

## Desenvolvimento de um dispositivo de análise de azeites e óleo vegetal de baixo custo

### Development of a low-cost olive oil and vegetable oil analysis device

#### RESUMO

**Matheus Bogo Polidório**  
[mpolidorio@alunos.utfpr.edu.br](mailto:mpolidorio@alunos.utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil.

**Alexandre Duarte de Souza**  
[alexandreduarti@gmail.com](mailto:alexandreduarti@gmail.com)  
Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil.

**Alberto Yoshihiro Nakano**  
[nakano@utfpr.edu.br](mailto:nakano@utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil.

**Ricardo Schneider**  
[rschneider@utfpr.edu.br](mailto:rschneider@utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil.

O azeite de oliva extravirgem é considerado hoje um produto de alto valor comercial e nutricional e, a esses fatores, soma-se a impossibilidade de um consumidor comum detectar possíveis fraudes, fazendo com que esse alimento se torne alvo de adulterações. Atualmente, a detecção dessas fraudes é realizada por métodos de análise laboratoriais demandando tempo e mão de obra especializada. Este trabalho tem por objetivo desenvolver um dispositivo de baixo custo para a aquisição e processamento de imagens de óleos/azeites vegetais de maneira rápida sem a violação da embalagem, a fim de mensurar possível adulteração do azeite. Para validação das técnicas de análise, foram utilizados como amostras, recipientes de óleo de soja de diferentes marcas e lotes, visto que seu valor comercial é inferior ao azeite de oliva. A aquisição das imagens digitais foi realizada com a plataforma de desenvolvimento Raspberry Pi 3, uma câmera digital e um software de captura elaborado em Python. Após o processamento digital, utilizou-se de dois métodos de classificação para análise dos dados: *Principal Componente Analysis* - PCA e o *Linear Discriminant Analysis* - LDA. Como resultado, observou-se que o equipamento desenvolvido é estável e que os classificadores conseguiram diferenciar parte dos dados.

**PALAVRAS-CHAVE:** Azeite de oliva. Adulteração. Classificadores.

#### ABSTRACT

Today, extra virgin olive oil is considered to be a product of high commercial and nutritional value and the inability of an ordinary consumer to detect possible fraud makes this food a target for adulteration. The detection of these frauds is usually carried out by laboratory analysis, demanding time and specialized labor. This work aims to develop a low cost device for the acquisition and processing of images of vegetable oils /olive oils quickly, without the violation of the container, in order to measure possible adulteration of the olive oil with other vegetable oils. To validate the analysis techniques, soybean oil containers of different brands and lots were analyzed. The acquisition of digital images was made with the development platform Raspberry Pi 3 and a digital camera, with a capture software elaborated in Python. After the digital processing, two classification methods were used: PCA and LDA. As a result, it was observed that the equipment developed is stable and that the PCA and LDA were able to classify part of the data.

**KEYWORDS:** Olive oil. Adulteration. Classifiers. Pattern recognition.

**Recebido:** 19 ago. 2020.

**Aprovado:** 01 out. 2020.

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



## INTRODUÇÃO

O azeite de oliva extravirgem é considerado hoje um produto de alto valor comercial e nutricional e, a esses fatores, soma-se a impossibilidade de um consumidor comum detectar possíveis fraudes, fazendo com que esse alimento se torne alvo de adulterações. De acordo com Karbasian *et. al* (2015), o azeite de oliva extra virgem, em particular, é muito suscetível a mistura com óleos vegetais mais baratos e de qualidade inferior por causa de seu estado líquido, prática que tem por objetivo aumentar o seu rendimento. A falsificação é um problema que acontece no mundo todo e de acordo com Lipp (2012), os riscos de fraude aumentam com o crescimento da globalização e o aumento na competição de mercado, fazendo com que tanto os azeites produzidos nacionalmente, quanto os importados, possam ser adulterados.

Como possível solução para tornar mais eficiente a análise de azeites falsificados, foram propostas alternativas baseadas em técnicas de processamento digital que vem sendo aplicadas na análise de alimentos, inclusive de azeites de oliva (MILANEZ *et. al*; 2015) e óleos vegetais (GODOY *et. al*; 2020), principalmente pelo fato de serem técnicas não invasivas, rápidas e baratas quando comparadas aos métodos tradicionais. A maior parte desses estudos se utiliza das informações contidas nas cores das imagens em conjunto com algoritmos de reconhecimento de padrão/classificadores (GODOY *et. al*; 2020).

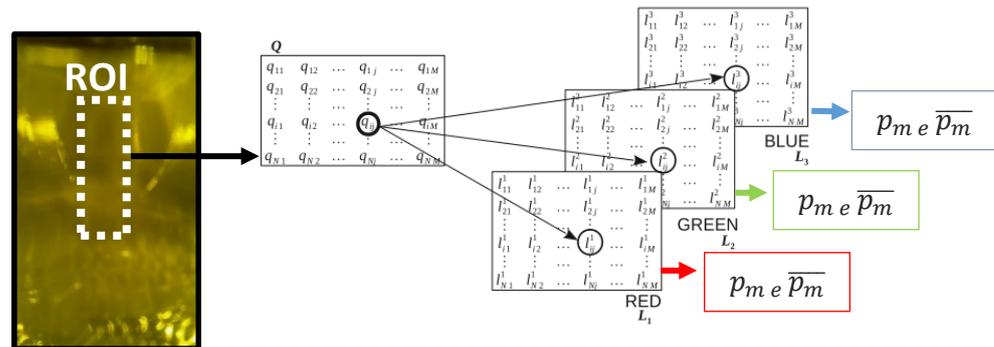
Deste modo, neste trabalho será apresentada a metodologia utilizada no desenvolvimento de um equipamento de baixo custo de produção cuja finalidade é realizar a análise rápida do azeite de oliva/óleos vegetais *in loco*, ou seja, sem a violação da embalagem. O azeite suspeito de uma possível adulteração será analisado através do processamento digital de imagens, sendo aplicada a combinação de métodos de classificação com PCA e LDA. O objetivo de se utilizar o PCA é reduzir a dimensionalidade, transformando um conjunto de dados em um subespaço com dimensões reduzidas. De acordo com Tharwat (2016), essa redução de dimensionalidade se dá pela identificação de variáveis que não são informativas, ou seja, que não acrescentam informações ao estudo e também reduzindo variáveis que apresentam informações redundantes. Os resultados obtidos do PCA, foram utilizados como parâmetros de entrada no LDA, que é um “classificador utilizado para encontrar uma combinação linear entre características que agrupem ou separem duas ou mais classes de objetos ou eventos” (THARWAT *et. al*; 2018).

## MATERIAIS E MÉTODOS

O princípio básico de funcionamento do dispositivo consiste na aquisição de imagens no espectro visível, que posteriormente são processadas na tentativa de se obter informações contidas na cor dessa imagem. Essa imagem pode ser vista como uma matriz (Q) de ordem  $N \times M$ , em que N representa as linhas e M as colunas (GODOY *et. al*; 2018). Cada elemento da matriz, expresso por  $q_{ij}$  com  $1 \leq i \leq N$ , e  $1 \leq j \leq M$ , é definido como um pixel e assume valores positivos entre 0 e 255 para cada canal de cor. A matriz que representa essa imagem pode ser apresentada através de camadas independentes  $L_k$ , com  $1 \leq k \leq k$ , em que cada camada indica uma matriz do sistema de cor e cada elemento pode ser expresso

por  $I_{ij}^k$  para cada canal de cor, como pode ser observado na Figura 1, que apresenta também o conceito de região de interesse (region of interest - ROI). No sistema *Red-Green-Blue* (RGB), faz-se  $k = \{1, 2, 3\}$ , assim tem-se  $L_1$  para R,  $L_2$  para G e  $L_3$  para B. Existem ainda diversos outros sistemas de cores como alternativas ao RGB, como por exemplo o *hue-saturation-value* (HSV) e *gray scale*.

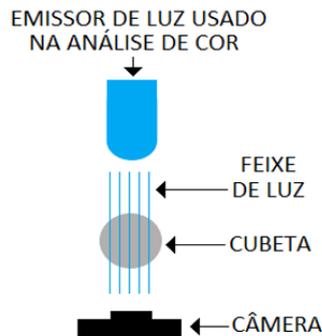
Figura 1: Exemplo de ROI e decomposição da matriz bidimensional(Q), que representa uma imagem digital, em camadas  $L_1$  para Red,  $L_2$  para Green e  $L_3$  para Blue



Fonte: Godoy *et al.* (2018)

O princípio de funcionamento desses sistemas pode ser observado na Figura 2, onde se tem um emissor de luz, a amostra a ser analisada e a câmera para realizar a aquisição das imagens. No dispositivo apresentado neste trabalho, a amostra a ser analisada fica em um ambiente fechado protegida de interferência externa, onde somente a luz proveniente do emissor incide nessa amostra.

Figura 2: Princípio de funcionamento do dispositivo



Fonte: Autoria própria (2020)

O dispositivo desenvolvido, que pode ser observado na figura 3, é composto por um Raspberry Pi 3B, no qual foi desenvolvido um software em Python 2.7 para a captura das imagens, uma Pi Câmera com resolução de 3280x2464 pixels na primeira seção da estrutura, uma fita de LED de 2 W controlada a partir de um controlador de corrente em conjunto com um sensor de luminosidade TSL2561, que se comunica com o Raspberry via protocolo I2C (*Inter-Integrated Circuit*) e apresenta em sua saída o valor convertido em Lux, unidade de intensidade de luz.

Figura 3: Dispositivo finalizado



Fonte: Autoria própria (2020)

Com o equipamento finalizado, o próximo passo foi realizar a captura das imagens para análise. A fim de validar a metodologia de classificação de óleos *in loco* e para facilitar a aquisição das amostras para análise, decidiu-se por utilizar recipientes de óleo de soja, visto que tem um valor comercial reduzido quando comparados a outros óleos vegetais. Foram então adquiridas diferentes marcas nacionais de óleo de soja de diferentes lotes no decorrer das semanas do mês de julho de 2020. A relação das marcas de óleo e a quantidade de amostras podem ser observadas na Tabela 1.

Tabela 1: Relação das marcas de óleo de soja e quantidade de lotes diferentes

Marcas	Quantidade de Lotes Diferentes
Liza	5
Leve	4
Cocamar	3
Soya	2
Coamo	2

Fonte: Autoria própria (2020)

Para cada amostra de óleo, foram capturadas e armazenadas 30 imagens digitais de 1600x900 pixels com intervalo de 1 segundo entre as capturas. Realizada a aquisição, foram calculados a partir da região de interesse (ROI), o pixel médio,

$$p_m = \frac{1}{N \times M} \times \sum_{i=1}^N \sum_{j=1}^M l_{ij}^k, \quad (1)$$

e o vetor de pixels médio,

$$\overline{p}_m = \frac{1}{N} \times \sum_{i=1}^N l_i^k, \quad (2)$$

sendo que N é o número de linhas e M o número de colunas da matriz. O termo k representa as componentes dos sistemas de cores utilizados, neste caso são os sistemas RGB, HSV e *Gray-Scale*, fazendo com que  $k = \{1, 2, 3, 4, 5, 6, 7\}$  para cada componente de cor {R, G, B, H, S, V e *GrayScale*}.

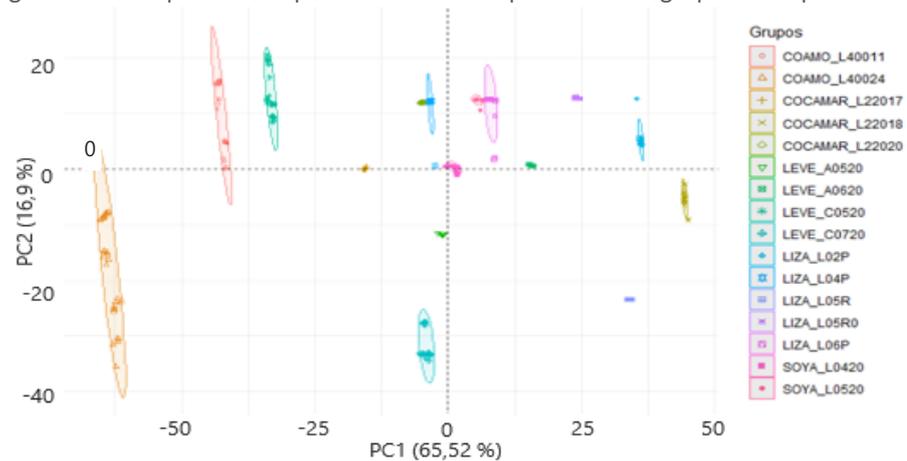
## RESULTADOS E DISCUSSÕES

Após o processamento das imagens, os valores das variáveis (pixel médio e vetor de pixels médio) foram utilizados como parâmetro de entrada no PCA e LDA. Os dados dos óleos vegetais foram analisados de duas maneiras: por marca,

independente do lote e por lote, ou seja, cada lote foi tratado como uma amostra diferente. Essa divisão teve por objetivo verificar se o método de classificação conseguiria diferenciar tanto os lotes quanto as marcas utilizadas. Como metodologia, utilizou-se os resultados do PCA como parâmetro de entrada no LDA e desses dados, 70 % foram utilizados para treinamento do classificador e os 30% restantes como teste. Essa metodologia foi utilizada nas análises tanto com o pixel médio quanto para o vetor de pixels médio.

Percebeu-se que utilizando a separação por lotes no PCA, independente de se utilizar o pixel médio ou o vetor de pixels médio, resultou em 100% de separação das amostras, como pode ser observada na Figura 4. Esperava-se que as amostras de uma mesma marca ficassem mais próximas umas das outras, dada as semelhanças entre lotes de uma mesma marca. Ao utilizar esses dados como parâmetro de entrada no LDA, obteve-se acerto global de 100% para todas as amostras, tanto com o pixel médio quanto utilizando o vetor médio. Essa taxa de acerto se deve ao fato de que foi utilizada somente uma amostra de um mesmo lote.

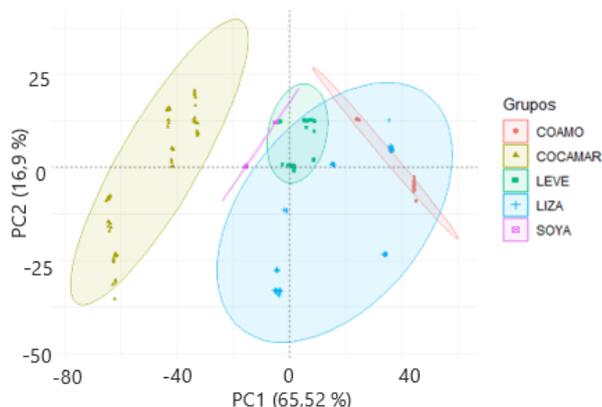
Figura 4: Scores produzidos pelo PCA - vetor de pixels médio- grupamento por lotes



Fonte: Autoria própria (2020)

Ao utilizar a separação dos dados por marca, sem considerar os lotes, percebeu-se que o PCA conseguiu definir os grupamentos para as marcas, porém como pode ser observado no gráfico de scores (Figura 5), houve sobreposição dos dados. Com a separação das amostras apenas por marca, a taxa de acerto global do LDA diminuiu, principalmente na classificação de marcas que se sobrepuseram no PCA. A teoria mostra que quanto mais informação melhor se dá o processo de classificação, mas neste caso, a inserção de novas marcas com baixa variabilidade entre si pode ter confundido o classificador, tornando necessário a inserção de um maior número de amostras (i.e., mais recipientes de cada marca de óleo).

Figura 5: Scores produzidos pelo PCA - vetor de pixels médio- grupamento por marcas



Fonte: Autoria própria (2020)

Observou-se certa semelhança entre os gráficos de *scores* dos grupamentos produzidos pelo PCA para o pixel médio separado por marcas e o vetor médio separado por marcas, e também entre a taxa de acerto global observado com o LDA. Isso significa que se for para se escolher entre os dois métodos para ser utilizado no equipamento utilizando-se do Raspberry para o processamento dos dados, a escolha mais eficiente seria utilizar-se do pixel médio para a classificação, visto que a sua implementação teria uma menor complexidade computacional.

## CONCLUSÃO

Pode-se concluir que o funcionamento do equipamento, tanto na parte de software quanto de hardware estão de acordo com o esperado. Algumas mudanças precisam ser realizadas em relação a dissipação de calor e na estrutura física, compactando os diferentes componentes do sistema. Em relação a análise dos óleos sem a violação das embalagens, ainda não se pode concluir muito, dada a pequena quantidade de amostras empregada. Outro ponto importante é saber identificar qual diferença entre as amostras está sendo considerada, visto que a geometria de cada embalagem de óleo, de cada marca, é diferente, fazendo com que as amostras possam estar sendo diferenciadas apenas pela distorção da luz incidente nas diferentes geometrias das garrafas.

Devido a pandemia de Covid-19, parte do trabalho teve de ser realizado na modalidade de home office, o que dificultou o processo de modificação do hardware. Em relação a aquisição de amostras, inicialmente foram firmadas algumas parcerias com pequenos mercados da região, para serem realizadas as aquisições das imagens no próprio estabelecimento sem a necessidade da compra de recipientes de óleo, mas devido as restrições de movimentação dentro das cidades e estabelecimentos não foi possível realizar esta etapa, diminuindo consideravelmente a quantidade de amostras a serem analisadas. Mais testes deverão ser realizados para conseguir estudar quais diferenças estão sendo consideradas pelo método.

## AGRADECIMENTOS

O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq – Brasil.

## REFERÊNCIAS

KARBASIAN, M; GIVIANRAD, M.H.; RAMEZAN, Y. A rapid method for detection of refined olive oil as adulterant in extra virgin olive oil by differential scanning calorimetry; **Oriental Journal of Chemistry**, 2015. p. 1735-1739.

MILANEZ, T. M.; DANIELLE, K.; PONTES, M. J. C. Classification of extra virgin olive oil and verification of adulteration using digital images and discriminant analysis **Food Analytical Methods**, vol 7, p. 8839-8846, 2015.

LIPP, M. Ingredient adulteration undermines food safety. **FoodSafety Magazine**, Glendale, California. Março 2012 Disponível em:  
<https://www.foodsafetymagazine.com/magazine-archive1/februarymarch-2012/ingredient-adulteration-undermines-food-safety/> Acesso em Agosto de 2020.

GODOY, A. C. *et al.* Snapshots analyses for turbidity measurements in water. **Water, Air, & Soil Pollution**, v. 229, n. 12, p. 378, Nov 2018.

GODOY, A.C., DOS SANTOS, P.D.S., NAKANO, A.Y. *et al.* Analysis of Vegetable Oil from Different Suppliers by Chemometric Techniques to Ensure Correct Classification of Oil Sources to Deal with Counterfeiting. **Food Analytical Methods**, vol. 13, p. 1138–1147, 2020.

THARWAT, A. Principal component analysis - a tutorial. **International Journal of Applied Pattern Recognition**. .vol. 3. p. 197, 2016.

THARWAT, A. G.; IBRAHIM, T; ABDELHAMEED, H.; ABOUL, E. Linear discriminant analysis: A detailed tutorial. **Ai Communications**. vol. 30. p. 169-190, 2018.