

# Análise comparativa de algoritmos de aprendizagem profunda na classificação de hortaliças

Eduardo L. de Andrade<sup>1</sup>, Danielli A. Lima<sup>2</sup>, Jefferson Rodrigo de Souza<sup>1</sup>, Shiguo Nomura<sup>1</sup>

<sup>1</sup>Faculdade de Computação - Universidade Federal de Uberlândia  
Uberlândia, MG - Brasil

<sup>2</sup>Faculdade de Engenharia Elétrica - Universidade Federal de Uberlândia  
Uberlândia, MG - Brasil

{eduaandrade, danielli, jrsouza, shigueonomura}@ufu.br

**Abstract.** *This paper presents a comparative analysis of the performance of four Convolutional Neural Network (CNN) architectures in the binary classification of table tomatoes: InceptionV3, ResNet-50, Xception, and MobileNetV3 Large. For the purposes of the study, three distinct datasets were used, including images collected in a controlled environment and those collected in an industrial and real setting. The models have been trained using transfer-learning and evaluated based on their accuracy, highlighting the potential of CNNs for industrial application within processing units.*

**Resumo.** *Este trabalho apresenta uma análise comparativa do desempenho de quatro arquiteturas de Redes Neurais Convolucionais (CNNs) na classificação binária de tomates de mesa, sendo elas: InceptionV3, ResNet-50, Xception e MobileNetV3 Large. Para fins do estudo, foram utilizados três conjuntos diferentes de dados, incluindo imagens coletadas em ambiente controlado e coletas em ambiente real industrial. Os modelos foram treinados via transfer-learning e avaliados quanto à acurácia, indicando o potencial das CNNs para aplicação industrial em unidades de beneficiamento.*

## 1. Introdução

O Brasil figura entre os dez maiores produtores de tomates do mundo [FAO 2025], sendo o estado de Goiás o maior produtor nacional em quantidade produzida e Minas Gerais o terceiro maior estado produtor do fruto [IBGE 2024]. Nesse cenário de produção em larga escala, a pós-produção do fruto apresenta diversos desafios, como a separação daqueles com anomalias realizada na fase inicial do beneficiamento que considera variáveis como a forma irregular do fruto, cor, aderências ou irregularidades superficiais evidentes e presença de corpos estranhos em processo [Becker et al. 2016]. Dessa forma, a qualidade do produto a ser disponibilizado para o cliente, seja ele indústria ou consumidor final, depende dos métodos empregados, sendo hoje predominante a classificação manual [Ferreira et al. 2004], que está sujeita à subjetividade do operador e estabelece limitações à competitividade do produtor.

O presente trabalho aplicou e avaliou métodos de visão computacional e machine learning, mais especificamente Redes Neurais Convolucionais (CNNs), para identificar atributos de tomates considerados “bons” e “ruins” e classificar as amostras dos datasets

definidos, comparando os resultados de quatro arquiteturas selecionadas. Este trabalho também possibilitou embasamento teórico e experimental preliminares buscando viabilizar a evolução para modelos de aplicação industrial, permitindo o desenvolvimento de tecnologia nacional voltada ao fortalecimento do agronegócio e indústria nacionais.

## 2. Fundamentação Teórica

As CNNs são uma classe de redes neurais artificiais profundas do tipo *feed-forward* amplamente utilizadas na agricultura [Kamilaris and Prenafeta-Boldú 2018], demonstrando excelente performance em diversos problemas de visão computacional e machine learning e utilizadas amplamente na classificação de tomates, como visto em [Mputu et al. 2024] e em [Eloge 2025]. Este estudo comparou o desempenho de quatro arquiteturas de CNNs com pesos pré-treinados na ImageNet (*transfer-learning*), sendo elas InceptionV3, Resnet-50, Xception e MobileNetV3 Large, avaliando o desempenho de cada modelo nos datasets selecionados.

Quanto às arquiteturas selecionadas para o estudo, essas arquiteturas vêm sendo usadas como benchmarking na literatura e cada uma delas apresenta características particulares que justificam a sua comparação. A arquitetura InceptionV3 consiste em um modelo desenvolvido para escalar redes neurais com eficiência computacional, mantendo um número baixo de parâmetros, através da fatoração de convoluções [Szegedy et al. 2016]. A Resnet-50, por sua vez, propõe uma arquitetura de aprendizado residual, que busca facilitar o treinamento de redes neurais profundas, permitindo atingir maior profundidade que redes anteriores, como a VGG, e apresentando menor complexidade [He et al. 2016].

A arquitetura Xception, por outro lado, deriva da premissa inicial da Inception, utilizando camadas convolucionais separáveis (*depthwise separable convolutions*), contando com 36 camadas convolucionais em sua base de extração de atributos [Chollet 2017]. Por outro lado, a arquitetura das MobileNets apresenta modelos computacionalmente eficientes, possibilitando o desenvolvimento de sistemas em dispositivos móveis ou sistemas embarcados [Howard et al. 2017] [Howard et al. 2019].

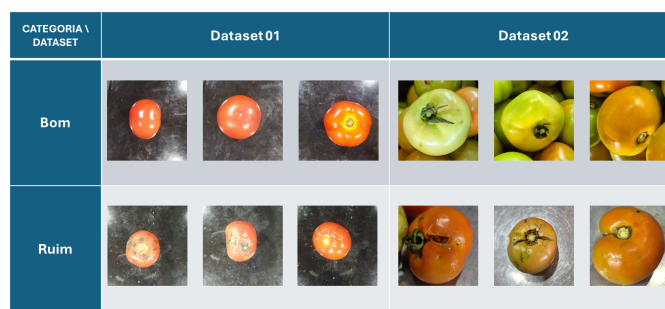
## 3. Metodologia

Para fins de comparação do desempenho de cada uma das arquiteturas, foram utilizados três datasets diferentes, com classes binárias de tomates (“bom” e “ruim”), conforme definidas pelos autores do Dataset 01 [Rahaman and Sarkar 2024] e conforme classificação realizada por especialistas para o Dataset 02, sendo detalhados na tabela 1.

**Tabela 1. Resumo dos datasets utilizados**

| <b>Dataset</b>                             | <b>Total de imagens</b> | <b>“Bom”</b> | <b>“Ruim”</b> |
|--------------------------------------------|-------------------------|--------------|---------------|
| Dataset 01 [Rahaman and Sarkar 2024]       | 1000                    | 500          | 500           |
| Dataset 02 (autores [Andrade et al. 2026]) | 349                     | 168          | 181           |
| Dataset 03 (junção dos Datasets 01 e 02)   | 1349                    | 668          | 681           |

A escolha de diferentes datasets busca entender como os modelos se comportam diante de quantidades diferentes de dados, bem como entender a capacidade de generalização do modelo para diferentes tipos de defeitos em tomates de mesa. Os datasets 01 e 02 podem ser visualizados na Figura 1 e estão disponíveis junto aos códigos utilizados em repositório do GitHub [Andrade et al. 2026].



**Figura 1. Visualização de amostras dos datasets**

Para realização dos experimentos, os códigos utilizados foram gerados em linguagem Python (3.14) utilizando a API Keras e as bibliotecas Pandas, Numpy, Matplotlib, Seaborn, TensorFlow e Scikit-Learn com o uso da LLM Gemini 3.1 Pro, sendo o código revisado pelos autores em termos de lógica do pipeline, separação treino/teste e ausência de *data-leakage*. O ambiente de execução utilizado foi o ambiente T4 do Google Colab, devido às limitações do ambiente local.

Além disso, cada um dos datasets foi previamente separado na proporção 80% treino, 20% teste, de forma aleatória e mantendo o equilíbrio entre as classes com o uso da biblioteca ScikitLearn, sendo gerado um gabarito em .csv para cada um dos datasets para fins de reprodutibilidade. Na etapa de treinamento, 20% das amostras do conjunto de treinamento foram utilizadas para validação. Foram também utilizadas técnicas de *data augmentation* para aumentar a diversidade dos dados de treinamento, com os seguintes parâmetros: “rotation\_range=20”, “width\_shift\_range=0.2”, “height\_shift\_range=0.2”, “horizontal\_flip=True”, “zoom\_range=0.2”.

Quanto aos hiperparâmetros, foram definidas 200 épocas de treinamento e *batch-size* = 32, utilizando *early stopping* com *patience* = 5, evitando o *overfitting* dos modelos. Todos os datasets foram processados pelas quatro arquiteturas selecionadas separadamente, totalizando doze experimentos realizados.

#### 4. Resultados e Discussão

Para fins de avaliação da performance dos modelos, a métrica de acurácia é a mais popularmente utilizada na literatura [Kamilaris and Prenafeta-Boldú 2018], sendo esta adotada e apresentada na Tabela 2.

**Tabela 2. Acurácia apresentada pelas arquiteturas**

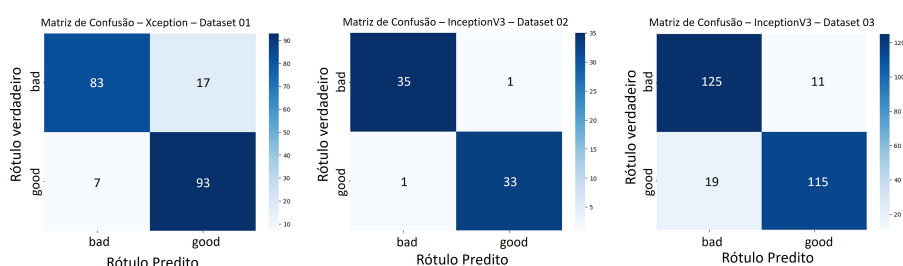
| CNN               | Dataset 01 | Dataset 02 | Dataset 03 |
|-------------------|------------|------------|------------|
| InceptionV3       | 86,50%     | 97,14%     | 88,89%     |
| ResNet-50         | 87,50%     | 94,29%     | 88,52%     |
| Xception          | 88,00%     | 94,29%     | 87,41%     |
| MobileNetV3 Large | 86,00%     | 95,71%     | 88,52%     |

A análise comparativa mostrou que as quatro arquiteturas apresentaram maior acurácia quando submetidas ao Dataset 02, comportamento que pode ser atribuído ao baixo número de amostras (70 amostras no conjunto de teste) ou à grande diferença visível

entre as classes apresentadas neste dataset, conforme observado na Figura 1. Dessa forma, para avaliar essas hipóteses, bem como viabilizar eventual modelo para aplicação em ambiente de beneficiamento, a coleta de mais amostras em ambiente de produção se faz fundamental, buscando constar no dataset maior número de amostras dos defeitos considerados na classificação manual.

Outrossim, quanto aos resultados observados para o dataset 01, em que pese ter sido identificadas acurácias próximas a 90%, considerando que a acurácia apresentada para esse conjunto foi significativamente menor se comparado ao dataset 02 (variando de 6,29% a 10,64%) deve-se destacar as dificuldades que podem ter sido enfrentadas pelos modelos quando da classificação deste dataset. A principal delas está atrelada às características intrínsecas do fruto que ocasionam brilho especular quando da incidência de iluminação, podendo comprometer a extração de características pelo modelo, incluindo a ocultação de defeitos (gerando falsos negativos) ou a classificação errônea de um tomate “bom” como defeituoso (falso positivo). Esse fator dificilmente poderá ser amenizado pelas características do ambiente industrial, devendo ser considerado quando do *fine-tuning* dos modelos, caso o aumento do volume amostral não melhore seu desempenho.

Os resultados experimentados no dataset 03, por sua vez, indicam que os problemas notados na classificação dos datasets 01 e 02 poderão ser contornados/amenizados com o aumento do espaço amostral através da coleta de mais amostras das classes estudadas. Isso pode ser observado pela performance apresentada pelas arquiteturas no dataset 03, que se situaram, para a maioria dos modelos, entre o apresentado nos demais datasets. Para fins de análise qualitativa, na Figura 2 são apresentadas as matrizes de confusão para o modelo de melhor performance para cada dataset apresentado.



**Figura 2. Matriz de confusão com maior acurácia por dataset**

## 5. Conclusão

Com base no observado, conclui-se que as CNNs apresentam grande potencial para aplicação na classificação de tomates em ambiente de beneficiamento. Como trabalhos futuros, destaca-se a necessidade de, além da acurácia, comparar métricas como velocidade de classificação e custo computacional, dada a necessidade de aplicação em ambiente industrial (esteira de classificação) e na potencial necessidade de sistema embarcado integrado ao equipamento.

Nesse sentido, em que pese a MobileNet não ter sido a arquitetura de melhor desempenho nos datasets propostos (demonstrado pela Tabela 2), esta apresentou a segunda melhor performance no dataset 02 e ainda se demonstra relevante para aplicação industrial, dada sua arquitetura compatível e otimizada para sistemas embarcados e possibilidade de adaptação de parâmetros, devendo ter seu desempenho testado com a coleta

de mais amostras, bem como com o *fine-tuning*, com ênfase no equilíbrio entre acurácia e custo computacional.

## Referências

- Andrade, E. L. d., Lima, D. A., Souza, J. R. d., and Nomura, S. (2026). brazilian-tomato-dataset-v1. <https://github.com/EduardoLimas/brazilian-tomato-dataset-v1>. Acesso em: 29 de maio de 2026.
- Becker, W. F., Wamser, A. F., Feltrim, A. L., Suzuki, A., Santos, J. d., Valmorbidia, J., Hahn, L., Marcuzzo, L. L., and Mueller, S. (2016). Sistema de produção integrada para o tomate tutorado em santa catarina. *Florianópolis: Epagri*, 1:151.
- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1251–1258.
- Eloge, K. W. (2025). Integrating resnet-50 and vision transformer architectures for robust and efficient tomato fruit ripeness classification. *Advances in Engineering Research: Possibilities and Challenges*, 1(2):102–112.
- FAO (2025). Faostat: Crops and livestock products. <https://www.fao.org/faostat/en/#data/QCL/visualize>. Acesso em: 21 abr. 2026.
- Ferreira, S. M. R., Freitas, R. J. S. d., and Lazzari, E. N. (2004). Padrão de identidade e qualidade do tomate (*lycopersicon esculentum* mill.) de mesa. *Ciência Rural*, 34(1):329–335.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.
- Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., and Adam, H. (2019). Searching for mobilenetv3.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- IBGE (2024). Produção agropecuária: Tomate - brasil. <https://www.ibge.gov.br/explica/producao-agropecuaria/tomate/br>. Acesso em: 21 abr. 2026.
- Kamilaris, A. and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147:70–90.
- Mputu, H. S., Abdel-Mawgood, A., Shimada, A., and Sayed, M. S. (2024). Tomato quality classification based on transfer learning feature extraction and machine learning algorithm classifiers. *IEEE Access*, 12:8283–8295.
- Rahaman, T. and Sarkar, T. (2024). TOMATO. <https://data.mendeley.com/datasets/hfyth5t3gg/1>. Acessado em: 18 de abril de 2026.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826.