

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA DE CIÊNCIAS EXATAS E DA COMPUTAÇÃO
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



**UTILIZAÇÃO DE IMAGENS HIPERESPECTRAIS PARA CONTROLE DE
QUALIDADE DE AMOSTRAS DE *CURCUMA LONGA***

RAFAEL JUSTINO DE BRITO

GOIÂNIA
2020

RAFAEL JUSTINO DE BRITO

**UTILIZAÇÃO DE IMAGENS HIPERESPECTRAIS PARA CONTROLE DE
QUALIDADE DE AMOSTRAS DE *CURCUMA LONGA***

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

Orientador(a):

Prof. Dr. Arlindo Rodrigues Galvão Filho

Banca examinadora:

Prof. Dr. Rafael Viana de Carvalho

Prof. Dr. Clarimar José Coelho

Msc. Isaac Yves Lopes de Macêdo

GOIÂNIA
2020

RAFAEL JUSTINO DE BRITO

**UTILIZAÇÃO DE IMAGENS HIPERESPECTRAIS PARA CONTROLE DE
QUALIDADE DE AMOSTRAS DE *CURCUMA LONGA***

Trabalho de Conclusão de Curso aprovado em sua forma final pela Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, para obtenção do título de Bacharel em Ciência da Computação, em 03 de dezembro de 2020.

Prof.^a Ma. Ludmilla Reis Pinheiro dos Santos
Coordenadora de Trabalho de Conclusão de Curso

Banca examinadora:

Orientador: Prof. Dr. Arlindo Rodrigues Galvão Filho

Prof. Dr. Rafael Viana de Carvalho

Prof. Dr. Clarimar José Coelho

Msc. Isaac Yves Lopes de Macêdo

GOIÂNIA
2020

AGRADECIMENTOS

Ao Professor Arlindo Galvão, orientador acadêmico, pelas aulas durante o decorrer do curso, por todo o apoio prestado durante o desenvolvimento deste trabalho e pela confiança concedida a mim.

Ao Isaac Yves, coorientador, e ao Laboratório de Análise Farmacêutica e Ambiental (LAFAM) pela disponibilização dos materiais utilizados neste estudo e pelo apoio a este trabalho.

Ao Laboratório de Computação Científica (LCC) por disponibilizar os equipamentos que tornaram este trabalho possível.

À banca de TCC 1 e todos os meus colegas do grupo de estudos pelos conselhos e sugestões proferidas ao meu trabalho.

Aos meus amigos Airton Chagas e Walacy Xavier por terem trilhado essa jornada acadêmica comigo, pelo apoio, incentivo e por todos os momentos de descontração durante o curso.

À minha família que sempre me incentivou e apoiou em todos os momentos e decisões tomadas.

À Deus por ter permitido e corroborado para que este trabalho pudesse ser realizado, mesmo com todos os imprevistos e dificuldades enfrentadas durante este ano de 2020.

RESUMO

O uso de imagens hiperespectrais tem se mostrado promissor entre as técnicas utilizadas no controle de qualidade de produtos alimentícios. A indústria tende a utilizar métodos não destrutivos e mais rápidos para as análises dos padrões de seus produtos. Este estudo tem como objetivo explorar a utilização de imagens hiperespectrais no controle de qualidade de amostras de *Curcuma longa*. Antes da análise, foi realizado um pré-processamento das imagens hiperespectrais a fim de remover regiões ruidosas, outliers e pixels que não são referentes à região da amostra na imagem. Foram utilizadas amostras comerciais de Açafrão-da-terra e Amido misturadas nas proporções 2:8, 4:6, 6:4, 8:2 (Amido:Cúrcuma) para criar uma curva de calibração utilizando regressão linear múltipla. Os resultados apontaram um modelo de regressão com erro de raiz quadrático médio 6,42, erro médio absoluto de 5,95 e coeficiente de determinação de 0,94.

Palavras-Chave: Imagens Hiperespectrais (HSI), Controle de Qualidade, Curcuma longa, Análise de Dados Multivariada, Regressão Linear Múltipla (RLM)

ABSTRACT

The use of hyperspectral imaging has proved promising among the techniques used in the quality control of food products. The industry tends to use non-destructive and faster methods for the analysis of their product standards. This study aims to explore the use of hyperspectral imaging in the quality control of *Curcuma longa* samples. Prior to the analysis, a pre-processing of hyperspectral images was performed in order to remove noisy regions, outliers and pixels that do not refer to the sample region in the image. Commercial samples of turmeric and starch mixed in the proportions 2:8, 4:6, 6:4, 8:2 (Starch:Turmeric) were used to create a calibration curve using multiple linear regression. The results showed a regression model of root-mean-square error 6.42, mean absolute error 5.95 and coefficient of determination of 0.94.

Keywords: *Hyperspectral Imaging (HSI), Quality Control, Curcuma longa, Multivariate Data Analysis, Multiple Linear Regression (MLR)*

LISTA DE ILUSTRAÇÕES

Figura 1 - Ilustração do hipercubo de uma amostra de Açafrão-da-terra	12
Figura 2 - Representação do espectro de um pixel de uma amostra	13
Figura 3 - Flor de Curcuma longa	16
Figura 4 - Raiz do Açafrão-da-terra	16
Figura 5 - Amostras de Açafrão puro	17
Figura 6 - Amostra de Amido puro	17
Figura 7 - Amostras de Açafrão Puro A	18
Figura 8 - Autor Rafael Justino preparando uma amostra	19
Figura 9 - Batoque com amostra de Açafrão	19
Figura 10 - SisuCHEMA, câmera hiperespectral utilizada	20
Figura 11 - Ilustração do desdobramento do hipercubo	21
Figura 12 - Imagem da amostra A 40:60	22
Figura 13 - Amostras do tipo A após a remoção de fundo com K-means	24
Figura 14 - Espectros de cada pixel da amostra A Puro – filtro mediana	24
Figura 15 - Espectro médio das amostras A e de Amido - Savitzki-Golay	25
Figura 16 - Espectro médio das amostras A e Amido - SNV	26
Figura 17 - Gráficos da PCA apresentando (a)PC1xPC2 e (b)PC1xPC3	27
Figura 18 - Gráficos das predições para cada tipo de amostra	31

LISTA DE TABELAS

Tabela 1 - Quantidade de clusters atribuída como parâmetro K-means	23
Tabela 2 - Valores preditos e valores esperados para cada amostra	30

LISTA DE SIGLAS

ANVISA	Agência Nacional de Vigilância Sanitária
FDA	<i>Food and Drug Administration</i>
HSI	<i>Hyperspectral Imaging</i>
LAFAM	Laboratório de Análise Farmacêutica e Ambiental
LCC	Laboratório de Computação Científica
MAE	<i>Mean Absolute Error</i>
MLR	<i>Multiple Linear Regression</i>
PCA	<i>Principal Components Analysis</i>
PLS	<i>Partial Least Squares</i>
PUC-GO	Pontifícia Universidade Católica de Goiás
RMSE	<i>Root Mean Square Error</i>
SNV	<i>Standard Normal Variate</i>
SVM	<i>Support Vector Machine</i>
UFG	Universidade Federal de Goiás
USDA	<i>United Environmental Protection Agency</i>

SUMÁRIO

1.	INTRODUÇÃO	10
2.	FUNDAMENTAÇÃO TEÓRICA	12
2.1.	Imagens Hiperespectrais (HSI).....	12
2.2.	Regressão Linear Múltipla (RLM).....	13
3.	MATERIAIS E MÉTODOS	15
3.1.	Estudo de caso	15
3.2.	Preparo das amostras	17
3.3.	Aquisição de Imagens Hiperespectrais	20
3.4.	Pré-processamento	20
3.4.1.	<i>Remoção da região espectral ruidosa.....</i>	21
3.4.2.	<i>Remoção de fundo com K-means.....</i>	21
3.4.3.	<i>Remoção de outliers com filtro de mediana</i>	24
3.4.4.	<i>Filtro de suavização Savitzki-Golay.....</i>	25
3.4.5.	<i>Filtro de Variação Normal Padrão (SNV)</i>	26
3.4.6.	<i>Análise de Componentes Principais (PCA).....</i>	27
3.4.7.	<i>Métricas de avaliação dos modelos.....</i>	27
4.	RESULTADOS.....	29
5.	CONCLUSÃO	32
	REFERÊNCIAS BIBLIOGRÁFICAS	33

1. INTRODUÇÃO

Para garantir uma boa posição no mercado de alimentos devido às exigências das legislações e órgãos governamentais de saúde sobre a segurança e autenticidade dos produtos, a indústria alimentícia adota diversas metodologias e estratégias. Uma delas é o controle de qualidade, que faz uso de parâmetros químicos, físicos, microbiológicos, nutricionais e sensoriais para mensurar propriedades como a cor, o sabor, o aroma, a quantidade de proteínas e fibras, entre outros atributos dos alimentos (ORNELAS-SOTO; BARBOSA-GARCÍA; LOPEZ-DE-ALBA, 2011). A padronização das métricas de qualidade e de procedimentos realizados, que vão desde a linha de produção até a entrega e armazenamento dos produtos, é controlada por leis e órgãos fiscalizadores. Nos Estados Unidos, por exemplo, órgãos como o FDA (*Food and Drug Administration*) e o USDA (*United Environmental Protection Agency*) são responsáveis por regulamentar a composição e a qualidade dos alimentos (MEGHWAL; GOYAL, 2018). No Brasil, a ANVISA (Agência Nacional de Vigilância Sanitária) e o Ministério da Agricultura são exemplos de órgãos fiscalizadores (CONTROLARE, 2020).

Apesar de serem considerados precisos e eficazes, os métodos clássicos utilizados em análises de produtos alimentícios, como a cromatografia líquida (ORNELAS-SOTO; BARBOSA-GARCÍA; LOPEZ-DE-ALBA, 2011), são onerosos. Por esse e por outros motivos, há uma busca por técnicas que geram uma economia de tempo e de custo, como as que utilizam espectroscopia. Também há uma procura por novos métodos práticos que possam ser aplicados ao longo da linha de produção (ORNELAS-SOTO; BARBOSA-GARCÍA; LOPEZ-DE-ALBA, 2011).

Métodos que utilizam imagens digitais em conjunto com modelos computacionais surgiram como alternativas aos métodos clássicos de análise no controle de qualidade. Existem estudos que propõem, por exemplo, a detecção de adulteração de leite a partir de imagens adquiridas por smartphone (SILVA; ROCHA, 2020) e o monitoramento do processo de maturação de abacates através de imagens digitais e redes neurais artificiais (JARAMILLO-ACEVEDO et al., 2020); Essas técnicas se mostraram bastante promissoras quando utilizadas em análises relacionadas à aparência dos alimentos. Entretanto, em análises que visam mensurar

características moleculares das substâncias, as imagens hiperespectrais apresentaram resultados melhores, conforme apresenta Taghizadeh, Gowen e O'Donnell (2011) em seu estudo de comparação.

Além dos métodos alternativos utilizando imagens digitais, as técnicas com imagens hiperespectrais (*Hyperspectral Imaging, HSI*) têm sido utilizadas como mais uma ferramenta não destrutiva de análise de controle de qualidade. Como apresenta Sandasi et al. (2014), é possível distinguir espécies diferentes de *Echinacea* através da análise de imagens HSI de sua raiz e folhas. Outros exemplos de estudos realizados são: detecção precoce de imperfeições em maçãs (ELMARSY et al., 2008), identificação de infecções fúngicas em morangos (SIEDLISKA et al., 2018) e determinação da proporção de cada matéria prima utilizada em chá de ervas (SANDASI et al., 2018).

Um problema enfrentado mundialmente pela indústria alimentícia e o controle de qualidade, é a existência da produção de alimentos adulterados (ORNELAS-SOTO; BARBOSA-GARCÍA; LOPEZ-DE-ALBA, 2011). A adulteração pode ser realizada de diversas maneiras como a adição, remoção ou a substituição de compostos químicos. Uma revisão científica que reuniu vários estudos relacionados ao tema, apresentou uma lista dos produtos alimentícios que mais são adulterados no mundo (MOORE; SPINK; LIPP, 2012). Entre esses alimentos está o açafrão, um tipo de planta herbácea de raiz tuberosa que possui o amido como um de seus principais componentes (ALMEIDA, 2006; FILHO et al., 2000; BRAGA et al., 2005) e como seu principal contaminante, conforme apresenta Moore, Spink e Lipp (2012).

Dentre as espécies de açafrão existe a *Curcuma longa*, comumente conhecida como Açafrão-da-terra. Com sua adulteração, vários de seus benefícios proporcionados como a prevenção de inúmeras doenças como artrite, diabetes, colesterol e malária podem ser prejudicados (CECILIO FILHO et al., 2000).

Sob tais circunstâncias, este trabalho propõe um estudo exploratório usando análise multivariada de dados para determinar a concentração de amido em imagens hiperespectrais de amostras de Açafrão-da-terra.

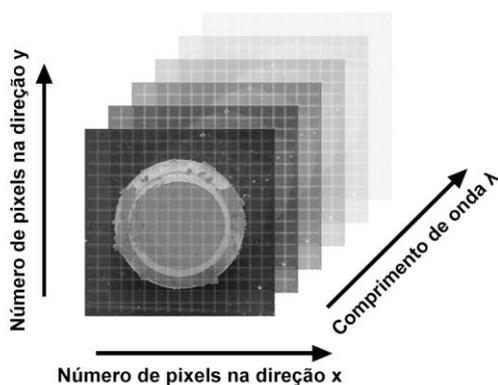
2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, o conceito de imagens hiperespectrais será apresentado, seguido de uma introdução à regressão linear múltipla.

2.1. Imagens Hiperespectrais (HSI)

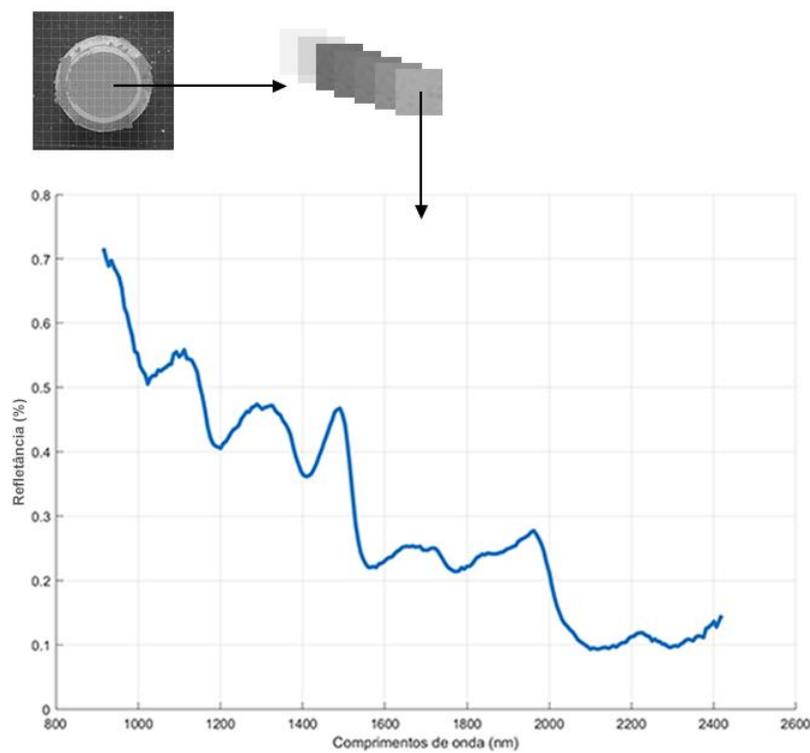
Uma imagem hiperespectral é formada por uma matriz de dados de três dimensões, sendo duas delas coordenadas espaciais que representam a largura (eixo x) e a altura (eixo y) da imagem. A dimensão da coordenada espectral é representada por λ e seus valores são obtidos através de técnicas como espectroscopias em várias regiões do espectro eletromagnético. Cada pixel da imagem contém um espectro de valores de refletância obtidos em comprimentos de ondas diferentes (GOWEN et al., 2007). A Figura 1 é uma representação da matriz de dados de uma imagem hiperespectral, denominada hipercubo e a Figura 2 apresenta o espectro de um pixel de uma amostra de Açafreão-da-terra.

Figura 1 - Ilustração do hipercubo de uma amostra de Açafreão-da-terra



Fonte: Elaborado pelo autor, 2020.

Figura 2 - Representação do espectro de um pixel de uma amostra de Açafão-da terra



Fonte: Elaborado pelo autor, 2020.

2.2. Regressão Linear Múltipla (RLM)

Dentre os algoritmos frequentemente utilizados em problemas de determinação envolvendo imagens HSI, a Regressão Linear Múltipla (*Multiple Linear Regression*, MLR) e a Regressão Parcial de Mínimos Quadrados (*Partial Least Squares Regression*, PLSR) estão presentes em diversos estudos envolvendo controle de qualidade de produtos alimentícios (PAN et al., 2016). A MLR tem como objetivo estabelecer uma equação que relaciona uma variável dependente à múltiplas variáveis independentes (BROWN, 2009). Aplicada a esse trabalho, a regressão visa relacionar a concentração de amido aos valores de refletância obtidos em cada comprimento de onda.

O modelo estatístico para uma regressão linear múltipla é definido por:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i \quad i = 1, 2, \dots, n. \quad (1)$$

Onde y_i é a variável dependente ou explicada; x_{ik} as variáveis independentes ou explicativas; β_k são os coeficientes de regressão e e_i são os erros aleatórios.

Primeiramente, os coeficientes de regressão são gerados. Em seguida, utiliza-se a matriz dos dados a serem testados para realizar uma soma de multiplicações de cada uma de suas variáveis pelo coeficiente de regressão correspondente e, ao final do processo, um valor de predição é obtido.

Algebricamente, o vetor de coeficientes de regressão pode ser expresso através da seguinte notação:

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (2)$$

De maneira que X^T é a matriz de dados transposta; X a matriz de dados original e Y é a matriz unidimensional dos valores observados.

3. MATERIAIS E MÉTODOS

3.1. Estudo de caso

Um dos alimentos mais adulterados no mundo, conforme apresentado anteriormente, é o açafrão (MOORE; SPINK; LIPP, 2012), um tipo de planta herbácea com raiz tuberosa. Dentre suas espécies, existe a *Curcuma longa*, comumente conhecida como Açafrão-da-terra ou Tumérico. Possui origem na Índia e no Sudeste da Ásia e é pertencente à mesma família do gengibre (*Zingiberaceae*). Além de ser bastante utilizada como condimento, o Açafrão-da-terra também é utilizado em áreas da indústria têxtil, em cosméticos, alimentos e em usos medicinais devido possuir substâncias antimicrobianas, antioxidantes, corantes, entre outras (FILHO et al., 2000). A Figura 3 apresenta a flor do Açafrão-da-terra. Sua raiz é comercializada em forma de pó após um processo de secagem e moagem. A Figura 4 apresenta a raiz do Açafrão-da-terra.

Sobre a composição centesimal da *Curcuma longa*, estudos indicam que os polissacarídeos, mais precisamente o amido, é considerado o maior componente dentre os demais (ALMEIDA, 2006; FILHO et al., 2000; BRAGA et al., 2005). Segundo o estudo, o amido compõe cerca de 25 a 50% do açafrão; as proteínas de 10 a 25%; os lipídeos de 6 a 12% e cinzas e fibras de 8 a 11%. Essa variação pode acontecer devido a forma que o açafrão foi cultivado, as práticas agrícolas adotadas, o local de plantio, a forma de utilização de fertilizantes e a forma que foi processada para sua comercialização. A quantidade de amido presente no açafrão pode estar diretamente ligada ao grau de maturação da planta pois, como o amido é a reserva energética das plantas, ele é consumido por meio da atividade metabólica que ocorre durante a maturação do açafrão. (BRAGA et al., 2003; LEONEL; CEREDA, 2002). Existe um outro estudo que estipula que a cada 100g de pó de açafrão, 30g são de amido (BRASIL, 2015).

Figura 3 - Flor de Curcuma longa



Fonte: H. Zell, 2009.

Figura 4 - Raiz do Açafrão-da-terra (à esquerda) e seu pó após processo de secagem e moagem (à direita)



Fonte: Simon A. Eugster, 2014.

Para alcançar os objetivos deste trabalho, três etapas precisaram ser realizadas para que os dados pudessem ser analisados: o preparo das amostras; a aquisição das imagens e o pré-processamento delas utilizando técnicas e algoritmos relatados como promissores em estudos envolvendo HSI (RINNAN; BERG; ENGELSEN, 2009).

3.2. Preparo das amostras

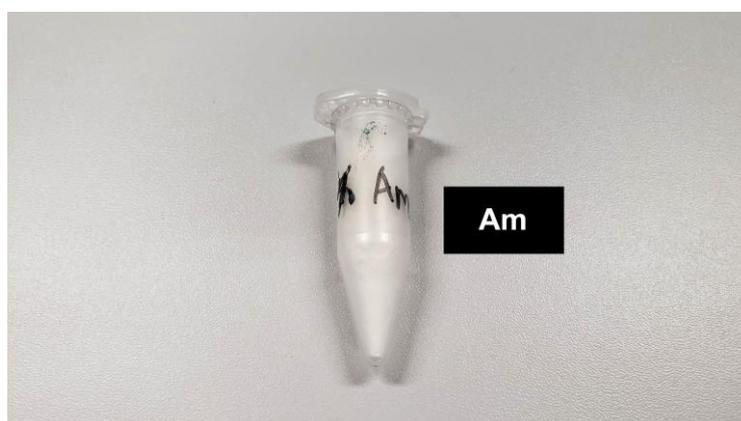
Para o desenvolvimento deste estudo, amostras de Açafração-da-terra e de Amido foram utilizadas. Elas foram gentilmente cedidas pelo Laboratório de Análise Farmacêutica e Ambiental (LAFAM) da Universidade Federal de Goiás (UFG). São 10 tipos de açafração (Figura 5) e um tipo de Amido Puro (Figura 6). A amostra de Amido Puro está rotulada como Am e as amostras de cada tipo de açafração estão rotuladas como A, B, C, D, E, F, G, H, I e J. As amostras de açafração com cerca de 35% de Amido presente de forma natural, estão rotuladas neste trabalho como “Açafração-da-terra Puro”. Ao total, 51 amostras foram utilizadas.

Figura 5 - Amostras A Puro (considerando cerca de 35% de Amido presente de forma natural)



Fonte: Elaborado pelo autor, 2020.

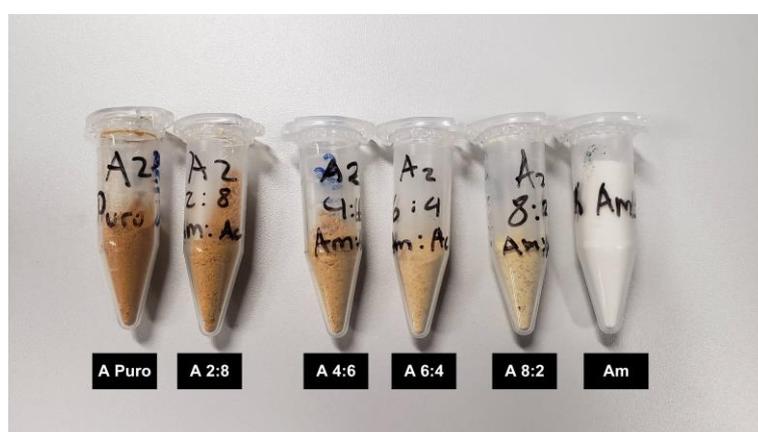
Figura 6 - Amostra de Amido puro



Fonte: Elaborado pelo autor, 2020.

Para cada tipo de açafrão, há 1 amostra de Açafrão-da-terra Puro e 4 amostras que consistem em misturas daquele tipo de açafrão com acréscimos de Amido nas proporções 2:8, 4:6, 6:4, 8:2 / Amido:Cúrcuma. Isso significa que, na amostra A 2:8, existe uma adição de 20% de Amido Puro para 80% de Açafrão-da-terra Puro. Em outras palavras, se a amostra for dividida em 100 partes, 80 serão de açafrão e 20 serão de Amido Puro.

Figura 7 - Amostras de A Puro, suas frações de misturas (da esquerda para a direita 2:8, 4:6, 6:4, 8:2) e a amostra de Amido Puro



Fonte: Elaborado pelo autor, 2020.

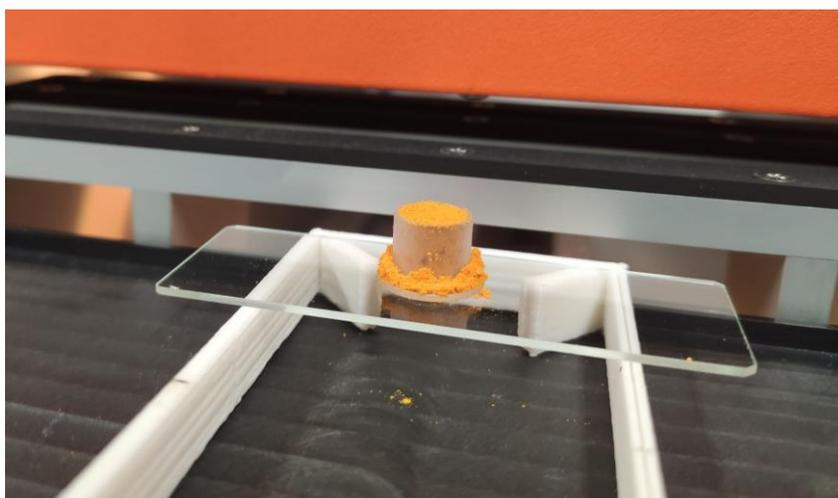
Com o auxílio de pinças e espátulas de metal, uma porção de cada amostra foi colocada em recipientes chamados batoques de tamanho M. Esse recipiente possui formato cilíndrico e tem 1.3cm de diâmetro por 1.1cm de altura. Cada batoque foi preenchido até a sua borda de forma com que a superfície ficasse o mais plana possível (Figura 8). Cada batoque foi utilizado exclusivamente por uma amostra com o objetivo de evitar que resquícios dela fossem misturados com outra. A Figura 9 apresenta como esse arranjo foi montado.

Figura 8 - Autor Rafael Justino preparando uma amostra com o auxílio de uma pinça e uma espátula de metal



Fonte: Elaborado pelo autor, 2020.

Figura 9 - Batoque com amostra de Açafião-da-terra posicionado em cima de uma lâmina de microscopia



Fonte: Elaborado pelo autor, 2020.

3.3. Aquisição de Imagens Hiperespectrais

A câmera hiperespectral utilizada foi o modelo SisuCHEMA da marca Specim (Specim, *Spectral Imaging Ltd.*, Oulu, Finlândia), a qual apresenta a Figura 10. Ela foi cedida para este estudo por meio do Laboratório de Computação Científica (LCC) da Pontifícia Universidade Católica de Goiás (PUC-GO). Essa câmera foi configurada para capturar imagens dentro do intervalo de comprimento de onda de 915nm e 2419nm, obtendo 255 bandas diferentes. Cada imagem HSI ficou com largura de 320 pixels e altura média de 946 pixels.

Figura 10 - SisuCHEMA, câmera hiperespectral utilizada para a aquisição das imagens das amostras



Fonte: Elaborado pelo autor, 2020.

3.4. Pré-processamento

As imagens HSI foram submetidas à uma sequência de etapas para remover ou amenizar itens indesejáveis como ruídos ou pixels que formam o plano de fundo da imagem.

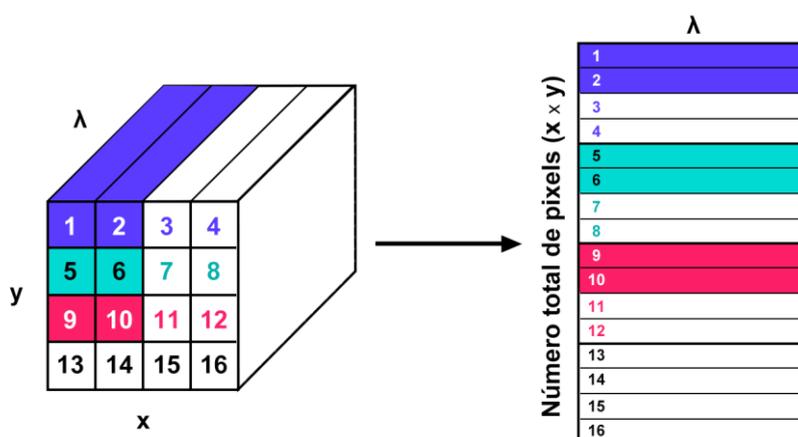
3.4.1. Remoção da região espectral ruidosa

Dos 255 comprimentos de onda das HSI originais, apenas os primeiros 240 foram utilizados devido ao alto nível de ruído dos demais. De acordo com estudos, é recomendável realizar este procedimento quando há uma região com bastante ruído na qual é inviável tentar corrigi-los. (VIDAL; AMIGO, 2012).

3.4.2. Remoção de fundo com K-means

O algoritmo de *K-means* tem se mostrado bastante promissor na clusterização de dados de imagens hiperespectrais (RANJAN et al., 2017; (REDDY; SARASCHANDRIKA; V. REDDY, 2020). Neste trabalho, ele foi utilizado para agrupar ou clusterizar os pixels referentes a região da amostra na imagem, permitindo a sua extração para ser utilizada na análise. Para isso, foi necessário realizar um desdobramento do hipercubo de cada amostra em uma matriz bidimensional. Esse desdobramento é ilustrado através da Figura 11. Nela, o exemplo utilizado foi um hipercubo de dimensões $4 \times 4 \times \lambda$, tendo como resultado uma matriz bidimensional de $16 \times \lambda$ após o desdobramento.

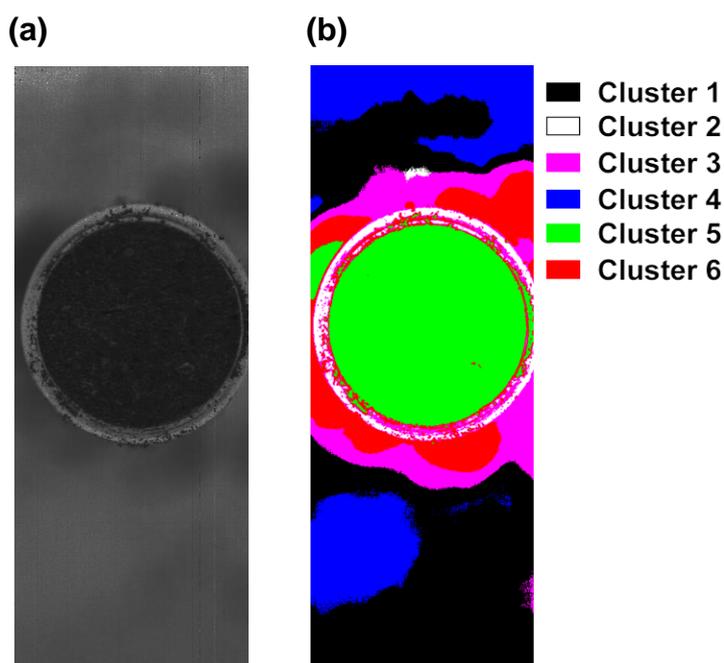
Figura 11 - Ilustração do desdobramento do hipercubo (esquerda) para uma matriz bidimensional (direita)



A partir das matrizes bidimensionais, o algoritmo de *K-means* foi aplicado individualmente em cada uma delas. Como resultado, obteve-se uma matriz de mesma largura e altura da imagem hiperespectral, de forma que, em cada *pixel* dessa matriz, foi atribuído um número referente ao cluster que aquele *pixel* passou a pertencer. Na amostra A 40:60, por exemplo, o número de *clusters*, que é um parâmetro do algoritmo, foi definido de forma empírica como 6.

Para obter uma noção visual deste processo de clusterização, foi atribuída uma cor distinta para cada *cluster*. A Figura 12 apresenta uma imagem em escala de cinza da amostra A 40:60 com cada *pixel* colorido da cor referente ao *cluster* ao qual foi classificado. É possível identificar que o agrupamento que contém os *pixels* referentes à região da amostra está na cor verde nessa figura.

Figura 12 - Imagem da amostra A 40:60 (a) em escala de cinza e (b) clusterizada



Fonte: Elaborado pelo autor, 2020.

A quantidade de *clusters* definida para cada amostra é apresentada na Tabela 1. Esses valores foram definidos a partir de testes de forma empírica, buscando o melhor valor que fizesse com que os *pixels* referentes à região da amostra ficassem, em sua maioria, separados em um *cluster*.

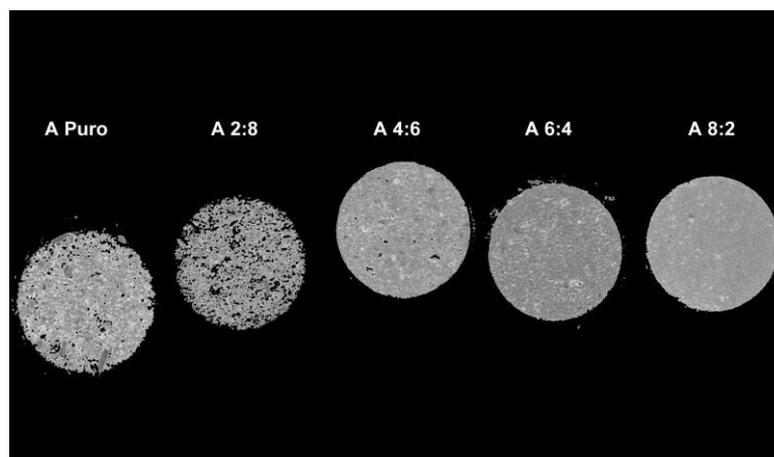
Tabela 1 - Quantidade de clusters atribuída como parâmetro do algoritmo de K-means para cada amostra.

Amostra	N° de Clusters	Amostra	N° de Clusters
A Puro	11	F Puro	13
A 2:8	9	F 2:8	10
A 4:6	10	F 4:6	11
A 6:4	6	F 6:4	10
A 8:2	8	F 8:2	9
Am	12	G Puro	7
B Puro	7	G 2:8	6
B 2:8	18	G 4:6	6
B 4:6	8	G 6:4	10
B 6:4	8	G 8:2	6
B 8:2	12	H Puro	8
C Puro	8	H 2:8	6
C 2:8	8	H 4:6	8
C 4:6	6	H 6:4	8
C 6:4	15	H 8:2	8
C 8:2	6	I Puro	7
D Puro	8	I 2:8	6
D 2:8	8	I 4:6	8
D 4:6	8	I 6:4	6
D 6:4	8	I 8:2	8
D 8:2	10	J Puro	7
E Puro	9	J 2:8	6
E 2:8	8	J 4:6	6
E 4:6	11	J 6:4	8
E 6:4	10	J 8:2	8
E 8:2	11		

Fonte: Elaborado pelo autor, 2020.

Com a matriz de *clusters*, foi possível extrair da imagem hiperespectral somente os pixels correspondentes ao cluster que contém a região da amostra. Dessa forma, os demais pixels não foram utilizados nas etapas seguintes do pré-processamento. A Figura 13 apresenta imagens em escala de cinza das amostras do tipo A. Após esse processo, foi realizada uma normalização dos valores de cada pixel e, em seguida, foi atribuído o valor zero para os pixels correspondentes à região do fundo da imagem.

Figura 13 - Amostras do tipo A após a remoção de fundo com K-means

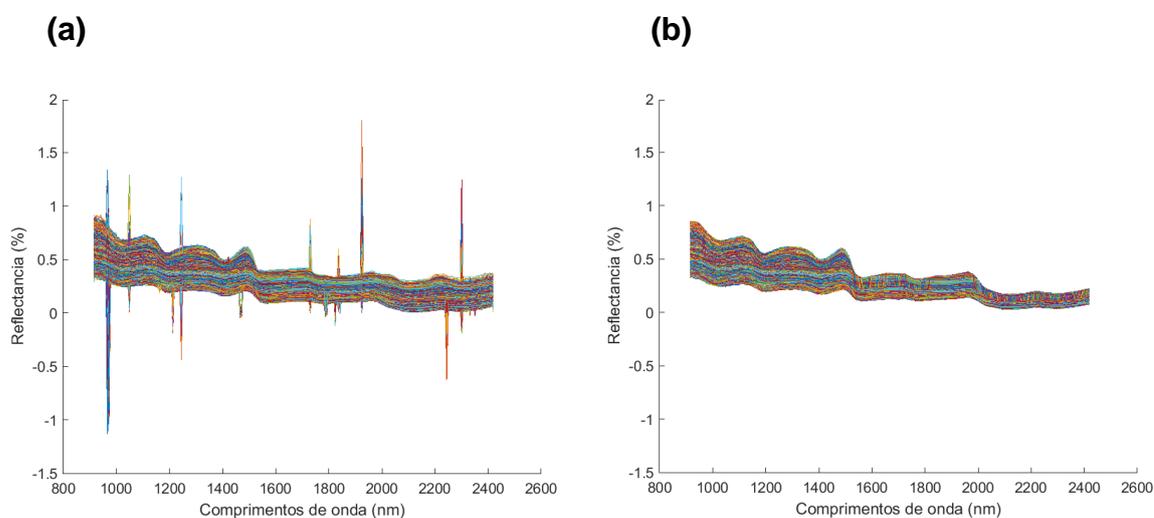


Fonte: Elaborado pelo autor, 2020.

3.4.3. Remoção de outliers com filtro de mediana

A remoção de *outliers* como *spikes* e *dead pixels* foi realizada utilizando o filtro de mediana. Para cada pixel da HSI, o filtro calcula a mediana de todos os valores de refletância dos espectros e substitui os valores os *outliers* pelo valor da mediana. O filtro considera um valor como *outlier* quando este é três vezes maior que a mediana. A Figura 14 apresenta um gráfico com os valores do espectro de cada pixel da amostra A Puro antes e depois da remoção dos *outliers*.

Figura 14 - Espectros de cada pixel da amostra A Puro (a) com outliers e após a aplicação do filtro de mediana



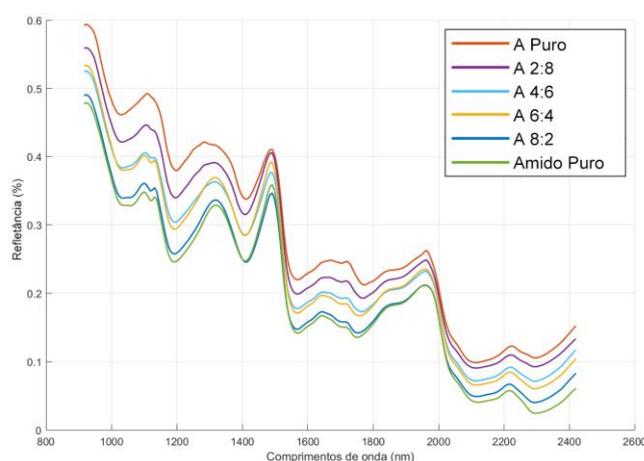
Fonte: Elaborado pelo autor, 2020.

3.4.4. Filtro de suavização Savitzki-Golay

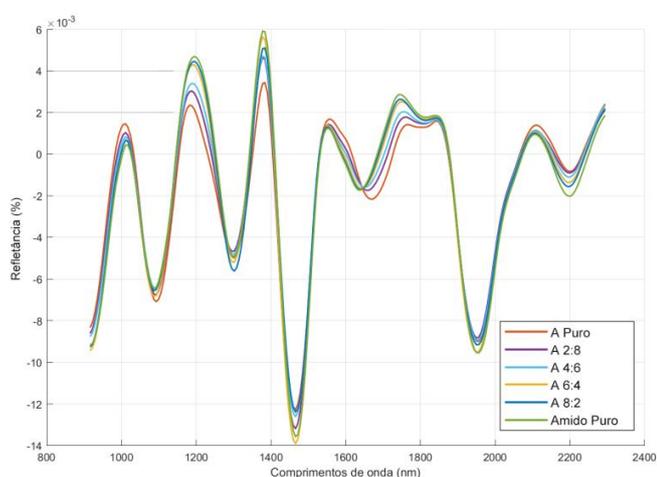
Dentre as técnicas mais citadas em artigos científicos relacionados ao pré-processamento de imagens hiperespectrais (LUO et al., 2005), o filtro derivativo de suavização Savitzki-Golay foi utilizado com o objetivo de enfatizar características espectrais dos dados. Os parâmetros utilizados foram: tamanho da janela de 21 pontos, 2ª ordem polinomial e derivada de 1ª ordem. Estes valores foram definidos seguindo o estudo apresentado por Galvão Filho, Galvão e Araújo (2011). A Figura 15 apresenta o espectro médio dos pixels das amostras do A e do Amido antes e depois da aplicação do filtro.

Figura 15 - Espectro médio das amostras A e de Amido (a) antes e depois (b) da filtragem de Savitzki-Golay

(a)



(b)



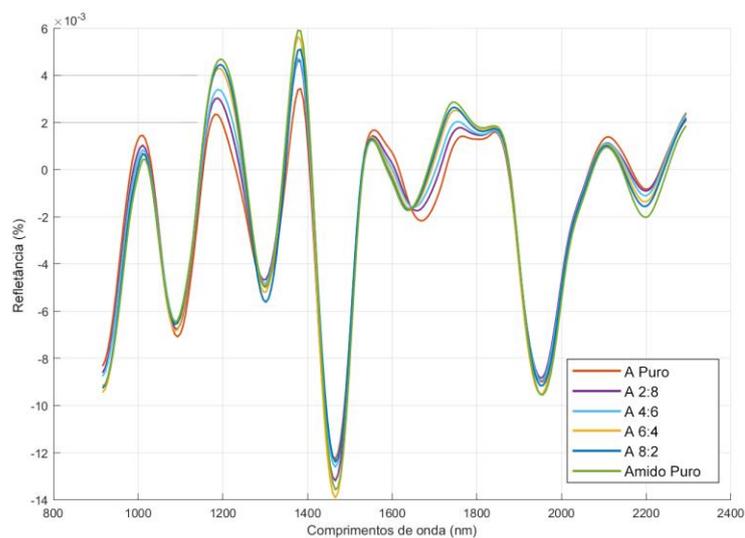
Fonte: Elaborado pelo autor, 2020.

3.4.5. Filtro de Variação Normal Padrão (SNV)

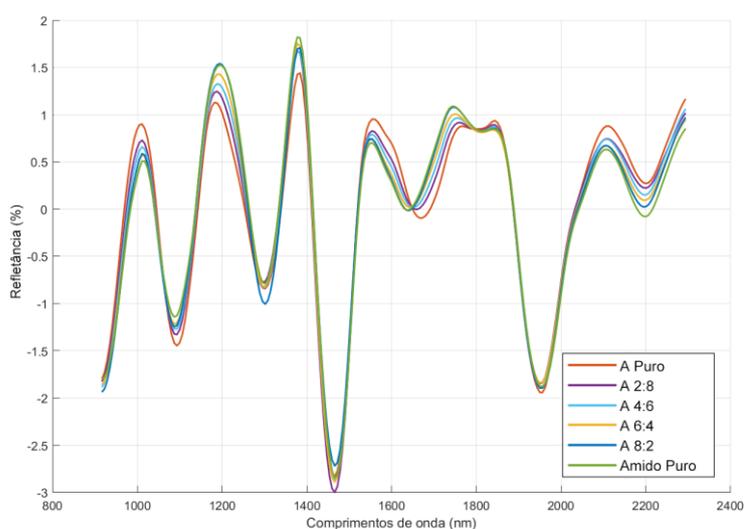
Comumente utilizado com o filtro de Savitzki-Golay, a variação normal padrão (*Standard Normal Variate*, SNV) foi aplicada para auxiliar na redução da variabilidade entre as amostras devido à dispersão (RINNAN; BERG; ENGELSEN, 2009). A Figura 16 apresenta o espectro médio dos pixels das amostras do A e do Amido antes e depois da aplicação do filtro.

Figura 16 - Espectro médio das amostras A e Amido (a) antes e após a aplicação do SNV

(a)



(b)

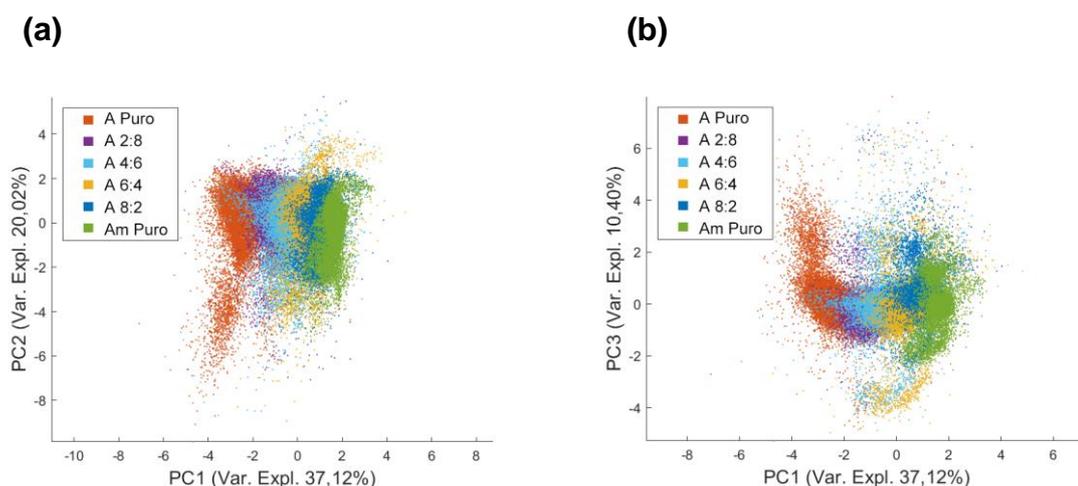


Fonte: Elaborado pelo autor, 2020.

3.4.6. Análise de Componentes Principais (PCA)

Uma análise exploratória utilizando a PCA (*Principal Components Analysis*) foi realizada. A Figura 17 apresenta os gráficos dos scores da PC1, PC2 e PC3, as quais possuem uma variância explicada de, respectivamente, 37,12%, 20,02% e 10,40% totalizando 67,54%. Os dados referentes à amostra A Puro (que possui cerca de 30% de amido naturalmente) estão agrupados mais à esquerda, enquanto os referentes à amostra de amido estão mais à direita. Dessa forma, é possível perceber que os dados possuem uma certa tendência a se espalharem da esquerda para a direita, conforme a adição de amido aumenta. Assim, é bem provável que a concentração de amido seja o principal motivo dessa separação gradual entre os dados das PCs.

Figura 17 - Gráficos da PCA apresentando (a)PC1xPC2 e (b)PC1xPC3



Fonte: Elaborado pelo autor, 2020.

3.4.7. Métricas de avaliação dos modelos

Para este estudo que faz uso de um modelo de regressão, foram utilizadas métricas que visam obter a raiz quadrada da média das diferenças entre as previsões e as observações reais, como o RMSE (*Root Mean Square Error*); o erro médio

absoluto (*Mean Absolute Error*, MAE) e o coeficiente de determinação (R^2). No RMSE, N é a quantidade valores totais, \hat{y}_i são os valores preditos e y_i os observados. Quanto mais próximo de zero o RMSE for, melhor é o modelo. O MAE e o R^2 também utilizam N para o total de valores, \hat{y}_i para os valores preditos e y_i para os observados. Entretanto, diferente do RMSE e MAE, quanto mais próximo de um o R^2 for, melhor será o modelo.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (3)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (5)$$

4. RESULTADOS

Para a obtenção dos resultados, foram realizadas duas etapas: testes das amostras do tipo A e testes das amostras do tipo B, C, D, E, F, G, H, I e J. Para a primeira, foi feita uma regressão utilizando 70% dos pixels totais (escolhidos sequencialmente) das amostras A Puro, A 2:8, A 4:6, A 6:4, A 8:2 e Am. Em seguida, as mesmas amostras, exceto a Am, foram testadas utilizando os 30% de pixels restantes. Nesse teste, foram obtidos 6,42 de RMSE, 5,95 de MAE e um R^2 de 0,94. O menor desvio padrão obtido foi 0,74 na amostra A Puro.

Na segunda etapa, foi realizada uma nova regressão, porém, utilizando todos os pixels das amostras A Puro, A 2:8, A 4:6, A 6:4, A 8:2 e Am. Nessa fase, as amostras B, C, D, E, F, G, H, I e J foram testadas, onde o melhor R^2 obtido foi -0,10 na amostra H; menor RMSE foi 10,76 na amostra J e menor MAE foi 8,53 também na amostra J.

A Tabela 2 apresenta uma comparação entre os valores observados (esperados) e os preditos para cada amostra, assim como também o RMSE, MAE e R^2 para cada tipo. A apresentação dos valores de desvio padrão foi inclusa para seguir o mesmo padrão de tabela apresentado no trabalho de Macêdo et. Al. (2020). A Figura 18 mostra gráficos de cada um dos testes. O eixo horizontal de cada gráfico comporta os valores preditos de concentração de amido, o eixo vertical os valores esperados e cada pixel foi representado como um sinal de (+).

Tabela 2 - Valores preditos e valores esperados para cada amostra

Tipo	Esperado					Predito					Acurácia		
	X-P ± DP ^a	X-2 ± DP ^b	X-4 ± DP ^b	X-6 ± DP ^b	X-8 ± DP ^b	X-P ± DP ^c	X-2 ± DP ^d	X-4 ± DP ^d	X-6 ± DP ^d	X-8 ± DP ^d	RMSE	MAE	R ²
A ^e	29,82 ± 0,56	43,86 ± 0,14	66,32 ± 0,12	86,53 ± 0,08	97,31 ± 0,07	26,94 ± 0,74	34,44 ± 2,71	58,33 ± 3,09	81,51 ± 2,68	92,88 ± 1,47	6,42	5,95	0,94
B	66.05 ± 0.36	72.84 ± 0.12	83.7 ± 0.14	93.48 ± 0.09	98.7 ± 0.05	88.84 ± 5.39	89.66 ± 4.14	92.15 ± 5.15	94.42 ± 4.71	96.66 ± 4.06	13.26	10.21	-0.82
C	33.99 ± 2.41	47.19 ± 0.27	68.31 ± 0.13	87.33 ± 0.07	97.47 ± 0.11	67.7 ± 6.55	73.34 ± 6.79	79.18 ± 7.55	87.16 ± 6.02	93.47 ± 7.17	19.77	14.98	-0.47
D	36.9 ± 0.27	49.52 ± 0.11	69.71 ± 0.1	87.88 ± 0.07	97.58 ± 0.02	63.81 ± 6.49	70.69 ± 5.97	79.48 ± 6.36	85.63 ± 6.32	94.63 ± 4.84	16.01	12.61	-0.12
E	22.14 ± 0.44	37.71 ± 0.14	62.63 ± 0.05	85.05 ± 0.1	97.01 ± 0.03	60.98 ± 4.16	70.71 ± 5.99	78.05 ± 4.29	86.72 ± 4.15	93.37 ± 4.15	23.88	18.51	-0.35
F	27.56 ± 0.08	42.05 ± 0.1	65.23 ± 0.07	86.09 ± 0.02	97.22 ± 0.04	75.12 ± 3.79	80.23 ± 5.34	85.78 ± 4.5	90.25 ± 4.52	94.67 ± 5.34	28.86	22.6	-0.62
G	39.3 ± 0.61	51.44 ± 0.12	70.87 ± 0.1	88.35 ± 0.03	97.7 ± 0.05	81.95 ± 7.83	84.46 ± 8.29	90.12 ± 7.07	93.06 ± 5.82	96.55 ± 6.3	25.7	20.15	-0.58
H	24.21 ± 1.25	39.37 ± 0.21	63.62 ± 0.11	85.45 ± 0.04	97.09 ± 0.03	56.73 ± 6.11	63.68 ± 9.77	75.89 ± 8.69	85.37 ± 8.18	91.78 ± 8.52	19.12	14.9	-0.10
I	53.98 ± 1.67	63.19 ± 0.23	77.91 ± 0.15	91.16 ± 0.07	98.23 ± 0.09	78.86 ± 6.66	83.06 ± 8.01	88.91 ± 6.81	92.39 ± 7.17	97.14 ± 5.8	15.08	11.61	-0.36
J	70.32 ± 1.11	76.26 ± 0.22	85.75 ± 0.18	94.3 ± 0.08	98.86 ± 0.08	88.18 ± 8.66	90.39 ± 8.95	93.2 ± 6.74	96.14 ± 7.13	97.49 ± 5.25	10.76	8.53	-0.53

Fonte: Elaborado pelo autor, 2020.

^a Amido por 100 gramas de pó de açafrão sem mistura ± desvio padrão.

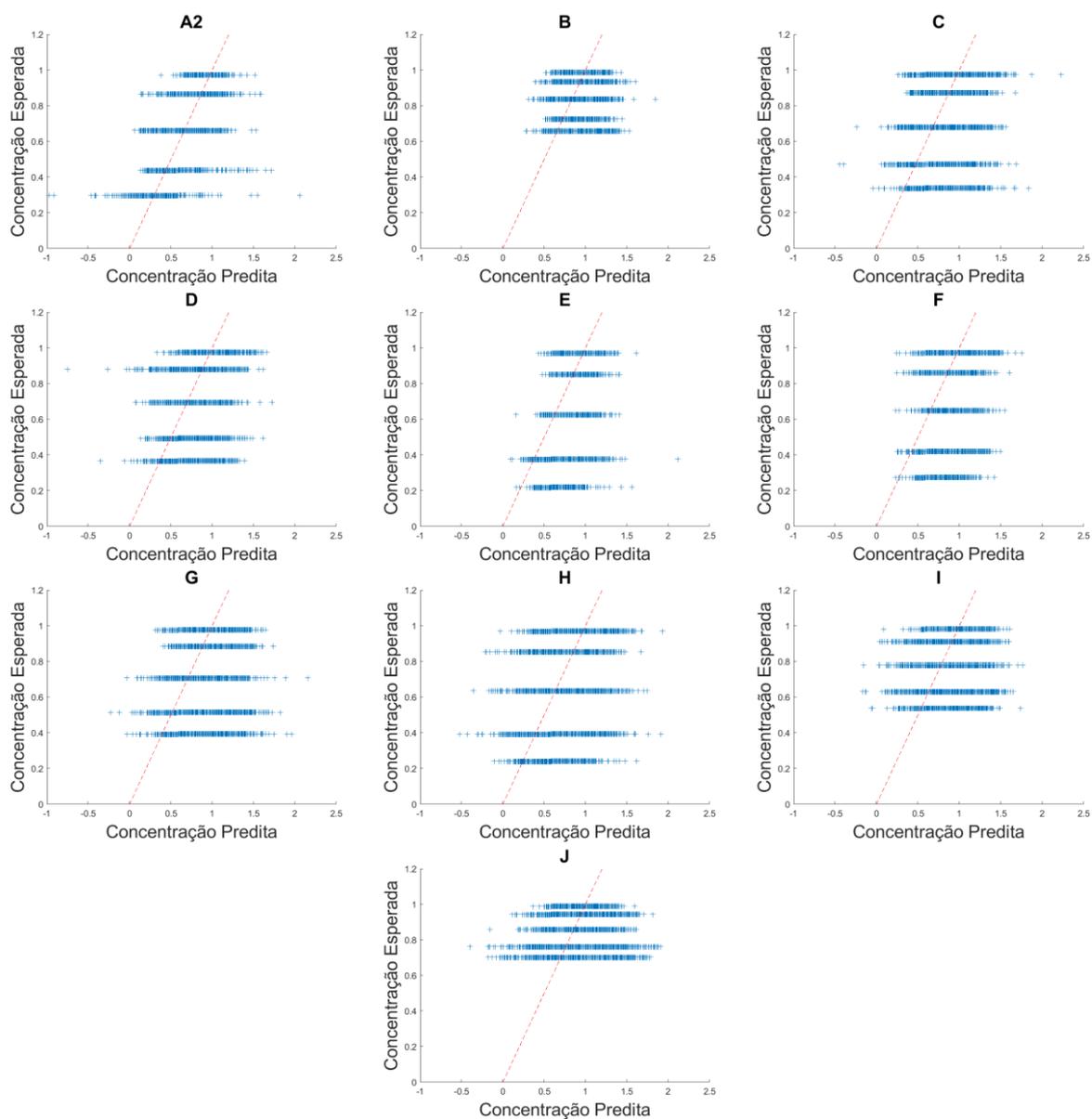
^b Mistura de X-2 até X-8 correspondente a 2:8 até 8:2 amido/cúrcuma, respectivamente.

^c Média do resultado da predição da concentração de amido para amostra de cúrcuma sem mistura ± desvio padrão.

^d Média do resultado da predição da concentração de amido para amostra de açafrão com mistura de X-2 até X-8 correspondente a 2:8 até 8:2 amido/açafrão, respectivamente ± desvio padrão.

^e Teste realizado a partir de uma regressão de 70% dos pixels das amostras, sendo 30% dos pixels totais utilizados no teste.

Figura 18 - Gráficos das previsões para cada tipo de amostra



Fonte: Elaborado pelo autor, 2020.

Nos gráficos, é possível observar que houve uma variação muito grande entre os valores preditos para cada pixel de cada amostra. Considerando os resultados obtidos através dos testes, é possível concluir que, ao utilizar as amostras A como referência, obteve-se bons resultados apenas para as próprias amostras A, visto que o R^2 dos demais tipos de Açafão-da-terra foram todos negativos.

5. CONCLUSÃO

Este estudo foi realizado objetivando à exploração do uso de imagens hiperespectrais para determinar a quantidade de amido em amostras de Açafração-da-terra. A abordagem utilizada compreendeu métodos recomendados para o pré-tratamento dos dados de imagens hiperespectrais. O resultado desta etapa pôde ser observado na análise de componentes principais, a qual apresentou uma relação entre a adição de amido e os valores de refletâncias em cada comprimento de onda. A utilização da regressão linear múltipla se mostrou bastante promissora quando testada com as amostras A, que foram utilizadas como referência. Já nos testes com os demais tipos de Açafração-da-terra, os resultados não foram satisfatórios, gerando um valor máximo de R^2 de -0,10. Os testes com as amostras do tipo B foi onde ocorreu o pior caso, com um R^2 de -0,82.

Como sugestão para trabalhos futuros, a utilização de algoritmos de seleção de características (*feature selection*) deve ser considerada visto que, diversos estudos relatam que na maioria dos casos não é necessário utilizar todos os comprimentos de onda das imagens HSI para se obter bons resultados. Além disso, o seu uso é recomendável quando se trata problemas utilizando abordagens de regressão linear. Com a utilização de um bom algoritmo, é possível obter apenas as bandas que mais diferenciam uma amostra da outra. Uma outra sugestão é o uso de outros algoritmos de regressão linear como o PLSR que também tem se mostrado bastante promissor em estudos envolvendo HSI ou até mesmo algoritmos não lineares como o SVM (*Support Vector Machine*). Um passo mais adiante, seria o uso de aprendizado de máquina (*Machine Learning*), tendo em vista que seu uso tem se tornado bastante popular quando devido o potencial de identificar padrões em uma quantidade enorme de dados multivariáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, L. P. de. **Caracterização de pigmentos da *Curcuma longa* L., Avaliação da atividade antimicrobiana, morfogênese in vitro na produção de curcuminóides e óleos essenciais.** 2006. 120 p. Tese (Ciência de Alimentos) — Universidade Federal de Minas Gerais. Disponível em: <<https://repositorio.ufmg.br/handle/1843/MBSA-6X4M39>>. Acesso em: 08/11/2020.

BRAGA, M. E. M. **Obtenção de compostos bioativos de *Curcuma longa* L. e *Lippia alba* M. por tecnologia supercrítica:** rendimento global, cinética de extração, composição química e aproveitamento do resíduo amiláceo. 2005. 200 p. Tese (Engenharia de Alimentos) — Universidade Estadual de Campinas. Disponível em: <<http://repositorio.unicamp.br/jspui/handle/REPOSIP/254897>>. Acesso em: 10/11/2020.

BRASIL. **MONOGRAFIA DA ESPÉCIE *Curcuma longa* L. (CURCUMA).** [S.l.]: Ministério da Saúde, 2015.

BROWN, S. H. Multiple Linear Regression Analysis: A Matrix Approach with MATLAB. **Alabama Journal of Mathematics**, p. 1 – 3, 2009.

CECILIO FILHO, A. B. et al. *Cúrcuma*: planta medicinal, condimentar e de outros usos potenciais. **Ciência Natural**, SciELO, Santa Maria, v. 30, n. 1, p. 171 – 175, Março 2000. ISSN 0103-8478.

CONTROLARE. **Vigilância Sanitária é um entre os mais de 10 órgãos reguladores.** 2020. Disponível em: <<https://www.controlare.com.br/blog/seguranca-alimentar/vigilancia-sanitaria-e-um-entre-os-mais-de-10-orgaos-reguladores-de-restaurantes/#:~:text=A%20ANVISA%2C%20Ag%C3%Aancia%20Nacional%20de,a%20fiscaliza%C3%A7%C3%A3o%20da%20ind%C3%BAstria%20aliment%C3%ADcia>>. Acesso em: 02/11/2020.

ELMARSY, G. et al. Early detection of apple bruises on different background colors using hyperspectral imaging. **LWT - Food Science and Technology**, Elsevier, v. 41, p. 337 – 345, Março 2008.

GOWEN, A. et al. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. **Trends in Food Science & Technology**, Elsevier, Dublin, v. 18, n. 0924-2244, p. 590 – 598, Dezembro 2007.

HUCK, C. W. Novel Analytical Tools for Quality Control in Food Science. In: AKYAR, I. (Ed.). **Latest Research into Quality Control.** 1. ed. In- tech, 2012. v. 1, cap. 5, p. 63 – 90. ISBN 978-953-307-682-9. Disponível em:

<<https://www.intechopen.com/books/latest-research-into-quality-control/novel-analytical-tools-for-quality-control-in-food-science>>. Acesso em: 10/11/2020.

HURST, W. C. et al. **Quality Control: A Model Program for the Food Industry**. 2010.

MEGHWAL, M.; GOYAL, M. R. Methods for Food Analysis and Quality Control. In: GOYAL, M. M. M. R. (ed.). **State-of-the-Art Technologies in Food Science: Human health, emerging issues and specialty topics**. 1. ed. [S.l.]: Apple Academic Press, 2018. v. 1, cap. 10, p. 299 – 320. ISBN 9781771886161.

JARAMILLO-ACEVEDO, C. A. et al. Hass avocado ripeness classification by mobile devices using digital image processing and ANN methods. **International Journal of Food Engineering**, De Gruyter, v. 1, n. 2019-0161, p. 1 – 8, Setembro 2020.

LUO, J.; YING, K.; BAI, J. Savitzky–Golay smoothing and differentiation filter for even number data. **Signal Processing** **85**, Elsevier, Beijing, v. 7, p. 1429 – 1434, Julho 2005.

MOORE, J. C.; SPINK, J.; LIPP, M. Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. **Journal of Food Science**, Journal of Food Science, v. 77, n. 4, p. 118 – 126, 01 2012.

ORNELAS-SOTO, N.; BARBOSA-GARCÍA, O.; LOPEZ-DE-ALBA, P. L. Procedures of Food Quality Control: Analysis Methods, Sampling and Sample Pretreatment. In: SHOYAMA, P. Y. (Ed.). **Quality Control of Herbal Medicines and Related Areas**. 1. ed. Rijeka: InTech, 2011. v. 1, cap. 5, p. 63 – 90. ISBN 978-953-307-682-9. Disponível em: <https://www.intechopen.com/books/quality-control-of-herbal-medicines-and-related-areas/procedures-of-food-quality-control-analysis-methods-sampling-and-sample-pretreatment>.

PAN, T. et al. Regression Algorithms in Hyperspectral Data Analysis for Meat Quality Detection and Evaluation. **Comprehensive Reviews in Food Science and Food Safety**, Institute of Food Technologists, Guangzhou, v. 15, p. 529 – 541, Março 2016.

RANJAN, S. et al. Hyperspectral Image Classification: A k-means Clustering Based Approach. In: TECHNOLOGY, N. I. of (Ed.). **4th International Conference on Advanced Computing and Communication Systems (ICACCS)**. Rourkela, 2017. p. 1 – 7.

REDDY, A. C.; SARASCHANDRIKA, A.; V. REDDY, A. Study of the Clustering Algorithms for Hyper Spectral Remote Sensing Images. **Journal of Hyperspectral Remote Sensing**, Kavali, v. 10, n. 2, p. 117 – 121, October 2020. ISSN 2237-2202. Disponível em: <<https://periodicos.ufpe.br/revistas/jhrs/article/view/246329>>. Acesso em: 11/11/2020.

- RINNAN, A.; BERG, F. van den; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **Trends in Analytical Chemistry**, Elsevier, Frederiksberg C, v. 28, p. 1201 – 1222, November 2009.
- SANDASI, M. et al. Hyperspectral Imaging and Chemometric Modeling of Echinacea — A Novel Approach in the Quality Control of Herbal Medicines. **Molecules**, Pretoria, v. 19, n. 1420-3049, p. 13104 – 13121, Setembro 2014.
- SANDASI, M. et al. Non-destructive quality assessment of herbal tea blends using hyperspectral imaging. **Pytochemistry Letters**, Elsevier, Pretoria, v. 24, p. 94 – 101, Abril 2018.
- SHAFIEE, M. N.; DRABU, S. Assessment of Quality Control Systems in Food Processing Units in Khunmoh Food Park Kashmir, India. **Indian Horticulture Journal**, Indian Society of Advanced Horticulture, Awantipora, v. 1, n. 403-16-IHJ-2911-16, p. 79 – 84, Fevereiro 2019.
- SIEDLISKA, A. et al. Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. **Postharvest Biology and Technology**, Elsevier, Lublin, v. 139, p. 115 – 156, Maio 2018.
- SILVA, A. F. S.; ROCHA, F. R. A novel approach to detect milk adulteration based on the determination of protein content by smartphone-based digital image colorimetry. **Food Control**, Elsevier, Piracicaba, p. 1 – 6, 04 2020. ISSN 0956-7135.
- SUKTANARAK, S.; TEERACHAICHAYUT, S. Non-destructive quality assessment of hens' eggs using hyperspectral images. **Journal of Food Engineering**, Elsevier, Bangkok, v. 215, p. 97 – 103, Dezembro 2017.
- TAGHIZADEH, M.; GOWEN, A. A.; O'DONNELL, C. P. Comparison of hyperspectral imaging with conventional RGB imaging for quality evaluation of *Agaricus bisporus* mushrooms. **Biosystems Engineering**, p. 191 – 194, January 2011.
- VIDAL, M.; AMIGO, J. M. Pre-processing of hyperspectral images. Essential steps before image analysis. **Chemometrics and Intelligent Laboratory Systems**, Elsevier, San Sebastián, v. 117, p. 138 – 148, Agosto 2012.