



INTERPRETABILIDADE PARA CLASSIFICAÇÃO PÓS-COLHEITA DE FRUTOS BASEADO EM DEEP-LEARNING

Palavras-Chave: LIME, VISION TRANSFORMER, EFFICIENTNET

Autores:

JOÃO VITOR DE MELO OLIVEIRA – FEAGRI, UNICAMP

Prof^ª. Dr^ª. JULIANA APARECIDA FRACAROLLI (orientadora) – FEAGRI, UNICAMP

INTRODUÇÃO:

A detecção de danos e doenças em frutos tem sido uma fonte de pesquisas, principalmente com aplicação de aprendizado profundo. Com aprendizado de máquina é possível até mesmo substituir a atividade de seleção e classificação de frutos feita de forma visual, que muitas vezes representa uma atividade cansativa aos trabalhadores. Porém, pesquisas para obter altas acurácias em modelos de aprendizado profundo têm gerado soluções complexas, cujas tomadas de decisão nem sempre são entendidas por humanos. Com o objetivo de apresentar informações para viabilizar a interpretação dos resultados, são adicionadas as técnicas de interpretabilidade. Entendendo a lógica e os recursos usados nas decisões dos modelos, é possível verificar e validar melhor os resultados, melhorar o algoritmo e os dados de treinamento e extração de informação. Dessa forma, neste trabalho foi aplicada a técnica de interpretabilidade LIME (Local Interpretable Model-Agnostic Explanations) em três redes neurais artificiais, sendo elas: Vision Transformer, EfficientNet-B5 e Graph Neural Network (GNN) na modelagem da classificação de frutos frescos e frutos danificados. Foi possível validar o uso da técnica LIME para essa aplicação e proporcionar a visualização dos principais aspectos, características e variáveis que contribuem para a classificação de imagens de frutos obtidas do Repositório 'An extensive dataset for successful recognition of fresh and rotten fruits' (SULTANA; JAHAN; UDDIN, 2022).

METODOLOGIA:

BANCO DE IMAGENS: O banco de imagens utilizado foi desenvolvido por (SULTANA; JAHAN; UDDIN, 2022) intitulado "Fresh and Rotten Fruits Dataset" e está disponível para ser utilizado na construção, treinamento, análise e para comparação de diversos algoritmos de deep learning para reconhecimento de frutos frescos e danificados. É composto por 3.200 imagens originais e 12.335 imagens aumentadas, cada imagem composta por 512x512 pixels, e 16 classes de frutos.

PRÉ-PROCESSAMENTO: Os autores (SULTANA; JAHAN; UDDIN, 2022) fizeram (a) augmentation, (b) redimensionamento das imagens, (c) divisão do conjunto de dados.

Foram selecionadas sem critério, em todas as 16 classes, 70% do total de cada uma para a fase de treino e 30% para a fase de teste (DA SILVA FERREIRA et al., 2024).

Para treinamento do modelo GNN foram usados os recursos computacionais do “Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD-SP)”. Para o Vision Transformer (DOSOVITSKIY, et al. 2020) foram usadas usadas as GPU’s T4 e L4 e por fim, para o EfficientNet-B5 (TAN e LE, 2019) foi usada a TPU v6e1, ambas do Google Colaboratory.

METODOLOGIA DE INTERPRETABILIDADE: Foi aplicada a técnica LIME (RIBEIRO, et al. 2016), através da qual foi possível ver por meio de demarcações nas frutas a forma como o Vision Transformer e o EfficientNet-B5 trabalharam nas predições. Foram geradas imagens por meio do LIME com demarcações e mapas de calor.

RESULTADOS E DISCUSSÃO:

DESEMPENHO DOS MODELOS: O desempenho de cada um dos modelos foi medido já na fase de treino por meio da acurácia (accuracy) e perda (loss). O Vision Transformer ao longo do treino não conseguiu reduzir o número de perda e nem aumentar a acurácia Figura 1, mostrando que o modelo não foi eficiente e já era previsto com base nestes dados que faria uma predição errônea (Figura 3). Ao contrário do ViT, o EfficientNet-B5 conseguiu durante a fase de treino reduzir os números de perda e aumentar a acurácia de forma significativa e satisfatória (Figura 2), refletindo essa eficiência na predição correta da Figura 6. Além disso, é possível supor que com maior número de épocas, o EfficientNet aperfeiçoaria ainda mais sua acurácia, porém foi determinado 10 épocas para este modelo e para o ViT para que não houvesse vantagem de um sobre o outro neste parâmetro.

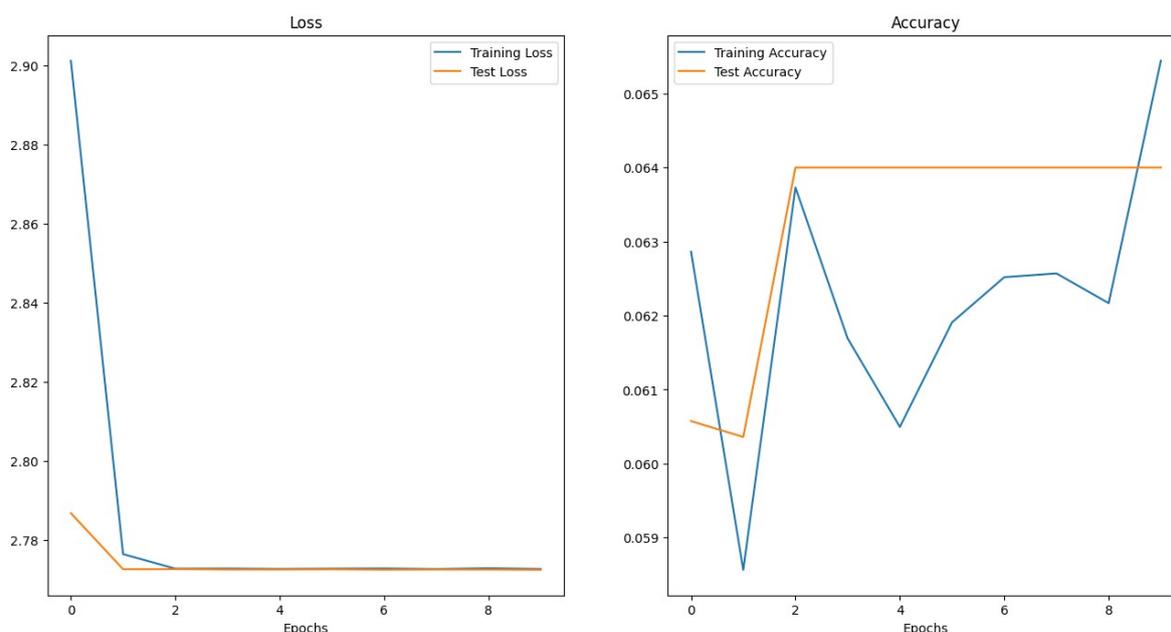


Figura 1: Gráfico de Perda e Acurácia (ViT)

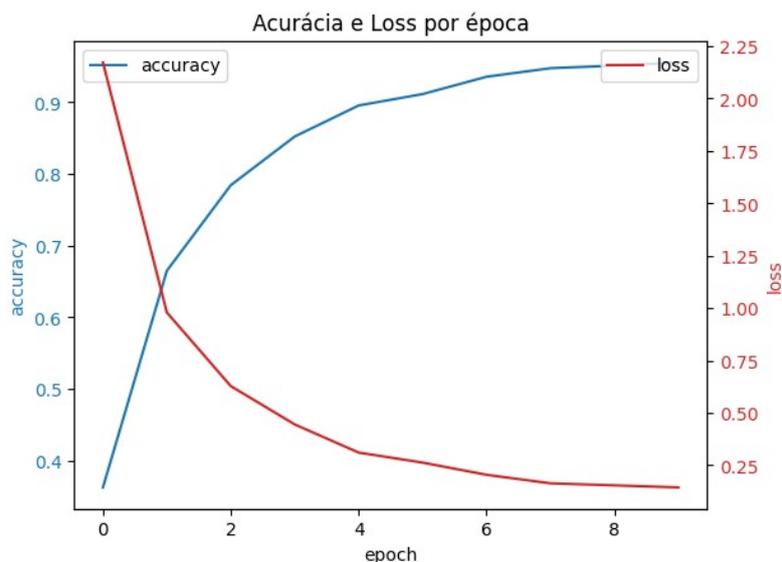


Figura 2: Gráfico de Perda e Acurácia (EfficientNet-B5)

Vision Transformer e LIME: O Vision Transformer, como foi apresentado nas métricas de acurácia e perda, não obteve uma boa predição, como é possível ver na Figura 3, em que errou a classificação da banana, definindo-a como uma laranja podre com uma baixa probabilidade de 6,5%. Na Figura 4 é apresentada uma imagem gerada pelo LIME, que mostrou como o ViT interpretou a imagem original no momento da predição, sendo que é possível notar as marcações no fundo da imagem, mostrando que isso foi considerado no momento de classificar a imagem e poucas marcações na banana, ou seja, o modelo está considerando os elementos de fundo para a classificação, sendo possível visualizar nesta imagem elementos rosáceos que influenciaram negativamente no momento da predição. Na Figura 5 está o mapa de calor, sendo que as partes em vermelho indicam as regiões que contribuíram negativamente na previsão, e as azuis as regiões que contribuíram positivamente. E por este mapa se confirma a hipótese de que o Vision Transformer está considerando os elementos de fundo e não somente a fruta a ser analisada, inclusive considerando partes da própria fruta como elementos negativos no momento da análise.



Figura 3: Teste de reconhecimento de imagem (ViT)

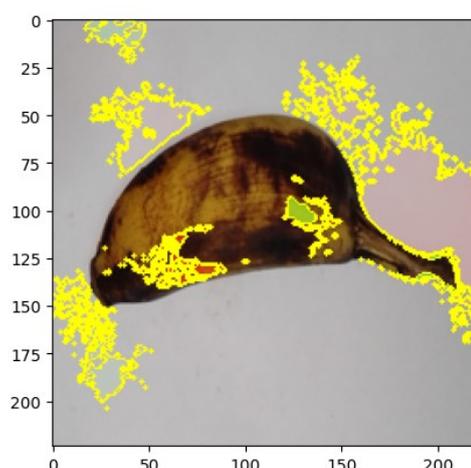


Figura 4: Marcações LIME (ViT)

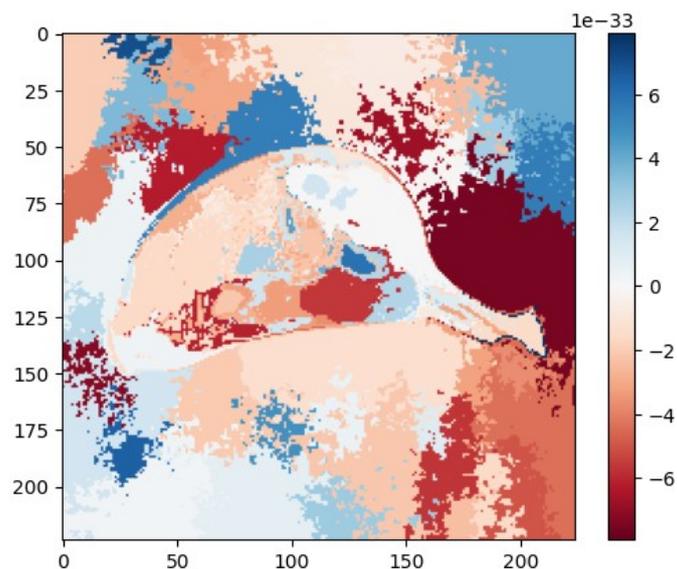


Figura 5: Mapa de Calor LIME (ViT)

EfficientNet-B5 e LIME: O EfficientNet se mostrou um modelo mais robusto desde o seu treino em relação ao Vision Transformer, como se pode visualizar em sua predição correta com probabilidade de 100% na Figura 6. Na Figura 7 o LIME evidenciou as marcações que contribuíram positivamente na classificação, em que se nota que há apenas elementos no morango e nenhum elemento de fundo sendo considerado. Na Figura 8 estão marcadas as regiões positivas e negativas, em comparação com a Figura 7, é possível determinar que as regiões negativas foram duas de fundo e uma folha na parte superior, que não influenciaram a predição. No mapa de calor (Figura 9) é confirmado que o modelo conseguiu identificar corretamente o morango, em que em sua totalidade está marcado em tons de azul, já os elementos de fundo em sua maioria estão marcados em vermelho.

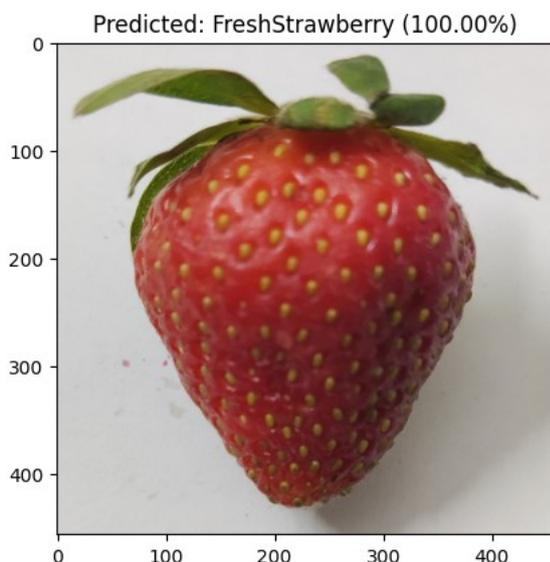


Figura 6: Teste de reconhecimento de imagem (EfficientNet-B5)



Figura 7: Contorno de regiões positivas no fruto (EfficientNet-B5)

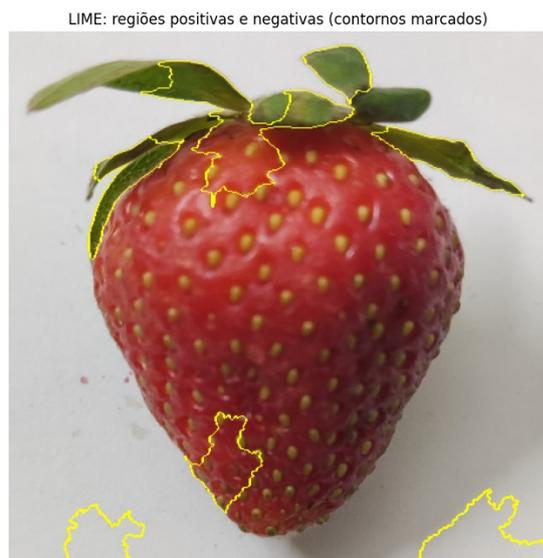


Figura 8: Contorno de regiões positivas e negativas no fruto (EfficientNet-B5)

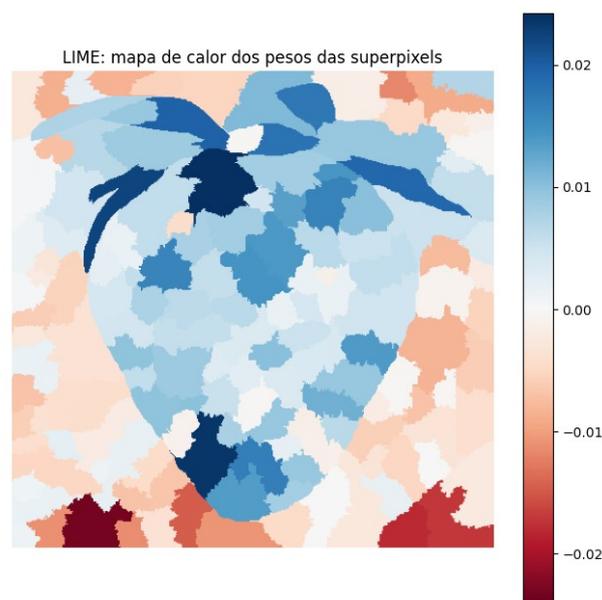


Figura 9: Mapa de Calor (EfficientNet-B5)

CONCLUSÕES:

Foi possível validar o uso da técnica de interpretabilidade LIME para a classificação pós-colheita de frutos baseada em deep-learning e proporcionar a visualização dos principais aspectos, características e variáveis que contribuem para a classificação de imagens de frutos obtidas do Repositório ‘An extensive dataset for successful recognition of fresh and rotten fruits’ (SULTANA; JAHAN; UDDIN, 2022). Pode-se identificar as partes do objeto que mais influenciam na decisão dos modelos, promovendo informações importantes para avanços na classificação, aplicando a interpretabilidade às redes de classificação de frutos.

BIBLIOGRAFIA

SULTANA, N.; JAHAN, M.; UDDIN, M. S. **An extensive dataset for successful recognition of fresh and rotten fruits**. Data in Brief, v. 44, p. 108552, out. 2022.

DOSOVITSKIY, A. et al. **An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale**. 2020.

TAN, M.; LE, Q. V. **EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks**. 2019

DA SILVA FERREIRA, M. V. et al. **Deep computer vision system and explainable artificial intelligence applied for classification of dragon fruit (Hylocereus spp.)**. Scientia Horticulturae, v. 338, p. 113605, dez. 2024.

RIBEIRO, Marco Tulio; SINGH, Sameer; GUESTRIN, Carlos. **“Why Should I Trust You?”: explaining the predictions of any classifier**. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 22., 2016, San Francisco. Proceedings... New York: Association for Computing Machinery, 2016. p. 1135–1144. Disponível em: <https://doi.org/10.1145/2939672.2939778>.