

Medição de qualidade de sementes de canola com visão computacional e aprendizado de máquina

Wellington Rangel dos Santos¹, Rosana Falcão²

Resumo

Produtos originados dos grãos da canola, como óleo, farinha e biodiesel, são diretamente afetados pela qualidade desses grãos. O uso da tecnologia tem contribuído para a identificação dos grãos impuros ou em estágios diferentes de maturidade. A visão computacional em conjunto com o aprendizado de máquina pode gerar ferramentas capazes de avaliar a qualidade das sementes de forma não invasiva, sem destruir amostras e com baixo custo, uma vez que utiliza imagens digitais como insumo. O presente estudo propõe o uso de visão computacional e aprendizado não supervisionado de máquina para análise de sementes de canola a partir de imagens digitais, com o objetivo de contar, identificar cada semente e calcular o percentual de grãos maduros. Os resultados mostraram que o método K-Means pode ser usado para contar e identificar sementes de canola em fotografias com alta precisão. Na amostra, foram identificadas noventa e três sementes maduras, quatro verdes e três secas.

Termos para indexação: Python, Google Colaboratory, sementes de canola.

Introdução

A canola é uma planta oleaginosa que produz pequenas sementes de onde é extraído um óleo com traços inferiores a 2% de ácido erúico e sem gordura trans. O processamento das sementes para produzir o óleo de cozinha e produtos alimentícios passa por processamento adicional de branqueamento (Gaber et al., 2018), e a biomassa, com alto valor proteico, pode ser destinada como torta para ração animal (<<https://www.canolacouncil.org/about-canola/oil/#fact-and-fiction>>). A canola também é usada como matéria-prima para combustíveis renováveis em diversas áreas de cultivo pelo Brasil (Tomm, 2007; Rocha et al., 2019; Silva, 2019). Fatores edafoclimáticos e fitossanitários afetam a qualidade das sementes, o que interfere na qualidade dos subprodutos: óleo de canola, farelo e farinha (Sabbahi et al., 2023). O uso de grãos uniformes e dentro do padrão sugerido pela Comissão Canadense de Grãos (Oilseeds Methods and Tests Used to Measure Quality – 2019, <<https://grainscanada.gc.ca/en/grain-research/export-quality/oilseeds/methods-tests.html>>) e dentro das Regras para análise de sementes (RAS) do Ministério da Agricultura e Pecuária (Mapa) (Brasil, 2009) visa a garantir a qualidade dos subprodutos da canola quando qualifica a densidade, o tamanho e o estágio de maturidade dos grãos com o objetivo de manter, ou até aumentar, a qualidade, visando a boa produtividade (Amaral et al., 2012).

Tecnologias emergentes geram ferramentas cada vez mais rápidas e precisas na tomada de decisão sobre a qualidade em lotes de sementes (Santana et al., 2022). Dentre essas tecnologias, duas se destacam: visão computacional e aprendizado de máquina. A visão computacional vem sendo utilizada em várias aplicações na agricultura já que análises em imagens reduzem custos e evitam manipulações diretas das amostras. Algumas bibliotecas de programação em visão computacional se destacam, como KerasCV (<https://github.com/keras-team/keras-cv>), PlantCV (<https://>

¹ Cientista da computação, analista da Embrapa Agroenergia, wellington.santos@embrapa.br

² Bióloga, mestre em Ciências Genômicas e Biotecnologia, analista da Embrapa Agroenergia, rosana.falcao@embrapa.br

plantcv.readthedocs.io/) e OpenCV (<https://opencv.org/>). Esta última é a mais utilizada, com ampla documentação e código aberto. O aprendizado de máquina amplia as aplicações baseadas em visão computacional na medida em que a extração de informações em imagens pode ser submetida a algoritmos capazes de encontrar padrões, detectar anomalias, classificar objetos e manipular inúmeras variáveis ao mesmo tempo. Nesse campo, o algoritmo K-Means (MacQueen, 1967) é referência no aprendizado de máquina não supervisionado, ou seja, que não precisa de dados históricos. K-Means é uma técnica de agrupamento de dados por K-médias em que, inicialmente, os grupos são formados em torno de um centro aleatório. Em seguida, os pontos de dados são ajustados ao centro. A cada novo ponto, esse centro é revisado, até que a distância entre elementos do mesmo grupo seja minimizada e maximizada entre os grupos diferentes (Jin & Han, 2011). Dessa forma, o objetivo deste trabalho foi avaliar se o método de agrupamentos (K-Means) pode ser capaz de contar, identificar cada semente de canola em uma fotografia e calcular o percentual de grãos maduros contribuindo com a verificação da qualidade da amostra.

Materiais e métodos

O material biológico foi coletado nos campos experimentais da Embrapa Agroenergia. As sementes foram derramadas sobre a régua perfurada, coladas em papel adesivo transparente e esmagadas em folha plástica transparente. A imagem principal foi obtida por digitalização em Impressora Samsung C 406x no formato JPEG e resolução 300dpi. Em seguida, foi editada para ter fundo transparente e incluir sementes verdes e secas para obtermos uma heterogeneidade na imagem de teste (Figura 1).

O software foi programado usando a linguagem Python (3.10.12) com os seguintes pacotes: OpenCV (4.7.0.72), imutils (0.5.4), matplotlib (3.7.1), numpy (1.22.4), pandas (1.5.3), scikit-learn (1.2.2). O ambiente de programação foi o Google Colaboratory.

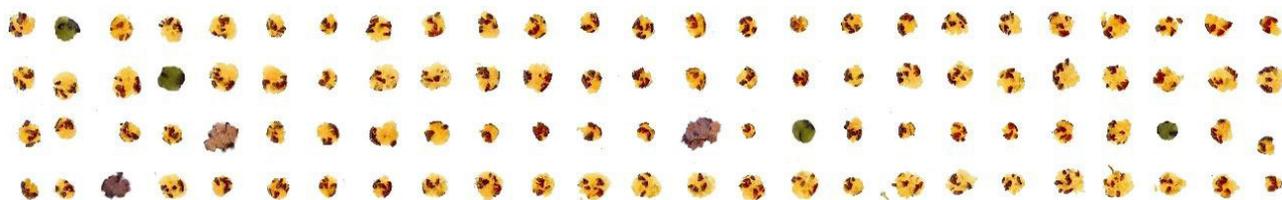


Figura 1. Imagem de teste com amostras de sementes de canola esmagadas.

Resultados e discussão

Segmentação, contagem e classificação das sementes

Um sistema especialista foi desenvolvido em duas etapas para, primeiro, extrair cada semente isoladamente da imagem de teste com o objetivo geral de contar sementes em uma imagem. Em seguida, comparar cada semente segmentada com as demais a partir de suas cores mais determinantes.

A ferramenta computacional que identifica e conta a quantidade de sementes presentes na imagem foi construída usando pacotes de softwares livres e, principalmente, a biblioteca OpenCV. Esse pacote foi escolhido por ser referência em visão computacional em sistemas código aberto, com vasta documentação, estável e de comunidade ativa na manutenção do código-fonte. A imagem de teste (Figura 1) foi, primeiramente, convertida para escala cinza com a função `cvtColor()` (Figura 2a), para, em seguida, ser convertida em uma imagem binária, isto é, reduzindo tons de

cinza em preto e branco. (Figura 2b). A partir das funções `getStructuringElement()` e `morphologyEx()` (Figura 2c), foram determinados quais os pixels relevantes e usados para fechar pequenos orifícios dentro dos objetos.

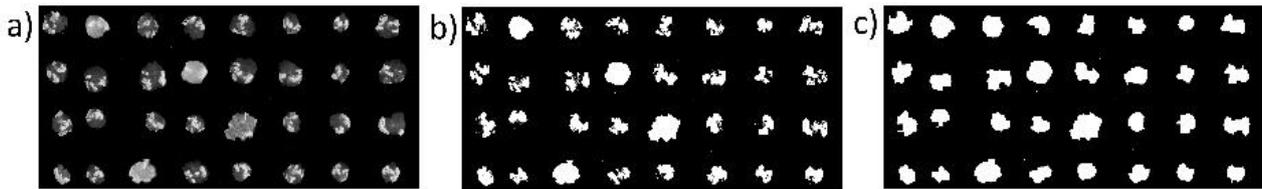


Figura 2. Imagem de teste após as primeiras etapas: (a) transformação em tons de cinza. (b) processo de binarização. (c) fechamento da imagem.

Com a imagem, agora simplificada, foi aplicada uma função `cvFindContours()` para encontrar todos os contornos existentes em uma imagem binária. Os contornos são as bordas externas de objetos presentes na imagem e são úteis em tarefas como detecção de bordas, reconhecimento de formas e contagem de objetos. O número de contornos encontrados se aproxima da quantidade de sementes presentes na imagem, exceto quando há pequenos objetos originados por ruídos na imagem. Nesse caso, foram descartados contornos com dimensões irrelevantes com o auxílio da função `boundingRect()`, que é usada para criar um retângulo aproximado em torno de uma área de interesse. Com isso, o número de sementes na imagem pode ser calculado. Adicionalmente, cada semente foi segmentada e exportada como um arquivo de imagem no formato PNG isoladamente. O tamanho médio resultante foi de 35 KB, com dimensões de, aproximadamente, 120 pixels x 110 pixels. Para a classificação das sementes, o algoritmo de agrupamento K-Means foi utilizado para destacar as dez principais cores de cada semente (Figura 3).



Figura 3. Três amostras de sementes e suas dez cores mais significativas.

O próximo passo foi calcular a média de cores de cada semente desconsiderando a cor branca. Essa cor foi comum a todas as amostras, mas com diferentes percentuais de ocorrência, podendo causar uma distorção na comparação entre sementes. Pelo padrão de cores RGB, cada cor é formada por três valores de Vermelho (R), Verde (G) e Azul (B) limitados de 0 a 255. Calculou-se a média de cada canal RGB separadamente, resultando em uma única cor para cada semente. O resultado dessa operação foi uma tabela com cada semente e sua cor determinante (Tabela 1).

Tabela 1. Exemplo de três sementes e sua média das cores mais determinantes.

Identificador da semente	Vermelho (R)	Verde (G)	Azul (B)	COR
1	176	123	60	
42	154	105	55	
50	180	124	60	

O último passo foi o agrupamento das sementes a partir da cor determinante obtida na etapa anterior. Para isso, novamente, foi aplicada a técnica de agrupamento K-Means agrupado em grupos de sementes mais semelhantes a partir de suas cores. A amostra de sementes foi separada em quatro agrupamentos (Figura 4) como produto do K-Means. Esse resultado demonstrou que houve diferenças entre os grupos.

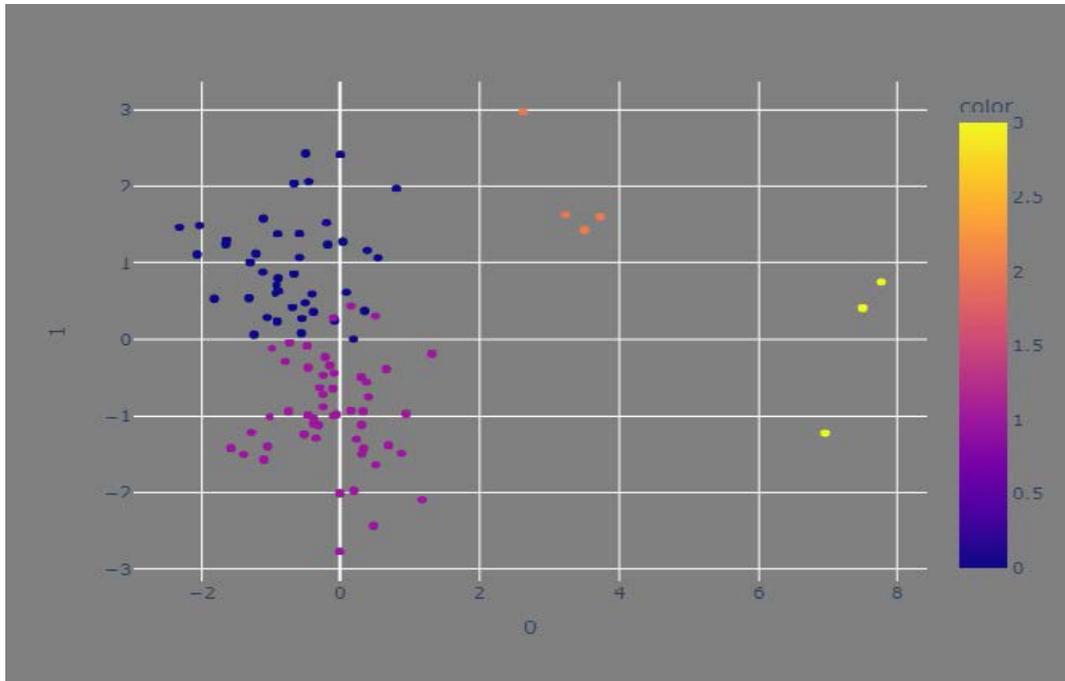


Figura 4. Gráfico de componentes principais (PCA) dos quatro agrupamentos de sementes.

No total, foram contadas 100 sementes. Dessas, três foram identificadas como do tipo “secas”; quatro, como do tipo “verdes”; e 93, “maduras” (Figura 5). Essas últimas foram distribuídas em dois agrupamentos indicando que houve diferenças na relação entre cores da casca e da polpa apesar de estarem no mesmo grau de maturidade.

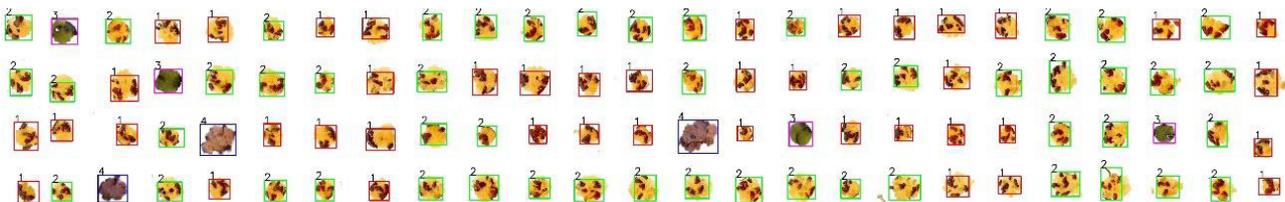


Figura 5. Imagem gerada pelo software com a identificação e classificação das sementes de acordo com os agrupamentos encontrados.

Considerando que as sementes maduras são as que podem apresentar maior produtividade, essa amostra contém 93% desse tipo de sementes. Essa ferramenta, embarcada em aparelhos celulares, pode auxiliar órgãos de controle de qualidade na inspeção de lotes de sementes comerciais. Produtores também podem se beneficiar, avaliando lotes de sementes no momento da aquisição e até mesmo antes do plantio. Na fase da colheita, as plantas estarão prontas para serem colhidas quando há mudanças de coloração dos grãos, do verde para a cor marrom (Embrapa, 2022). Assim, pode-se estender essa técnica para avaliar o ponto de maturação natural de grãos sem o esmagamento, tornando o processo não destrutivo, com menor custo e maior precisão. Finalmente, essa solução pode ser ampliada para inferir o percentual do teor de óleo e de proteínas de amostras de sementes esmagadas, avaliando não somente a cor da casca, mas também seu interior.

Conclusão

Esse resultado mostrou que o algoritmo K-Means associado a análises de cores em imagens pode ser aplicado para identificar sementes de tipos diferentes em uma fotografia digital. Foram contadas 100 sementes, com o percentual 93% maduras. Estudos posteriores serão necessários para medir o desempenho em amostras com sementes mais heterogêneas, com sementes sobrepostas e com imagens de menor qualidade.

Referências bibliográficas

- AMARAL, A. D. do; MEDEIROS, S. L. P.; MENEZES, N. L. de; LUZ, G. L. da; PIVOTO, D.; BIALOZOR, A. Qualidade de sementes de canola classificadas por densidade. **Revista Brasileira de Sementes**, v. 34, n. 2, p. 302-309, 2012. DOI: 10.1590/S0101-31222012000200016.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. **Regras para análise de sementes**. Brasília, DF, 2009.
- EMBRAPA. **Mais canola**. 2022. Disponível em: <https://play.google.com/store/apps/details?id=com.embrapa.canola&pli=1>. Acesso em: 26/09/2023
- GABER, M. A. F. M.; TUJILLO, F. J.; MANSOUR, M. P.; JULIANO, P. Improving oil extraction from canola seeds by conventional and advanced methods. **Food Engineering Reviews**, v. 10, p. 198-210, 2018. DOI: 10.1007/s12393-018-9182-1.
- JIN, X.; HAN, J. K-means clustering. In: SAMMUT, C.; WEBB, G. I. (Ed.). **Encyclopedia of machine learning**. Boston: Springer, 2011. p. 563-564. DOI: 10.1007/978-0-387-30164-8_425.
- MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 5., 1967, Berkeley CA. **Proceedings...** Berkeley, CA: University of California, 1967.
- ROCHA, L. de S.; GOMES, E. S.; CARNEIRO, A. de S.; FERREIRA, A. G. C.; COMBY, A. C. O.; CARVALHO, J. L. R. de; LAVIOLA, B. G.; SANTOS, A. dos; MARANA, J. C.; COSTA, L. T. da; RODRIGUES, E. V. Análise de trilha para produtividade de grãos em canola no Cerrado. In: CONGRESSO DA REDE BRASILEIRA DE TECNOLOGIA E INOVAÇÃO DE BIODIESEL, 7., 2019, Florianópolis. **Empreendedorismo e inovação: construindo um futuro competitivo para o biodiesel: anais**. Florianópolis, SC: Rede Brasileira de Tecnologia e Inovação de Biodiesel, 2019. p. 530-531.
- SABBAHI, R.; AZZAQUI, K.; RHAZI, L.; AYERDI-GOTOR, A.; AUSSENAC, T.; DEPEINT, F.; TALEB, M.; HAMMOUTI, B. Factors affecting the quality of canola grains and their implications for grain-based foods. **Foods**, v. 12, n. 11, 2023. 2219. DOI: 10.3390/foods12112219.
- SANTANA, D. C.; OLIVEIRA CUNHA, M. P. de; SANTOS, R. G. dos; COTRIM, M. F.; TEODORO, L. P. R.; SILVA JUNIOR, C. A. da; BAIO, F. H. R.; TEODORO, P. E. High-throughput phenotyping allows the selection of soybean genotypes for earliness and high grain yield. **Plant Methods**, v. 18, 2022. 13. DOI: 10.1186/s13007-022-00848-4.
- SILVA, R. J. da. **Qualidade de grãos de canola produzidos no cerrado mineiro**. 2019. Monografia (Graduação em Agronomia) - Universidade Federal de Uberlândia-MG.
- TOMM, G. O. **Indicativos tecnológicos para produção de canola no Rio Grande do Sul**. Passo Fundo: Embrapa Trigo, 2007. (Embrapa Trigo. Sistemas de Produção, 4).