJUNTO FINAL DE AMOSTRAS
CLASSE	QUANTIDADE	QUANTIDADE	QUANTIDADE
SILICATO	95	22	73
OXIDO	27	8	19
FOSFATO	7	0	7
CARBONATO	3	1	2
FLUORITA	3	0	3
TOTAL	135	31	104

#### 4.1. Potencial de Diferenciação de uma Classe Mineral Específica *versus* Restante das Classes de Minerais

Os 104 (cento e quatro) espectros SNV-LIBS das amostras minerais foram avaliados, no *range* espectral de 180 a 961 nm. Os espectros médios SNV-LIBS para cada classe são mostrados na figura 10. De acordo com o banco de dados do Instituto Nacional de Padrões e Tecnologia (NIST – acrônimo do inglês *National Institute of Standard and Technology*) (NIST - NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 2021), os principais elementos químicos observados nas classes foram: Silício (Si), Alumínio (Al), Calcio (Ca), Magnésio (Mg), Potássio (K), Fósforo (P), Oxigênio (O), Argônio (Ar) e Sódio (Na).

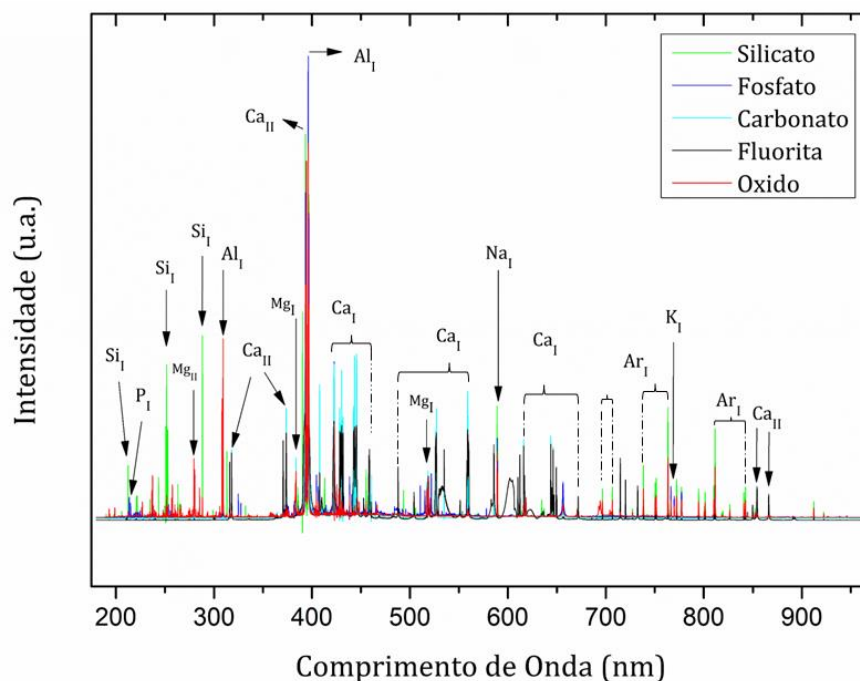


Figura 10 – Espectro médio SNV-LIBS das amostras das classes Silicato (verde), Fosfato (azul), Carbonato (ciano), Fluorita (preto) e óxido (vermelho). Em destaque, alguns elementos químicos observados em cada classe.

De acordo com a figura 10, os espectros SNV-LIBS apresentaram emissão espectral de um mesmo elemento químico para mais de uma classe mineral. As classes Silicato, Fosfato e Óxido apresentaram emissão para o alumínio (Al), argônio (Ar), cálcio (Ca), potássio (K), magnésio (Mg), sódio (Na) e oxigênio (O), enquanto as classes Carbonato e Fluorita apresentaram emissões, principalmente, para o cálcio (Ca) e sódio (Na).

A partir da análise da figura 10, se nota que os espectros médios SNV-LIBS não apresentaram picos de emissões espectrais que possam, a priori, diferenciar as classes minerais uma das outras. Assim, a utilização de métodos quimiométricos se torna adequada a fim de se explorar matematicamente essas informações espectrais, para se obter a diferenciação entre as classes minerais. A Análise das Componentes Principais (PCA) foi utilizada para verificar o potencial de separação/agrupamento (clusterização) das amostras pertencentes as classes minerais Silicato, Fosfato, Carbonato, Fluorita e Óxido. A figura 11 mostra a dispersão 3D dos *scores* da PCA das amostras, nas componentes principais PC1, PC2 e PC3.

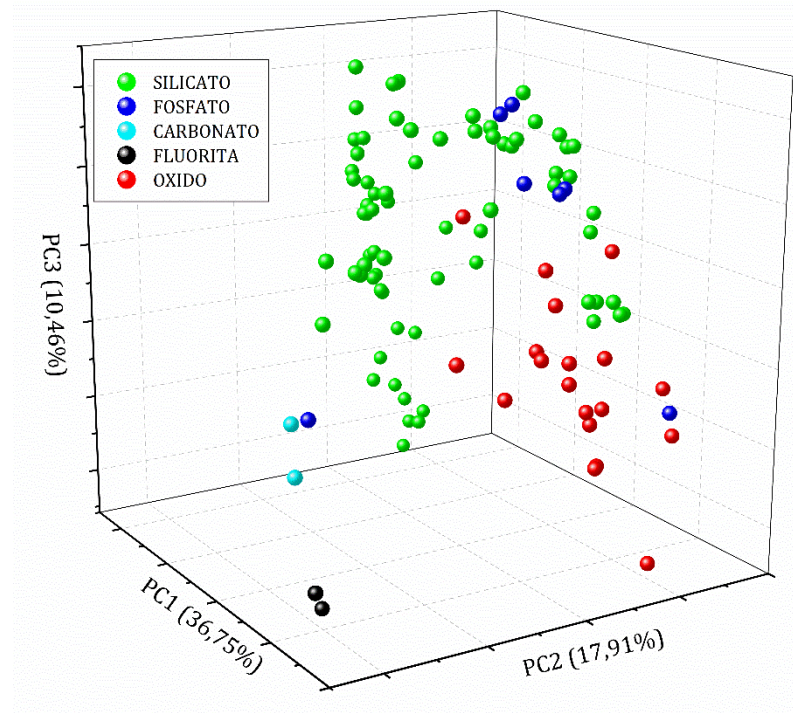


Figura 11 - Dispersão PCA 3D das amostras minerais. Classe mineral: Silicato (esferas em verde), Fosfato (esferas em azul), Carbonato (esferas em cian), Fluorita (esferas em preto) e Oxido (esferas em vermelho). Cada esfera colorida representa uma amostra diferente do conjunto de dados.

A análise da figura 11 mostra que a variância na PC1, PC2 e PC3 foi de 36,75%, 17,91% e 10,46%, respectivamente. Se nota que três classes apresentaram agrupamento entre si, não mostrando tendência de diferenciação, havendo uma sobreposição das amostras pertencentes as classes minerais Silicato (verde), Fosfato (azul escuro) e Oxido (vermelho). Isso pode ser atribuído ao fato de que essas amostras apresentam composição e assinaturas elementar espectral similares (RIBEIRO et al., 2020). Já as classes Carbonato (cian) e Fluorita (preto), apresentaram agrupamento entre suas amostras e separação do restante do conjunto. A análise dos *loadings* das três primeiras componentes principais, figura 12, mostra que as maiores contribuições para a variância dos dados amostrais podem ser atribuídas aos elementos químicos Si<sub>I</sub> (212,4 nm, 251,6 nm, 288,1 nm, 390,5 nm), Al<sub>I</sub> (308,2 nm, 309,3 nm, 394,4 nm, 396,1 nm), Be<sub>II</sub> (313,1 nm), Ca<sub>II</sub> (393,3 nm), Ca<sub>I</sub> (422,7 nm, 430,2 nm, 445,5 nm) e Na<sub>I</sub> (589 nm e 589,6 nm).

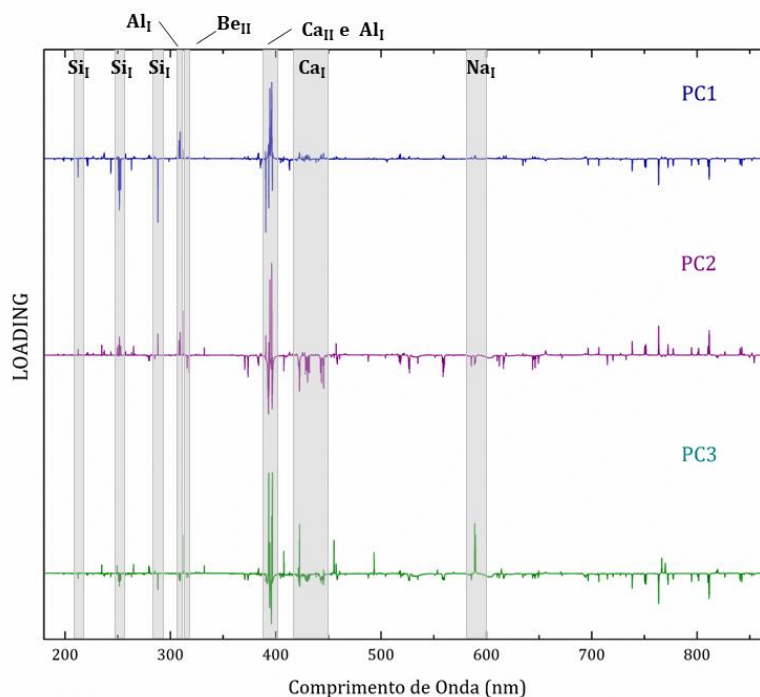


Figura 12 - Loadings da PCA dos espectros SNV-LIBS das amostras minerais, para PC1 (azul), PC2 (roxo) e PC3 (verde).

#### 4.1.1. Seleção de Regiões Espectrais LIBS para Separação de Classes

De acordo com a figura 11, não houve separação/diferenciação das amostras pertencentes as cinco diferentes classes minerais. Assim, visando obter uma separação visual das amostras pertencentes a uma classe mineral específica *versus* o restante das classes minerais, para PCA 3D (PC1 x PC2 x PC3), se foram selecionados e combinados diferentes regiões espectrais LIBS, que fossem capazes de obter tal diferenciação visual das classes. As análises foram realizadas de forma binária, ou seja, uma classe *versus* outra classe, devido as quantidades de amostras em cada classe mineral serem assimétricas.

### *Classe Silicato versus Restante das Classes*

Do *range* espectral dos espectros LIBS, de 180 a 961 nm, as regiões espectrais que, combinadas, forneceram uma melhor separação das amostras pertencentes a classe mineral Silicato *versus* as amostras pertencentes ao restante das classes minerais (Fosfato, Carbonato, Fluorita e Oxido), foram as seguintes, como mostra a tabela 4:

Tabela 4 - Regiões espectrais LIBS selecionadas, para separar amostras da classe mineral Silicato *versus* as amostras dos restantes das classes minerais. O comp. de onda inicial e o Comp. de onda final indicam, respectivamente, o comprimento de onda inicial e final, onde o pico central do elemento químico foi selecionado. O elemento químico correspondente a cada pico central é mostrado na última coluna da tabela.

<b>COMP. DE ONDA INICIAL (nm)</b>	<b>COMP. DE ONDA FINAL (nm)</b>	<b>PICO CENTRAL (nm)</b>	<b>ELEMENTO</b>
212	212,6	212,4	Si I
220,6	222	220,8; 221,1; 221,7	Si I
243,3	243,7	243,5	Si I
250,5	253	250,7; 251,4; 251,6; 251,9; 252,4; 252,9	Si I
262,9	263,4	263,1	Al II ou Si I
287,8	288,5	288,1	Si I
385	387	385,6; 386,2	Si II
390,4	390,8	390,5	Si I
412,4	413,6	413,1	Fe I

Aplicando o procedimento matemático de análise das componentes principais (PCA), no conjunto de dados formado somente pelas regiões selecionadas da tabela 4, se obteve a dispersão dos *scores* PCA em PC1 x PC2 x PC3, como mostra a figura 13.



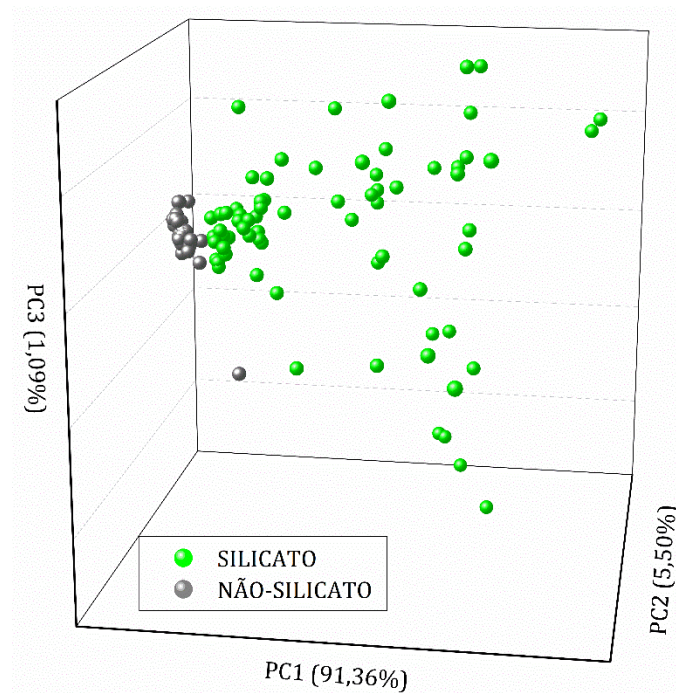


Figura 13 - Dispersão PCA 3D (PC1 x PC2 x PC3), após a seleção das regiões espectrais LIBS da Tabela 4. Em verde, são os *scores* das amostras pertencentes a classe mineral Silicato. Em cinza, são os *scores* das amostras pertencentes as classes minerais Fosfato, Carbonato, Fluorita e Oxido.

A análise da figura 13 indica que, com a seleção das regiões espectrais da tabela 4, as amostras pertencentes a classe mineral Silicato foram agrupadas juntas, com uma boa separação do agrupamento das amostras pertencentes a classe mineral não-silicato (Oxido, Carbonato, Fluorita e Fosfato). Os principais elementos químicos responsáveis pela diferenciação das classes minerais silicato e não-silicato foram o silício (Si), alumínio (Al) e ferro (Fe), de acordo com a tabela 4. A análise dos *loadings* da PCA, após a seleção das regiões LIBS da tabela 4, como mostra a figura 14, indicam que elemento silício (Si) foi o grande responsável pela variância das amostras da figura 13. De acordo com a composição elementar das amostras minerais, todas as amostras do grupo mineral silicato possuem silício em sua composição, o que não acontece com as amostras do grupo não-silicato. Isso justifica não haver grande dispersão das amostras não-silicato na figura 13 (*scores* em cinza), já que o elemento químico principal responsável pela dispersão das amostras, silício (Si), é ausente nessa classe.

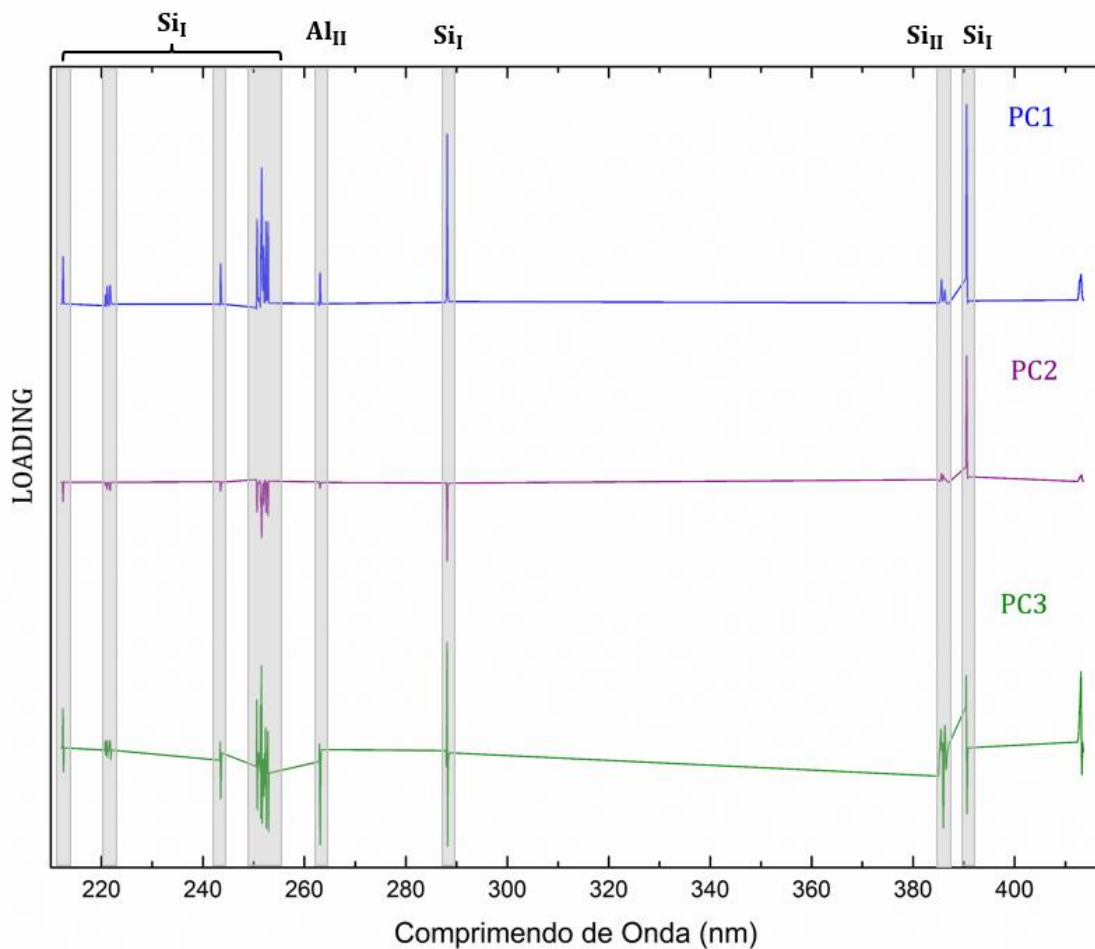


Figura 14 – Loadings das componentes principais 1 (em azul), 2 (em roxo) e 3 (em verde), para PCA das regiões LIBS selecionadas da tabela 4. Diferenciação das amostras silicato das amostras não-silicato. Em destaque, principais elementos químicos responsáveis pela variância em cada componente principal.

#### *Classe Oxido versus Restante das Classes*

Do *range* espectral dos espectros LIBS, de 180 a 961 nm, as regiões espectrais que, combinadas, forneceram uma melhor separação das amostras pertencentes a classe mineral Oxido *versus* as amostras pertencentes ao restante das classes minerais (Fosfato, Carbonato, Fluorita e Silicato), foram as seguintes, como mostra a tabela 5:

Tabela 5 - Regiões espectrais LIBS selecionadas, para separar amostras da classe mineral Oxido versus as amostras dos restantes das classes minerais. O comp. de onda inicial e o Comp. de onda final indicam, respectivamente, o comprimento de onda inicial e final, onde o pico central onde o elemento químico foi selecionado.

COMP. DE ONDA INICIAL (nm)	COMP. DE ONDA FINAL (nm)	PICO CENTRAL (nm)	ELEMENTO
225,9	227,6	226,9	Fe II
236,6	237,7	236,7; 237,3	Al II
265,9	266,3	266	Fe II
305,5	305,8	305,7	Fe I

Aplicando o procedimento matemático de análise das componentes principais (PCA), no conjunto de dados formado somente pelas regiões selecionadas da tabela 5, se obteve a dispersão dos *scores* da PCA em PC1 x PC2 x PC3, como mostra a figura 15.

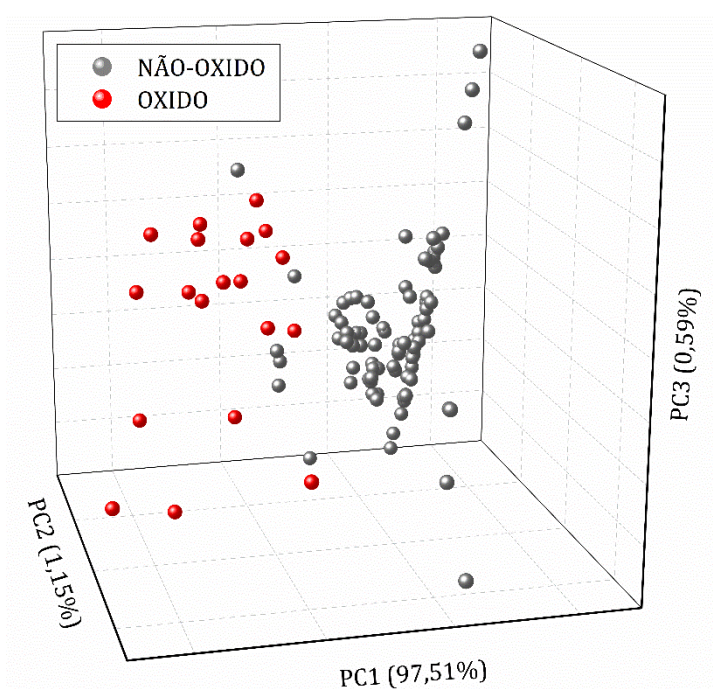


Figura 15 - Dispersão PCA 3D (PC1 x PC2 x PC3), após a seleção das regiões espectrais LIBS da Tabela 5. Em vermelho, são os *scores* das amostras pertencentes a classe mineral Oxido. Em cinza, são os *scores* das amostras pertencentes as classes minerais Fosfato, Carbonato, Fluorita e Silicato.

A análise da figura 15 indica que, mesmo com a seleção das regiões espectrais da tabela 5, as amostras pertencentes a classe mineral Oxido não conseguiram uma boa separação das amostras pertencentes a classe mineral não-óxido (Silicato, Carbonato, Fluorita e Fosfato), onde há algumas sobreposições de dispersão entre algumas amostras. Porém, a figura 15 mostra que há uma tendência de separação entre essas classes, e agrupamento das suas respectivas amostras.

Os principais elementos químicos responsáveis pela diferenciação das classes minerais óxido e não-óxido foram o alumínio (Al) e ferro (Fe), de acordo com a tabela 5. A análise dos *loadings* da PCA, após a seleção das regiões LIBS da tabela 5, como mostra a figura 16, indicam que os elementos alumínio (Al) e ferro (Fe) foram os grandes responsáveis pela variância das amostras da figura 15. De acordo com a composição elementar das amostras minerais, as amostras do grupo mineral óxido são constituídas, principalmente, por ferro e alumínio. Porém, amostras do grupo não-óxido, como as do grupo fosfato e silicato, também possuem a presença de um ou dos dois elementos químicos, alumínio ou ferro. Isso justifica não haver uma completa separação das amostras entre as classes óxido e não-óxido, já que existe alguma semelhança em suas composições elementares.

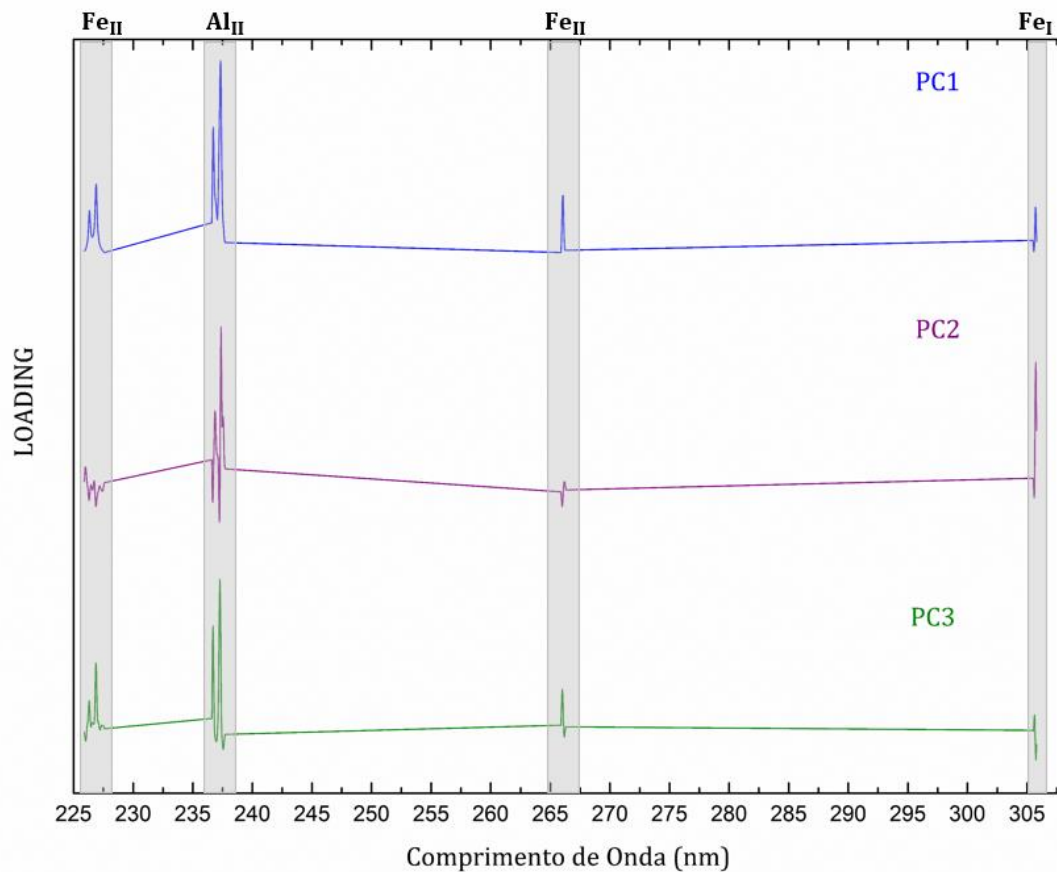


Figura 16 - Loadings das componentes principais 1 (em azul), 2 (em roxo) e 3 (em verde), para PCA das regiões LIBS selecionadas da tabela 5. Diferenciação das amostras oxido das amostras não-oxido. Em destaque, principais elementos químicos responsáveis pela variância em cada componente principal.

#### *Classe Fosfato versus Restante das Classes*

Do *range* espectral dos espectros LIBS, de 180 a 961 nm, as regiões espectrais que, combinadas, forneceram uma melhor separação das amostras pertencentes a classe mineral Fosfato *versus* as amostras pertencentes ao restante das classes minerais (Oxido, Carbonato, Fluorita e Silicato), foram as seguintes, como mostra a tabela 6:

Tabela 6 - Regiões espectrais LIBS selecionadas, para separar amostras da classe mineral Fosfato versus as amostras dos restantes das classes minerais. O comp. de onda inicial e o Comp. de onda final indicam, respectivamente, o comprimento de onda inicial e final, onde o pico central do elemento químico foi selecionado.

COMP. DE ONDA INICIAL (nm)	COMP. DE ONDA FINAL (nm)	PICO CENTRAL (nm)	ELEMENTO
213,4	213,8	213,6	P I
214,7	215,1	214,9	P I
253,3	253,7	253,5	P I ou Fe I ou Fe II
257,3	258	257,5	Al I ou Fe II
374,4	375,1	374,5; 374,9	Fe I ou Ca I
381,4	381,7	381,6	Fe I
381,8	382,2	382	Fe I
385,8	386,2	385,9	Fe I
414,2	414,5	414,3	Fe I
424,5	424,8	424,7	Fe I ou Cu I
425,9	426,2	426	Fe I
484	484,5	484,3	*

Aplicando o procedimento matemático de análise das componentes principais (PCA), no conjunto de dados formado somente pelas regiões selecionadas da tabela 6, se obteve a dispersão dos *scores* da PCA em PC1 x PC2 x PC3, como mostra a figura 17.

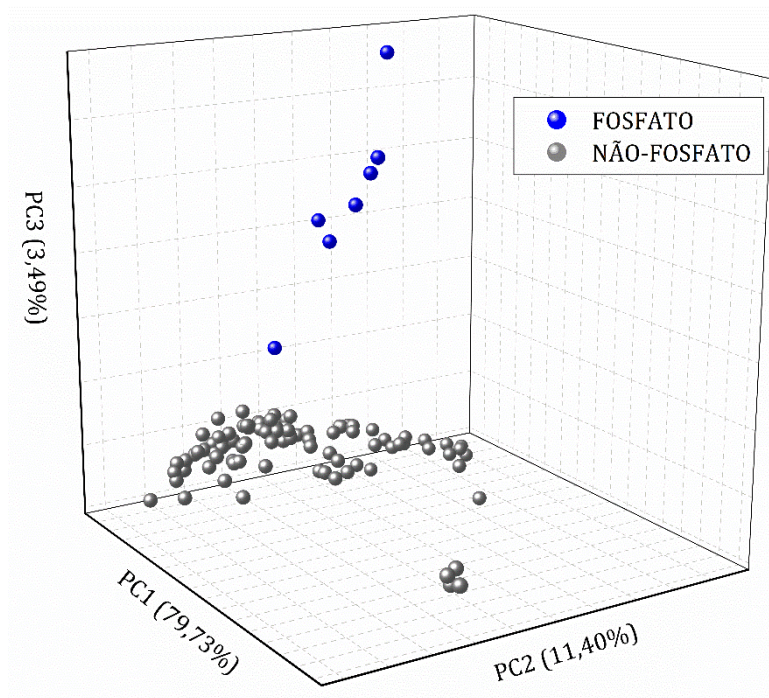


Figura 17 - Dispersão PCA 3D (PC1 x PC2 x PC3), após a seleção das regiões espectrais da Tabela 6. Em azul, são os *scores* das amostras pertencentes a classe mineral Fosfato. Em cinza, são os *scores* das amostras pertencentes as classes minerais Oxido, Carbonato, Fluorita e Silicato.

A análise da figura 17 indica que, com a seleção das regiões espectrais da tabela 6, as amostras pertencentes a classe mineral Fosfato foram agrupadas, com uma boa separação do agrupamento das amostras pertencentes as classes minerais não-fosfato (Oxido, Carbonato, Fluorita e Silicato). Assim, a seleção de regiões espectrais se mostrou factível para a separação e agrupamento dessas classes minerais.

Os principais elementos químicos responsáveis pela diferenciação das classes minerais fosfato e não-fosfato foram o alumínio (Al), ferro (Fe), cálcio (Ca) e fósforo (P) de acordo com a tabela 6. A análise dos *loadings* da PCA, após a seleção das regiões LIBS da tabela 6, como mostra a figura 18, indicam que o elemento ferro (Fe) foi o principal responsável pela variância dos dados nas PC1 e PC2, e que o elemento fósforo (P) foi o principal responsável pela variância na PC3. De acordo com a figura 17, as amostras do grupo mineral fosfato foram dispersadas das amostras não-fosfato, principalmente na direção da PC3, o que justifica a sua separação, já que o principal constituinte químico das amostras fosfato é o elemento fósforo (P).

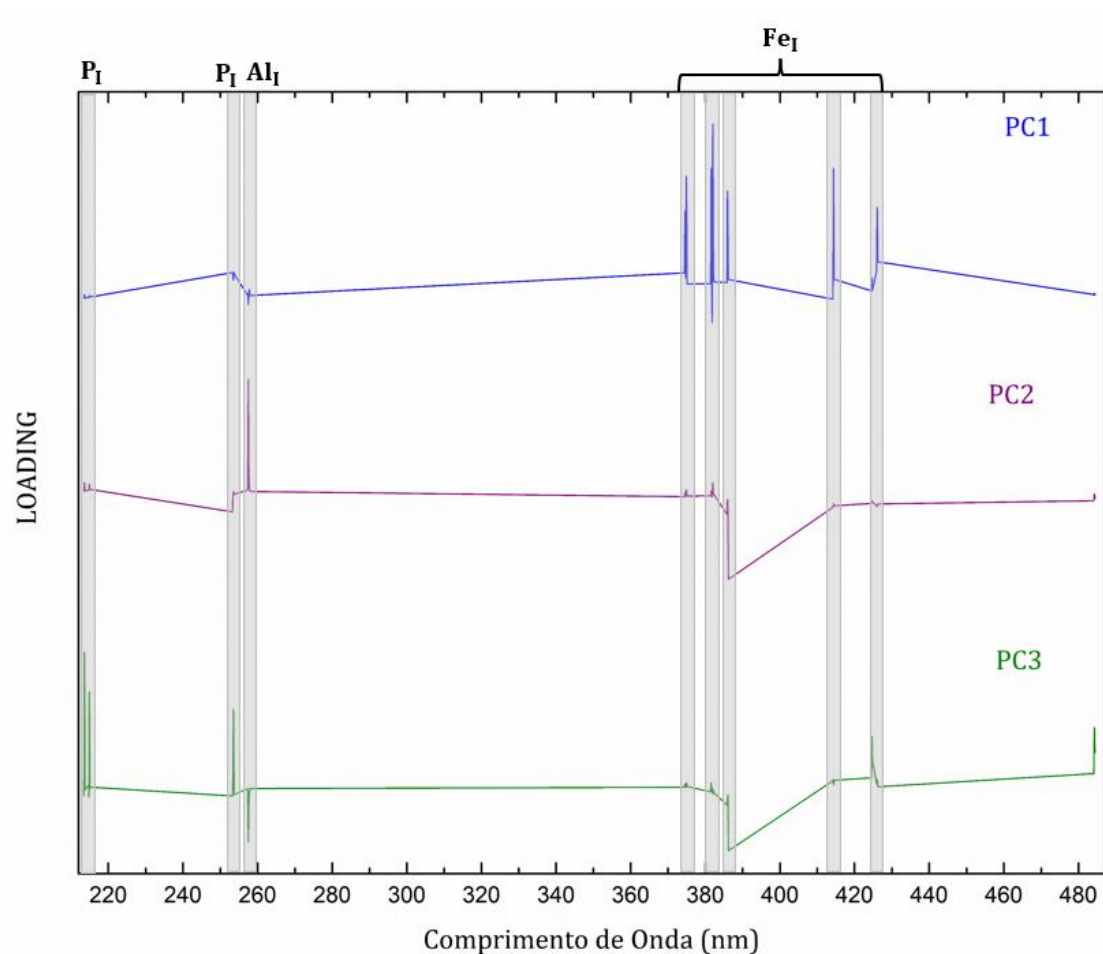


Figura 18 - Loadings das componentes principais 1 (em azul), 2 (em roxo) e 3 (em verde), para PCA das regiões LIBS selecionadas da tabela 6. Diferenciação das amostras fosfato das amostras não-fosfato. Em destaque, principais elementos químicos responsáveis pela variância em cada componente principal.

#### *Classe Carbonato versus Restante das Classes*

Do *range* espectral dos espectros LIBS, de 180 a 961 nm, as regiões espectrais que, combinadas, forneceram uma melhor separação das amostras pertencentes a classe Carbonato *versus* as amostras pertencentes ao restante das classes minerais (Óxido, Fosfato, Fluorita e Silicato), foram as seguintes, como mostra a tabela 7:



Tabela 7 - Regiões espectrais LIBS selecionadas, para separar amostras da classe mineral Carbonato versus as amostras dos restantes das classes minerais. O comp. de onda inicial e o Comp. de onda final indicam, respectivamente, o comprimento de onda inicial e final, onde o pico central do elemento químico foi selecionado.

COMP. DE ONDA INICIAL (nm)	COMP. DE ONDA FINAL (nm)	PICO CENTRAL (nm)	ELEMENTO
452,5	453	452,7	Ca I ou Fe I
516,5	517	516,7	Fe I ou Mg I
517,1	517,5	517,3	Mg I
518,1	518,6	518,4	Mg I
558	558,4	558,2	Ca I ou P II

Aplicando o procedimento matemático de análise das componentes principais (PCA), no conjunto de dados formado somente pelas regiões selecionadas da tabela 7, se obteve a dispersão dos *scores* da PCA em PC1 x PC2 x PC3, como mostra a figura 19.

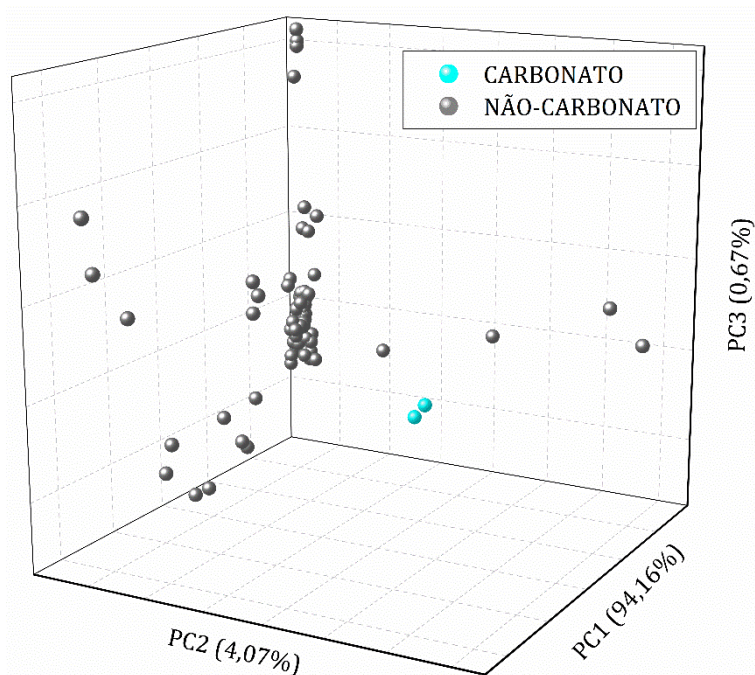


Figura 19 - Dispersão PCA 3D (PC1 x PC2 x PC3), após a seleção das regiões espectrais da Tabela 7. Em ciano, são os *scores* das amostras pertencentes a classe mineral Carbonato. Em cinza, são os *scores* das amostras pertencentes as classes minerais Oxido, Fosfato, Fluorita e Silicato.

A análise da figura 19 indica que, com a seleção das regiões espectrais da tabela 7, as amostras pertencentes a classe mineral Carbonato foram agrupadas, com uma boa separação da dispersão das amostras pertencentes a classe mineral não-carbonato (Óxido, Silicato, Fluorita e Fosfato).

Os principais elementos químicos responsáveis pela diferenciação das classes minerais carbonato e não-carbonato foram o magnésio (Mg), ferro (Fe), cálcio (Ca) e fósforo (P) de acordo com a tabela 7. A análise dos *loadings* da PCA, após a seleção das regiões LIBS da tabela 7, como mostra a figura 20, indicam que os elementos magnésio (Mg) e cálcio (Ca) foram os principais responsáveis pela variância dos dados nas PC1, PC2 e PC3. Houve um melhor agrupamento das amostras pertencentes a classe carbonato entre PC1 e PC2, e dispersão das amostras não-carbonato na direção da PC3. Como o principal constituinte das amostras minerais carbonato é o cálcio (Ca), e como há também a presença de cálcio (Ca), magnésio (Mg) e ferro (Fe) nas amostras pertencentes a classe não-carbonato, o agrupamento das amostras não-carbonato não ocorreu de maneira eficiente, havendo grande dispersão das amostras nas três componentes principais.

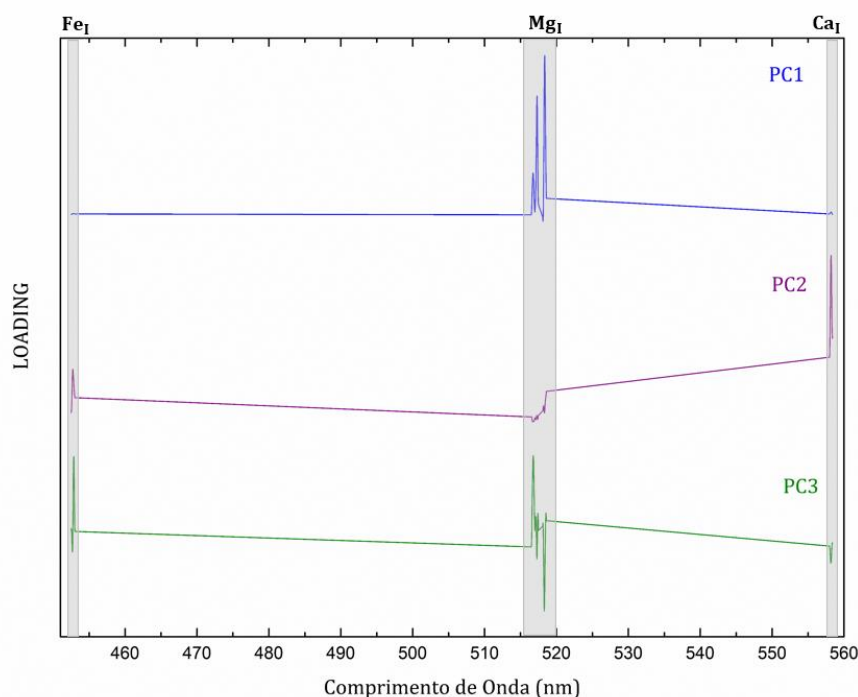


Figura 20 - Loadings das componentes principais 1 (em azul), 2 (em roxo) e 3 (em verde), para PCA das regiões LIBS selecionadas da tabela 7. Diferenciação das

amostras carbonato das amostras não-carbonato. Em destaque, principais elementos químicos responsáveis pela variância em cada componente principal.

*Classe Fluorita versus Restante das Classes*

Do *range* espectral dos espectros LIBS, de 180 a 961 nm, as regiões espectrais que, combinadas, forneceram uma melhor separação das amostras pertencentes a classe mineral Fluorita *versus* as amostras pertencentes ao restante das classes minerais (Óxido, Fosfato, Carbonato e Silicato), foram as seguintes, como mostra a tabela 8:

Tabela 8 - Regiões espectrais LIBS selecionadas para separar amostras da classe mineral Fluorita versus as amostras dos restantes das classes minerais. O comp. de onda inicial e o Comp. de onda final indicam, respectivamente, o comprimento de onda inicial e final, onde o pico central do elemento químico foi selecionado.

COMP. DE ONDA INICIAL (nm)	COMP. DE ONDA FINAL (nm)	PICO CENTRAL (nm)	ELEMENTO
580,5	583,8	583,1	*
645,4	645,8	645,5	Ca I ou Fe I
849,4	850,2	849,7	K I

Aplicando o procedimento matemático de análise das componentes principais (PCA), no conjunto de dados formado somente pelas regiões selecionadas da tabela 8, se obteve a dispersão dos *scores* da PCA em PC1 x PC2 x PC3, como mostra a figura 21.

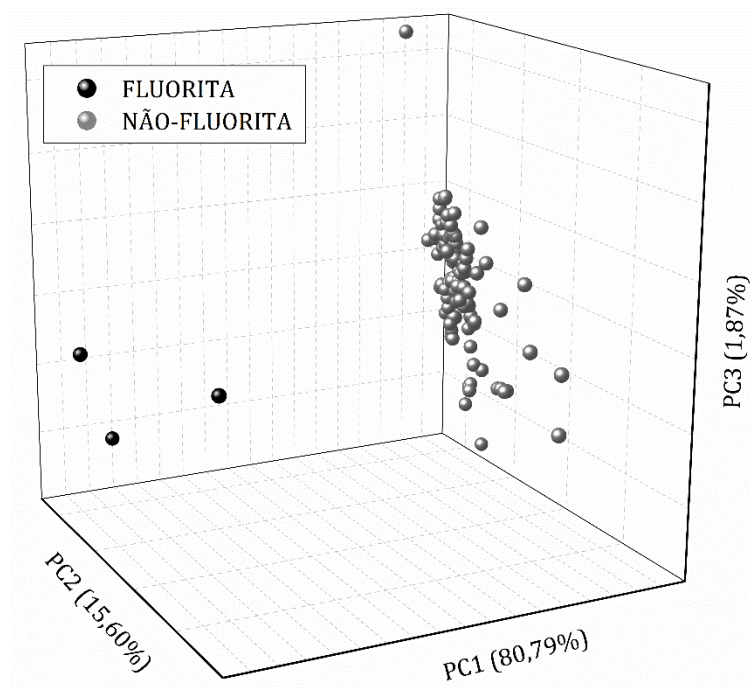


Figura 21 - Dispersão PCA 3D (PC1 x PC2 x PC3), após a seleção das regiões espectrais da Tabela 8. Em preto, são os *scores* das amostras pertencentes a classe mineral Fluorita. Em cinza, são os *scores* das amostras pertencentes as classes minerais Oxido, Fosfato, Carbonato e Silicato.

A análise da figura 21 indica que, com a seleção das regiões espectrais da tabela 8, as amostras pertencentes a classe mineral Fluorita foram agrupadas, com uma boa separação do agrupamento das amostras pertencentes as classes minerais não-fluorita (Oxido, Carbonato, Silicato e Fosfato). Assim, a seleção de regiões espectrais se mostrou ser necessária para a separação e agrupamento dessas classes minerais.

Os principais elementos químicos responsáveis pela diferenciação das classes minerais fluorita e não-fluorita foram o ferro (Fe), cálcio (Ca) e potássio (K) de acordo com a tabela 8. A análise dos *loadings* da PCA, após a seleção das regiões LIBS da tabela 8, como mostra a figura 22, indicam que os elementos cálcio (Ca) e potássio (K) foram os principais responsáveis pela variância dos dados nas PC1, PC2 e PC3. Houve um melhor agrupamento das amostras pertencentes a classe fluorita entre PC1 e PC2, e dispersão das amostras não-fluorita na direção da PC3.

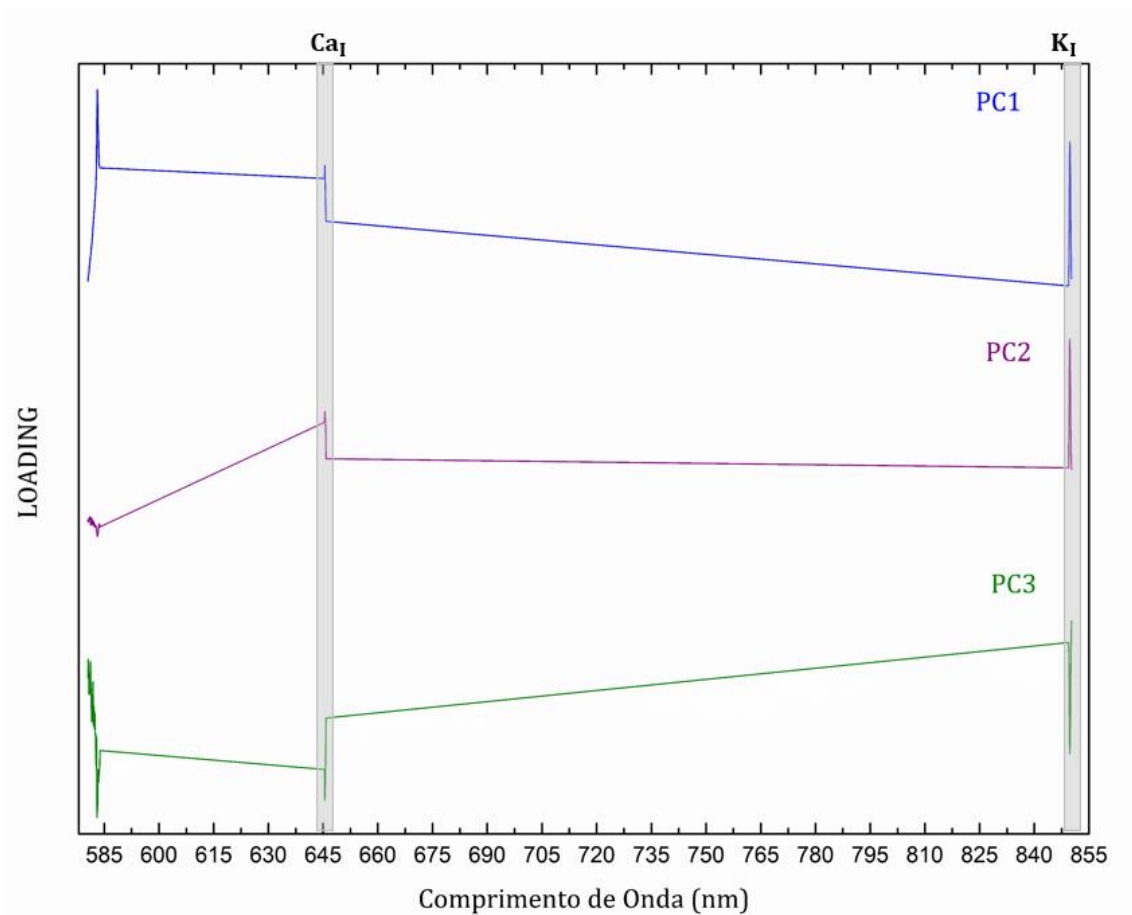


Figura 22 - Loadings das componentes principais 1 (em azul), 2 (em roxo) e 3 (em verde), para PCA das regiões LIBS selecionadas da tabela 8. Diferenciação das amostras fluorita das amostras não-fluorita. Em destaque, principais elementos químicos responsáveis pela variância em cada componente principal.

#### 4.1.2. Aplicação de Algoritmos Supervisionados de Aprendizado de Máquinas para Diferenciação e Classificação das Amostras Minerai

A seleção de regiões espectrais LIBS, onde os resultados foram mostrados na seção anterior, é um procedimento manual elaborado e complexo, onde para se chegar na melhor combinação de regiões espectrais que sejam capazes de separar as amostras nas classes desejadas, se exige muito tempo de análise. Comumente, em análise de espectroscopia óptica combinada com algoritmos de aprendizado de máquinas para separação e classificação de amostras, se é feito o treinamento dos classificadores (isto é, dos algoritmos de aprendizado de máquinas), para verificar qual é o classificador mais adequado para cada tipo de amostras, utilizando todas as regiões espectrais LIBS.

Assim, para verificar o potencial de diferenciação das amostras dentre as cinco classes de minerais, algoritmos de aprendizado de máquinas (*Machine Learning*) foram aplicados aos espectros SNV-LIBS, do *range* espectral de 180 a 961 nm. O potencial de diferenciação das amostras foi verificado para a separação de uma classe mineral específica *versus* o restante das classes minerais. Os algoritmos supervisionados de aprendizado de máquinas, também chamados de classificadores, k-vizinhos próximos (*k-nearest neighbour* - KNN), máquinas de vetores de suporte (*Support Vector Machines* - SVM) e análise discriminante (*Discriminant Analysis* - D.A.), foram treinados e validados internamente, tendo como dados de entrada os *scores* da PCA das amostras. O método de validação interna foi o *Leave-One-Out Cross-Validation* (LOO-CV) (LARIOS et al., 2021). Durante o treinamento e validação interna dos classificadores, o número de componentes principais (PC`s – do acrônimo em inglês *Principal Components*) adequado para cada classificador, se foi obtido através de um processo de otimização. Nesse processo de otimização, cada modelo é treinado e validado, inicialmente somente com a primeira componente principal, PC1, e a cada iteração, a próxima PC é adicionada ao conjunto de variáveis, até ser treinado e validado com todas as PC`s.

#### *Classe Mineral Silicato versus Restante das Classes Minerais*

Para a diferenciação da classe mineral Silicato *versus* não-silicato (fosfato, óxido, carbonato e fluorita), os classificadores que obtiveram as melhores acurácias foram a máquina de vetores de suporte (*Cubic SVM*) e o k-vizinhos próximos (*Weighted KNN*), com acurácia de 99,04%, utilizando 10 PC`s como variáveis. Através da matriz de confusão da figura 23, a Acurácia, a Sensitividade e a Especificidade dos dois classificadores podem ser avaliadas, onde a sensibilidade é a taxa de amostras verdadeiras positivas (não-silicato) que foram corretamente preditos pelo classificador, e a especificidade é a taxa de amostras verdadeiros negativos (silicato) que foram corretamente preditos pelo classificador.



Figura 23 - Matriz de confusão para classe mineral silicato versus não-silicato, mostrando os valores da acurácia de classificação, sensibilidade e especificidade. Matriz de confusão para os classificadores *Weighted KNN* e *Cubic SVM*.

A análise da matriz de confusão da figura 23, que é tanto para o classificador *Weighted KNN* quanto para o classificador *Cubic SVM*, indica que todas as 73 amostras da classe mineral Silicato foram corretamente classificadas (preditadas), onde a especificidade foi de 100%. As amostras não-silicato tiveram apenas 1 das 31 amostras incorretamente classificada, onde a sensibilidade foi de 96,77%.

#### *Classe Oxido versus Restante das Classes*

Para a diferenciação da classe mineral Oxido *versus* não-oxido (silicato, carbonato, fluorita e fosfato), os classificadores que obtiveram as melhores acurácias foram o k-vizinhos próximos (*Cosine KNN* e *Fine KNN*), com acurácia de 98,08%, utilizando 25 PC's como variáveis. Através da matriz de confusão da figura 24, a Acurácia, a Sensitividade e a Especificidade dos dois classificadores podem ser avaliadas, onde a sensibilidade é a taxa de amostras verdadeiras positivas (não-oxido) que foram corretamente preditos pelo classificador, e a especificidade é a taxa de amostras verdadeiros negativos (oxido) que foram corretamente preditos pelo classificador.

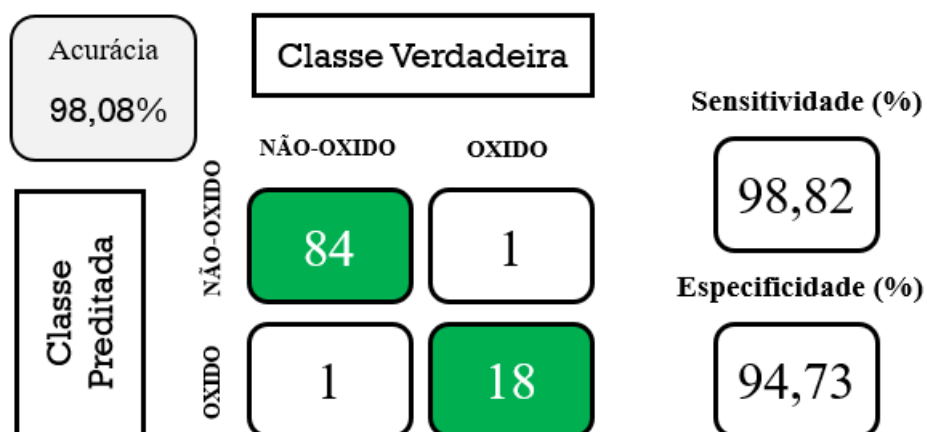


Figura 24 - Matriz de confusão para classe mineral oxido versus não-oxido, mostrando os valores da acurácia de classificação, sensibilidade e especificidade. Matriz de confusão para os classificadores *Fine KNN* e *Cosine KNN*.

A análise da matriz de confusão da figura 24, que é tanto para o classificador *Fine KNN* quanto para o classificador *Cosine KNN*, indica que 18 das 19 amostras da classe mineral Oxido foram corretamente classificadas (preditadas), onde a especificidade foi de 94,73%. As amostras não-oxido tiveram apenas 1 das 85 amostras incorretamente classificada, onde a sensibilidade foi de 98,82%.

#### *Classe Fosfato versus Restante das Classes*

Para a diferenciação da classe mineral Fosfato *versus* não-fosfato (silicato, oxido, carbonato e fluorita), os classificadores que obtiveram as melhores acurácias foram o k-vizinhos próximos (*Fine KNN* e *Weighted KNN*) e máquina de vetores de suporte (*Cubic SVM*, *Linear SVM* e *Quadratic SVM*), com acurácia de 99,04%, utilizando 15 PC`s como variáveis. Através da matriz de confusão da figura 25, a Acurácia, a Sensitividade e a Especificidade dos classificadores podem ser avaliadas, onde a sensibilidade é a taxa de amostras verdadeiras positivas (não-fosfato) que foram corretamente preditos pelo classificador, e a especificidade é a taxa de amostras verdadeiros negativos (fosfato) que foram corretamente preditos pelo classificador.





Figura 25 - Matriz de confusão para classe mineral fosfato versus não-fosfato, mostrando os valores da acurácia de classificação, sensibilidade e especificidade. Matriz de confusão para os classificadores *Fine KNN*, *Weighted KNN*, *Cubic SVM*, *Linear SVM* e *Quadratic SVM*.

A análise da matriz de confusão da figura 25, que é tanto para os classificadores *KNN* (*Fine* e *Weighted*), quanto para os classificadores *SVM* (*Cubic*, *Linear* e *Quadratic*), indica que 6 das 7 amostras da classe mineral Fosfato foram corretamente classificadas (preditadas), onde a especificidade foi de 85,71%. As 97 amostras não-fosfato foram corretamente classificadas, onde a sensibilidade foi de 100%.

#### *Classe Carbonato versus Restante das Classes*

Para a diferenciação da classe mineral Carbonato *versus* não-carbonato (silicato, oxido, fosfato e fluorita), os classificadores que obtiveram as melhores acurácias foram o k-vizinhos próximos (*Fine KNN* e *Weighted KNN*) e máquina de vetores de suporte (*Cubic SVM*, *Linear SVM* e *Quadratic SVM*), com acurácia de 100%, utilizando 2 (duas) PC's como variáveis. Através da matriz de confusão da figura 26, a Acurácia, a Sensitividade e a Especificidade dos classificadores podem ser avaliadas, onde a sensibilidade é a taxa de amostras verdadeiras positivas (carbonato) que foram corretamente preditos pelo classificador, e a especificidade é a taxa de amostras verdadeiros negativos (não-carbonato) que foram corretamente preditos pelo classificador.

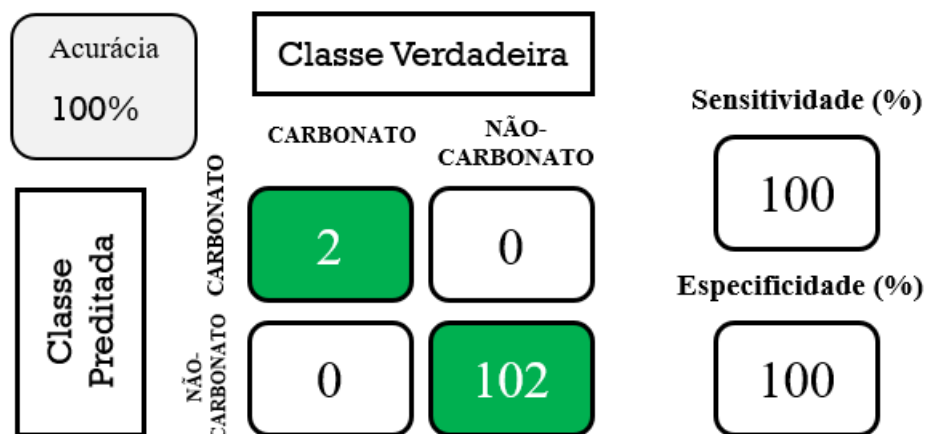


Figura 26 - Matriz de confusão para classe mineral carbonato versus não-carbonato, mostrando os valores da acurácia de classificação, sensibilidade e especificidade. Matriz de confusão para os classificadores *Fine KNN*, *Weighted KNN*, *Cubic SVM*, *Linear SVM* e *Quadratic SVM*.

A análise da matriz de confusão da figura 26, tanto para os classificadores *KNN* (*Fine* e *Weighted*), quanto para os classificadores *SVM* (*Cubic*, *Linear* e *Quadratic*), indica que as 2 (duas) amostras da classe mineral Carbonato foram corretamente classificadas (preditadas), e todas as 102 amostras não-carbonato também foram corretamente classificadas, onde a sensibilidade e a especificidade foram de 100%.

#### *Classe Fluorita versus Restante das Classes*

Para a diferenciação da classe mineral Fluorita *versus* não-fluorita (silicato, carbonato, oxido e fosfato), os classificadores que obtiveram as melhores acurácias foram o k-vizinhos próximos (*Fine KNN* e *Weighted KNN*) e máquina de vetores de suporte (*Cubic SVM*, *Fine Gaussian SVM* e *Medium Gaussian SVM*), com acurácia de 100%, utilizando 3 (três) PC's como variáveis. Através da matriz de confusão da figura 27, a Acurácia, a Sensitividade e a Especificidade dos classificadores podem ser avaliadas, onde a sensibilidade é a taxa de amostras verdadeiras positivas (fluorita) que foram corretamente preditos pelo classificador, e a especificidade é a taxa de amostras verdadeiros negativos (não-fluorita) que foram corretamente preditos pelo classificador.

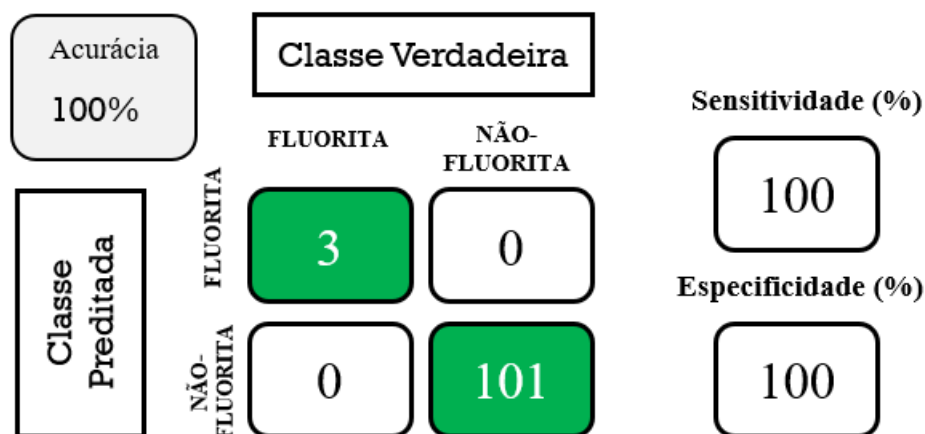


Figura 27 - Matriz de confusão para classe mineral fluorita versus não-fluorita, mostrando os valores da acurácia de classificação, sensibilidade e especificidade. Matriz de confusão para os classificadores *KNN* (*Fine* e *Weighted*) e *SVM* (*Cubic*, *Fine Gaussian* e *Medium Gaussian*).

A análise da matriz de confusão da figura 27, tanto para os classificadores *KNN* (*Fine* e *Weighted*), quanto para os classificadores *SVM* (*Cubic*, *Fine Gaussian* e *Medium Gaussian*), indica que as 3 (três) amostras da classe mineral Fluorita foram corretamente classificadas (preditadas), e todas as 101 amostras não-fluorita também foram corretamente classificadas, onde a sensibilidade e a especificidade foram de 100%.

#### 4.1.3. Algoritmo Seletor de Componentes Principais

##### *Classe Silicato versus não-silicato*

Analisando a matriz de confusão da figura 23, a acurácia dos classificadores *Cubic SVM* e *Weighted KNN* foi de 99,04%, onde a especificidade, que corresponde a taxa das amostras da classe Silicato que foram corretamente preditados pelos classificadores como sendo pertencentes a classe Silicato, foi de 100%. Já a sensibilidade, que corresponde a taxa das amostras não-silicato que foram corretamente preditados pelos classificadores, foi de 96,77%, indicando que apenas uma amostra foi incorretamente classificada. Visando melhorar a acurácia e a sensibilidade, se foi

aplicado um algoritmo seletor de componentes principais, que selecionou diferentes combinações de componentes principais, dentre as 10 PC's originais, que foram utilizadas como variáveis. Para o *Cubic SVM*, o algoritmo seletor de componentes principais obteve uma acurácia de 100%, ao se utilizar as 7 seguintes componentes principais como variáveis: PC1, PC2, PC4, PC5, PC7, PC9 e PC10. A figura 28 mostra a matriz de confusão, após a aplicação do algoritmo seletor de componentes principais, para esse classificador. Para o classificador *Weighted KNN*, o algoritmo seletor de componentes principais manteve a mesma acurácia de 99,04%, ao se utilizar as 9 seguintes componentes principais como variáveis dos classificadores: PC1, PC2, PC4, PC5, PC6, PC7, PC8, PC9 e PC10. A figura 29 mostra a matriz de confusão, após a aplicação do algoritmo seletor de componentes principais, para esse classificador.

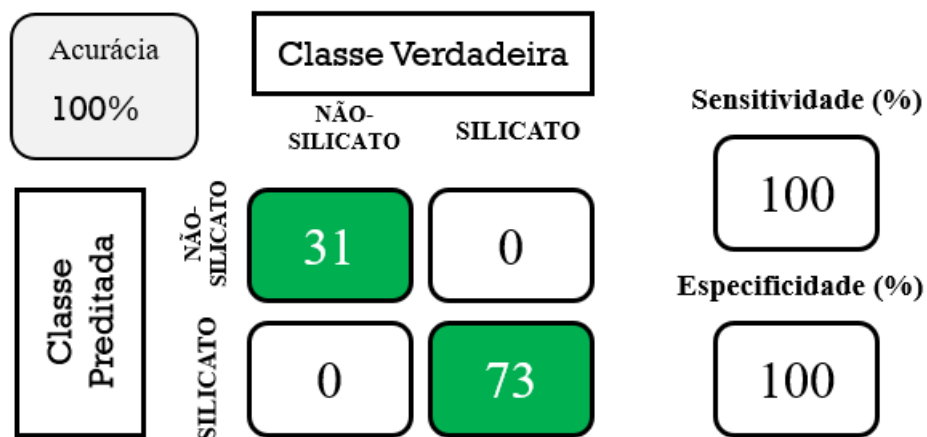


Figura 28 - Matriz de confusão para o classificador Cubic SVM, após a seleção de componentes principais. Classe mineral Silicato versus não-silicato. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

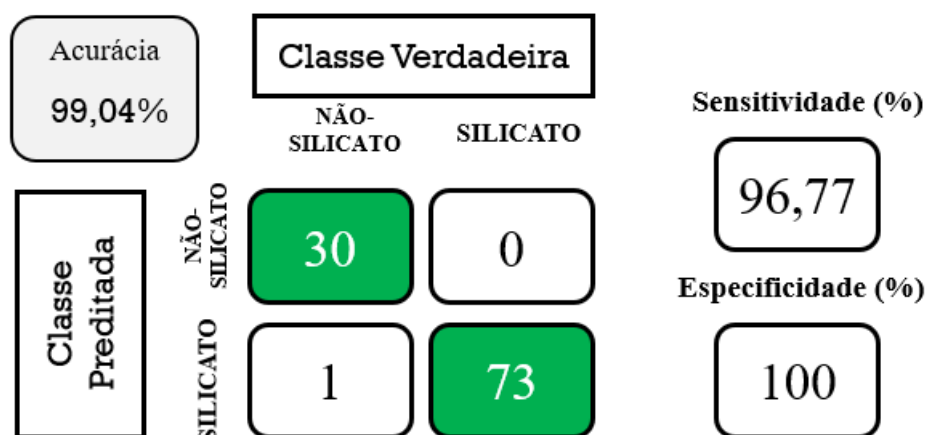


Figura 29 - Matriz de confusão para o classificador Weighted KNN, após a seleção de componentes principais. Classe mineral Silicato versus não-silicato. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

A análise da matriz de confusão da figura 28 indica que o classificador *Cubic SVM* apresentou melhora no seu poder de diferenciação e classificação das amostras pertencentes a classe não-silicato, onde a sensibilidade foi 96,77% para 100%. A especificidade se manteve a mesma de 100%, indicando que a seleção de componentes principais foi essencial para a melhora da acurácia do classificador *Cubic SVM*, além da utilização de uma menor quantidade de variáveis (PC`s) pelo classificador.

Já a análise da matriz de confusão da figura 29 indica que, mesmo após a seleção de componentes principais, o classificador *Weighted KNN* não apresentou melhora no seu poder de diferenciação entre as classes minerais silicato e não-silicato, indicando apenas que a mesma acurácia, sensibilidade e especificidade podem ser obtidas ao se utilizar uma quantidade menor de PC`s como variáveis, saindo de 10 PC`s para 9 PC`s.

#### *Classe Oxido versus não-oxido*

Analisando a matriz de confusão da figura 24, a acurácia dos classificadores *Cosine KNN* e *Fine KNN* foi de 98,08%, onde a especificidade, que corresponde a taxa das amostras da classe Oxido que foram corretamente preditos pelos classificadores como sendo pertencentes a classe Oxido, foi de 94,73%, indicando que apenas uma amostra foi incorretamente classificada. Já a sensibilidade, que corresponde a taxa das

amostras não-óxido que foram corretamente preditos pelos classificadores, foi de 98,82%, indicando que apenas uma amostra foi incorretamente classificada. Visando melhorar a especificidade e a sensibilidade, se foi aplicado um algoritmo seletor de componentes principais, que selecionou diferentes combinações de componentes principais, dentre as 25 PC's originais, que foram utilizadas como variáveis. Para o *Cosine KNN*, o algoritmo seletor de componentes principais obteve uma acurácia de 99,04%, ao se utilizar as 18 seguintes componentes principais como variáveis dos classificadores: PC1, PC3, PC4, PC8, PC9, PC10, PC12, PC13, PC14, PC15, PC18, PC19, PC20, PC21, PC22, PC23, PC24 e PC25. A figura 30 mostra a matriz de confusão, após a aplicação do algoritmo seletor de componentes principais, para esse classificador. Para o *Fine KNN*, o algoritmo seletor de componentes principais obteve uma acurácia de 99,04%, ao se utilizar as 16 seguintes componentes principais como variáveis dos classificadores: PC1, PC4, PC5, PC6, PC8, PC10, PC11, PC13, PC15, PC17, PC18, PC21, PC22, PC23, PC24 e PC25. A figura 31 mostra a matriz de confusão, após a aplicação do algoritmo seletor de componentes principais, para esse classificador.

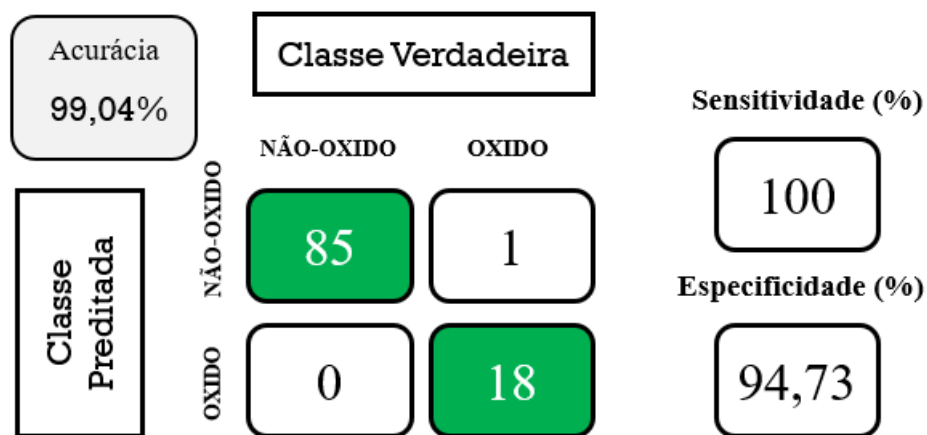


Figura 30 - Matriz de confusão para o classificador Cosine KNN, após a seleção de componentes principais. Classe mineral Oxido versus não-óxido. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

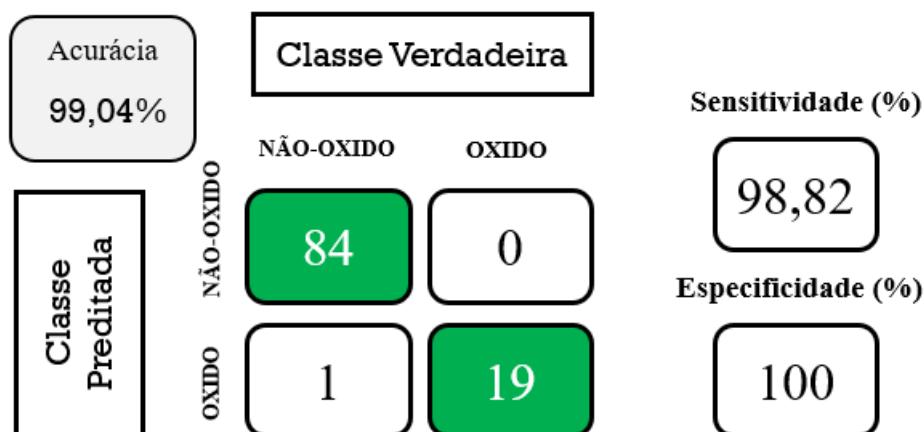


Figura 31 - Matriz de confusão para o classificador Fine KNN, após a seleção de componentes principais. Classe mineral Oxido versus não-oxido. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

A análise da matriz de confusão da figura 30 indica que o classificador *Cosine KNN* apresentou melhora no seu poder de diferenciação e classificação das amostras pertencentes a classe não-oxido, onde a sensibilidade foi 98,82% para 100%. A especificidade se manteve a mesma de 94,73%, indicando que a seleção de componentes principais foi essencial para a melhora da classificação das amostras não-oxido, e por consequência, da acurácia final do classificador, além da utilização de uma menor quantidade de variáveis (PC`s) pelo classificador.

Já a análise da matriz de confusão da figura 31 indica que o classificador *Fine KNN* não apresentou melhora no seu poder de diferenciação das amostras pertencentes a classe não-silicato, onde a sensibilidade se manteve a mesma de 98,82%. Já a especificidade apresentou uma melhora, atingindo o valor de 100%. Isso indica que a seleção de componentes principais foi essencial para a melhora da classificação das amostras oxido, e também da acurácia final do classificador, além da utilização de uma menor quantidade de variáveis (PC`s) pelo classificador.

#### *Classe Fosfato versus não-fosfato*

Analisando a matriz de confusão da figura 25, a acurácia dos classificadores *SVM (Cubic, Linear e Quadratic)* e *KNN (Fine e Weighted)* foi de 99,04%, onde a

especificidade, que corresponde a taxa das amostras da classe Fosfato que foram corretamente preditos pelos classificadores como sendo pertencentes a classe Fosfato, foi de 85,71%, indicando que apenas uma amostra foi incorretamente classificada. Já a sensibilidade, que corresponde a taxa das amostras não-fosfato que foram corretamente preditos pelos classificadores, foi de 100%. Visando melhorar a acurácia e a especificidade, se foi aplicado um algoritmo seletor de componentes principais, que seleciona diferentes combinações de componentes principais, dentre as 15 PC`s originais que foram utilizadas como variáveis. Para o *Cubic SVM*, o algoritmo seletor de componentes principais manteve a acurácia de 99,04%, ao se utilizar as 4 seguintes componentes principais como variáveis dos classificadores: PC6, PC9, PC12, PC14. Para o *Fine KNN*, o algoritmo seletor de componentes principais manteve a acurácia de 99,04%, ao se utilizar as 4 seguintes componentes principais como variáveis dos classificadores: PC9, PC12, PC14, PC15. Para o *Linear SVM*, o algoritmo seletor de componentes principais manteve a acurácia de 99,04%, ao se utilizar as 4 seguintes componentes principais como variáveis dos classificadores: PC6, PC9, PC14, PC15. Para o *Quadratic SVM*, o algoritmo seletor de componentes principais manteve a acurácia de 99,04%, ao se utilizar as 3 seguintes componentes principais como variáveis dos classificadores: PC9, PC10, PC15. Para o *Weighted KNN*, o algoritmo seletor de componentes principais manteve a acurácia de 99,04%, ao se utilizar as 2 seguintes componentes principais como variáveis dos classificadores: PC9, PC15. A figura 32 mostra a matriz de confusão, para todos esses classificadores *SVM* e *KNN*, após a aplicação do algoritmo seletor de componentes principais.





Figura 32 - Matriz de confusão para os classificadores *KNN* e *SVM*, após a seleção de componentes principais. Classe mineral Fosfato versus não-fosfato. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

A análise da matriz de confusão da figura 32 indica que os classificadores *SVM* (*Cubic*, *Linear* e *Quadratic*) e *KNN* (*Fine* e *Weighted*) não apresentaram melhora no seu poder de diferenciação e classificação das amostras pertencentes a classe fosfato, onde a especificidade se manteve em 85,71%. A sensibilidade se manteve a mesma de 100%, indicando que a seleção de componentes principais foi importante apenas na quantidade menor de variáveis (PC's) que os classificadores utilizaram para obter o mesmo valor de acurácia.

*Classe Carbonato versus não-carbonato e Classe Fluorita versus não-fluorita*

O algoritmo seletor de componentes principais não foi aplicado nos classificadores para a diferenciação das classes Carbonato e Fluorita *versus* o restante das classes, já que a acurácia de 100% foi obtida ao se fazer o processo de otimização de componentes principais, como mostram as figuras 26 e 27.

## 4.2. Classe Silicato *versus* não-silicato: Treinamento e Validação Externa

Analisando a tabela 3, se verifica que a classe mineral Silicato é a que possui a maior quantidade de amostras no conjunto de dados LIBS. Como os minerais silicato compõem grande parte dos minerais presentes na crosta terrestre, se de um foco maior no estudo dessa classe. Dessa maneira, se verificou o potencial de classificação/diferenciação das amostras silicato, através do treinamento dos algoritmos supervisionados de aprendizado de máquinas, e verificação do seu poder de classificação, utilizando amostras de validação externa.

A validação externa é um procedimento utilizado a fim de verificar o comportamento dos classificadores, que foram treinados e validados internamente, em prever as classes minerais das amostras que não fizeram parte do conjunto de treinamento e da validação interna do classificador. A validação externa simula e verifica o potencial de aplicação real do uso de espectroscopia óptica LIBS combinado com algoritmos de aprendizado de máquinas para diferenciar amostras minerais, com base em suas informações espectrais.

Para as análises de validação externa, dos 104 espectros SNV-LIBS das amostras minerais, 20% dos espectros da classe mineral Silicato e da classe não-silicato foram retirados dos conjuntos de dados LIBS, e utilizados para validação externa. Esse conjunto de amostras foi nomeado de amostras de validação externa. Assim, as amostras de validação externa foram divididas das seguintes maneiras:

- 20% de 73 amostras Silicato = 15 amostras Silicato para validação externa
- 20% de 31 amostras não-silicato = 6 amostras não-silicato para validação externa

Para o treinamento e validação interna dos classificadores, as amostras que não fizeram parte da validação externa foram nomeadas de amostras de treinamento, e foram divididas das seguintes maneiras:

- 58 amostras Silicato
- 25 amostras não-silicato

#### 4.2.1. Treinamento dos Algoritmos Supervisionados de Aprendizado de Máquinas

Para verificar o potencial de diferenciação das amostras entre a classe mineral Silicato e o restante das classes minerais (não-silicato), algoritmos supervisionados de aprendizado de máquinas (*Machine Learning*) foram inicialmente treinados com os espectros SNV-LIBS, do *range* de 180 a 961 nm, para as amostras de treinamento. Os algoritmos supervisionados de aprendizado de máquinas k-vizinhos próximos (*k-nearest neighbour* - KNN), máquinas de vetores de suporte (*Support Vector Machines* - SVM) e análise discriminante (*Discriminant Analysis* - D.A.) foram treinados e validados internamente, tendo como dados de entrada os *scores* da PCA das amostras de treinamento. O método de validação interna foi o *Leave-One-Out Cross-Validation* (LOO-CV) (LARIOS et al., 2021). Durante o treinamento e validação interna dos classificadores, o número de componentes principais (PC`s – do acrônimo em inglês *Principal Components*) adequado para cada classificador se foi obtido através de um processo de otimização. Nesse processo de otimização, cada modelo é treinado e validado, inicialmente somente com a primeira componente principal, PC1, e a cada iteração, a próxima PC é adicionada ao conjunto de variáveis, até ser treinado e validado com todas as PC`s. A figura 33 mostra a dispersão PCA 3D, para as amostras de treinamento das classes minerais silicato e não-silicato.

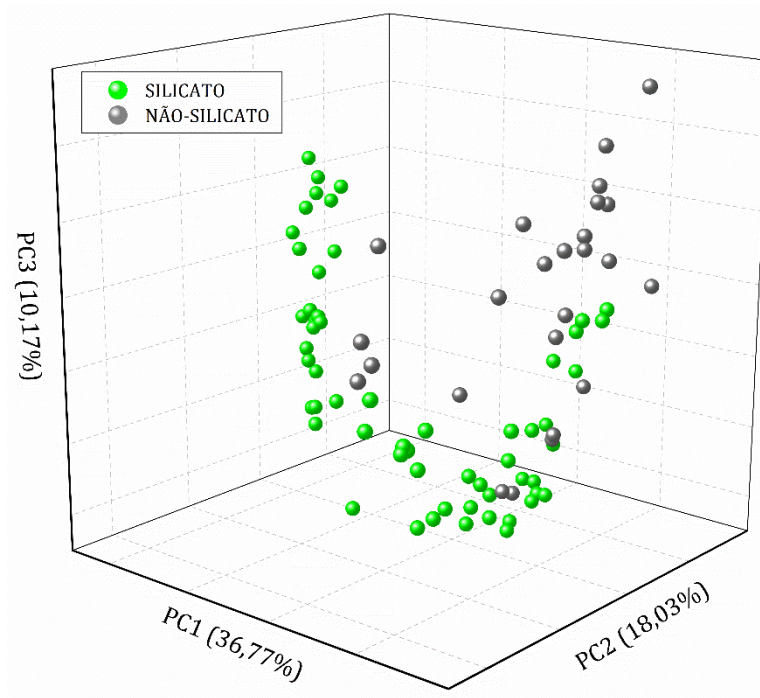


Figura 33 - Dispersão PCA 3D (PC1 x PC2 x PC3), para as amostras de treinamento. Em verde, são os *scores* das amostras pertencentes a classe mineral Silicato. Em cinza, são os *scores* das amostras pertencentes a classe não-silicato.

Para as amostras de treinamento, o classificador que obteve a melhor acurácia foi o k-vizinhos próximos (*Fine KNN*), com acurácia de 98,80%, utilizando 10 PC`s como variáveis. A figura 34 mostra a matriz de confusão para esse classificador, com a Acurácia, a Sensitividade e a Especificidade. A sensibilidade, que é a taxa de amostras verdadeiras positivas (não-silicato) que foram corretamente preditos pelo classificador, foi de 96%. Isso indica que apenas 1 (uma) das 25 amostras não-silicato foi incorretamente predita pelo classificador. A especificidade, que é a taxa de amostras verdadeiros negativos (silicato) que foram corretamente preditos pelo classificador, foi de 100%, indicando que todas as amostras de treinamento da classe silicato foram corretamente preditas pelo classificador.



Figura 34 - Matriz de confusão para o classificador *Fine KNN*, para as amostras de treinamento, classe mineral silicato versus classe mineral não-silicato. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

#### 4.2.2. Algoritmo Seletor de Componentes Principais – Amostras de Treinamento

Visando melhorar a sensibilidade e a acurácia do classificador *Fine KNN*, se foi aplicado o algoritmo seletor de componentes principais, selecionando componentes principais dentre as 10 PC`s originais, que foram utilizadas como variáveis. Para o *Fine KNN*, o algoritmo seletor de componentes principais manteve a acurácia de 98,80%, ao se utilizar as 7 seguintes componentes principais como variáveis dos classificadores: PC1, PC4, PC6, PC7, PC8, PC9 e PC10. A figura 35 mostra a matriz de confusão para o *Fine KNN*, após a aplicação do algoritmo seletor de componentes principais.

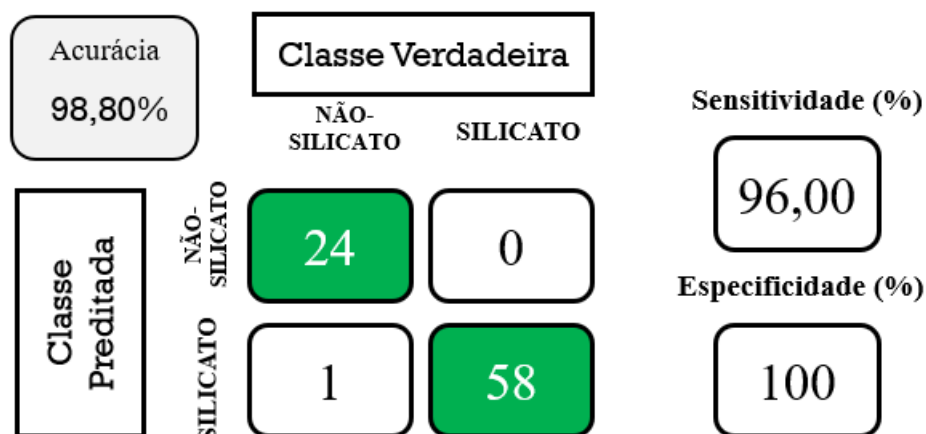


Figura 35 - Matriz de confusão para o classificador Fine KNN, após a seleção de componentes principais. Classe mineral Silicato versus não-silicato, para as amostras de treinamento. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

A análise da matriz de confusão da figura 35 indica que o classificador *Fine KNN* não apresentou melhora no seu poder de diferenciação e classificação das amostras pertencentes a classe não-silicato, onde a sensibilidade se manteve em 96,00%. A especificidade se manteve a mesma de 100%, indicando que a seleção de componentes principais foi importante apenas na quantidade menor de variáveis (PC's) que o classificador utiliza para diferenciar e classificar as amostras não-silicato das amostras silicato.

#### 4.2.3. Seleção de Regiões Espectrais LIBS - Amostras de Treinamento

As regiões espectrais LIBS da tabela 4 foram selecionadas e utilizadas, para verificar o potencial de diferenciação das amostras de treinamento entre a classe mineral Silicato da classe mineral não-silicato. Após a seleção das regiões da tabela 4, a análise das componentes principais (PCA) apresentou a seguinte dispersão mostrada na figura 36.

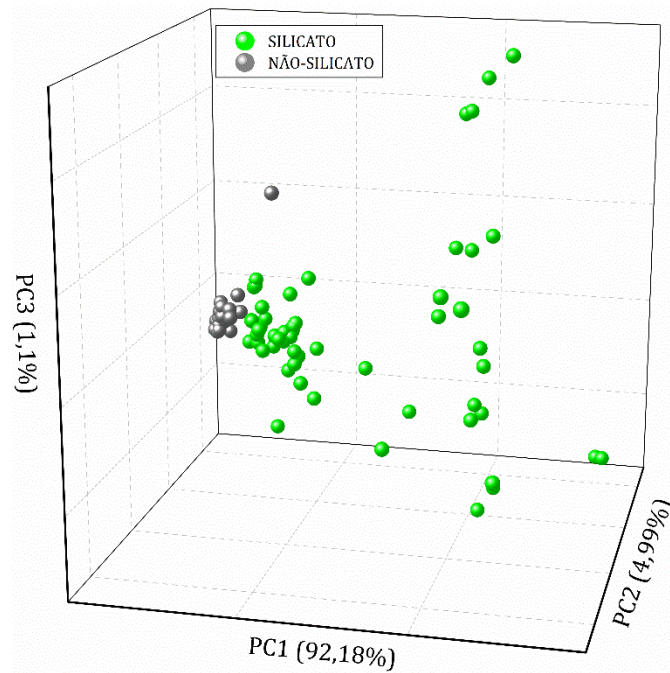


Figura 36 - Dispersão PCA 3D (PC1 x PC2 x PC3), para as amostras de treinamento, após a seleção de regiões espectrais LIBS da tabela 4. Em verde, são os *scores* das amostras pertencentes a classe mineral Silicato. Em cinza, são os *scores* das amostras pertencentes a classe mineral não-silicato.

Para as amostras de treinamento, após a seleção de regiões espectrais LIBS, os classificadores que obtiveram as melhores acurácias foram os k-vizinhos próximos (*Weighted KNN*) e máquinas de vetores de suporte (*Cubic, Linear e Medium Gaussian SVM*), com acurácia de 98,80%, com 3, 1, 2 e 1 componentes principais, respectivamente, como variáveis dos classificadores. A figura 37 mostra a matriz de confusão para esses classificadores, com a Acurácia, a Sensitividade e a Especificidade. A sensibilidade, que é a taxa de amostras verdadeiras positivas (não-silicato) que foram corretamente preditos pelo classificador, foi de 96%. A especificidade, que é a taxa de amostras verdadeiros negativos (silicato) que foram corretamente preditos pelo classificador, foi de 100%. Isso indica que a seleção de regiões espectrais LIBS não contribuiu para a melhora da acurácia, sensibilidade ou especificidade dos classificadores, para as amostras de treinamento, em comparação com a utilização do espectro completo, figura 34, e seleção de componentes principais, figura 35.

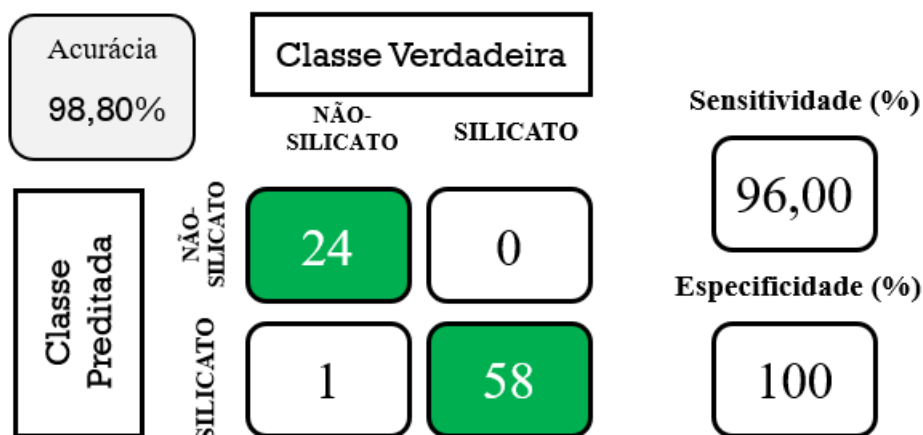


Figura 37 - Matriz de confusão para os classificadores KNN e SVM, para seleção de regiões espectrais LIBS da tabela 4. Classe mineral Silicato versus não-silicato, para as amostras de treinamento. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

#### 4.2.4. Validação Externa

As amostras do conjunto de validação externa foram utilizadas, a fim de se verificar o poder de diferenciação e classificação para os classificadores que foram treinados e validados com as amostras de treinamento. O conjunto de amostras de validação externa é constituído de 21 amostras, sendo 15 amostras da classe mineral Silicato e 6 amostras da classe não-silicato.

##### 4.2.4.1. Validação Externa – Melhor Classificador da Otimização de Componentes Principais

O classificador que obteve a melhor acurácia de treinamento durante o processo de otimização de PC`s foi o *Fine KNN*, como mostra a matriz de confusão da figura 34. Para esse classificador, 10 componentes principais (PC`s) foram utilizadas como variáveis. As amostras de validação externa analisadas e suas classes foram previstas por esse classificador, onde a figura 38 mostra a sua matriz de confusão.





Figura 38 - Matriz de confusão para o classificador *Fine KNN*, após o processo de otimização de PC's. Classe mineral Silicato versus não-silicato, para as amostras de validação externa. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

A análise da matriz de confusão da figura 38 mostra que todas as 6 amostras da classe não-silicato e todas as 15 amostras da classe silicato, para o conjunto de amostras de validação externa, foram corretamente classificadas/preditadas pelo classificador. Os valores de acurácia, sensibilidade e especificidade foram de 100%, respectivamente.

#### 4.2.4.2. Validação Externa – Algoritmo Seletor de Componentes Principais

As amostras de validação externa foram avaliadas pelo classificador *Fine KNN*, após a seleção de componentes principais. Para esse classificador, 7 (sete) componentes principais foram utilizadas como variáveis, onde a acurácia de classificação das amostras de validação externa é mostrada na matriz de confusão da figura 39.



Figura 39 - Matriz de confusão para o classificador *Fine KNN*, após a seleção de componentes principais. Classe mineral Silicato versus não-silicato, para as amostras de validação externa. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

A análise da matriz de confusão da figura 39 mostra que todas as 6 amostras da classe não-silicato e todas as 15 amostras da classe silicato, para o conjunto de amostras de validação externa, foram corretamente classificadas/preditadas pelo classificador. Os valores de acurácia, sensibilidade e especificidade foram de 100%, respectivamente.

#### 4.2.4.3. Validação Externa – Seleção de Regiões Espectrais LIBS

As amostras de validação externa foram avaliadas pelos classificadores *KNN* e *SVM*, que obtiveram as melhores acurácias após a seleção de regiões espectrais LIBS. De acordo com a matriz de confusão da figura 40, todas as amostras da classe não-silicato e todas as amostras da classe silicato foram corretamente classificadas, onde acurácia de 100% foi atingida, como também 100% de sensibilidade (classe não-silicato corretamente classificada como não-silicato) e 100% de especificidade (classe silicato corretamente classificada como silicato).

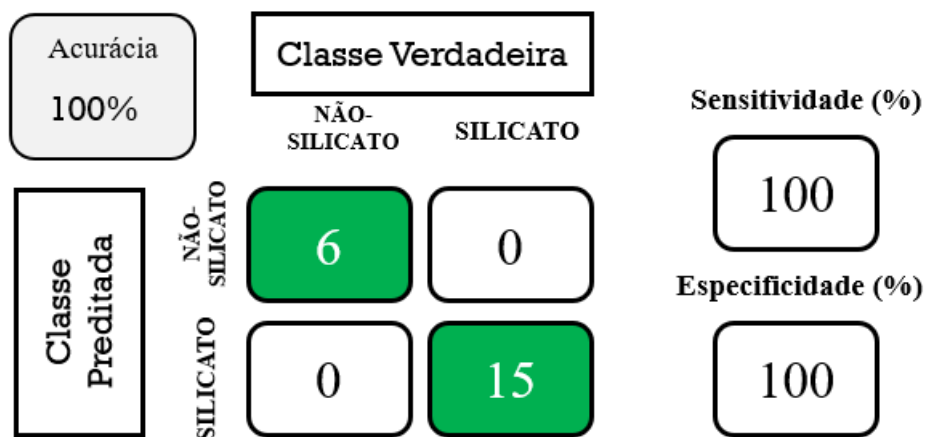


Figura 40 - Matriz de confusão para os classificadores *KNN* e *SVM*, após a seleção de regiões espectrais LIBS. Classe mineral Silicato versus não-silicato, para as amostras de validação externa. Os valores de acurácia, sensibilidade e especificidade são mostrados em porcentagem.

## 5. CONCLUSÃO

Neste trabalho, a aplicação de espectroscopia óptica combinada com métodos quimiométricos e algoritmos supervisionados de aprendizado de máquinas, foi avaliada para aplicação na diferenciação e classificação de amostras minerais. As amostras minerais avaliadas são pertencentes as classes minerais Silicato, Oxido, Fosfato, Carbonato e Fluorita. A técnica de espectroscopia de emissão óptica por plasma induzido por laser (LIBS) foi utilizada, fornecendo informações a respeito da composição elementar das amostras minerais, onde foram identificados como os principais constituintes, os elementos silício (Si), alumínio (Al), cálcio (Ca), magnésio (Mg), potássio (K), fósforo (P), oxigênio (O) e sódio (Na). Dois tipos diferentes de análises foram feitos e seus resultados avaliados. Para a análise 1, a seleção de regiões espectrais LIBS mostraram ser viáveis de serem aplicadas para separação visual das classes minerais, onde os gráficos de dispersão PCA 3D apresentaram boa tendência de separação para as amostras pertencentes as classes minerais silicato, carbonato, fosfato e fluorita, enquanto, por outro lado, houve uma sobreposição para algumas amostras pertencentes a classe mineral oxido, não obtendo boa separação dos restantes das classes minerais. Para o treinamento e validação interna dos algoritmos supervisionados de aprendizado de máquinas, acurácia de 100% foram obtidas para as classes carbonato e fluorita, e para as classes silicato, oxido e fosfato foram de, respectivamente, 99,04%, 98,08% e 99,04%. A seleção de componentes principais, através da aplicação do algoritmo seletor de componentes principais, diminuiu a quantidade de variáveis utilizada pelo classificador para a classificação da classe fosfato, onde sua acurácia de 99,04% se manteve, e apresentou melhora na acurácia de classificação, atingindo 100% para a classe silicato e 99,04% para a classe oxido. Da análise 2, as amostras da classe silicato foram separadas em amostras de treinamento e de validação externa. Para o treinamento e validação interna dos algoritmos supervisionados de aprendizado de máquinas, a melhor acurácia foi de 98,80%, para o classificador *Fine KNN*, utilizando 10 PC's como variáveis. A aplicação do algoritmo seletor de componentes principais não melhorou a acurácia final do classificador, mas diminuiu a quantidade de componentes principais que o classificador utilizou, caindo de 10 para 7 componentes principais. A seleção de regiões LIBS também influenciou somente na quantidade menor de componentes principais que os classificadores precisaram, mantendo a mesma

acurácia de 98,80%. Houve sucesso na diferenciação e classificação das amostras silicato de validação externa, atingindo acurácias de 100% pelos classificadores na otimização de componentes principais, na seleção de componentes principais e na seleção de regiões espectrais LIBS. Os resultados obtidos mostraram que a análise de diferenciação e classificação das amostras minerais foi um protocolo desafiador, pois além das amostras minerais apresentarem grande variabilidades internas em uma mesma classe, o que é característico desse tipo de amostra, as quantidades de amostras em cada classe mineral eram assimétricas, dificultando ainda mais a verificação da tendência de agrupamento ou separação entre elas. Com as três abordagens distintas feitas, otimização de componentes principais, seleção de componentes principais e seleção de regiões LIBS, e principalmente com os resultados da validação externa, se mostrou que há grande potencial de aplicação a ser desenvolvido e aplicado para a diferenciação e classificação de amostras pertencentes as outras classes minerais.

## 6. PERSPECTIVAS FUTURAS

Com base nos resultados obtidos, foi desenvolvido uma abordagem para a construção de uma máquina computacional que fosse capaz de identificar e classificar amostras minerais, com base em suas composições elementares. Essa abordagem seria um protocolo de etapas que devem ser realizadas, para a correta identificação da classe mineral das amostras em análise. As amostras seriam classificadas dentre as cinco classes minerais que o algoritmo foi treinado: silicato, oxido, fosfato, carbonato e fluorita. O fluxograma da figura 41 mostra como se dará o procedimento para a classificação da amostra externa:

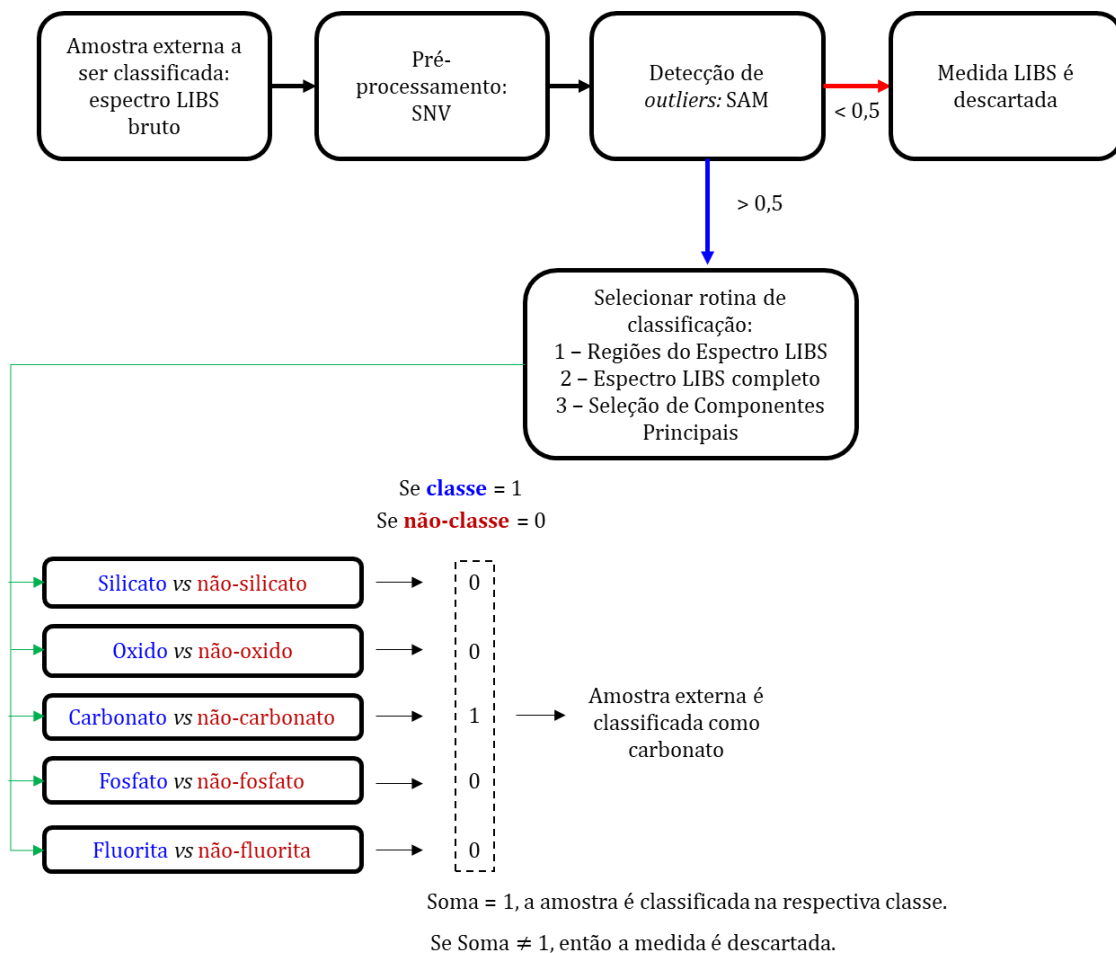


Figura 41 - Fluxograma mostrando como seriam as etapas para a classificação de uma amostra mineral externa dentre as classes silicato, oxido, carbonato, fosfato ou fluorita, na qual os algoritmos supervisionados de máquinas foram treinados.

O desenvolvimento dessas ferramentas que combinam espectroscopia óptica, com análises quimiométricas e algoritmos supervisionados de aprendizado de máquinas, permitiram a publicação de vários trabalhos em que participei como coautor, como *Metronidazole-loaded gold nanoparticles in natural rubber latex as potential wound dressing* (BORGES et al., 2022), *Geochemical identification and classification of cherts using handheld laser induced breakdown spectroscopy (LIBS) supported by machine learning algorithms* (SENESI et al., 2023), *Differentiation of latex biomembrane with collagen and non-collagen using laser induced breakdown spectroscopy* (NICOLODELLI et al., 2022), *A new strategy for canine visceral leishmaniasis diagnosis based on FTIR spectroscopy and machine learning* (LARIOS et al., 2021), *Soybean seed vigor discrimination by using infrared spectroscopy and machine learning algorithms* (LARIOS et al., 2020a), *Laser-induced breakdown spectroscopy as a powerful tool for distinguishing high- and low-vigor soybean seeds lots* (LARIOS et al., 2020b). Como autor principal, foi publicado o trabalho *Evaluation of rice varieties using LIBS and FTIR techniques associated with PCA and machine learning algorithms* (RIBEIRO et al., 2020), e também um pedido de depósito de patente em que sou coautor, processo número BR 10 2020 016368 0 de titulação Método Para Determinação De Vigor De Sementes Utilizando Técnicas De Espectroscopia Óptica Associadas Com Análise Por Aprendizado De Máquina.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

AKBARZADEH, S. et al. Plant discrimination by Support Vector Machine classifier based on spectral reflectance. **Computers and Electronics in Agriculture**, v. 148, n. August 2017, p. 250–258, 2018.

AKHTAR, M. et al. Magnetic field enhanced detection of heavy metals in soil using laser induced breakdown spectroscopy. **Spectrochimica Acta - Part B Atomic Spectroscopy**, v. 148, p. 143–151, 1 out. 2018.

AKRITAS, A. G.; MALASCHONOK, G. I. Applications of singular-value decomposition (SVD). **Mathematics and Computers in Simulation**, v. 67, n. 1–2, p. 15–31, 2004.

BELLOU, E. et al. Laser-induced breakdown spectroscopy assisted by machine learning for olive oils classification: The effect of the experimental parameters. **Spectrochimica Acta - Part B Atomic Spectroscopy**, v. 163, 1 jan. 2020.

BEUREN, G. M.; ANZANELLO, M. J. Variable selection using statistical non-parametric tests for classifying production batches into multiple classes. **Chemometrics and Intelligent Laboratory Systems**, v. 193, n. April, p. 1–8, 2019.

BORGES, F. A. et al. Metronidazole-loaded gold nanoparticles in natural rubber latex as a potential wound dressing. **International Journal of Biological Macromolecules**, v. 211, p. 568–579, 2022.

BUSTILLO, M. A. et al. Is the macroscopic classification of flint useful? A petroarchaeological analysis and characterization of flint raw materials from the Iberian neolithic mine of Casa Montero. **Archaeometry**, v. 51, n. 2, p. 175–196, 2009.

COIMBRA, R. et al. **Carbonate-hosted clay minerals: A critical re-evaluation of extraction methods and their possible bias on palaeoenvironmental information.** **Earth-Science Reviews** Elsevier B.V., , 1 mar. 2021.

DUFRESNE, W. J. B.; RUFLEDT, C. J.; MARSHALL, C. P. Raman spectroscopy of the eight natural carbonate minerals of calcite structure. **Journal of Raman Spectroscopy**, v. 49, n. 12, p. 1999–2007, 1 dez. 2018.

EL HADDAD, J. et al. Multiphase mineral identification and quantification by laser-induced breakdown spectroscopy. **Minerals Engineering**, v. 134, p. 281–290, 1 abr. 2019.

FENG, Q. et al. **Flotation of copper oxide minerals: A review.** **International Journal of Mining Science and Technology** China University of Mining and Technology, , 1 nov. 2022.

FRITSCH, E.; KARAMELAS, S.; MEVELLEC, J. Y. Raman spectra of gem-quality variscite and metavariscite. **Journal of Raman Spectroscopy**, v. 48, n. 11, p. 1554–1558, 1 nov. 2017.



GARCIA PINTO, F.; ESCALFONI, R.; SAINT'PIERRE, T. D. **Sample Preparation for Determination of Rare Earth Elements in Geological Samples by ICP-MS: A Critical Review**. *Analytical Letters*, ago. 2012.

GÖNEN, M.; TANUGUR, A. G.; ALPAYDIN, E. Multiclass posterior probability support vector machines. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, v. 19, n. 1, p. 130–139, 2008.

GOTTFRIED, J. L. et al. Multivariate analysis of laser-induced breakdown spectroscopy chemical signatures for geomaterial classification. *Spectrochimica Acta - Part B Atomic Spectroscopy*, v. 64, n. 10, p. 1009–1019, out. 2009.

HANDELMAN, G. S. et al. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. *AMERICAN JOURNAL OF ROENTGENOLOGY*, v. 212, n. 1, p. 38–43, 2019.

JIA, Z. Controlling the Overfitting of Heritability in Genomic Selection through Cross Validation. *Scientific Reports*, v. 7, n. 1, p. 1–9, 2017.

KAHLENBERG, V.; KAINDL, R.; KONZETT, J.  $\eta$ -Y<sub>2</sub>Si<sub>2</sub>O<sub>7</sub> - Structural investigations on a quenchable high-pressure mixed anion silicate. *Solid State Sciences*, v. 9, n. 6, p. 542–550, jun. 2007.

KIM, Y.; FABRE, C.; CAUZID, J. Access to quantitative analysis of carbonates using a portable LIBS instrument: First applications to single minerals and mineral mixtures. *Spectrochimica Acta - Part B Atomic Spectroscopy*, v. 191, 1 maio 2022.

KUSUMANINGRUM, D. et al. Non-destructive technique for determining the viability of soybean (*Glycine max*) seeds using FT-NIR spectroscopy. *Journal of the Science of Food and Agriculture*, v. 98, n. 5, p. 1734–1742, 2018.

LARIOS, G. et al. Soybean seed vigor discrimination by using infrared spectroscopy and machine learning algorithms. *Analytical Methods*, v. 12, n. 35, p. 4303–4309, 2020a.

LARIOS, G. et al. A new strategy for canine visceral leishmaniasis diagnosis based on FTIR spectroscopy and machine learning. *Journal of Biophotonics*, v. 14, n. 11, p. 1–9, 2021.

LARIOS, G. S. et al. Laser-Induced Breakdown Spectroscopy as a Powerful Tool for Distinguishing High- and Low-Vigor Soybean Seed Lots. *Food Analytical Methods*, v. 13, p. 1691–1698, 2020b.

LAWLEY, C. J. M.; SOMERS, A. M.; KJARSGAARD, B. A. Rapid geochemical imaging of rocks and minerals with handheld laser induced breakdown spectroscopy (LIBS). *Journal of Geochemical Exploration*, v. 222, 1 mar. 2021.

MARKIEWICZ-KESZYCKA, M. et al. **Laser-induced breakdown spectroscopy (LIBS) for food analysis: A review**. *Trends in Food Science and Technology* Elsevier Ltd, , 1 jul. 2017.

- MCMILLAN, N. J. et al. Laser-induced breakdown spectroscopy analysis of minerals: Carbonates and silicates. **Spectrochimica Acta - Part B Atomic Spectroscopy**, v. 62, n. 12, p. 1528–1536, dez. 2007.
- MOULIN, C. et al. Fisher Linear Discriminant Analysis for text-image combination in multimedia information retrieval. **Pattern Recognition**, v. 47, n. 1, p. 260–269, 2014.
- NAVAZO, M. et al. Raw stone material supply for Upper Pleistocene settlements in Sierra de Atapuerca (Burgos, Spain): flint characterization using petrographic and geochemical techniques. **Journal of Archaeological Science**, v. 35, n. 7, p. 1961–1973, 2008.
- NICOLODELLI, G. et al. **Recent advances and future trends in LIBS applications to agricultural materials and their food derivatives: An overview of developments in the last decade (2010–2019). Part I. Soils and fertilizers.** **TrAC - Trends in Analytical Chemistry** Elsevier B.V., , 1 jun. 2019.
- NICOLODELLI, G. et al. Differentiation of latex biomembrane with collagen and non-collagen using laser induced breakdown spectroscopy. **Materials Today Communications**, v. 30, p. 103099 (1–7), 2022.
- NIST - NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **National Institute of Standards and Technology - NIST.** Disponível em: <[https://physics.nist.gov/PhysRefData/ASD/lines\\_form.html](https://physics.nist.gov/PhysRefData/ASD/lines_form.html)>.
- PALÁSTI, D. J. et al. Qualitative Analysis of Glass Microfragments Using the Combination of Laser-Induced Breakdown Spectroscopy and Refractive Index Data. **Sensors**, v. 22, n. 8, 1 abr. 2022.
- PÉREZ-AGUILAR, C. D.; CUÉLLAR-CRUZ, M. **The formation of crystalline minerals and their role in the origin of life on Earth.** **Progress in Crystal Growth and Characterization of Materials** Elsevier Ltd, , 1 fev. 2022.
- PÉREZ-DIEZ, S. et al. Chemometrics and elemental mapping by portable LIBS to identify the impact of volcanogenic and non-volcanogenic degradation sources on the mural paintings of Pompeii. **Analytica Chimica Acta**, v. 1168, 11 jul. 2021.
- RANULFI, A. C. et al. Nutritional characterization of healthy and *Aphelenchoides besseyi* infected soybean leaves by laser-induced breakdown spectroscopy (LIBS). **Microchemical Journal**, v. 141, n. May, p. 118–126, 2018.
- RIBEIRO, M. C. S. et al. Evaluation of rice varieties using LIBS and FTIR techniques associated with PCA and machine learning algorithms. **Applied Optics**, v. 59, n. 32, p. 10043, 2020.
- ROLDÁN, C. et al. Identification of local and allochthonous flint artefacts from the Middle Palaeolithic site “Abrigo de la Quebrada” (Chelva, Valencia, Spain) by macroscopic and physicochemical methods. **X-Ray Spectrometry**, v. 44, n. 4, p. 209–216, 1 jul. 2015.
- RUI ALVES, M. Evaluation of the predictive power of biplot axes to automate the construction and layout of biplots based on the accuracy of direct readings from

common outputs of multivariate analyses: 1. application to principal component analysis. **Journal of Chemometrics**, v. 26, n. 5, p. 180–190, 2012.

SALT LAKE. **AN INTRODUCTION TO GEOLOGY: FREE TEXT BOOK**. Disponível em: </opengeology.org/textbook/3-minerals/>. Acesso em: 4 abr. 2023.

SAVORANI, F. et al. **Data handling in science and technology: Chemometrics in Food Chemistry**. [s.l.: s.n.].

SENESI, G. S. **Laser-Induced Breakdown Spectroscopy (LIBS) applied to terrestrial and extraterrestrial analogue geomaterials with emphasis to minerals and rocks**. **Earth-Science Reviews** Elsevier, , 1 dez. 2014.

SENESI, G. S. et al. Geochemical identification and classification of cherts using handheld laser induced breakdown spectroscopy (LIBS) supported by supervised machine learning algorithms. **Applied Geochemistry**, p. 105625, abr. 2023.

SEZER, B. et al. Coffee arabica adulteration: Detection of wheat, corn and chickpea. **Food Chemistry**, v. 264, p. 142–148, 30 out. 2018.

STEFAS, D. et al. Laser-induced breakdown spectroscopy assisted by machine learning for plastics/polymers identification. **Atoms**, v. 7, n. 3, p. 1–13, 1 set. 2019.

SWEETAPPLE, M. T.; TASSIOS, S. Laser-induced breakdown spectroscopy (LIBS) as a tool for in situ mapping and textural interpretation of lithium in pegmatite minerals. **American Mineralogist**, v. 100, n. 10, p. 2141–2151, 1 out. 2015.

THARWAT, A. Classification assessment methods. **Applied Computing and Informatics**, 2018.

WANG, F. et al. Kinetic evaluation of mineral carbonation of natural silicate samples. **Chemical Engineering Journal**, v. 404, 15 jan. 2021.

WANG, R. et al. Slow-release of fluorite and its effect on flotation separation of magnesite from calcite. **Minerals Engineering**, v. 185, 1 jul. 2022.

WIENS, R. C. et al. **Joint analyses by laser-induced breakdown spectroscopy (LIBS) and Raman spectroscopy at stand-off distances**. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*. **Anais...ago**. 2005.

XIONG, Y. et al. A spectra partition algorithm based on spectral clustering for interval variable selection. **Infrared Physics and Technology**, v. 105, n. February, 2020.

YANG, P. et al. Laser-induced breakdown spectroscopy assisted chemometric methods for rice geographic origin classification. **Applied Optics**, v. 57, n. 28, p. 8297–8302, 2018.

YANG, Z. et al. Fast determination of oxides content in cement raw meal using NIR-spectroscopy and backward interval PLS with genetic algorithm. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 223, p. 117327, 2019.

YANG, Z. et al. Contradictory tendency of As(V) releasing from Fe–As complexes: Influence of organic and inorganic anions. **Chemosphere**, v. 286, 1 jan. 2022.

ZHANG, L.; ZHOU, W.; JIAO, L. Wavelet Support Vector Machine. **IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics**, v. 34, n. 1, p. 34–39, 2004.

ZHOU, F. et al. Surface Electrical Behaviors of Apatite, Dolomite, Quartz, and Phosphate Ore. **Frontiers in Materials**, v. 7, 18 fev. 2020.

ZHU, Y. et al. Effect of phosphate on the adsorption of antibiotics onto iron oxide minerals: Comparison between tetracycline and ciprofloxacin. **Ecotoxicology and Environmental Safety**, v. 205, 1 dez. 2020.