

## Análise de amostras de fubá por processamento digital de imagens e aprendizado de máquina

### Analysis of thin corn powder samples by digital image processing, and machine learning

#### RESUMO

Neste trabalho, analisou-se por meio de imagens, a similaridade entre três diferentes marcas de fubás. No total, foram obtidas 180 amostras, as quais foram foto-registradas e recortadas no programa GIMP. Os 768 tons de cinza dos canais RGB (Red-Green-Blue) das imagens das matrizes foram extraídos e analisados no software ChemoStat, no qual foi aplicado a Análise por agrupamento hierárquico (HCA). O dendrograma mostrou claramente, a formação de três clusters, que representaram as três marcas de fubá. Cada marca de fubá mostrou-se separadamente uma da outra, com poucos erros de classificação, essas diferenças basearam-se em seus padrões RGB de cores. Como não podemos identificar sutis diferenças de cores a olho nu em matrizes, o método proposto nessa pesquisa é indicado para identificar pequenas diferenças nas cores de matrizes e rastreamento de marcas de produtos alimentícios.

**PALAVRAS-CHAVE:** Agrupamento hierárquico, Quimiometria, Visão computacional.

#### ABSTRACT

In this work, the similarity among three different brands of cornmeal was analyzed by images and computational vision. In total, 180 samples were obtained, which were photo-registered and cut in the GIMP software. The 768 grayscale of the RGB (Red-Green-Blue) channels of the matrix images were extracted and analyzed using the ChemoStat software, in which the hierarchical cluster analysis (HCA) was applied. The dendrogram clearly showed the formation of three clusters, which represented the three brands of cornmeal. Each cornmeal brand was shown separately from each other, with few classification errors, these differences were based on their RGB color standards. As we cannot identify subtle color differences with the naked eye in matrices, the method proposed in this research is indicated to identify small differences in matrix colors and tracking food product brands.

**KEYWORD:** Hierarchical cluster, Chemometric, Computational vision.

Patrick Guilherme Roza

[Patrickbarbosa742@gmail.com](mailto:Patrickbarbosa742@gmail.com)

Universidade Tecnológica Federal do Paraná, Pato Branco, Paraná, Brasil

Vanderlei Aparecido de Lima

[valima@utfpr.edu.br](mailto:valima@utfpr.edu.br)

Universidade Tecnológica Federal do Paraná, Pato Branco, Paraná, Brasil

Recebido:

Aprovado:

19 ago. 2020.

Direito autoral:

01 out. 2020.

Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



## INTRODUÇÃO

A ferramenta que resulta da união entre as disciplinas de química e estatística é conhecida como Quimiometria. Essa ferramenta, considerada hoje como disciplina regular dos cursos de graduação e pós-graduação em química, tem base no avanço da tecnologia e na necessidade de uma maior capacidade de tratamento de dados multivariados e extrair resultados confiáveis de maneira rápida. Atualmente os métodos Quimiométricos estão sendo disseminados em cursos de graduação e pós-graduação, pois está cada vez mais se mostrando necessário para tratamentos de dados em indústrias e para pesquisas nas mais diversas áreas da química. (SANTANA, BACHION, 2020. FERREIRA, MÁCIA, 1999.)

O software Chemostat foi criado para suprir a demanda na área de pesquisa em química, e foi gerado pelo avanço tecnológico para alunos, professores e pesquisadores, que abrange amplamente análise exploratória e confirmatória de dados. Diferente dos vários outros aplicativos e programas que precisam da compra de licença para seu uso, ou algum investimento para entendimento de suas sintaxes que surgiram em função dessa demanda, Chemostat é um software desktop gratuito de fácil instalação e manuseio. Esse software apresenta também janelas e menus autoexplicativos, múltiplas entradas de dados, gráficos e janelas com recursos de cores para melhor identificação das amostras, múltiplas saídas de dados, nos formatos de texto (Excel, “txt” e ASCII) e figuras tipo “bmp”, “png”, “jpg”, entre outras. (HELFER, 2015.)

A análise por Agrupamento Hierárquico (HCA) é uma das técnicas desenvolvidas para quimiometria e essa metodologia utiliza matrizes de distância a partir dos dados iniciais das amostras para formação de clusters. Quando aplicamos a análise por agrupamentos a um conjunto de dados heterogêneos, é fornecido subgrupos homogêneos, no qual, as amostras podem ser unidas e/ou separadas e sua representação é apresentada por um dendrograma. O HCA associa ou desassocia objeto a objeto, e esse processo de associação começa com cada objeto em um cluster que são ligados sequencialmente, diminuindo o número de clusters até que todos pertençam a apenas a um determinado cluster

Pode-se utilizar o Chemostat para análises de imagens por meio de seus histogramas RGB, ou cada componente de cor individuais, também R relativo, G relativo, B relativo, H, S, V, I e L apenas importando imagens nos formatos “bmp”, “jpg” e “png”.

Diante do exposto, esse trabalho visou utilizar aplicativos gratuitos e com o objetivo de avaliar a similaridade de três diferentes marcas de fubás.

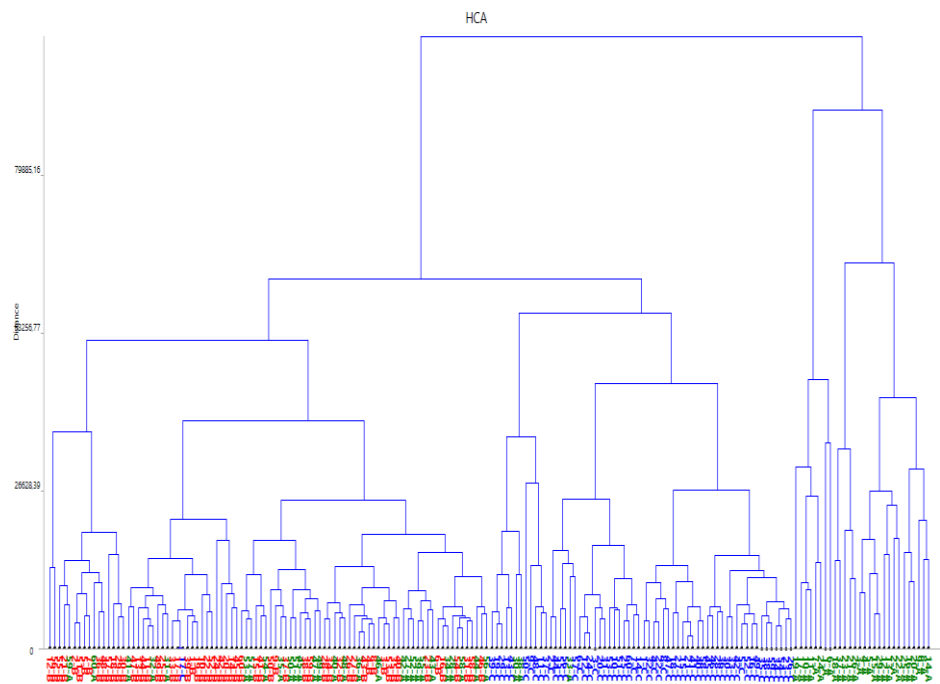
## MATERIAIS E MÉTODOS

Neste trabalho foi utilizado três marcas de fubás. As marcas receberam as denominações A, B e C. Amostras foram preparadas em copos plásticos, as quais foram foto-registradas com um smartphone da marca Motorola C. As imagens foram adquiridas sobre um fundo branco em ambiente claro para sua melhor aquisição e visualização, em seguida foram recortadas e salvas por meio do programa GIMP 2.10.8, e os padrões de cores RGB das amostras foram extraídos e analisados no software Chemostat. Nesse mesmo software, foi aplicado a Análise por agrupamento hierárquico (HCA) e gerado o dendrograma de similaridade.

## RESULTADOS E DISCUSSÕES

O padrão de cores RGB das três amostras de fubás apresentaram sutis diferenças entre si (Figura 1). Pode-se observar claramente, a formação de três clusters, que representaram as três diferentes marcas de fubás.

Figura 1. Dendrograma gerado por HCA. Padrão RGB das imagens das três diferentes marcas de fubás.



A marca de fubá A, foi representada pela cor verde, a marca B por vermelho e a marca de fubá C por azul. Esse padrão de cores foi atribuído por meio do *Sample ID* do próprio software. Pode-se observar que ocorreu a formação de três grupos de amostras no dendrograma. Cada marca de fubá apresentou-se separadamente uma da outra, com poucos erros de classificação. Algumas amostras da marca A que ficaram mais próximas de amostras da marca B e C, e de uma amostra da marca C que ficou mais próxima da marca B. Isso pode ter ocorrido devido a própria variação de cor da amostra A, pois pode-se observar que a distância de ligação entre as amostras dessa marca foi sempre maior entre si. Dessa maneira, essa marca A de fubá parece não apresentar um padrão bem definido de cores, quando comparados às outras marcas B e C de fubás.

## CONCLUSÕES

A análise de agrupamento hierárquico baseado na distância euclidiana nos mostrou diferenças entre as três marcas de fubás analisadas. Essas diferenças basearam-se em seus padrões RGB de cores. Como não podemos identificar diferenças a olho nu, esse método proposto nessa pesquisa é indicado para identificar pequenas diferenças nas cores de produtos que estão diretamente ligadas à sua composição química, facilitando a identificação de fraudes em produtos dessa linha alimentícia

### AGRADECIMENTOS

Agradecimentos à UTFPR pela bolsa de IC, ao meu orientador Dr. Vanderlei Aparecido de Lima pela oportunidade de aprendizagem e a orientação e também aos meus pais pela dedicação e toda ajuda para seguir com meus estudos.

### REFERÊNCIAS

SANTANA, F. C. et. al. Experimento didático de quimiometria para classificação de óleos vegetais comestíveis por espectroscopia no infravermelho médio combinado com análise discriminante por mínimos quadrados parciais: um tutorial, parte v. **Quím. Nova**, São Paulo, v. 43, n. 3, p. 371-381, mar. 2020.

FERREIRA, MÁRCIA M. C. et. al. Quimiometria I: calibração multivariada, um tutorial. **Quím. Nova**, São Paulo, v. 22, n. 5, p. 724-731, set. 1999.

HELPER, GILSON A. et. al. CHEMOSTAT, UM SOFTWARE GRATUITO PARA ANÁLISE EXPLORATÓRIA DE DADOS MULTIVARIADOS. **Quím. Nova**, São Paulo, v. 38, n. 4, p. 575-579, maio 2015.

ALMEIDA, J. A. S, et. al. Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering. **Chemometrics and Intelligent Laboratory Systems**. v. 87, p. 208–217, fev. 2007.