

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS  
ESCOLA POLITÉCNICA E DE ARTES  
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



**CLASSIFICAÇÃO DE VINHOS USANDO IMAGENS HIPERESPECTRAIS E  
K-VIZINHOS MAIS PRÓXIMOS**

FERNANDO COSTA OLIVEIRA

GOIÂNIA,

2024

FERNANDO COSTA OLIVEIRA

**CLASSIFICAÇÃO DE VINHOS USANDO IMAGENS HIPERESPECTRAIS E  
K-VIZINHOS MAIS PRÓXIMOS**

Trabalho de Conclusão de Curso apresentado à Escola Politécnica e de Artes, da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação.

Orientador(a):

Prof. Dr. Clarimar José  
Coelho

GOIÂNIA,  
2024

FERNANDO COSTA OLIVEIRA

**CLASSIFICAÇÃO DE VINHOS USANDO IMAGENS HIPERESPECTRAIS E  
K-VIZINHOS MAIS PRÓXIMOS**

Este Trabalho de Conclusão de Curso foi julgado adequado para a obtenção do título de Bacharel em Ciência da Computação, e aprovado em sua forma final pela Escola Politécnica, da Pontifícia Universidade Católica de Goiás em 21/06/2024.

---

Profa. Ma. Ludmilla Reis Pinheiro dos Santos  
Coordenadora de Trabalho de Conclusão de Curso

Banca Examinadora:

---

Orientador: Prof. Dr. Clarimar José Coelho

---

Prof. Dr. Rafael Viana de Carvalho

---

Eng. Walcy Santos Rezende Rios

GOIÂNIA

2024

## **AGRADECIMENTOS**

Agradeço primeiramente à minha família, em especial aos meus pais e minha irmã, por todo o apoio, amor e compreensão durante esta jornada. Vocês sempre estiveram ao meu lado, oferecendo encorajamento e suporte incondicional. Esta conquista é tão minha quanto de vocês.

Gostaria de expressar minha profunda gratidão ao meu orientador, o Prof. Dr. Clarimar, por sua orientação, paciência e ter aceitado esse desafio. Sua expertise foi fundamental para a realização deste projeto.

Agradeço também a LCC pela valiosa assistência técnica e suporte. Em especial, agradeço à Gabriela que contribuiu bastante me auxiliando nas soluções de desafios técnicos, ao Prof. Dr. Rafael Viana, ao Eng. Walcy e ao Paulo, cujas contribuições após minhas apresentações foram cruciais.

Por fim, meu sincero agradecimento à minha namorada, Iasmin, pelo carinho, compreensão e apoio durante todo esse percurso. Aos meus amigos Manoella, Hiago e Felype, que tornaram essa jornada acadêmica mais leve e divertida, agradeço pela amizade e pelos momentos compartilhados.

## RESUMO

O objetivo deste trabalho é classificar amostras de vinhos para melhorar o controle de qualidade dos produtos. As amostras são imagens hiperespectrais e o classificador utilizado é o K-Vizinhos Próximos. Como pré-processamento foi utilizado o K-means com o objetivo de remover o fundo e possíveis ruídos da amostra. O classificador alcançou uma acurácia de 62%, sendo possível concluir que o uso de imagens hiperespectrais combinado com técnicas de aprendizado de máquina, como KNN, mostra-se eficaz.

**Palavras-chave:** Classificação de Vinhos. Controle de Qualidade. Imagens Hiperespectrais. K-means. K-Vizinhos Próximos. Algoritmos de Classificação.

## ABSTRACT

The objective of this work is to classify wine samples to improve product quality control. The samples are hyperspectral images and the classifier used is K-Nearest Neighbors. K-means was used as pre-processing with the aim of removing the background and possible noise from the sample. The classifier achieved an accuracy of 62%, making it possible to conclude that the use of hyperspectral images combined with machine learning techniques, such as KNN, is effective.

**Keywords:** Wine Classification. Quality Control. Hyperspectral Images. K-means. K-Nearest Neighbors. Classification Algorithms.

## LISTA DE ABREVIATURAS E SIGLAS

FN - Falso Negativo

FP - Falso Positivo

HSI - Imagem Hiperespectral

IR - Espectroscopia de Infravermelho

KNN - K-Nearest Neighbors (K-Vizinhos Mais Próximos)

NM - Nanômetro

PUC-GO - Pontifícia Universidade Católica de Goiás

RGB - Red, Green, Blue

SWIR - Infravermelho de Ondas Curtas (short-wave infrared)

TN - Verdadeiro Negativo

TP - Verdadeiro Positivo

LCC - Laboratório de Computação Científica

LAFAM-UFG - Laboratório de Análise Farmacêutica e Ambiental -  
Universidade Federal de Goiás

## LISTA DE FIGURAS

Figura 1: Estação de trabalho SisuCHEMA [7].

Figura 2: Faixas do espectro eletromagnético, destacando a região SWIR (900-2500 nm) e sua posição relativa a outras faixas como NIR, MWIR e LWIR [9].

Figura 3: Exemplo de hipercubo extraído de uma imagem hiperespectral, mostrando a relação entre as dimensões espaciais e espectrais e as assinaturas espectrais de diferentes pixels (ELMASRY; SUN, 2010, p. 22).

Figura 4: Transformação do hipercubo para uma matriz 2D, mostrando a desdobragem dos dados hiperespectrais para facilitar a análise (ELMASRY; SUN, 2010, p. 21-22).

Figura 5: Exemplo das etapas do processo de clustering K-means [10].

Figura 6: Ilustração do K-Nearest Neighbors.

Figura 7: Distribuição inicial dos dados antes do KNN [15].

Figura 8: Distância entre A1 e B1 (Pontos de referência) [15].

Figura 9: Classificação após KNN [15].

Figura 10: Matriz de confusão para duas classes de problema [16].

Figura 11: Assinatura Espectral. Fonte: O autor.

Figura 12: Assinatura Espectral. Fonte: O autor.

Figura 13: Assinatura Espectral. Fonte: O autor.

Figura 14: Assinatura Espectral. Fonte: O autor.

Figura 15: Assinatura Espectral. Fonte: O autor.

Figura 16: Amostra A1\_231110-110915. Fonte: O autor.

Figura 17: Amostra A1C\_231110-114859. Fonte: O autor.

Figura 18: Amostra A2\_231110-111118. Fonte: O autor.

Figura 19: Amostra A2C\_231110-115152. Fonte: O autor.

Figura 20: Amostra A3\_231110-112129. Fonte: O autor.

Figura 21: Amostra A1\_231110-110915. Fonte: O autor.

Figura 22: Amostra A1C\_231110-114859. Fonte: O autor.

Figura 23: Amostra A2\_231110-111118. Fonte: O autor.

Figura 24: Amostra A2C\_231110-115152. Fonte: O autor.

Figura 25: Amostra A3\_231110-112129. Fonte: O autor.

Figura 26: Amostra A3C\_231110-115553. Fonte: O autor.

Figura 27: Amostra A4\_231110-112314. Fonte: O autor.

Figura 28: Amostra A4C\_231110-120859. Fonte: O autor.

Figura 29: Amostra A5\_231110-112525. Fonte: O autor.

Figura 30: Amostra A5C\_231110-120221. Fonte: O autor.

Figura 31: Amostra A6\_231110-112904. Fonte: O autor.

Figura 32: Amostra A6C\_231110-120459. Fonte: O autor.

Figura 33: Amostra A7C\_231110-121112. Fonte: O autor.

Figura 34: Amostra A8\_231110-113257. Fonte: O autor.

Figura 35: Amostra A8C\_231110-121302. Fonte: O autor.

Figura 36: Amostra A9\_231110-113609. Fonte: O autor.

Figura 37: Amostra A9C\_231110-121436. Fonte: O autor.

Figura 38: Amostra A10\_231110-113907. Fonte: O autor.

Figura 39: Amostra A10C\_231110-121657. Fonte: O autor.

Figura 40: Amostra A11\_231110-114218. Fonte: O autor.

Figura 41: Amostra A12\_231110-114432. Fonte: O autor.

Figura 42: Matriz de confusão. Fonte: O autor.

Figura 43: Amostra A6\_. Fonte: O autor.

Figura 44: Amostra A5\_. Fonte: O autor.

Figura 45: Amostra A3C. Fonte: O autor.

Figura 46: Amostra A2\_. Fonte: O autor.

Figura 47: Amostra A5C. Fonte: O autor.

Figura 48: Amostra A1C. Fonte: O autor.

Figura 49: Amostra A4C. Fonte: O autor.

Figura 50: Amostra A6C. Fonte: O autor.

Figura 51: Amostra A2C. Fonte: O autor.

Figura 52: Amostra A3\_. Fonte: O autor.

Figura 53: Amostra A7C. Fonte: O autor.

Figura 54: Amostra A4\_. Fonte: O autor.

Figura 55: Amostra A1\_. Fonte: O autor.

Figura 56: Amostra A8C. Fonte: O autor.

Figura 57: Amostra A8\_. Fonte: O autor.

## SUMÁRIO

1. INTRODUÇÃO	1
2. MATERIAIS E MÉTODOS	2
2.1 Aquisição das amostras	3
2.1.1 SisuCHEMA	3
2.2 HSI	4
2.3 Pré-Processamento dos dados	6
2.3.2 K-Means	6
2.4 K-Vizinhos Próximos	8
2.5 Matriz de confusão	10
3. RESULTADOS	12
3.1 HSI	12
3.2 Assinatura Espectral	13
3.3 Pré-processamento	15
3.3.1 K-Means	15
3.4 Classificação	29
4. CONCLUSÃO	38
5. REFERÊNCIAS	38

## 1. INTRODUÇÃO

O desenvolvimento de técnicas avançadas para a classificação de vinhos tem se tornado cada vez mais relevante devido à crescente demanda por métodos precisos e eficientes na indústria vinícola. A classificação correta dos vinhos em termos de qualidade e características específicas é crucial não apenas para garantir a excelência do produto final, mas também para o controle de qualidade durante o processo de produção. As técnicas tradicionais de análise são muitas vezes invasivas e demoradas, o que impulsiona a busca por métodos mais modernos e automatizados [1, 2].

Com o decorrer dos anos, há um aumento de requisitos nas produções dos alimentos. Cada vez mais os consumidores são mais seletivos e exigentes, desde o tipo de uva utilizada, a região e o ano que foi colhida. Essa escolha pode ser influenciada através do processo de elaboração dos vinhos, que variam de acordo com seu tipo branco, rosado ou tinto [20].

Como método utilizado para a produção de vinhos, a vinificação é a alteração da uva sazoadada em vinho. Dos diferentes tipos de uvas, necessita de uma região específica e de um clima ideal para uma boa qualidade. As variações climáticas influenciam diretamente nos componentes químicos da uva, desde o plantio até a fase de preparo do vinho, resultando na acidez, sabor e teor alcoólico [20, 21].

Imagens hiperespectrais (HyperSpectral Imaging, HSI) capturam informações em várias bandas espectrais, fornecendo dados detalhados sobre a composição dos objetos analisados. Essa técnica combina aspectos da espectroscopia e da imagem, permitindo a obtenção de um "hipercubo" de dados, onde cada pixel da imagem contém um espectro completo de reflectância em diferentes comprimentos de onda. Isso possibilita a identificação e a análise precisa de diversos componentes químicos e físicos presentes nas amostras [2]. As HSIs são amplamente utilizadas em várias áreas, incluindo agricultura, segurança alimentar, e análise de materiais, devido à sua capacidade de fornecer uma análise não destrutiva e rápida da composição dos objetos [30]. No contexto da classificação de vinhos, a HSI é utilizada para capturar informações detalhadas sobre as características dos vinhos, permitindo a distinção e a classificação com base em suas assinaturas espectrais únicas [31]. Essas imagens são particularmente úteis para identificar diferentes componentes e suas distribuições espaciais nas amostras, facilitando uma análise mais aprofundada e precisa [34].

K-vizinhos mais próximos (K-NN) é uma técnica de aprendizado de máquina utilizada para classificar objetos com base nas características dos vizinhos mais próximos no espaço de características. O valor de K representa

o número de vizinhos considerados no processo de classificação. O algoritmo funciona calculando a distância entre o ponto de dados que está sendo classificado e todos os outros pontos de dados no conjunto de treinamento, selecionando os K pontos mais próximos. Em seguida, a classe mais frequente entre esses vizinhos é atribuída ao ponto de dados de interesse [13]. O KNN é conhecido por sua simplicidade e eficácia em problemas de classificação, especialmente quando os dados são bem distribuídos e a relação entre as classes é clara [14].

Métodos tradicionais de controle de qualidade muitas vezes são subjetivos e dependem da experiência dos avaliadores humanos, o que pode resultar em variabilidade nos resultados. Acselrad (1994) destaca que a qualidade do produto na indústria de alimentos depende diretamente da qualidade de vida dos trabalhadores envolvidos no processo de controle de qualidade. Esta relação ressalta a importância de melhorar as condições de trabalho para reduzir a variabilidade causada por erros humanos. A aplicação de tecnologias objetivas, como a hiperespectral, oferece uma solução mais consistente e confiável, mitigando a influência de fatores subjetivos e melhorando a precisão dos resultados de controle de qualidade [5].

Este trabalho tem como objetivo desenvolver um método mais eficiente e preciso para a classificação de vinhos, através da aplicação de imagens hiperespectrais em conjunto com algoritmos de aprendizado de máquina [2]. As imagens hiperespectrais foram escolhidas por sua capacidade de capturar uma vasta quantidade de informações espectrais em cada pixel da imagem, permitindo uma análise detalhada e precisa das amostras de vinho. O KNN, um algoritmo de aprendizado supervisionado, será empregado para classificar essas amostras com base em suas características espectrais [16].

A estrutura deste trabalho está organizada da seguinte forma. Primeiramente, a seção de Materiais e Métodos descreve em detalhe os procedimentos experimentais, os equipamentos utilizados e os algoritmos aplicados. Em seguida, a seção de Resultados apresenta os achados do trabalho. Por fim, a seção de Conclusão discute as implicações dos resultados obtidos e sugere possíveis direções para pesquisas futuras.

Com base nessas abordagens teóricas e metodológicas, espera-se que este trabalho contribua para o avanço das técnicas de controle de qualidade, oferecendo um método mais robusto e preciso para a classificação de vinhos.

## **2. MATERIAIS E MÉTODOS**

As amostras de vinho fornecidas pela(o) LCC em cooperação com o LAFAM-UFG, submetidas a um processo de liofilização para preservação de suas características essenciais [23]. As análises foram realizadas utilizando a

estação de trabalho SisuCHEMA, que combina espectroscopia de infravermelho de ondas curtas (SWIR) com imagens de alta resolução. Para a captura das imagens hiperespectrais, a técnica de HSI foi empregada, permitindo uma análise detalhada das amostras [7, 8, 2]. No pré-processamento dos dados, o algoritmo K-Means foi utilizado para remover o fundo das imagens, separando os pixels das amostras dos pixels do fundo [11]. O KNN foi aplicado para classificar as amostras de vinho com base nas informações espectrais [13].

## **2.1 Aquisição das amostras**

Foram utilizadas 21 amostras de vinhos que foram submetidas a um processo de liofilização, um método avançado de desidratação que preserva as características essenciais do produto através da sublimação da água diretamente do estado sólido para o gasoso, sem passar pelo estado líquido [29].

A liofilização, também conhecida como secagem por congelamento, é um processo de três etapas: congelamento, secagem primária e secagem secundária. Durante a secagem primária, a temperatura do produto deve ser mantida abaixo da temperatura crítica para evitar colapsos, garantindo a qualidade do produto final [26]. A secagem secundária remove a água residual através da exposição do produto a temperaturas mais elevadas enquanto permanece em condições de vácuo [27].

As amostras de vinho foram coletadas e imediatamente congeladas para preservar suas características originais. O processo seguiu as etapas de congelamento inicial a  $-40^{\circ}\text{C}$ , seguido pela secagem primária sob vácuo e baixa temperatura, e finalizado com a secagem secundária a  $30^{\circ}\text{C}$  por 4 horas [28]. Este método avançado garante a preservação das características essenciais das amostras de vinho, permitindo uma análise precisa e confiável de suas propriedades espectrais para fins de classificação [24].

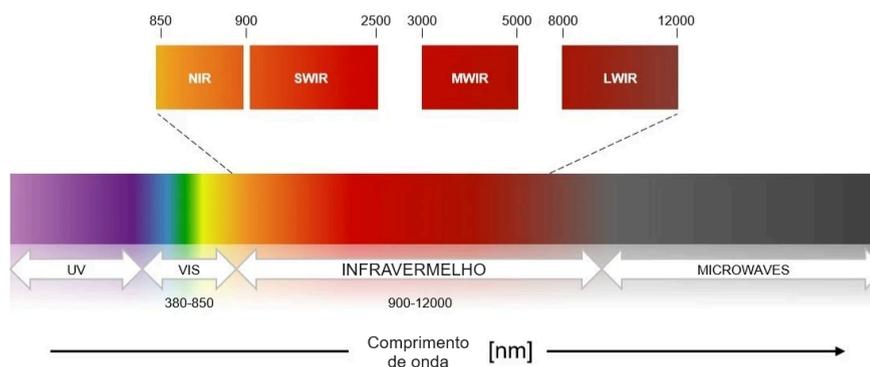
### **2.1.1 SisuCHEMA**

A SisuCHEMA é uma estação de trabalho completa de imagem química que combina espectroscopia no infravermelho de ondas curtas (Short Wave InfraRed, SWIR) com imagens de alta resolução. Oferecendo informações detalhadas sobre as características químicas, suas quantidades e distribuições dentro da amostra. Possuindo um sensor MCT de alto desempenho com 384 pixels espaciais e 256 bandas espectrais. A câmera suporta uma taxa de linha de 400 linhas por segundo. Um sistema de resfriamento Stirling proporciona um resfriamento profundo do sensor, que é abrigado em uma carcaça IP54 [7].



**Figura 1:** Estação de trabalho SisuCHEMA [7].

A radiação infravermelha de onda curta (SWIR) cobre a faixa do espectro eletromagnético entre aproximadamente 900 nm e 2500 nm, sendo crucial para a análise detalhada de componentes químicos em amostras complexas. Esta faixa permite a penetração em materiais biológicos e inorgânicos, mantendo alta resolução, o que é essencial para a caracterização em diversas aplicações [8].



**Figura 2:** Faixas do espectro eletromagnético, destacando a região SWIR (900-2500 nm) e sua posição relativa a outras faixas como NIR, MWIR e LWIR [9].

## 2.2 HSI

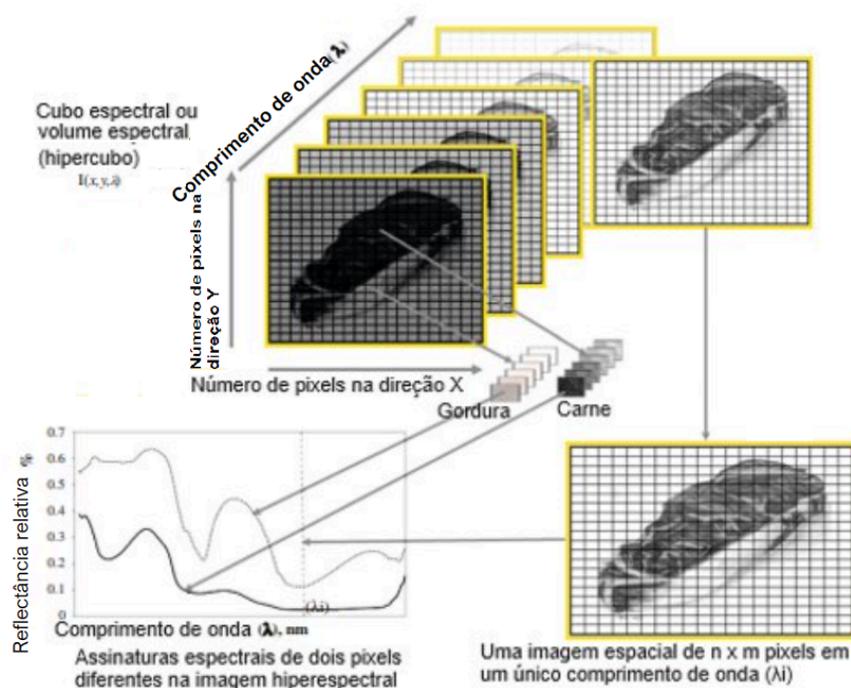
As imagens hiperespectrais (HSI) são uma técnica avançada que integra espectroscopia e imagem, permitindo uma análise detalhada e precisa das amostras. Esta técnica gera um mapa espacial da variação espectral, tornando-se uma ferramenta valiosa em diversas aplicações, como detecção automática de alvos e reconhecimento de composição analítica. A HSI foi desenvolvida para integrar técnicas espectroscópicas e de imagem,

permitindo a identificação direta de diferentes componentes e sua distribuição espacial na amostra testada [2].

O SWIR é frequentemente utilizado para obter imagens hiperespectrais, abrangendo comprimentos de onda na faixa de 900 a 2500 nm, oferecendo maior penetração em materiais biológicos e melhor resolução espectral. Este intervalo é particularmente útil, pois permite a detecção de componentes químicos específicos e a diferenciação entre diferentes materiais [30].

Os dados hiperespectrais são organizados em um cubo de dados ou hipercubo. Este hipercubo contém duas dimensões espaciais ( $x$  e  $y$ ) e uma dimensão espectral ( $\lambda$ ), que representa os diferentes comprimentos de onda. Para facilitar o processamento e a análise dos dados, o hipercubo pode ser transposto para uma matriz bidimensional, onde cada linha representa um pixel da imagem e cada coluna corresponde a um comprimento de onda específico. Esta transposição permite a aplicação de técnicas de análise multivariada [31].

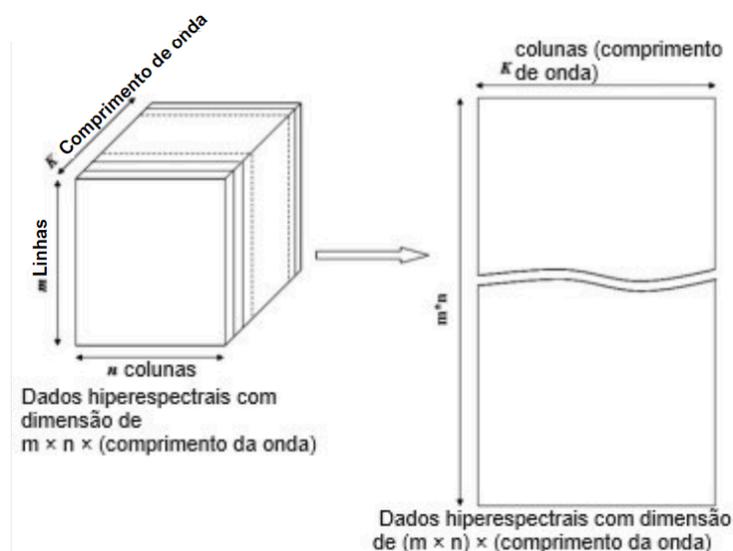
A Figura 3 ilustra um exemplo de hipercubo extraído de uma imagem hiperespectral adquirida para uma peça de carne. Esta figura demonstra como os dados hiperespectrais são organizados em uma matriz tridimensional, com duas dimensões espaciais e uma dimensão espectral. A imagem mostra como diferentes assinaturas espectrais podem ser extraídas de diferentes pixels no hipercubo, representando diferentes componentes da amostra, como gordura e carne magra.



**Figura 3:** Exemplo de hipercubo extraído de uma imagem hiperespectral, mostrando a relação entre as dimensões espaciais e espectrais e as assinaturas espectrais de diferentes pixels [2].

A assinatura espectral de um pixel é o espectro de reflectância correspondente a esse pixel, que serve como uma "impressão digital" única do material presente. Cada material possui uma assinatura espectral distinta, que pode ser utilizada para sua identificação e classificação. Ao analisar as assinaturas espectrais, é possível identificar e classificar os diferentes componentes de uma amostra com alta precisão [32].

Para facilitar o processamento e a análise dos dados, a matriz tridimensional (3D) do hipercubo pode ser transposta para uma matriz bidimensional (2D). A Figura 4 ilustra esta transformação, onde os dados hiperespectrais brutos com dimensões  $m \times n \times (\text{Comprimento da onda})$  são desdobrados em uma matriz 2D com dimensões  $(m \times n) \times (\text{Comprimento da onda})$ . Cada linha da matriz 2D representa um pixel da imagem e cada coluna corresponde a um comprimento de onda específico, permitindo a aplicação de técnicas de análise multivariada [33].



**Figura 4:** Transformação do hipercubo para uma matriz 2D, mostrando a desdobragem dos dados hiperespectrais para facilitar a análise (ELMASRY; SUN, 2010, p. 21-22).

As principais vantagens das imagens hiperespectrais sobre os métodos tradicionais incluem: mínima preparação da amostra, método não destrutivo, aquisição rápida e a capacidade de visualizar a distribuição espacial de diversas composições químicas simultaneamente [34]. Estas características tornam as imagens hiperespectrais uma ferramenta poderosa para a análise de qualidade e classificação em diversas áreas [35].

## 2.3 Pré-Processamento dos dados

### 2.3.2 K-Means

O K-means é amplamente utilizado para a clusterização de dados, permitindo a organização de grandes conjuntos de dados em clusters

distintos, onde os dados dentro de cada cluster são mais similares entre si do que em relação a dados de outros clusters [11]. Esta técnica é particularmente útil no pré-processamento de dados, auxiliando na identificação de padrões e estruturas subjacentes que podem ser posteriormente analisados de forma mais detalhada.

Segundo Han, Kamber e Pei (2012, p. 452), o processo de K-means inicia-se com a definição do número de clusters  $K$ , seguido pela escolha inicial dos centros dos clusters, que são selecionados aleatoriamente. Após a inicialização, o algoritmo procede iterativamente através dos seguintes passos:

1. Escolha arbitrária de  $k$  objetos de  $D$  como centros iniciais dos clusters;
2. Repetir:
  - 2.1. (Re)atribuir cada objeto ao cluster cujo centroide é o mais próximo, baseado no valor médio dos objetos no cluster;
  - 2.2. Atualizar os centros dos clusters, ou seja, calcular a média dos valores dos objetos de cada cluster;
3. Até que não haja mais mudanças.

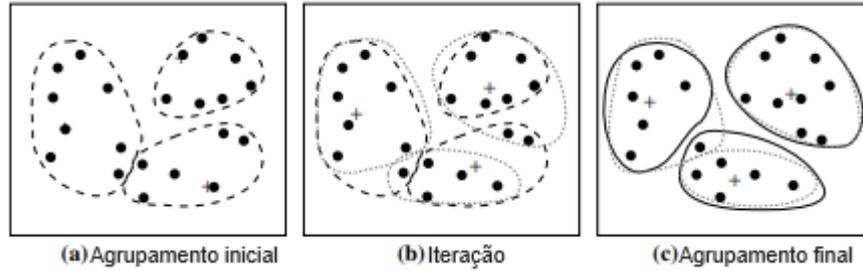
Para medir a similaridade entre os pontos de dados, utiliza-se a distância euclidiana. A fórmula da distância euclidiana é:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

Onde  $p$  e  $q$  são vetores de dados e  $n$  é a dimensionalidade dos dados [10]. Esta métrica calcula a raiz quadrada da soma das diferenças quadradas entre as coordenadas correspondentes dos pontos de dados, proporcionando uma medida direta da distância entre dois pontos no espaço de dados.

Com isso, este agrupamento facilita a identificação de regiões homogêneas na amostra, melhorando a eficiência de técnicas subsequentes de análise [10].

Este processo é repetido até que as atribuições dos pontos de dados não mudem mais [11]. A Figura 5 ilustra as etapas do processo de clustering K-means, desde a inicialização dos clusters até a convergência final dos centros dos clusters.

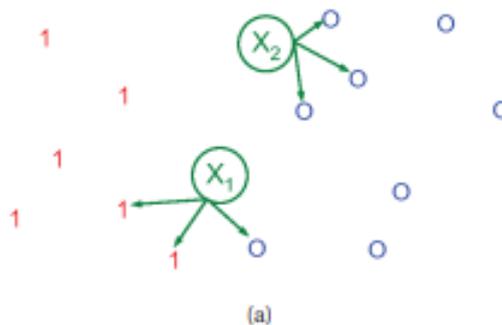


**Figura 5:** Exemplo das etapas do processo de clustering K-means [10].

## 2.4 K-Vizinhos Próximos

O K-Vizinhos Próximos (KNN) é um algoritmo de classificação usado em aprendizado supervisionado que classifica novos pontos de dados com base na proximidade a pontos de dados existentes no conjunto de treinamento. Este método é amplamente utilizado devido à sua simplicidade e eficácia em resolver problemas de classificação. A ideia central do KNN é que objetos semelhantes estão próximos uns dos outros, portanto, a classificação de um novo ponto de dados é determinada pelas classes dos seus vizinhos mais próximos no espaço das características [13].

Operando através da atribuição de um rótulo ao ponto de teste mais comum entre os  $k$  pontos de treinamento mais próximos, a escolha do valor de  $k$  é crítica. Um valor muito pequeno pode tornar o modelo sensível ao ruído nos dados de treinamento, enquanto um valor muito grande pode suavizar demais as fronteiras de decisão, resultando em menor precisão do modelo [14].



**Figura 6:** Ilustração do K-Nearest Neighbors

A Figura 6, adaptada de Murphy (2012), ilustra o funcionamento do algoritmo KNN em um espaço bidimensional com  $k = 3$ . Os pontos  $x_1$  e  $x_2$  são pontos de teste, e cada ponto é classificado com base nos três vizinhos mais próximos. Os círculos verdes ao redor dos pontos de teste representam a distância até os três vizinhos mais próximos.

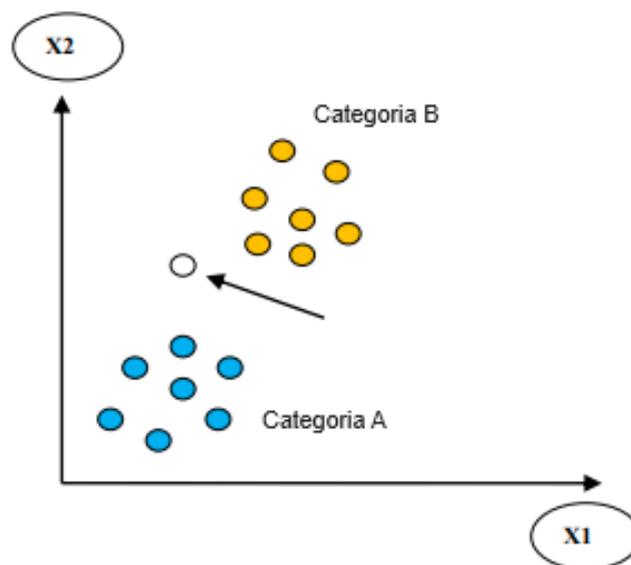
Para medir a proximidade entre os pontos é utilizado a distância euclidiana conforme definido na equação (1). Método particularmente útil em

algoritmos como o KNN, pois permite identificar os vizinhos mais próximos de um ponto de dados no espaço das características [13].

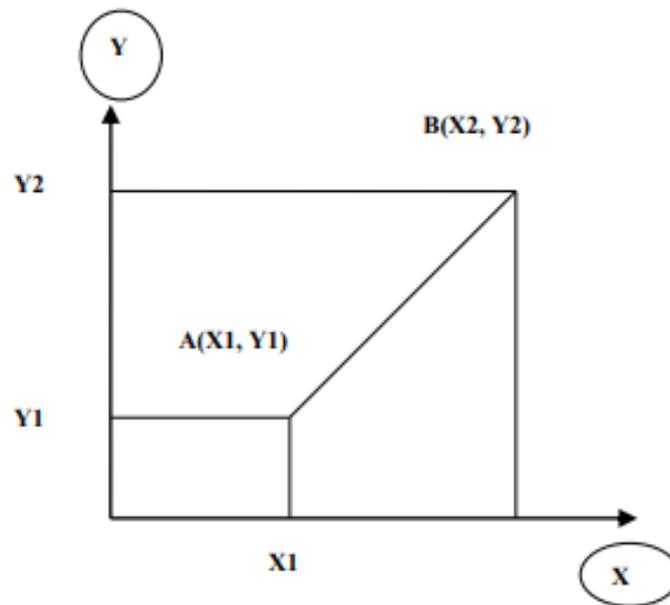
Segundo Suyal e Goyal (2022), o passo a passo do KNN é da seguinte forma:

1. **Escolha do número de vizinhos (k):** Primeiramente, seleciona-se o número  $k$  de vizinhos.
2. **Cálculo da distância:** A distância Euclidiana deve ser calculada para o número  $k$  de vizinhos.
3. **Identificação dos vizinhos mais próximos:** Seleciona-se os  $k$  vizinhos mais próximos com base na distância Euclidiana calculada.
4. **Contagem das classes dos vizinhos:** Entre os  $k$  vizinhos mais próximos, conta-se o número de pontos de dados em cada categoria.
5. **Classificação:** Os novos pontos de dados serão atribuídos à categoria para a qual o número de vizinhos é máximo.
6. **Modelo KNN pronto:** Desta forma, o modelo de classificação KNN estará pronto

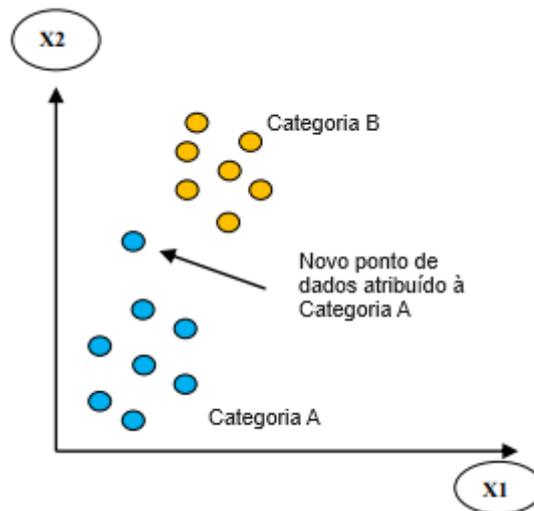
Para exemplificar o passo a passo temos as figuras 7, 8 e 9. A Figura 7 ilustra a distribuição inicial dos dados antes da aplicação do algoritmo KNN. Os pontos azuis representam as amostras pertencentes à Categoria A, enquanto os pontos amarelos representam as amostras da Categoria B. O objetivo é classificar um novo ponto (representado pelo ponto branco) com base nas categorias dos pontos mais próximo. Já a Figura 8 demonstra o cálculo da distância entre dois pontos de referência,  $A(X_1, Y_1)$  e  $B(X_2, Y_2)$ . A distância euclidiana entre os pontos A e B é representada pela linha diagonal que conecta os dois pontos no gráfico. A Figura 10 apresenta a classificação dos dados após a aplicação do KNN. O novo ponto de dados, inicialmente representado na Figura 9, foi classificado na Categoria A com base nos pontos mais próximos [15].



**Figura 7:** Distribuição inicial dos dados antes do KNN [15].



**Figura 8:** Distância entre A1 e B1 (Pontos de referência) [15].



**Figura 9:** Classificação após KNN [15].

### 2.5 Matriz de confusão

Uma matriz de confusão é uma matriz de tamanho  $n \times n$  que quando associada a um classificador mostra a classificação prevista e real, onde  $n$  é o número de classes diferentes [16].

	PREVISAO NEGATIVA	PREVISAO POSITIVA
REAL NEGATIVO	$a$	$b$
REAL POSITIVO	$c$	$d$

**Figura 10:** Matriz de confusão para duas classes de problema [16].

Segundo VISA, e exemplificado na figura 10, a tabela contém quatro componentes principais, que representam a comparação entre as previsões do modelo e os valores reais:

- **a:** Verdadeiros Negativos (TN) - O número de instâncias que foram corretamente previstas como negativas.
- **b:** Falsos Positivos (FP) - O número de instâncias que foram incorretamente previstas como positivas, mas que são negativas.
- **c:** Falsos Negativos (FN) - O número de instâncias que foram incorretamente previstas como negativas, mas que são positivas.
- **d:** Verdadeiros Positivos (TP) - O número de instâncias que foram corretamente previstas como positivas.

A matriz de confusão permite calcular diversas métricas de avaliação que são fundamentais para analisar o desempenho de um modelo de classificação. Entre as principais métricas estão a acurácia, a precisão, o recall e a pontuação F1.

$$Accuracy = \frac{a+b}{a+b+c+d} \quad (2)$$

A equação 2 indica como calcular a acurácia. A acurácia mede a proporção de previsões corretas (tanto verdadeiros positivos quanto verdadeiros negativos) em relação ao total de previsões [16].

$$precision = \frac{TP}{TP+FP} \quad (3)$$

A equação 3 indica como calcular a precisão, que é responsável por medir a proporção de previsões positivas corretas em relação ao total de previsões positivas [16].

$$recall = \frac{TP}{TP+FN} \quad (4)$$

A equação 4 indica como calcular o *recall*, que mede a proporção de verdadeiros positivos em relação ao total de instâncias que são realmente positivas.

Para proporcionar uma visão mais completa da performance do modelo, mesmo em cenários desbalanceados, foi adotado o *F1 - score*, oferecendo um equilíbrio entre os falsos positivos e falsos negativos, sendo indicado na equação 5. Sendo crucial em situações onde tanto a precisão (a capacidade de evitar falsos positivos) quanto o recall (a capacidade de detectar todos os verdadeiros positivos) são importantes [17].

$$F1\ score = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (5)$$

Para aprimorar nossa análise utilizamos o *support*, uma métrica que conta o número de ocorrências de cada classe no conjunto de dados, sendo crucial para entender a distribuição das classes e avaliar a justiça do modelo ao tratar cada classe, especialmente em conjuntos de dados desbalanceados [19].

Por fim, consideramos a *Weighted Avg* que calcula métricas agregadas *precisão/recall/F1-score*, atribuindo pesos com base no número de instâncias de cada classe, oferecendo uma visão mais realista do desempenho do modelo em conjuntos de dados desbalanceados [18].

### 3. Resultados

#### 3.1 HSI

Para facilitar a interpretação dos resultados obtidos a partir das imagens hiperespectrais, definimos abreviações específicas para cada amostra de vinho. Essas abreviações serão utilizadas para casos onde o nome da amostra não fica claro. A Tabela 1 abaixo apresenta as abreviações correspondentes a cada imagem de amostra de vinho analisada.

#	Nome da amostra	Abreviação
1	A1_231110-110915	A1_
2	A1C_231110-114859	A1C
3	A2_231110-111118	A2_
4	A2C_231110-115152	A2C
5	A3_231110-112129	A3_
6	A3C_231110-115553	A3C
7	A4_231110-112314	A4_
8	A4C_231110-120859	A4C
9	A5_231110-112525	A5_
10	A5C_231110-120221	A5C
11	A6_231110-112904	A6_
12	A6C_231110-120459	A6C
13	A7C_231110-121112	A7C
14	A8_231110-113257	A8_

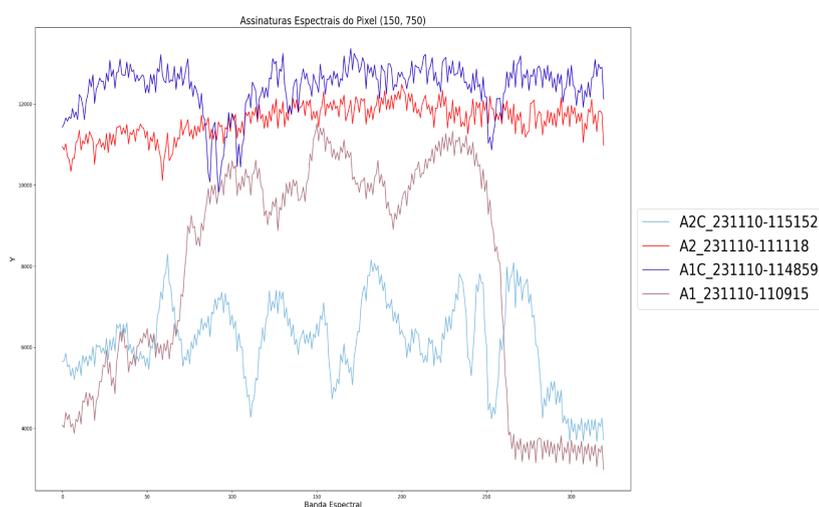
15	A8_231110-113257	A8C
16	A9_231110-113609	A9_
17	A9C_231110-121436	A9C
18	A10_231110-113907	A10
19	A10C_231110-121657	A10C
20	A11_231110-114218	A11
21	A12_231110-114432	A12

**Tabela 1:** Abreviações das amostras

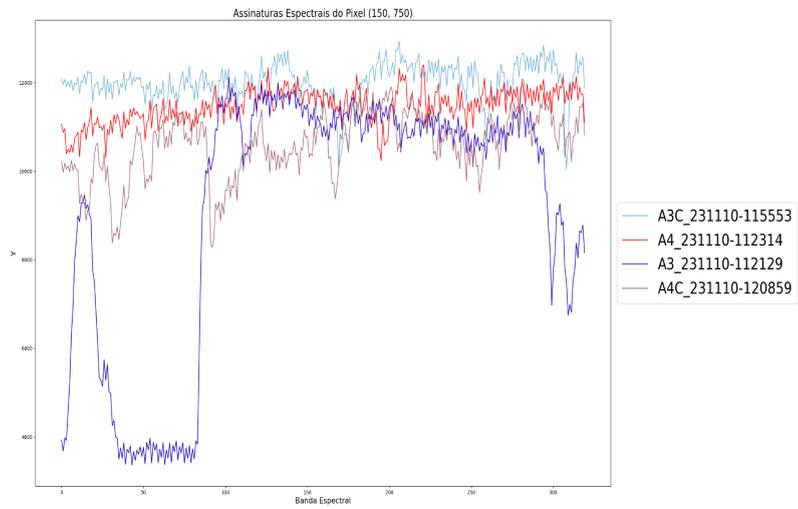
### 3.2 Assinatura Espectral

As Figuras 11 a 15 apresentam as assinaturas espectrais das amostras de vinho, permitindo uma análise detalhada das diferenças espectrais entre elas. A Figura 11 inclui as amostras A2C, A2\_, A1C e A1\_. A Figura 12 apresenta as amostras A3C, A4\_, A3\_ e A4C. A Figura 13 exhibe as amostras A5\_, A5C, A6C e A6\_. A Figura 14 contém as amostras A7C, A19\_, A9C, A8\_ e A8C. Por fim, a Figura 15 mostra as amostras A12, A11, A10C, A10 e A13. Todas essas amostras foram analisadas no pixel [150, 750], onde se observa uma concentração maior de vinho.

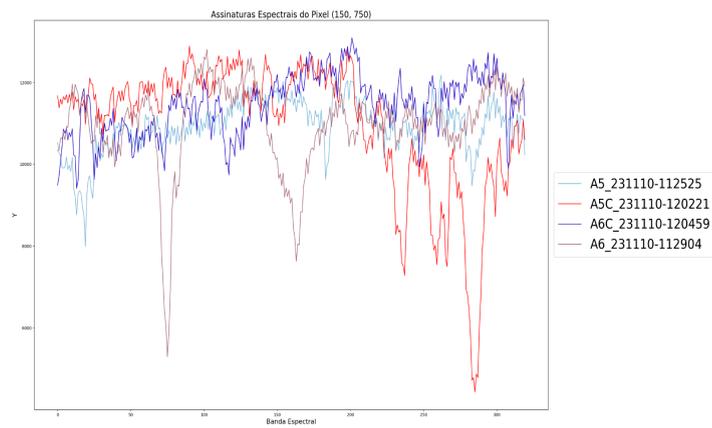
O objetivo dessas figuras é demonstrar que, por meio das assinaturas espectrais, é possível identificar diferenças significativas entre as amostras. Essas diferenças são fundamentais para a classificação, pois cada tipo de vinho apresenta uma assinatura espectral única. Com essas assinaturas espectrais distintas, podemos avançar com confiança no processo de classificação.



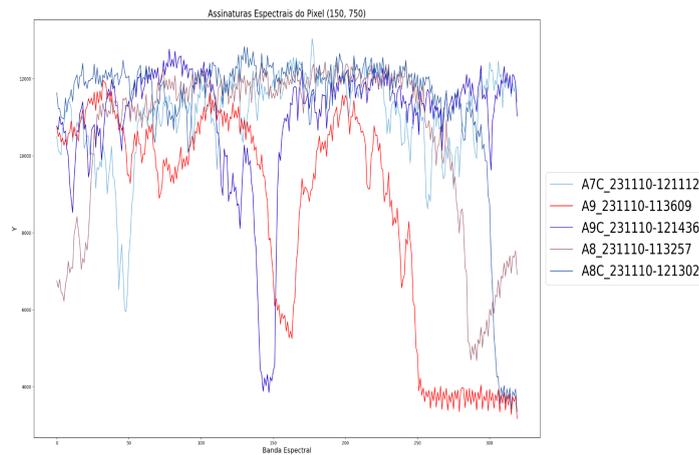
**Figura 11: Assinatura Espectral. Fonte: O autor**



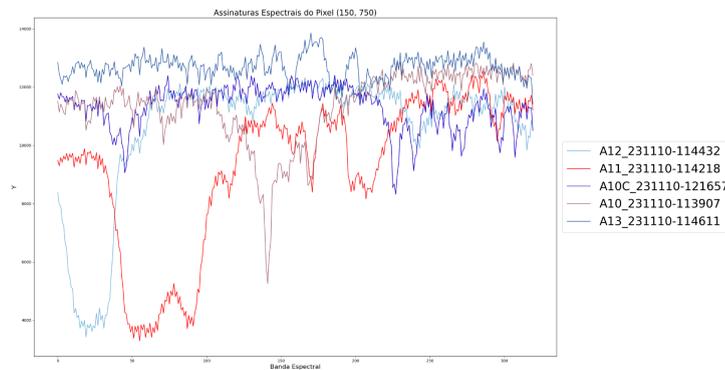
**Figura 12: Assinatura Espectral. Fonte: O autor**



**Figura 13: Assinatura Espectral. Fonte: O autor**



**Figura 14:** Assinatura Espectral. Fonte: O autor



**Figura 15:** Assinatura Espectral. Fonte: O autor

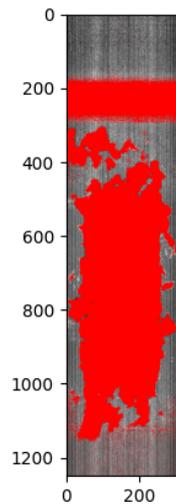
### 3.3 Pré-processamento

#### 3.3.1 K-Means

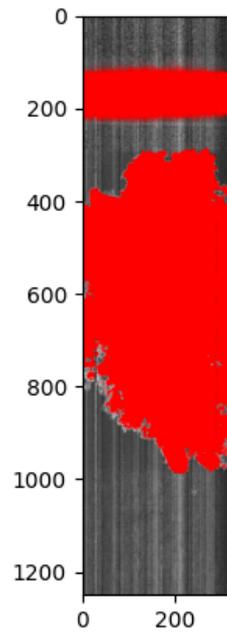
Para a etapa de pré-processamento dos dados, utilizei o K-Means com o objetivo de remover o fundo e possíveis ruídos das amostras de vinho. A presença do fundo e possíveis ruídos poderiam interferir na precisão da classificação, tornando essencial a sua remoção.

Após algumas experimentações, determinei que a utilização de  $k = 2$  seria a abordagem mais eficaz. As experimentações ocorreram de modo onde tentei utilizar  $k = 3$  para tentar fazer uma maior remoção de ruídos, mas o resultado acabou não sendo satisfatório. Um dos centróides foi reconhecido como fundo, enquanto os outros dois centróides dividiram a amostra de vinho. Isso resultou em um resultado indesejado, pois era necessário que um único centróide representasse toda a amostra de vinho. Essa configuração permitiu a distinção clara entre os pixels pertencentes ao vinho e os pixels pertencentes ao fundo da prancha/possíveis ruídos.

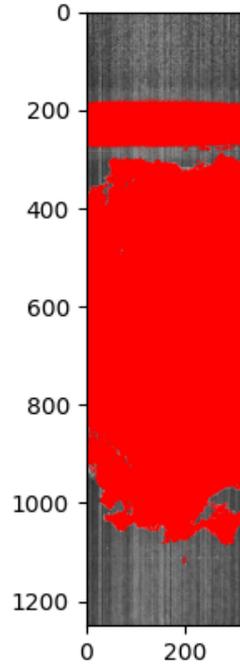
As Figuras 16 a 20 ilustram a coloração de um centróide após a aplicação do algoritmo K-means. Se o centróide colorido em vermelho for identificado como o fundo, ele será removido, mantendo apenas o cluster correspondente à amostra de vinho. Da mesma forma, se o centróide colorido for identificado como a amostra de vinho, ele será mantido, e o centróide sem coloração será removido, sendo identificado como o fundo. É importante ressaltar que a coloração em vermelho representa a tentativa de isolar o centróide que contém a amostra de vinho, e não a confirmação do centróide que contém o vinho. A escolha do centróide correto é feita através da observação cuidadosa da imagem gerada.



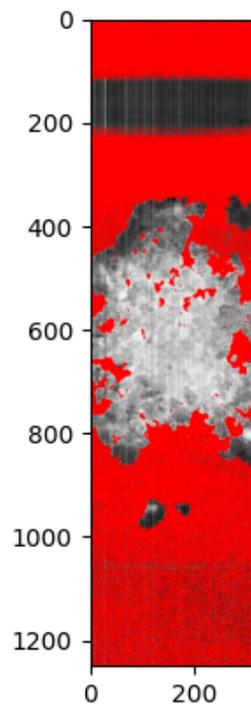
**Figura 16:** Amostra A1\_231110-110915. Fonte: O autor.



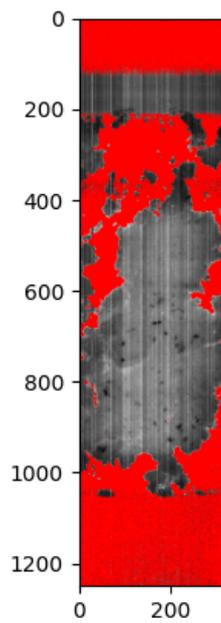
**Figura 17:** Amostra A1C\_231110-114859. Fonte: O autor.



**Figura 18:** Amostra A2\_231110-111118 Fonte: O autor.

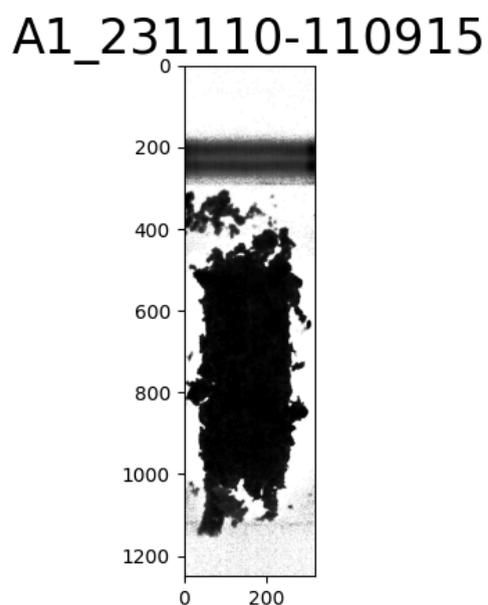


**Figura 19:** Amostra A2C\_231110-115152 Fonte: O autor.

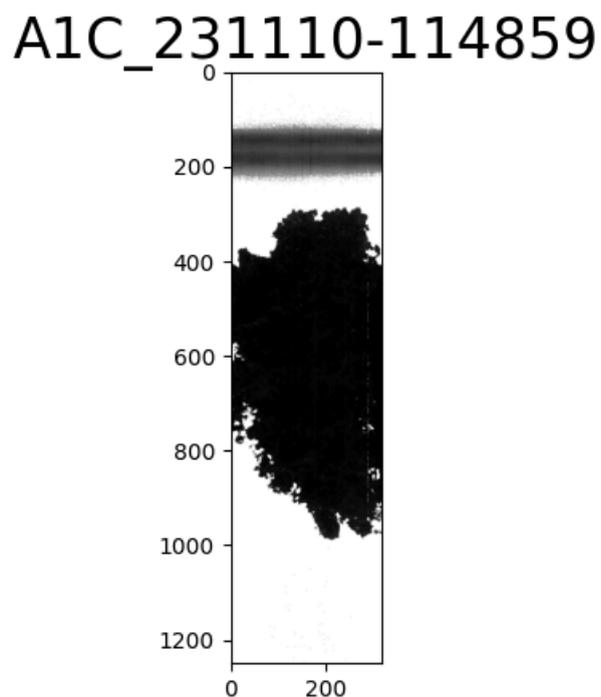


**Figura 20:** Amostra A3\_231110-112129 Fonte: O autor.

Após a identificação do cluster que representa a amostra de vinho foi feita remoção do cluster desnecessário, resultando na separação eficiente dos dados, podendo ser evidenciado nas figuras 21 a 41. Com a remoção do fundo, as assinaturas espectrais das amostras de vinho puderam ser analisadas de forma mais precisa, garantindo a integridade dos dados para as etapas subsequentes de classificação.



**Figura 21:** Amostra A1\_231110-110915 Fonte: O autor.



**Figura 22:** Amostra A1C\_231110-114859 Fonte: O autor.

A2\_231110-111118

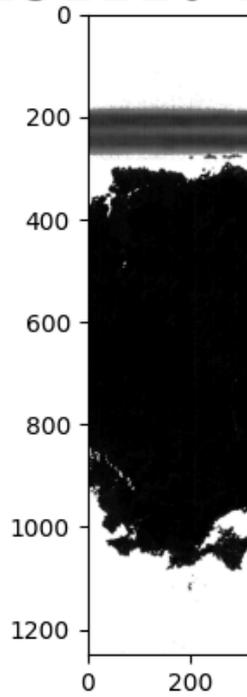


Figura 23: Amostra A2\_231110-111118 Fonte: O autor.

A2C\_231110-115152

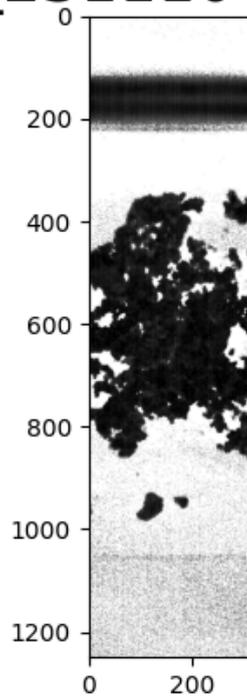


Figura 24: Amostra A2C\_231110-115152 Fonte: O autor.

A3\_231110-112129

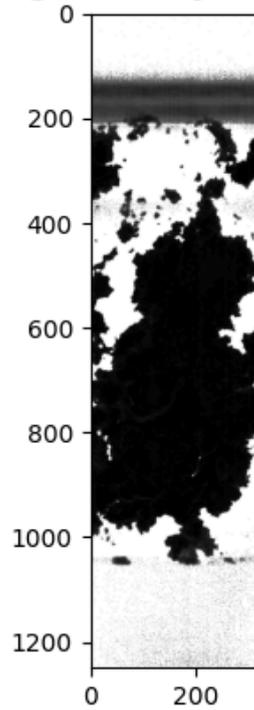


Figura 25: Amostra A3\_231110-112129 Fonte: O autor.

A3C\_231110-115553

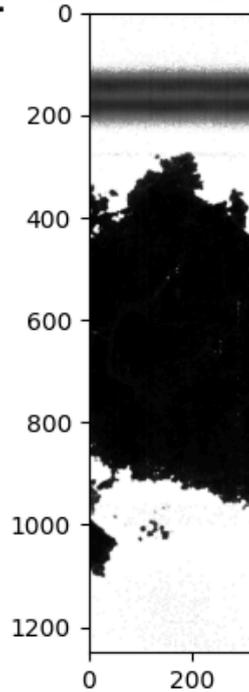
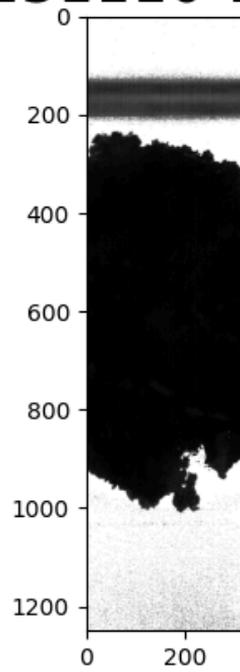


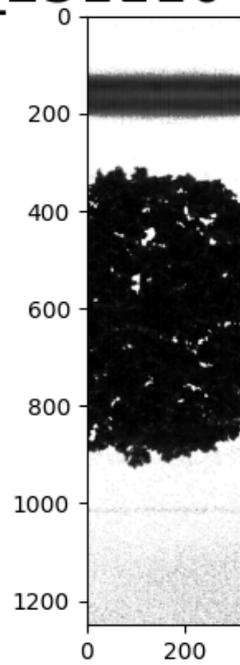
Figura 26: Amostra A3C\_231110-115553 Fonte: O autor.

A4\_231110-112314



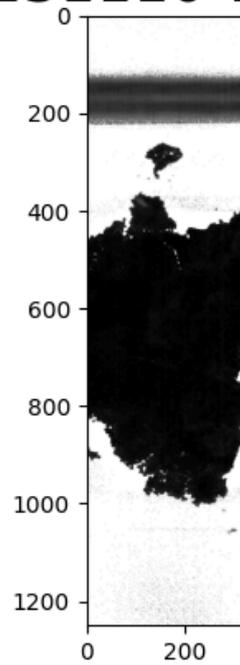
**Figura 27:** Amostra A4\_231110-112314 Fonte: O autor.

A4C\_231110-120859



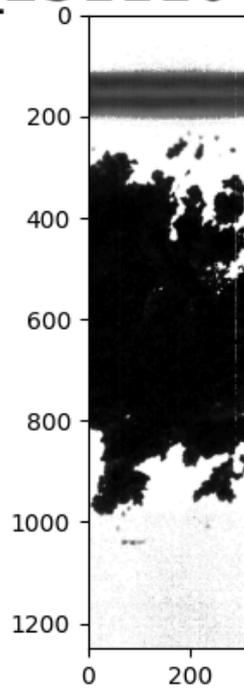
**Figura 28:** Amostra A4C\_231110-120859 Fonte: O autor.

A5\_231110-112525



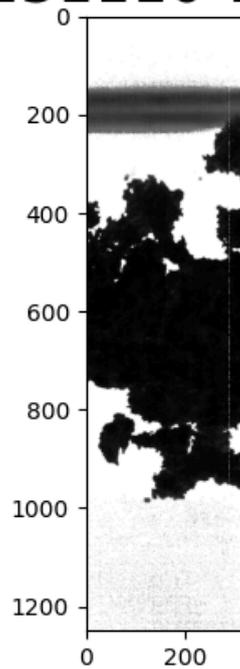
**Figura 29:** Amostra A5\_231110-112525 Fonte: O autor.

A5C\_231110-120221



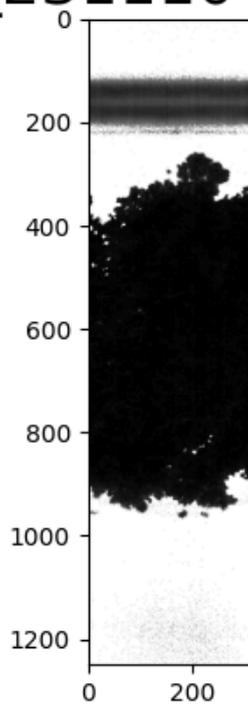
**Figura 30:** Amostra A5C\_231110-120221 Fonte: O autor.

A6\_231110-112904



**Figura 31:** Amostra A6\_231110-112904 Fonte: O autor.

A6C\_231110-120459



**Figura 32:** Amostra A6C\_231110-120459 Fonte: O autor.

A7C\_231110-121112

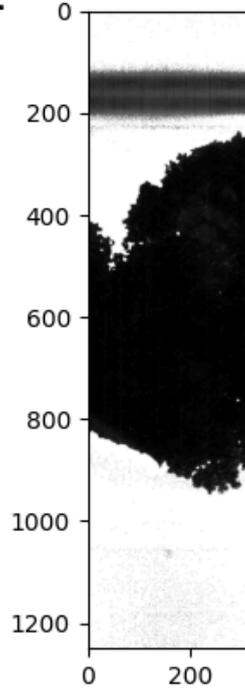


Figura 33: Amostra A7C\_231110-121112 Fonte: O autor.

A8\_231110-113257

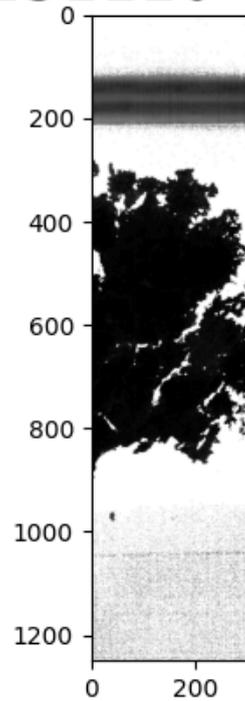
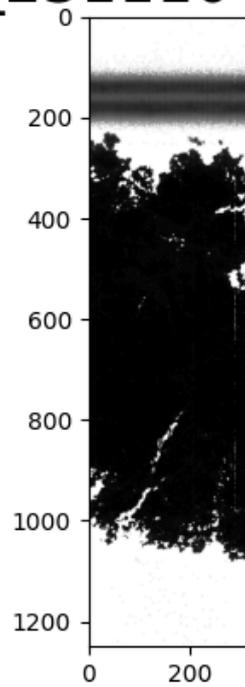


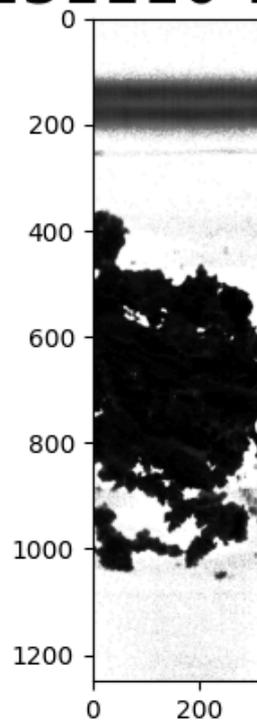
Figura 34: Amostra A8\_231110-113257 Fonte: O autor.

A8C\_231110-121302



**Figura 35:** Amostra A8C\_231110-121302 Fonte: O autor.

A9\_231110-113609



**Figura 36:** Amostra A9\_231110-113609 Fonte: O autor.

A9C\_231110-121436

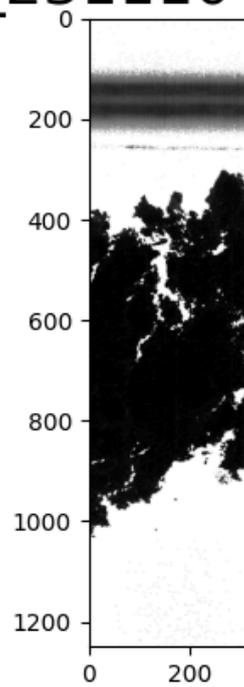


Figura 37: Amostra A9C\_231110-121436 Fonte: O autor.

A10\_231110-113907

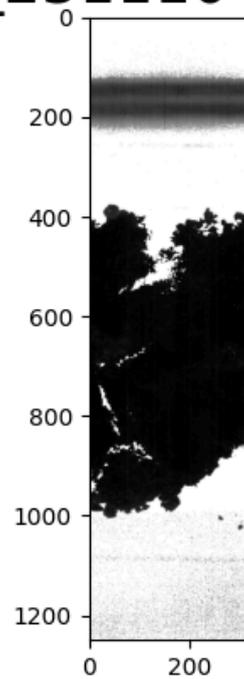


Figura 38: Amostra A10\_231110-113907 Fonte: O autor.

A10C\_231110-121657

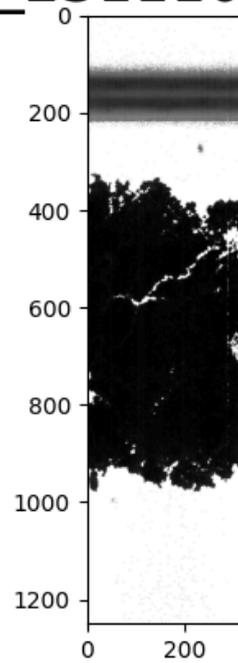


Figura 39: Amostra A10C\_231110-121657 Fonte: O autor.

A11\_231110-114218

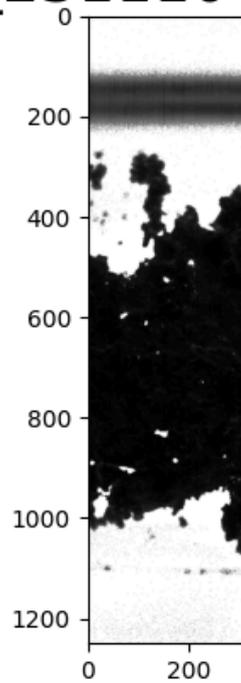
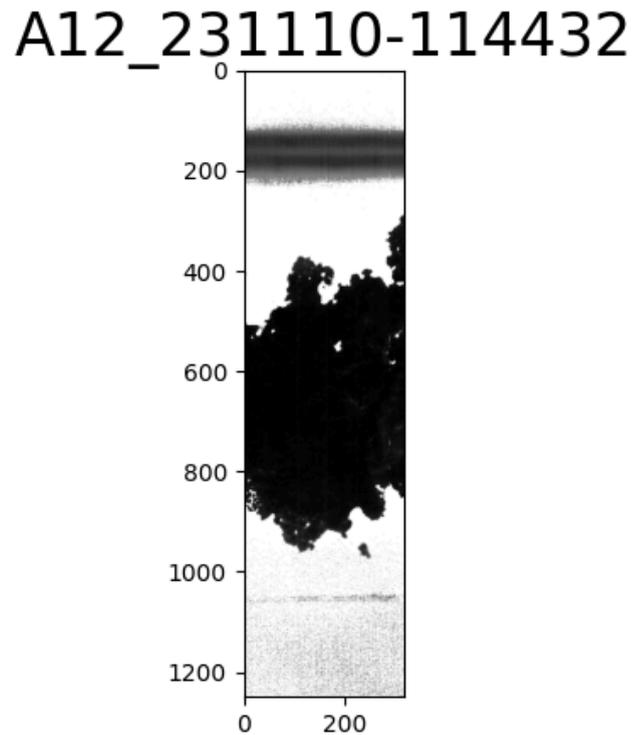


Figura 40: Amostra A11\_231110-114218 Fonte: O autor.



**Figura 41:** Amostra A12\_231110-114432 Fonte: O autor.

### 3.4 Classificação

Primeiramente, os dados foram tratados para serem passados ao KNN. A preparação dos dados iniciou-se com a divisão em conjuntos de treinamento e teste. Em seguida, os dados foram organizados e os rótulos associados, garantindo uma divisão adequada entre treinamento e teste. Para finalizar, os dados foram embaralhados para evitar qualquer viés na aprendizagem. Como resultado, obtivemos as matrizes de dados espectrais e os rótulos, prontos para serem utilizados na fase de classificação, juntamente com uma lista de nomes das classes alvo. Isso assegura que os dados estejam estruturados de maneira eficiente para a aplicação do KNN no processo de classificação.

Durante o início do treinamento, enfrentei um grande desafio de processamento computacional com 21 amostras. Inicialmente, precisei trabalhar na resolução dessa questão. A solução encontrada foi trabalhar apenas com 15 amostras, escolhidas aleatoriamente. As amostras escolhidas são: A3C, A7C, A5\_, A2C, A2\_, A5C, A4\_, A6C, A3\_, A1C, A4C, A1\_, A6\_, A8\_, A8C.

Para a aplicação do KNN, utilizei  $K = 5$ , valor determinado por experimentações prévias que demonstraram ser a melhor escolha para o conjunto de dados em questão. Este valor de  $K$  equilibra a complexidade do

modelo e a precisão da classificação. Para definir o valor de K, realizei um processo de experimentação utilizando os valores 3 e 7, porém os resultados não foram tão satisfatórios quanto com K = 5.

A Tabela 1 detalha os resultados da aplicação do KNN na classificação das amostras de vinho. Os resultados são apresentados em termos de precisão (precision), revocação (recall) e a métrica f1-score, além do suporte (support) para cada classe.

	precision	recall	f1-score	support
A3C	0.53	0.76	0.62	113499
A7C	0.52	0.34	0.41	108279
A5_	0.41	0.55	0.47	106544
A2C	0.39	0.91	0.54	95162
A2_	0.98	0.86	0.92	129921
A5C	0.82	0.52	0.64	110698
A4_	0.89	0.58	0.70	137928
A6C	0.92	0.60	0.73	117031
A3_	0.48	0.76	0.59	114894
A1C	0.72	0.84	0.78	107525
A4C	0.77	0.66	0.71	115849
A1_	0.79	0.73	0.76	105200
A6_	0.52	0.49	0.50	102959
A8_	0.51	0.29	0.37	106351
A8C	0.63	0.47	0.53	135440
accuracy			0.62	1707280
macro avg	0.66	0.62	0.62	1707280
weighted avg	0.67	0.62	0.62	1707280

**Tabela 2:** Relatório de Classificação para Teste

O desempenho do KNN foi avaliado através de várias métricas de desempenho, incluindo precisão, recall, *F1-score* e *support*. Essas métricas

são essenciais para compreender a eficácia do modelo em categorizar corretamente as amostras de vinho [16, 17, 19].

Analisando os resultados podemos confirmar que a precisão varia entre as classes de vinho. A classe A2\_ apresentou a maior precisão com um valor de 0.98, indicando que o modelo foi altamente eficaz em prever corretamente esta classe específica. Em contraste, a classe A2C teve uma precisão de 0.39, sugerindo desafios na identificação correta dessa categoria.

A classe A2C apresentou um recall de 0.91, o mais alto entre todas as classes, indicando que quase todas as instâncias dessa classe foram corretamente identificadas pelo modelo. Por outro lado, a classe A7C teve um recall de 0.34, demonstrando uma dificuldade significativa do modelo em recuperar todas as instâncias dessa classe.

As classes com maior F1-score incluem A2\_ (0.92) e A1C (0.78), refletindo um bom equilíbrio entre precisão e recall para estas categorias. Classes como A7C e A8\_ tiveram F1-scores de 0.41 e 0.37, respectivamente, indicando que há espaço para melhorias na classificação dessas categorias.

Analisando o *support* é possível confirmar que as classes variaram em termos de número de instâncias, com a classe A4\_ tendo o maior *support* de 137928, enquanto a classe A2C teve o menor *support* de 95162.

A acurácia geral do modelo foi de 0.62, indicando que 62% das previsões feitas pelo modelo estavam corretas. A *macro avg* também foi 0.62 para precisão, *recall* e *F1-score*, destacando que, em média, o desempenho do modelo foi consistente. A *weighted avg* das métricas foi ligeiramente superior, com um valor de 0.67 para precisão, refletindo a influência do *support* no cálculo dessas médias.

A Figura 42 ilustra a performance do KNN no conjunto de treinamento. Cada linha da matriz representa as instâncias reais de uma classe específica, enquanto cada coluna representa as instâncias previstas pelo modelo para essas classes.

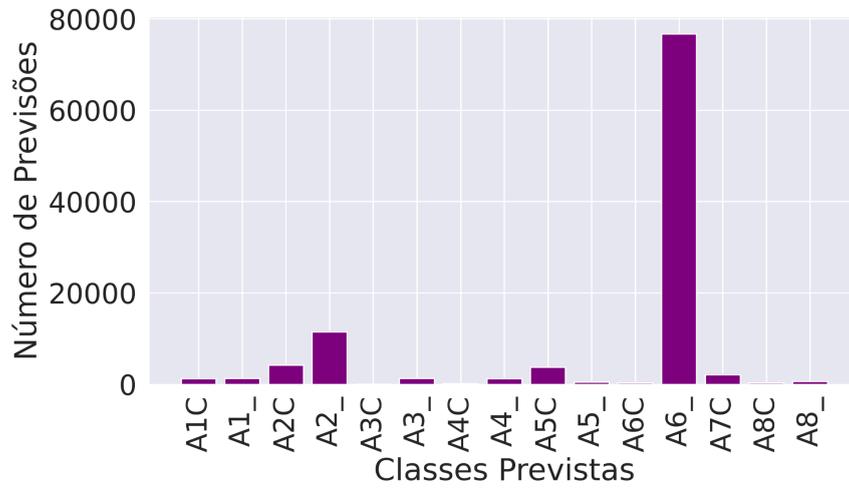


**Figura 42:** Matriz de confusão. Fonte: O autor.

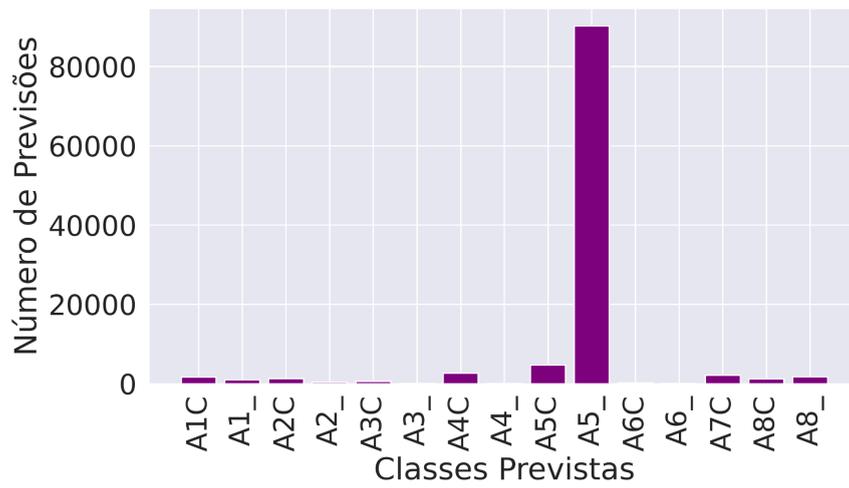
Fazendo uma análise geral é possível perceber que a classe A1C teve um total de 69,911 instâncias reais, das quais todas foram corretamente classificadas pelo modelo. No entanto, houve confusões significativas, como 8,690 instâncias classificadas como A1\_, 4,740 instâncias classificadas como A3\_, 6,030 instâncias classificadas como A4\_ e 6,530 instâncias classificadas como A8\_. A classe A2C, com um total de 95,162 instâncias reais, teve 86,131 instâncias corretamente classificadas, destacando-se por ter menos confusões, sendo 7,416 instâncias classificadas como A2\_ e 1,110 instâncias classificadas como A5\_. Já a classe A5\_, que teve 106,544 instâncias reais, apresentou 90,195 instâncias corretamente classificadas, com confusões notáveis onde 4,690 instâncias foram classificadas como A5C e 3,435 instâncias como A2C.

Classes como A1C e A2C demonstraram alta acurácia, com uma classificação quase perfeita para A1C e poucas confusões significativas para A2C. Por outro lado, classes como A3\_ e A4\_ tiveram um desempenho moderado, com 58,005 e 79,139 instâncias corretamente classificadas, respectivamente, mas com confusões dispersas entre várias outras classes. A classe A8\_ teve 63,162 instâncias corretamente classificadas, mas apresentou muitas confusões com A8C e A6\_, indicando uma maior dificuldade do modelo em distinguir essas classes. Da mesma forma, a classe A7C, com 50,253 instâncias corretamente classificadas, teve confusões significativas com A7\_ e A8\_.

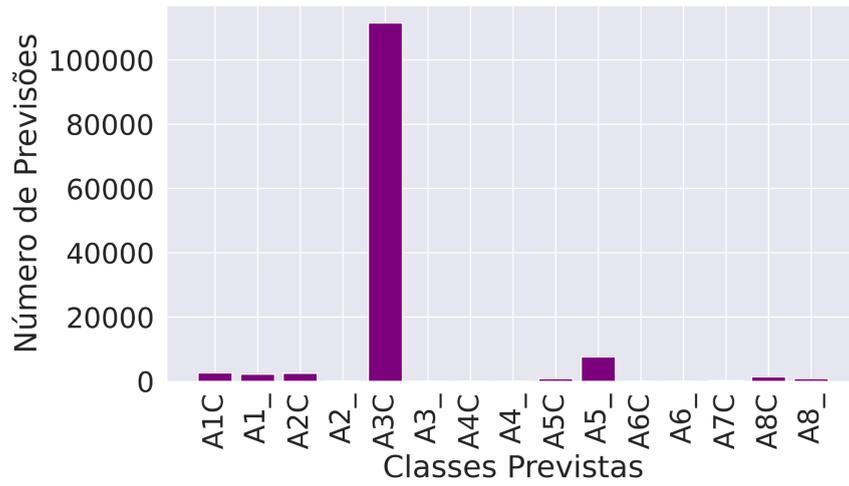
Outro desafio encontrado foi criar a matriz de confusão devido a quantidade de amostras, a solução encontrada foi criar um gráfico de barras que representa de forma visual as previsões do modelo. As figuras 43 a 57 busca destacar de maneira clara e concisa como as instâncias de classes foram classificadas, evidenciando tanto as classificações corretas quanto as confusões com outras classes.



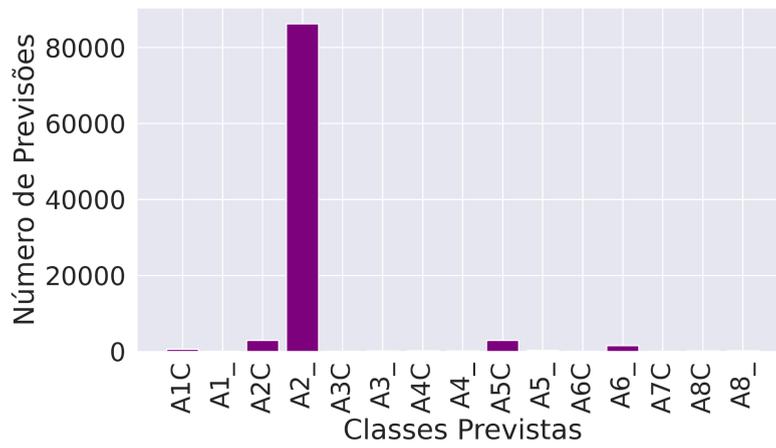
**Figura 43:** Amostra A6\_. Fonte: O autor.



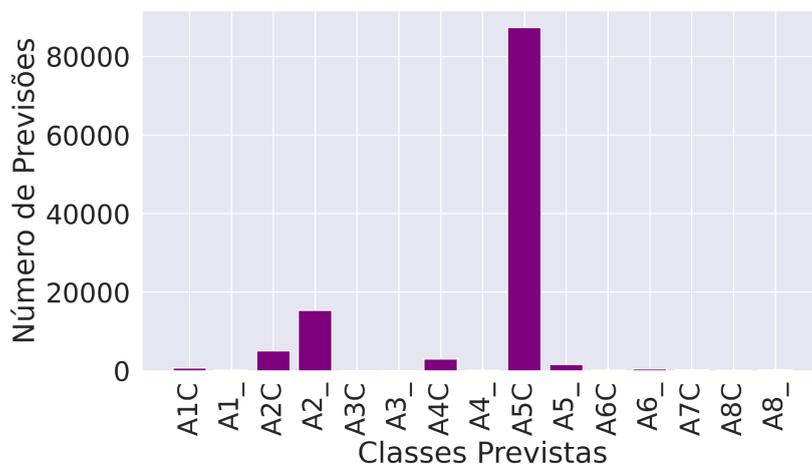
**Figura 44:** Amostra A5\_. Fonte: O autor.



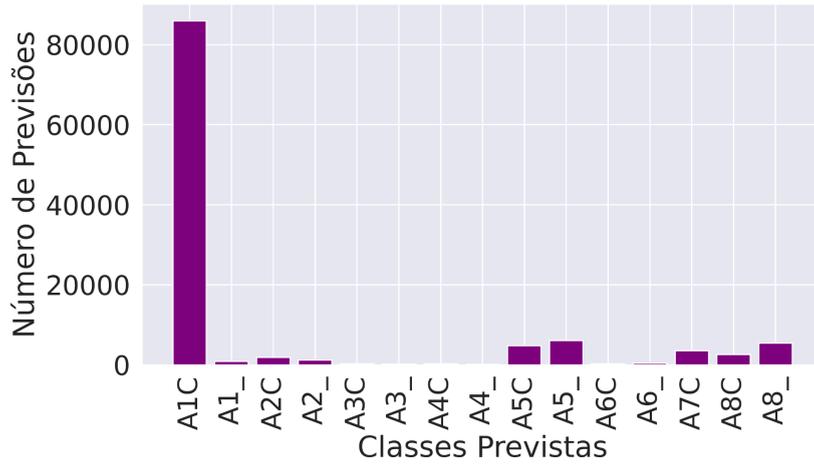
**Figura 45:** Amostra A3C. Fonte: O autor.



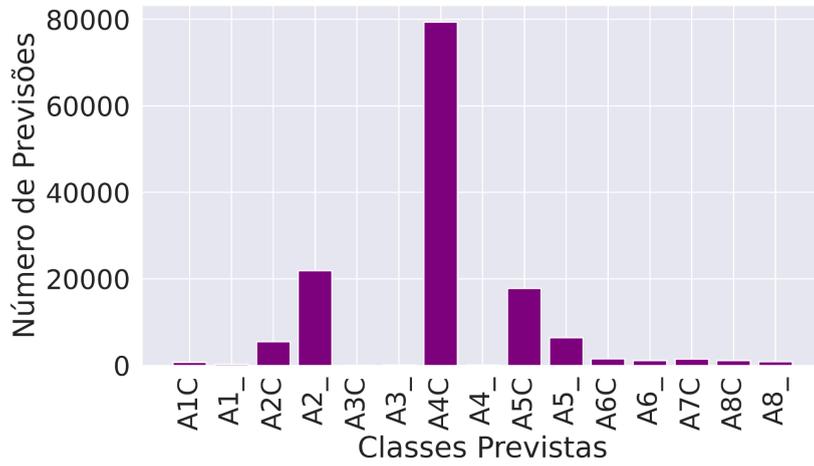
**Figura 46:** Amostra A2\_. Fonte: O autor.



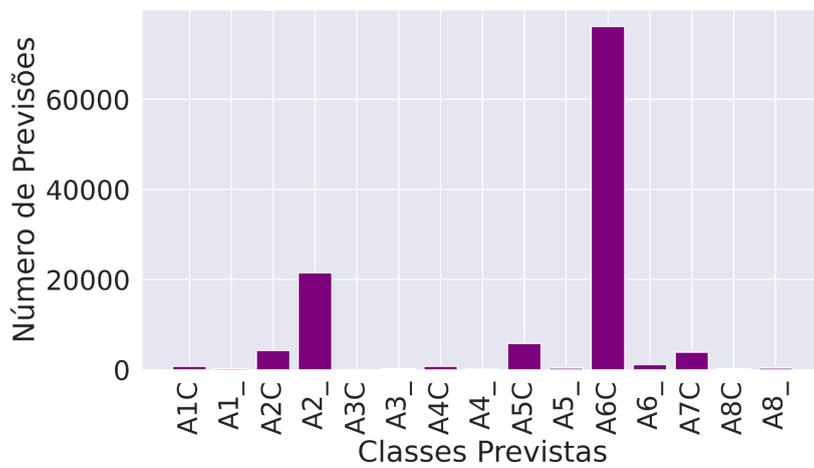
**Figura 47:** Amostra A5C. Fonte: O autor.



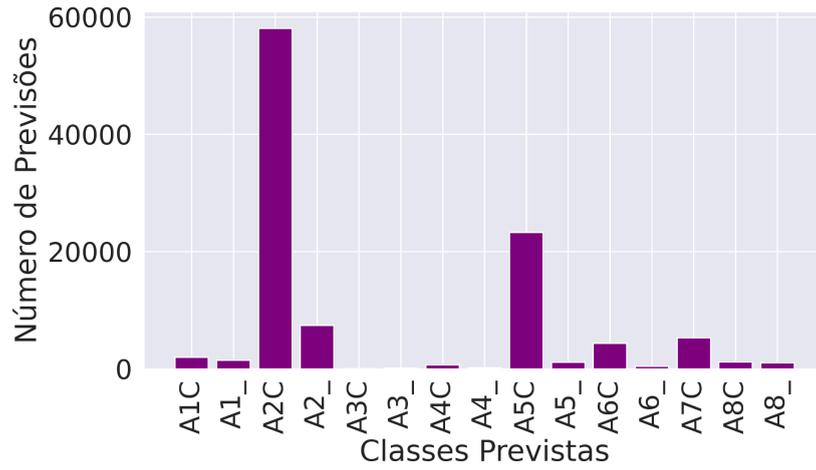
**Figura 48:** Amostra A1C. Fonte: O autor.



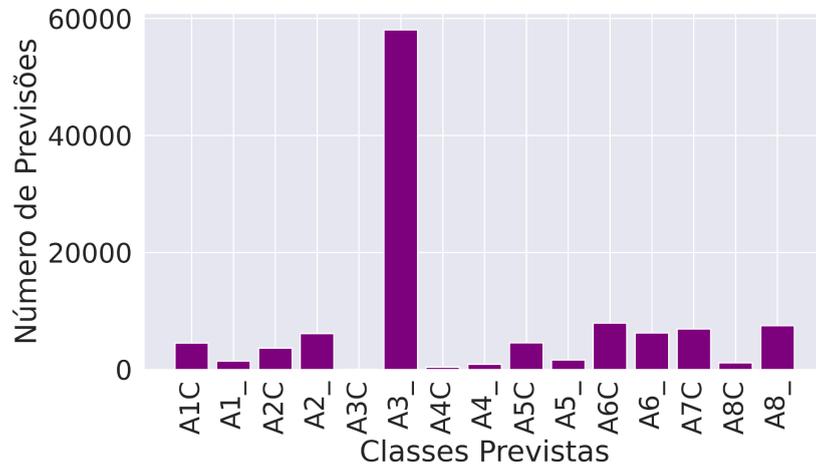
**Figura 49:** Amostra A4C. Fonte: O autor.



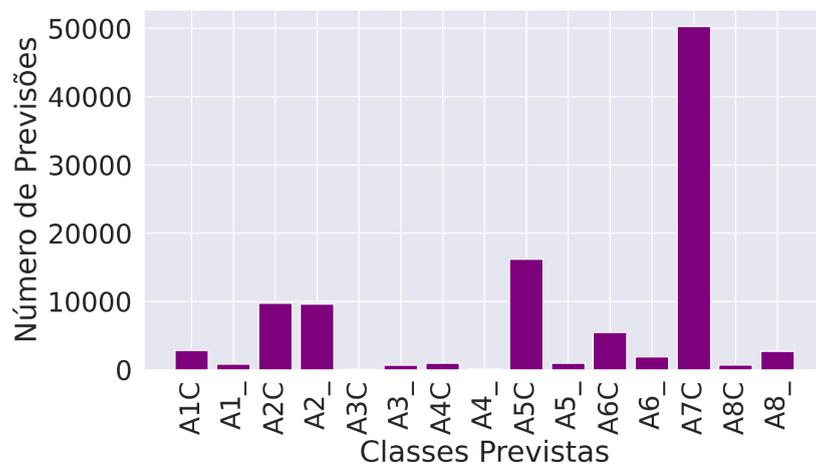
**Figura 50:** Amostra A6C. Fonte: O autor.



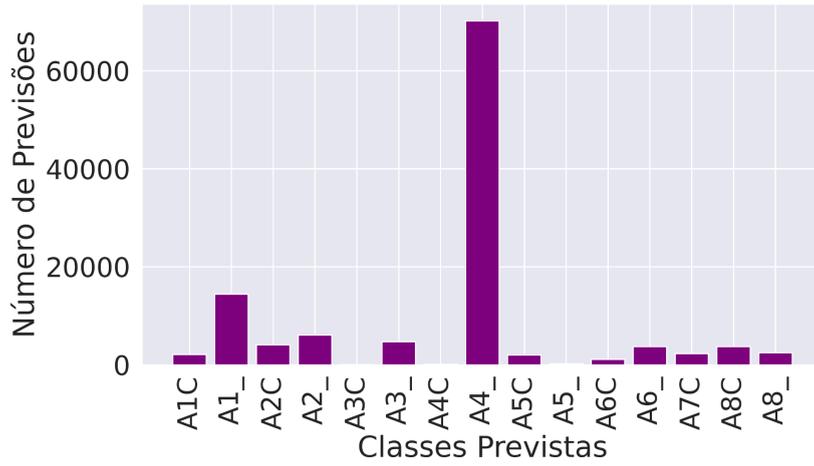
**Figura 51:** Amostra A2C. Fonte: O autor.



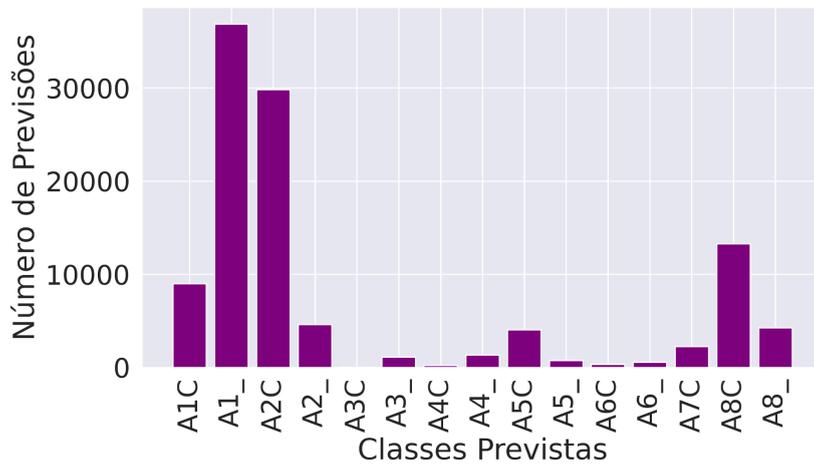
**Figura 52:** Amostra A3\_. Fonte: O autor.



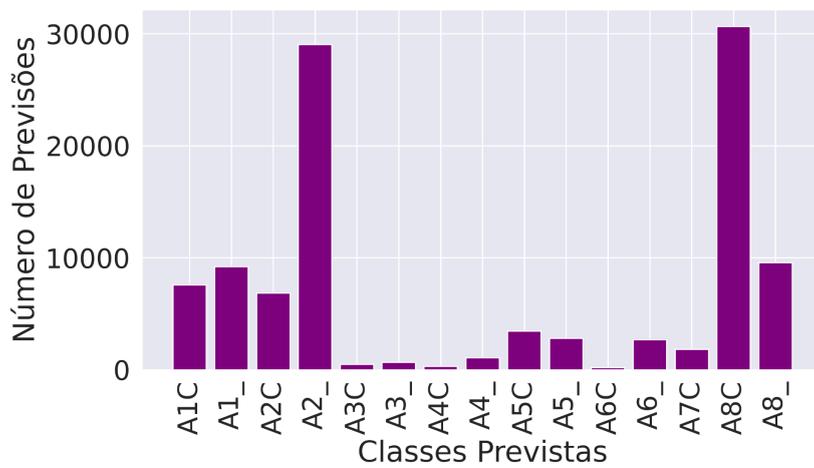
**Figura 53:** Amostra A7C. Fonte: O autor.



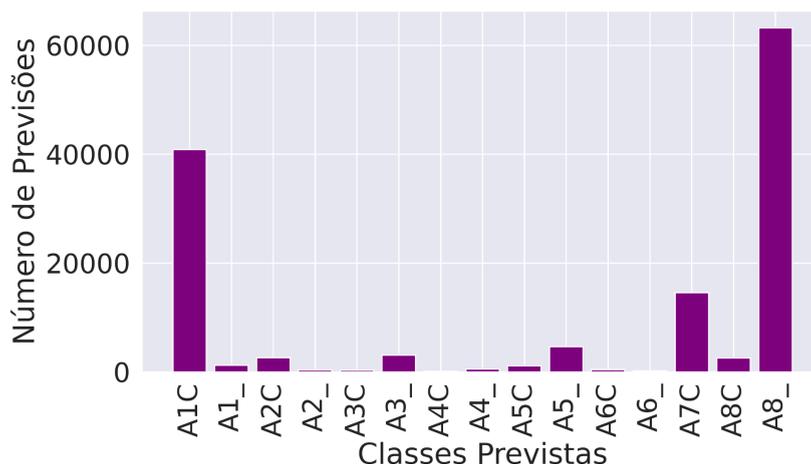
**Figura 54:** Amostra A4\_. Fonte: O autor.



**Figura 55:** Amostra A1\_. Fonte: O autor.



**Figura 56:** Amostra A8C. Fonte: O autor.



**Figura 57:** Amostra A8\_. Fonte: O autor.

## 4 CONCLUSÃO

Este trabalho apresenta um método para a classificação de vinhos utilizando HSI e o KNN, com o objetivo de aprimorar o controle de qualidade. A metodologia empregada demonstrou potencial para diferenciar com precisão as amostras de vinho com base em suas assinaturas espectrais.

Comparando os resultados obtidos com os de outros estudos mencionados, observa-se que o uso de HSI combinado com técnicas de aprendizado de máquina, como KNN, mostra-se eficaz em diversas áreas. Por exemplo, **Knauer et al. (2017)** utilizaram análise espacial-espectral de imagens hiperespectrais para melhorar a acurácia na classificação de níveis de infecção por míldio em uvas para vinho, destacando a capacidade da HSI de detectar variações sutis nos componentes químicos das amostras. Da mesma forma, **Dale et al. (2013)** revisaram várias aplicações da HSI na agricultura e controle de qualidade de produtos agroalimentares, ressaltando a versatilidade e precisão da técnica para diferentes tipos de análise.

Em comparação com outros métodos tradicionais de controle de qualidade, que muitas vezes são subjetivos e dependem da experiência dos avaliadores humanos (**Acsehrad, 1994**), a abordagem baseada em HSI e KNN oferece uma solução mais objetiva e confiável, mitigando a variabilidade causada por fatores subjetivos.

Para trabalhos futuros, seria interessante explorar outras técnicas de pré-processamento e outras técnicas de treinamento com o KNN, por exemplo, **MURPHY (2012)** indica usar 70% dos dados para treino e 30% para teste.

## 5. REFERÊNCIAS

- [1] MITCHELL, T. M. *Machine Learning*. New York: McGraw-Hill, 1997. p. 18.
- [2] ELMASRY, G.; SUN, D.-W. Principles of Hyperspectral Imaging Technology. In: *Hyperspectral Imaging for Food Quality Analysis*. Elsevier, 2010.
- [3] MURPHY, K. P. *Machine Learning: A Probabilistic Perspective*. Cambridge, Massachusetts: The MIT Press, 2012. p. 354-355.
- [4] DUDA, Richard O.; HART, Peter E.; STORK, David G. *Pattern Classification*. 2. ed. New York: Wiley-Interscience, 2001. p. 27.
- [5] ACSELRAD, Henri. Trabalho e controle de qualidade na indústria de alimentos. *Revista de Administração de Empresas*, São Paulo, v. 34, n. 2, p. 33-45, Mar./Abr. 1994.
- [6] ELMASRY, Gamal; SUN, Da-Wen. *Hyperspectral Imaging for Food Quality Analysis and Control*. 1. ed. Dublin: Elsevier Inc., 2010.
- [7] MIDDLETON SPECTRAL VISION. SisuCHEMA. Disponível em: <https://www.middletonspectral.com/product/sisuchema/>. Acesso em: 21 maio 2024.
- [8] THIMSEN, Elijah; SADTLER, Bryce; BEREZIN, Mikhail Y. Shortwave-infrared (SWIR) emitters for biological imaging: a review of challenges and opportunities. *Nanophotonics*, 2017.
- [9] CORTEX ROBOTICS. What is SWIR? Application & AOI. Disponível em: <https://cortexrobotics.my/what-swir-application-aoi/>. Acesso em: 20 maio 2024.
- [10] ADAMS, Ryan P. *K-Means Clustering and Related Algorithms*. 2014.
- [11] HAN, Jiawei; KAMBER, Micheline; PEI, Jian. *Data Mining: Concepts and Techniques*. 3. ed. San Francisco: Morgan Kaufmann, 2012.
- [12] BISHOP, Christopher. *Pattern Recognition and Machine Learning*. 1. ed. New York: Springer, 2006, p. 402-403.
- [13] HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. New York: Springer, 2009, p. 463-464.
- [14] MURPHY, Kevin P. *Machine Learning: A Probabilistic Perspective*. Cambridge, Massachusetts: MIT Press, 2012, p. 212-213.

- [15] SUYAL, Manish; GOYAL, Parul. A Review on Analysis of K-Nearest Neighbor Classification Machine Learning Algorithms based on Supervised Learning. *International Journal of Engineering Trends and Technology*, v. 70, n. 7, p. 43-48, 2022. Disponível em: <https://doi.org/10.14445/22315381/IJETT-V70I7P205>. Acesso em: 2 jun. 2024.
- [16] VISA, Sofia; RAMSAY, Brian; RALESCU, Anca; VAN DER KNAAP, Esther. Confusion Matrix-based Feature Selection. In: *Proceedings of the Twenty-second Midwest Artificial Intelligence and Cognitive Science Conference, MAICS 2011*. p. 120-123
- [17] WHALEN, Samuel et al. TargetFinder: assessment of predicted enhancer-promoter interactions using F1 score. *BMC Genomics*, v. 20, n. 1, p. 1-14, 2019. Disponível em: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-019-6413-7>. Acesso em: 15 maio. 2024.
- [18] VUJOVIC, Željko Đ. Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*, v. 12, n. 6, 2021.
- [19] Multi-class Model Evaluation with Confusion Matrix and Classification Report. Towards AI. Disponível em: <https://towardsai.net/p//multi-class-model-evaluation-with-confusion-matrix-and-classification-report>. Acesso em: 17 maio jun. 2024.
- [20] EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (EMBRAPA). Processo de elaboração. Disponível em: <https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/uva-para-processamento/pos-producao/processamento-da-uva/aspectos-agroindustriais/vinho/processo-de-elaboracao>. Acesso em: 18 maio. 2024.
- [21] Theodoros Markopoulos Dimitra Stougiannidou Stavros Kontakos and Christos Staboulis, Wine Quality Control Parameters and Effects of Regional Climate Variation on Sustainable Production, *Sustainability* 2023, 15, 3512. <https://doi.org/10.3390/su15043512>.
- [22] Knauer, U., Matros, A., Petrovic, T., Zanker, T., Scott, E. S., & Seiffert, U. (2017). Improved classification accuracy of powdery mildew infection levels of wine grapes by spatial-spectral analysis of hyperspectral images. *Plant Methods*, 13(47). DOI: 10.1186/s13007-017-0198-y
- [23] GADRE, Harish et al. Lyophilization: Process, Methods and Application. *European Journal of Pharmaceutical and Medical Research*, v. 7, n. 10, 2020.

- [24] GARCÍA-AGUILERA, Martha E. et al. Study of the Stability of Wine Samples for <sup>1</sup>H-NMR Metabolomic Profile Analysis through Chemometrics Methods. *Molecules*, v. 28, n. 16, p. 5962, 2023. Disponível em: <https://doi.org/10.3390/molecules28165962>.
- [25] LUECKEL, B.; BODMER, D.; HELK, B.; LEUENBERGER, H. Formulations of sugars with amino acids or mannitol: influence of concentration ratio on the properties of the freeze-concentrate and the lyophilizate. *Pharmaceutical Development and Technology*, v. 3, n. 2, p. 325-336, 1998.
- [26] CHRISTENSON, Sara. An introduction to lyophilization: process, benefits & possibilities. *Promega Connections*, 2024. Disponível em: <https://www.promegaconnections.com/an-introduction-to-lyophilization-process-benefits-possibilities>. Acesso em: 24 jun. 2024.
- [27] TRIMLEAF. Understanding lyophilization: what makes it special? *Trimleaf*, 2024. Disponível em: <https://www.trimleaf.com/blogs/news/understanding-lyophilization-what-makes-it-special>. Acesso em: 24 jun. 2024.
- [28] LYOPHILIZATION: process design, robustness, and risk management. *SpringerLink*, 2024. Disponível em: <https://link.springer.com/article/10.1007/s11095-009-9911-2>. Acesso em: 24 jun. 2024.
- [29] JAMEEL, F. Principles and Practices of Lyophilization in Product Development and Manufacturing. In: SWARBRICK, J.; BOYLAN, J. (eds.). *Encyclopedia of Pharmaceutical Technology*. New York: Springer, 2024. p. 1299-1326.
- [30] QUATERNARY. Theoretical Principles and Perspectives of Hyperspectral Imaging Applied to Sediment Core Analysis. *MDPI*, 2022. DOI: 10.3390/quat5020028.
- [31] CUI, Y.; JI, X.; XU, K.; WANG, L. A double-strategy-check active learning algorithm for hyperspectral image classification. *Photogrammetric Engineering & Remote Sensing*, v. 85, n. 11, p. 841-851, 2019. DOI: 10.14358/PERS.85.11.841.
- [32] OTEGBAYO, B.; MAZIYA-DIXON, B. A Review of the Use of Near-Infrared Hyperspectral Imaging (NIR-HSI) Techniques for the

Non-Destructive Quality Assessment of Root and Tuber Crops. *Applied Sciences*, v. 13, n. 9, p. 5226, 2023. DOI: 10.3390/app13095226.

[33] DIANA, M. Intraoperative Guidance Using Hyperspectral Imaging: A Review for Surgeons. *Diagnostics*, v. 11, n. 11, p. 2066, 2021. DOI: 10.3390/diagnostics11112066.

[34] GERLACH, R. W.; KOWALSKI, B. R.; WOLD, H. O. A. Partial least-squares path modeling with latent-variables. *Analytica Chimica Acta*, v. 3, n. 4, p. 417-421, 1979. DOI: 10.1016/S0003-2670(79)80024-3.

[35] DIANA, M.; HUIDOBRO, A.; PASTOR, A.; LOPEZ-CABALLERO, M. E.; TEJADA, M. Washing effect on the quality index method (QIM) developed for raw gilthead seabream (*Sparus aurata*). *European Food Research and Technology*, v. 212, n. 4, p. 408-412, 2001. DOI: 10.1007/s002170100285.