

## MÉTODO NÃO DESTRUTIVO PARA ESTIMAR TEOR DE COMPOSTOS BIOATIVOS POR IMAGENS EM MICROVERDES DE CULTIVO *INDOOR*

VITOR G. DA SILVA<sup>1</sup>, THAIS Q. ZORZETO CESAR<sup>2</sup>, LUIS F. V. PURQUERIO<sup>3</sup>,  
LUIZ H. A. RODRIGUES<sup>4</sup>, RAFAEL A. DE OLIVEIRA<sup>5</sup>

<sup>1</sup>Eng. Agrícola, Doutorando em Engenharia Agrícola, FEAGRI/UNICAMP, Campinas – SP, vsilva.agri@gmail.com.

<sup>2</sup>Eng. Agrícola, Professora Doutora, FEAGRI/UNICAMP, Campinas – SP.

<sup>3</sup>Eng. Agrônomo, Centro de Horticultura, Instituto Agronômico, Campinas – SP.

<sup>4</sup>Eng. Agrícola, Professor Livre Docente, FEAGRI/UNICAMP, Campinas – SP.

<sup>5</sup>Eng. Agrícola, Professor Livre Docente, FEAGRI/UNICAMP, Campinas – SP.

Apresentado no  
LII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2023  
18 a 21 de outubro de 2023 – Ribeirão Preto - SP, Brasil

**RESUMO:** O cultivo *indoor* de microverdes de alta qualidade requer iluminação artificial e controle do ambiente, desempenhando um papel vital na biossíntese de compostos bioativos, principalmente pigmentos naturais. No entanto, os métodos tradicionais para determinar pigmentos são caros e destrutivos, exigindo alternativas de baixo custo e não destrutivas. Este estudo teve como objetivo desenvolver métodos não destrutivos de baixo custo utilizando técnicas de aprendizado de máquina em conjuntos de dados gerados a partir de imagens, a fim de estimar a concentração de compostos bioativos em microverdes de cultivo *indoor*. Algoritmos de aprendizado de máquina supervisionado foram usados para modelar a relação entre atributos de cor e teores de pigmentos. Os resultados mostraram que os atributos de cor vermelho/verde/azul (RGB) e matiz/saturação/valor (HSV), juntamente com dados de espécies e densidade de fluxo de fótons fotossintéticos, foram atributos preditivos eficazes para a construção de modelos de regressão do teor de pigmentos e *Random Forest* se destaca como melhor algoritmo para o problema. Nossa abordagem, combinada com dados do ambiente, oferece uma ferramenta valiosa para otimizar o cultivo *indoor* de microverdes.

**PALAVRAS-CHAVE:** pigmentos, *random forest*, aprendizado de máquina

### NON-DESTRUCTIVE METHOD FOR BIOACTIVE COMPOUNDS CONTENT ESTIMATION BY IMAGING IN INDOOR GROWING MICROGREENS

**ABSTRACT:** Indoor cultivation of high-quality microgreens requires artificial lighting and environmental control, playing a vital role in the biosynthesis of bioactive compounds, mainly natural pigments. However, traditional methods for determining pigments are expensive and destructive, requiring low-cost and non-destructive alternatives. This study aimed to develop low-cost non-destructive methods using machine learning techniques on datasets generated from images, in order to estimate the content of bioactive compounds in indoor microgreens. Supervised machine learning algorithms were used to model the relationship between color features and pigment contents. The results showed that the color attributes red/green/blue (RGB) and hue/saturation/value (HSV), along with species data and photosynthetic photon flux density, were effective predictive attributes for building regression models of pigment content and *Random Forest* stands out as the best algorithm for the problem. Our approach, combined with environmental data, offers a valuable tool for optimizing indoor microgreen cultivation.

**KEYWORDS:** pigments, random forest, machine learning

**INTRODUÇÃO:** A produção de microverdes (plantas cultivadas poucos dias após a emergência, colhidas ainda jovens e imaturas, logo após a completa expansão das folhas cotiledonares), tem despertado interesse por suas características físicas, químicas, sensoriais e nutraceuticas únicas (XIAO et al., 2012). Elas são ricas em compostos bioativos, como pigmentos naturais (antocianinas e carotenoides), que conferem propriedades antioxidantes (GOFF; KLEE, 2006). Sua produção ocorre principalmente em ambientes controlados com iluminação artificial, induzindo alterações fisiológicas e metabólicas nas plantas para obter maior qualidade, uma vez que a biossíntese de seus compostos está intimamente ligada à iluminação artificial (PESCARINI et al., 2023; SAMUOLIENĖ et al., 2017). No entanto, as análises tradicionais de determinação desses compostos são caras, demoradas e destrutivas, levando à necessidade de técnicas não destrutivas e acessíveis. Embora técnicas baseadas em imagens e aprendizado de máquina tenham mostrado eficácia na estimativa da qualidade de produtos hortícolas, sua aplicação ainda é dispendiosa (KOYAMA et al., 2021). Portanto, explorar as propriedades ópticas dos pigmentos para estimar compostos bioativos por meio de técnicas não destrutivas, como o aprendizado de máquina baseado em imagens, oferece uma alternativa promissora e acessível para a produção de microverdes. Este estudo teve como objetivo desenvolver métodos não destrutivos de baixo custo utilizando técnicas de aprendizado de máquina em conjuntos de dados gerados a partir de imagens, a fim de estimar a concentração de compostos bioativos em microverdes de cultivo *indoor*.

**MATERIAL E MÉTODOS:** Nove espécies de microverdes — amaranto (*Amaranthus caudatus*), rúcula (*Eruca sativa*), beterraba (*Beta vulgaris esculenta*), canola (*Brassica napus* L. var. oleífera), repolho verde (*Brassica oleracea* L. var. capitata), repolho roxo (*Brassica oleracea* L. var. capitata f. rubra), mostarda (*Brassica juncea*), rabanete (*Raphanus sativus*) e girassol (*Helianthus annuus*) -, foram cultivadas em ambiente *indoor* com temperatura controlada a 25 °C e sob densidade de fluxo de fótons fotossinteticamente ativos (DFFFA) de 50, 100, 150 e 200  $\mu\text{mol m}^{-2} \text{s}^{-1}$ , com fotoperíodos de 16 h. Para cada espécie, foram produzidas 10 repetições em cada tratamento de PPF, totalizando 40 amostras por espécie e 360 amostras no total. Antes da colheita, imagens RGB (.png) de cada amostra foram capturadas em laboratório. O sistema consistia em um minicomputador Raspberry Pi Zero W (Broadcom de 1 GHz, processador BCM2835 e 512 MB de RAM) equipado com uma câmera digital de 8 MP (Módulo v2), colocada dentro de uma câmara de inspeção quadrada com sistema de iluminação composto por *ring light* LED (diâmetro 26 cm) posicionada no topo, com luz vermelha (R), verde (G), azul (B) e branca (W) (6500 K). Um total de 40 imagens (10 para cada fonte de luz BGRW) foram obtidas para cada amostra, girando a amostra para uma posição diferente após cada captura. A aquisição resultou em 1.600 imagens por espécie, totalizando 14.400 imagens. Após a aquisição de imagens, as plantas foram colhidas em suas partes aéreas e prensadas para obtenção do extrato líquido do qual determinou-se os teores de pigmentos (clorofila a, clorofila b e antocianinas (SIMS; GAMON, 2002)). Realizou-se o processamento das imagens em Python (versão 3.9.16) via Google Collaboratory utilizando a biblioteca OpenCV. As imagens foram redimensionadas (de 8MP para 2MP) e a parte vegetal foi separada do fundo por *thresholding*. Em seguida, as regiões de interesse (ROIs) de cada imagem foram extraídas como quadrados de 256 x 256 pixels centrados, das quais foram extraídos sete atributos de cor com base na distribuição média de pixels nos canais de cores RGB e HSV. Esses atributos de cor foram transformados em 62 novos atributos por combinações matemáticas (YANG et al, 2016), gerando o conjunto de dados utilizado na modelagem. Utilizou-se a proporção de 70/30 para conjunto de treino e teste, respectivamente. A Figura 1 traz um exemplo de imagem de microverde.



FIGURA 1. Imagem de amostra de microverde com fonte de iluminação branca (W).

O problema de regressão, tendo os teores de clorofila a, clorofila b, antocianinas e carotenoides como alvos, foi abordado com o uso de cinco técnicas de aprendizado de máquina supervisionado (Regressão Linear, K-vizinhos mais próximos (KNN), Árvore de Regressão, *Random Forest* e *Extreme Gradient Boosting* (XGB)), implementados por meio das bibliotecas scikit-learn e XGBoost. Os hiperparâmetros foram ajustados para os diferentes algoritmos no conjunto de treino, com validação cruzada *5-fold*. Os modelos obtidos foram validados no conjunto de teste, utilizando o erro absoluto percentual médio (MAPE) e o erro quadrático médio (RMSE) como métricas de avaliação.

**RESULTADOS E DISCUSSÃO:** Para a modelagem, utilizou-se a Regressão Linear como *baseline*, pois considera apenas uma relação linear entre o conjunto de variáveis independentes com uma variável dependente, neste caso, o teor de pigmentos. Observa-se que os modelos obtidos por *Random Forest*, XGB e Árvore de Regressão, na etapa de validação, apresentaram menores valores de MAPE e RMSE (Tabela 1). Todos os modelos KNN tiveram desempenho pior que o basal, mostrando-se uma técnica inapropriada para a regressão avaliada neste trabalho.

TABELA 1. Erro percentual médio absoluto (MAPE) e erro quadrático médio (RMSE) (mg 100 g<sup>-1</sup>) na validação dos modelos de regressão para o conjunto de dados.

Variável Predita	Erro	Regressão Linear	KNN	Árvore de Regressão	Random Forest	XGB
Clorofila a	MAPE	0,18	0,45	0,11	0,09	0,11
	RMSE	1,67	1,39	0,70	0,56	0,61
Clorofila b	MAPE	0,21	0,45	0,15	0,12	0,14
	RMSE	0,89	1,42	0,73	0,60	0,66
Antocianina	MAPE	0,48	2,06	0,13	0,10	0,13
	RMSE	1,96	6,65	1,15	0,90	0,97
Carotenoide	MAPE	0,24	0,34	0,16	0,14	0,16
	RMSE	0,17	0,24	0,15	0,12	0,14

Dentre os modelos de regressão adequados, as melhores previsões foram obtidas por *Random Forest*, com valores de MAPE inferiores a 0,15 para todos os modelos e significativamente inferiores à Regressão Linear. Isso reforça o potencial que *Random Forest* tem demonstrado em outros estudos com problemas de regressão aplicados à qualidade dos alimentos e à presença de compostos bioativos (ASKEY et al., 2019; TAN et al., 2021). Embora a interpretabilidade seja um dos critérios importantes para a construção de modelos de previsão - que é uma das vantagens das árvores de regressão (BREIMAN, 2001) -, ela é limitada pelos parâmetros de complexidade encontrados no ajuste. Portanto, embora os modelos de Árvore

de Regressão também tenham se mostrado adequados, com valores de MAPE ligeiramente superiores aos de *Random Forest*, o elevado número de nós nas árvores torna implausível a interpretação dos modelos.

**CONCLUSÕES:** Métodos não destrutivos, usando técnicas de aprendizado de máquina em conjuntos de dados gerados a partir de imagens, são eficazes para estimar o teor de compostos bioativos em microverdes de cultivo *indoor*. Os modelos construídos com base em parâmetros de cor, especialmente os atributos de cor RGB e HSV, associados à densidade de fluxo de fótons fotossinteticamente ativos e dados de espécies, mostram-se promissores na predição de teores de pigmentos, como clorofila a e b, antocianinas e carotenoides. O algoritmo *Random Forest* destaca-se como uma técnica útil para esse propósito. As descobertas têm o potencial de serem valiosas para o cultivo de microverdes *indoor*, contribuindo para alcançar altos rendimentos e alta qualidade, sendo ferramenta para otimização de ambientes *indoor* na horticultura.

**AGRADECIMENTOS:** Agradecemos ao Programa de Bolsas DeepMind (Convênio FUNCAMP nº 5629) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (processo nº 88887.841689/2023-00) pela bolsa ao primeiro autor.

#### **REFERÊNCIAS:**

- ASKEY, B. C.; DAI, R.; LEE, W. S.; KIM, J. A noninvasive, machine learning–based method for monitoring anthocyanin accumulation in plants using digital color imaging. **Applications in Plant Sciences**, v.7, n.11, e11301, 2019.
- BREIMAN, L. Statistical modeling: The two cultures. **Statistical Science**, v.16, n.3, p.199-215, 2001.
- GOFF, S. A.; KLEE, H. J. Plant volatile compounds: Sensory cues for health and nutritional value? **Science**, v.311, n.5762, p.815-819, 2006.
- KOYAMA, K.; TANAKA, M.; CHO, B.; YOSHIKAWA, Y.; KOSEKI, S. Predicting sensory evaluation of spinach freshness using machine learning model and digital images. **PLoS ONE**, v.16, n. 3, p.1-16, 2021.
- PESCARINI, H. B.; SILVA, V. G.; MELLO, S. C.; PURQUERIO, L. F. V.; SALA, F. C.; ZORZETO-CESAR, T. Q. Updates on Microgreens Grown under Artificial Lighting: Scientific Advances in the Last Two Decades. **Horticulturae**, v.9, n.8, p.864, 2023.
- SAMUOLIENĖ, G.; VIRŠILĖ, A.; BRAZAITYTĖ, A.; JANKAUSKIENĖ, J.; SAKALAIUSKIENĖ, S.; VAŠTAKAITĖ, V.; NOVICKOVAS, A.; VIŠKELIENĖ, A.; SASNAUSKAS, A.; DUCHOVSKIS, P. Blue light dosage affects carotenoids and tocopherols in microgreens. **Food Chemistry**, v.228, p.50-56, 2017.
- SIMS, D. A.; GAMON, J. A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. **Remote Sensing of Environment**, v.81, n.2-3, p.337-354, 2002.
- TAN, W. H.; IBRAHIM, H.; CHAN, D. J. C. Estimation of mass, chlorophylls, and anthocyanins of *Spirodela polyrhiza* with smartphone acquired images. **Computers and Electronics in Agriculture**, v.190, p.106449, 2021.
- XIAO, Z.; LESTE, G. E.; LUO, Y.; WANG, Q. Assessment of vitamin and carotenoid concentrations of emerging food products: Edible microgreens. **Journal of Agricultural and Food Chemistry**, v.60, n.31, p.7644-7651, 2012.
- YANG, X.; ZHANG, J.; GUO, D.; XIONG, X.; CHANG, L.; NIU, Q.; HUANG, D. Measuring and evaluating anthocyanin in lettuce leaf based on color information. **IFAC-PapersOnLine**, v.49, n.16, p.96-99, 2016.