

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

## Transferência de aprendizado na avaliação da qualidade de maçãs

**João Paulo Herrera**

Monografia - MBA em Inteligência Artificial e Big Data



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**João Paulo Herrera**

## **Transferência de aprendizado na avaliação da qualidade de maçãs**

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Fernando Pereira dos Santos

**Versão original**

**São Carlos**

**2023**

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTA TRABALHO,  
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E  
PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi, ICMC/USP, com os dados  
fornecidos pelo(a) autor(a)

S856m	Herrera, João Paulo Transferência de aprendizado na avaliação da qualidade de maçãs / João Paulo Herrera ; orientador Fernando Pereira dos Santos. – São Carlos, 2023. 41 p. : il. (algumas color.) ; 30 cm.  Monografia (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universi- dade de São Paulo, 2023.  1. LaTeX. 2. abnTeX. 3. Classe USPSC. 4. Editoração de texto. 5. Normalização da documentação. 6. Tese. 7. Disserta- ção. 8. Documentos (elaboração). 9. Documentos eletrônicos. I. dos Santos, Fernando Pereira, orient. II. Título.
-------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**João Paulo Herrera**

# **Transferência de aprendizado na avaliação da qualidade de maçãs**

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

**Original version**

**São Carlos  
2023**



## RESUMO

Herrera, J. P. **Transferência de aprendizado na avaliação da qualidade de maçãs**. 2023. 41p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

Dentre as diversas técnicas de END empregadas na indústria, a espectroscopia é uma das mais significativas no âmbito da visão computacional. Através dela, é possível observar detalhes que não podem ser vistos em imagens RGB. No entanto, a aquisição de equipamentos capazes de capturar imagens espectrais é cara e de difícil acesso, o que resulta em bases de dados com poucas amostras em outras faixas além da visível. Neste trabalho, é explorado o papel da transferência de aprendizado a partir de uma pequena base de dados no infravermelho próximo de maçãs para potencializar a classificação de maçãs em RGB, que são de muito mais fácil acesso. A faixa espectral utilizada é capaz de evidenciar os defeitos em maçãs que são difíceis de serem vistos nas imagens RGB, dificultando, assim, a identificação dos principais padrões que caracterizam uma maçã de má e boa qualidade. Para comparação, são realizadas transferências de aprendizado a partir de bases de dados de diferentes frutos frente às imagens espectrais. Os resultados indicaram que uma base composta de imagens RGB de limões apresentou maior impacto do que as próprias imagens de maçãs no infravermelho próximo.

**Palavras-chave:** Transferência de aprendizado. Imagens espectrais. Maçãs. Classificação.



## **ABSTRACT**

Herrera, J. P. **Transfer learning in apple quality assessment**. 2023. 41p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

Among the various Non-Destructive Evaluation (NDE) techniques employed in the industry, spectroscopy holds significant importance in the field of computer vision. It allows us to observe details that are not visible in RGB images. However, acquiring equipment capable of capturing spectral images is expensive and often challenging, resulting in databases with limited samples beyond the visible range. In this study, we explore the potential of transfer learning from a small database of near-infrared apple images to improve the classification of RGB images, which are more readily available. The selected spectral range can highlight defects in apples that are challenging to see in RGB images, making it difficult to identify the key patterns characterizing good or bad quality apples. For comparison, transfer learning is performed using databases of different fruits against spectral images. The results indicate that a database composed of RGB images of lemons had a more significant impact than images of apples in the near-infrared themselves.

**Keywords:** Transfer Learning. Spectral images. Apples. Classification.



## LISTA DE FIGURAS

Figura 1 – Faixas do espectro eletromagnético. . . . .	21
Figura 2 – Exemplificação da operação de Max pooling em uma matrix de 3x3 considerando uma janela de tamanho 2x2 e stride 1. . . . .	23
Figura 3 – Esquematização de um neurônio artificial . . . . .	24
Figura 4 – Arquitetura de modelo de rede residual de 34 camadas proposta por He <i>et al.</i> (2016). FONTE: He <i>et al.</i> (2016) . . . . .	25
Figura 5 – Exemplo de amostras da base de dados "Lemon Quality Dataset". As imagens a) e c) correspondem a limões de boa qualidade, e as imagens b) e d), limões de má qualidade . . . . .	28
Figura 6 – Exemplo de amostras da base de dados "Mangas Embrapa". Em a), uma imagem inteira com identificação da fruta. Em b), c) e d), exemplares de mangas em diferentes estágios de maturação. . . . .	29
Figura 7 – Exemplo de amostras da base de dados "Maçãs 1.0". Na primeira coluna, formada pelas amostras a) e c), é possível visualizar duas maçãs capturadas com a câmera HQ Cam. Em b) e d) as mesmas frutas obtidas utilizando a câmera espectral Mako G-125B. . . . .	30



## LISTA DE TABELAS

Tabela 1	– Algoritmos e parâmetros utilizados para aumento de dados durante a etapa de treinamento . . . . .	32
Tabela 2	– Treinamentos realizados. A coluna "Base de dados" indica a base de dados utilizada para treinar o classificador e a coluna "Pesos iniciais", o conjunto de pesos utilizado como ponto de partida. . . . .	33
Tabela 3	– Resultados dos experimentos de classificação de qualidade de frutas . .	34
Tabela 4	– Resultados dos experimentos de classificação de qualidade de frutas . .	35



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
<b>1.1</b>	<b>Organização do texto</b>	<b>19</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>21</b>
<b>2.1</b>	<b>Luz e o espectro eletromagnético</b>	<b>21</b>
<b>2.2</b>	<b>Conceitos básicos de aprendizado de máquina</b>	<b>22</b>
2.2.1	Redes Neurais Convolucionais	22
2.2.2	Pooling	23
2.2.3	Camadas Densas	23
2.2.4	Normalização por lotes	24
2.2.5	Dropout	24
2.2.6	ResNets	25
2.2.7	Transferência de Aprendizado	26
<b>3</b>	<b>AVALIAÇÃO EXPERIMENTAL</b>	<b>27</b>
<b>3.1</b>	<b>Aquisição de Dados</b>	<b>27</b>
3.1.1	Lemon Quality Dataset	27
3.1.2	Mangas Embrapa	27
3.1.3	Maças 1.0	28
<b>3.2</b>	<b>Anotação</b>	<b>30</b>
<b>3.3</b>	<b>Pré-processamento</b>	<b>31</b>
<b>3.4</b>	<b>Configuração Experimental</b>	<b>32</b>
<b>3.5</b>	<b>Avaliação do aprendizado</b>	<b>33</b>
<b>3.6</b>	<b>Resultados e Discussões</b>	<b>34</b>
<b>4</b>	<b>CONCLUSÕES</b>	<b>37</b>
	<b>Referências</b>	<b>39</b>



## 1 INTRODUÇÃO

Nos últimos anos, programas de computador baseados em técnicas de visão computacional ganharam destaque no setor industrial global. Com o advento da indústria 4.0, modelos de aprendizado de máquina têm sido explorados em larga escala dado seu alto potencial em realizar tarefas humanamente exaustivas, custosas e até mesmo perigosas (LIU *et al.*, 2023). Aliados aos benefícios que a tecnologia baseada em inteligência artificial tem a oferecer, avanços realizados na microeletrônica permitiram não apenas o aprendizado de modelos profundos em tempo viável, mas também a otimização de processos industriais por meio de ensaios não destrutivos (END), isto é, inspeções realizadas sem a necessidade de danificar a estrutura de um objeto.

Dentre as diversas técnicas de END empregadas na indústria, a espectroscopia é uma das mais significativas no âmbito da visão computacional. Isso se deve à capacidade dos sensores espectrais de obter informações em um nível mais amplo do espectro eletromagnético, além da faixa perceptível ao olho humano. Essas informações, também chamadas de assinatura espectral, são únicas e estão estritamente associadas à composição química dos objetos. (YAN; REN; SUN, 2022; MONAKHOVA *et al.*, 2020; BACCA; MARTINEZ; ARGUELLO, 2023). A partir de uma análise minuciosa da assinatura espectral, é possível identificar diferentes tipos de materiais, objetos e processos físicos (BACCA; MARTINEZ; ARGUELLO, 2023). Em vista disso, são amplas as aplicações envolvendo sensores espectrais que podem ser empregadas na indústria, como por exemplo na área de alimentos (XIANG *et al.*, 2022; THIELE *et al.*, 2023), metalurgia (YAN; REN; SUN, 2022), agricultura (LI; ZHANG; SHEN, 2020; AGYEMAN *et al.*, 2022; WU *et al.*, 2023) e médica (HERRERA *et al.*, 2023; SEIDLITZ *et al.*, 2022).

As câmeras multi e hiperespectrais são as principais fontes de aquisição de imagens espectrais. A principal diferença entre esses dois tipos de câmeras está relacionada com a quantidade de bandas que cada uma é capaz de capturar. As câmeras multiespectrais, em média, capturam de uma a quatro bandas distintas, enquanto as câmeras hiperespectrais podem capturar dezenas ou até centenas delas (BACCA; MARTINEZ; ARGUELLO, 2023; VIVONE, 2023). Apesar do grande potencial que a tecnologia oferece, as câmeras espectrais possuem elevado custo, necessidade de alto poder computacional para processamento de suas imagens e baixa resolução espacial, isto é, incapacidade de distinguir objetos muito próximos ou detalhes finos. Devido a isso, a disponibilidade de bases de dados públicas de diferentes aplicações envolvendo imagens espectrais é limitada. Dentre as disponíveis, poucas se apresentam como alternativas viáveis para produzir conhecimento em modelos de deep learning para tarefas como detecção, classificação e segmentação.

Visando mitigar tais problemas, técnicas já difundidas na literatura podem ser

apontadas como alternativas para auxiliar no processo de adaptação do conhecimento a partir de uma base de dados com poucas amostras. Xishuai *et al.* (2016) sugerem o uso de transferência de aprendizado em imagens RGB para classificação de cenas em imagens NIR. Os resultados indicaram uma eficácia no uso da técnica em comparação com outros métodos já existentes. Anos depois, os mesmos autores novamente propuseram o uso de transferência de aprendizado para aliviar a escassez de imagens nas bandas RGB-NIR, além da destilação de conhecimento para realizar a fusão das principais características extraídas para tarefa de classificação de cenas (PENG *et al.*, 2018).

Estratégias alternativas, como a geração de imagens sintéticas em bandas NIR a partir de imagens RGB por meio de transferência de estilo, também têm apresentado grande destaque na comunidade científica. Stump *et al.* (2022) explora técnicas de transferência de estilos para, a partir de datasets diversos compostos por imagens RGB, gerar representações virtuais na banda NIR para treinamento de detectores de objetos. Os autores também propuseram um novo modelo baseado em meta-learning que se mostrou mais robusto dentre as técnicas abordadas. Xu, Tang and Pang (2022) propõem uma nova abordagem para detecção de falhas em motores de indução utilizando técnicas de few-shot learning em imagens térmicas infravermelhas. A partir de um módulo de atenção desenvolvido denominado coordinate attention feature extraction module (CAFEM), foi possível explorar as características espaciais de forma eficaz, permitindo, assim, que o modelo CAPNet, variação proposta do modelo ProtoNet, atingisse acurácia expressiva frente à outros modelos difundidos na literatura.

Apesar dos significativos avanços relatados nos últimos anos no treinamento de modelos profundos com um número limitado de amostras, poucos estudos exploraram o tema tratando-se de imagens além do espectro visível. Esse cenário pode ser atribuído aos custos elevados dos dispositivos para aquisição dessas imagens e, em alguns casos, à necessidade de conhecimento prévio das propriedades físicas e químicas dos objetos ao escolher os sensores e filtros ideais para que suas principais características sejam evidenciadas na imagem. Além disso, é sabido que uma porcentagem significativa dos modelos profundos necessitam, além de um grande número de amostras, um alto poder computacional. Esses fatores restringem o acesso à tecnologia fora dos centros acadêmicos e empresas de tecnologia, capazes de realizar altos investimentos em infraestrutura. Dentre esses e outros motivos, é que a transferência de aprendizado desempenha um importante papel no acesso à tecnologia, haja visto que o conhecimento adquirido anteriormente pode ser reaproveitado para treinamento em diferentes contextos.

Dito isso, algumas questões podem ser levantadas em relação ao papel da transferência de aprendizado no treinamento de redes neurais profundas a partir de imagens espectrais, como "Um pré-treinamento de objetos iguais em bandas diferentes possui maior impacto do que diferentes objetos em bandas iguais?", ou então "Como sucessivas

transferências de aprendizado afetam na robustez de um modelo de classificação?". Neste sentido, este trabalho tem como objetivo explorar o papel da transferência de aprendizado em diferentes cenários envolvendo imagens no espectro visível e infravermelho próximo. Para tais experimentos, foi desenvolvido uma base de dados composta por 70 imagens RGB e NIR de maçãs tiradas em ambiente com iluminação controlada. Os experimentos foram conduzidos utilizando o modelo de redes neurais convolucionais ResNet50.

## **1.1 Organização do texto**

No Capítulo 2 são apresentadas noções básicas relacionadas à radiação eletromagnética e sua interação com objetos do mundo real, base para o entendimento das imagens infravermelhas utilizadas nos experimentos. Além disso, são percorridos os principais conceitos envolvendo o modelo ResNet50, utilizado neste trabalho para classificação de frutas. No Capítulo 3 descreve a metodologia utilizada para aquisição das imagens de maçãs e dos experimentos realizados para avaliar o impacto da transferência de aprendizado. Por fim, os resultados obtidos são apresentados na Seção 3.6, seguidos pelas conclusões, percorridas no Capítulo 4.



## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são percorridos os principais conceitos que fundamentam a natureza do problema e da solução proposta.

### 2.1 Luz e o espectro eletromagnético

A luz possui uma natureza dual, comportando-se ora como onda - pois possui um comprimento de onda eletromagnético - ora como partícula. Embora os seres humanos possam interpretar informações apenas em uma pequena faixa do espectro eletromagnético, câmeras e outros dispositivos especializados têm a capacidade de processar uma gama muito mais ampla de ondas, incluindo faixa de luz ultravioleta e infravermelha. (IKEUCHI *et al.*, 2020)

O ultravioleta abrange comprimentos de onda entre 200 e 400 nanômetros, com ondas abaixo de 320 nanômetros sendo consideradas prejudiciais aos seres humanos. Acima desse limite, é utilizado como fonte de luz para excitar substâncias fluorescentes. A faixa visível do espectro se estende de 380 a 760 nanômetros. Além disso, temos o infravermelho, que pode ser subdividido em três faixas principais: infravermelho próximo (760 a 1400 nm), infravermelho médio (1400 a 15000 nm) e infravermelho distante (15 a  $1 \times 10^6$  nm). A 1 ilustra o espectro eletromagnético e as diferentes faixas de ondas. À esquerda, concentram-se raios de menor comprimento de onda. Mais à direita, os comprimentos de onda aumentam.

O infravermelho próximo é amplamente utilizado em aplicações de escaneamento

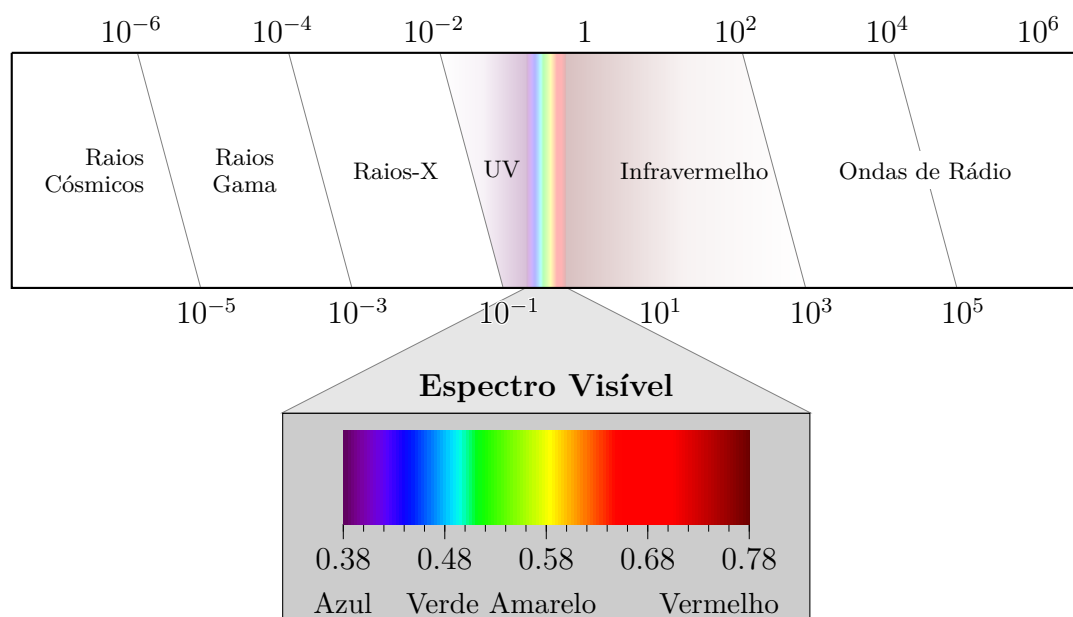


Figura 1 – Faixas do espectro eletromagnético.

3D, controle remoto e comunicações. Embora invisíveis aos olhos humanos, sensores de silício podem detectar esses comprimentos de onda. Isso difere das ondas do infravermelho distante, que são geralmente absorvidas por moléculas de água na atmosfera. A região do infravermelho próximo (NIR) é especialmente útil na biofotônica devido à sua capacidade de penetrar mais profundamente nos tecidos em comparação com a faixa visível (VIS), permitindo estudos com especificidade química (POPESCU, 2019).

## 2.2 Conceitos básicos de aprendizado de máquina

Até o tardar dos anos 2000, as técnicas de machine learning eram frequentemente associadas ao termo de engenharia de características. Isso porque as técnicas do estado-da-arte até então compreendiam-se em reconhecer padrões de características extraídas por um especialista. Cabia à este especialista, então, conhecer profundamente as características relevantes dos dados, os descritores e seus parâmetros ideais, além do modelo ideal para aprender a discriminar com base nas informações extraídas.

De forma oposta, o aprendizado profundo têm sido empregado de uma forma mais ampla, de modo que a etapa de extração de características também fosse incumbida ao próprio modelo. Cabia a este, através de operações matemáticas, descobrir quais e como extrair as características mais relevantes do conjunto de dados de forma autônoma. Para dados não estruturados, como imagens e texto, tais modelos contribuíram significativamente para o avanço da inteligência artificial na área (STEVENS; ANTIGA; VIEHMANN, 2020).

### 2.2.1 Redes Neurais Convolucionais

Redes Neurais Convolucionais, como o próprio nome sugere, são majoritariamente formadas por camadas convolucionais. As camadas convolucionais usualmente recebem como entrada uma imagem representada como um tensor, aplica um determinado número de kernels, bias, funções de ativações para introduzir não-linearidade ao resultado da convolução e produzem como resultado uma matriz. Uma convolução pode ser definida como:

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} * \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix} = \sum_{i=1}^N \sum_{j=1}^M \alpha_{ij} k_{ij},$$

onde  $\alpha_{11}, \alpha_{12}, \dots, \alpha_{mn}$  representam os respectivos índices de um tensor de entrada bidimensional e  $k_{11}, k_{12}, \dots, k_{mn}$  valores do kernel utilizado para convolução. Tipicamente, uma imagem tende a ser muito maior que um kernel, por isso, durante a convolução, realiza-se um deslocamento da esquerda para a direita, de cima para baixo:

$$B_{ij} = (A * K)_{ij} = \sum_{f=0}^{n_K-1} \sum_{h=0}^{n_K-1} A_{i+f,j+h} K_{i+f,j+h}$$

A medida em que a variabilidade de uma base de dados aumenta, torna-se cada vez mais complexo definir manualmente quais são os kernels ideais para convoluir as imagens de entrada. Por isso é necessário aprender os valores através das imagens de entrada e saída. Primeiramente, o kernel é construído com valores aleatórios, em seguida, iterativamente, aplica-se o erro quadrático médio para comparar uma saída  $Y$  com a saída da camada convolucional, calculando, por fim, o gradiente para atualização dos pesos (ZHANG *et al.*, 2021).

### 2.2.2 Pooling

As camadas pooling são muito importantes para permitir que características sejam extraídas em diferentes níveis, desde pequenos detalhes na imagem até traços mais grosseiros que podem remeter ao objeto como um todo (ZHANG *et al.*, 2021). Geralmente, essas camadas sempre acompanham uma camada convolucional, haja visto que elas possuem um propósito bem específico que é de reduzir a dimensionalidade dos tensores. Por isso, não possuem pesos, apesar de possuírem parâmetros (MICHELUCCI, 2019).

As camadas de pooling possuem dois principais parâmetros: o tamanho da janela e o deslocamento na imagem. Em suma, uma janela de tamanho pré-definido percorre toda a imagem, e, assim como a convolução, sua operação resulta em apenas um escalar. As operações mais comuns são o cálculo do valor máximo e do valor médio dos pixels que estiverem dentro da janela.

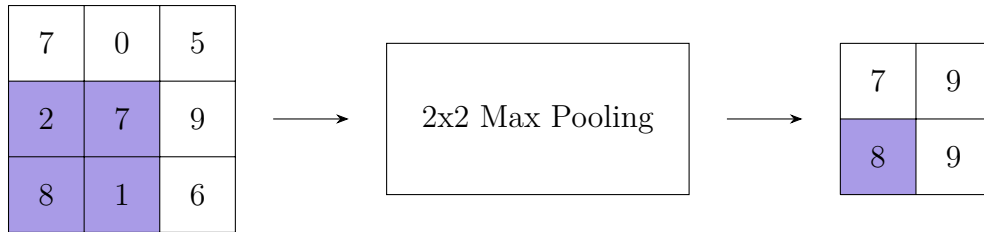


Figura 2 – Exemplificação da operação de Max pooling em uma matrix de 3x3 considerando uma janela de tamanho 2x2 e stride 1.

### 2.2.3 Camadas Densas

Camadas densas são camadas onde há a conectividade completa entre os neurônios de uma rede neural. Isso significa que cada neurônio em uma camada está conectado a todos os neurônios na camada anterior e na camada seguinte (BALLARD, 2018). Essa conectividade permite a captura de padrões complexos nos dados de entrada, por isso, são geralmente empregadas entre as camadas finais de uma rede neural profunda para realizar

tarefas específicas como classificação de padrões, regressão, entre outras (ROSEBROCK, 2017).

Cada neurônio da camada densa recebe um estímulo representado pelas variáveis de entrada. Cada valor de entrada é multiplicado pelo seu respectivo peso sináptico - cujo valores são iniciados aleatoriamente e vão se ajustando ao longo do treinamento por meio do algoritmo de *Backpropagation*. Dessa multiplicação, obtém-se o potencial de ativação através da soma ponderada dos sinais de entrada, que, ao passar por uma função de ativação, produz uma saída normalizada (SILVA; SPATTI; FLAUZINO, 2010). A Figura 3 ilustra um neurônio artificial, principal componente das camadas densas.

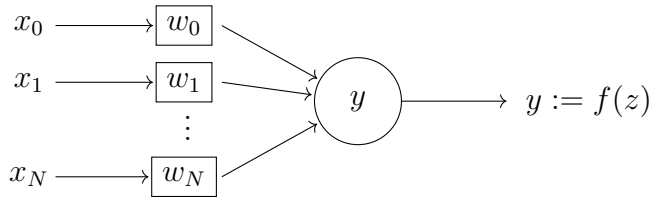


Figura 3 – Esquemática de um neurônio artificial

#### 2.2.4 Normalização por lotes

Normalização por lotes é uma eficiente técnica para estabilização e aceleração da convergência do aprendizado durante a etapa de treinamento, permitindo, assim, que o aprendizado de representações complexas dos dados ocorra de forma mais eficaz (ZHANG *et al.*, 2021; ROSEBROCK, 2017). Isso se dá pois, à medida em que os dados percorrem as camadas da rede durante o processo de aprendizado, a magnitude das camadas podem variar significativamente, podendo resultar em gradientes muito pequenos (*gradient vanishing*) ou gradientes muito grandes (*gradient exploding*). A partir da normalização por lotes, os dados passam a ficar centralizados em 0 e possuem um desvio padrão unitário (CHOLLET, 2021; GOODFELLOW; BENGIO; COURVILLE, 2016), *i.e.*:

$$H' = \frac{H - \mu}{\sigma},$$

onde  $H$  corresponde à um mini lote de ativações da camada a ser normalizada,  $\mu$  é um vetor de médias de cada unidade do lote e  $\sigma$  é um vetor com o desvio-padrão de cada unidade do lote.

#### 2.2.5 Dropout

Camadas de dropout são técnicas de regularização do estado da arte e correspondem basicamente em desconectar aleatoriamente uma quantidade finita de ativações em cada ciclo de treinamento. Em outras palavras, alguns nós da rede são multiplicados por zero, fazendo com que apenas uma pequena subseção seja treinada por vez. A eficácia dessa técnica baseia-se no fato de que pequenas redes tendem a gerar menos *overfitting*,

pois possuem uma menor capacidade de se apegar aos detalhes mais granulares, isto é, ruídos (TRASK, 2019; CHOLLET, 2021; GOODFELLOW; BENGIO; COURVILLE, 2016; BALLARD, 2018).

### 2.2.6 ResNets

Ainda que a técnica de normalização por lotes (Subseção 2.2.4) possa mitigar o problema de *gradient vanishing* e *gradient exploding* ao normalizar as informações que fluem pela rede durante a etapa de treinamento e ao reduzir a dependência de uma inicialização cuidadosa dos pesos, ela não garante a total solução do problema (BALLARD, 2018). Isto posto, He *et al.* (2016) propuseram um novo módulo de arquitetura capaz de propagar camadas para regiões mais profundas da rede, criando, assim, um caminho alternativo para o gradiente fluir (ELGENDY, 2020). Tais camadas são chamadas de *Skip Connections*.

O modo com que as *Skip Connections* solucionam o problema de *gradient vanishing* e *gradient exploding*, permitindo que redes muito mais profundas possam convergir, está relacionado com o fato de que ao propagá-las, as saídas das funções de ativação das primeiras camadas são somadas à operação de multiplicação dos pesos pelas entradas de camadas mais profundas (RAMO, 2019). Na arquitetura do modelo ResNet34, ilustrada na Figura 4, é possível visualizar as *skip connections* entre cada bloco (também chamado de unidade ou bloco residual) de duas camadas convolucionais com kernel 3x3, stride 1 e função de ativação ReLU.

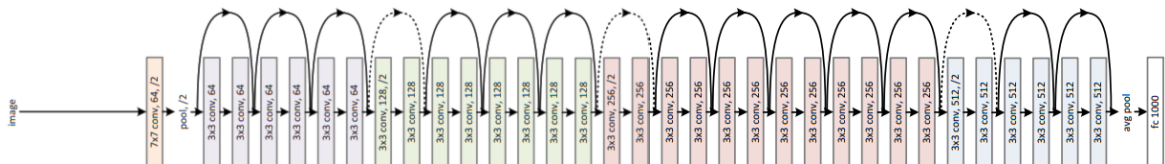


Figura 4 – Arquitetura de modelo de rede residual de 34 camadas proposta por He *et al.* (2016). FONTE: He *et al.* (2016)

O advento das *Skip Connections* permitiu com que mais camadas pudessem ser adicionadas (e, conseqüentemente, maior complexidade) sem degradação de performance. Nesse sentido, diversas variações da ResNet foram propostas, como a ResNet50, ResNet101 e ResNet152. O modelo da ResNet50, como o próprio nome sugere, possui 50 camadas de profundidade. Além disso, a partir desse modelo, os blocos residuais passaram a compor 3 camadas: 1 camada convolucional com kernel 1x1 para diminuir a dimensão da pilha de mapas de características de entrada, 1 camada convolucional com kernel de 3x3 para extração de características e uma outra camada convolucional de kernel 1x1 para diminuir a dimensão da pilha de mapas de características de saída. Todas as camadas são seguidas de normalização por lotes e função de ativação ReLU (ELGENDY, 2020). Essa abordagem

se mostrou útil para controlar o número de mapas de características nas camadas mais profundas da rede, reduzindo, assim, a complexidade e a quantidade de pesos.

### 2.2.7 Transferência de Aprendizado

Com o advento de modelos cada vez mais profundos e computacionalmente custosos, houve uma crescente demanda por técnicas capazes de acelerar o processo de treinamento e otimizar a forma com que tais modelos realizam o aprendizado. Esta é justamente a proposta das técnicas de transferência de aprendizado. A transferência de aprendizado, em poucas palavras, corresponde em transferir o aprendizado realizado por um modelo para solucionar um problema para a generalização de outro problema (GOODFELLOW; BENGIO; COURVILLE, 2016; ELGENDY, 2020; MICHELUCCI, 2019).

Para que a transferência de aprendizado seja realmente efetiva, Goodfellow, Bengio and Courville (2016) mencionam que fatores que explicam as variações no conjunto de dados usado para aprendizado prévio devem ser, de algum modo, relevantes para o aprendizado atual. Tratando-se de visão computacional, sabe-se que muitas características visuais compartilham as mesmas representações básicas, como por exemplo de formas, mudanças geométricas, iluminação, textura, etc. Transferir aprendizado, neste contexto, significaria utilizar as camadas de um modelo aptas a reconhecer tais representações para um novo propósito.

Segundo Ramo (2019), existem dois principais fatores que interferem o uso de transferência de aprendizado. O primeiro é a quantidade e qualidade dos dados que temos em mãos, e o segundo, a capacidade de poder computacional disponível para treinamento do modelo. Quando não há uma grande quantidade de dados disponíveis e há recursos computacionais limitados, uma possível solução é procurar um modelo já treinado previamente para um problema similar, que envolva, por exemplo, classes e contextos iguais. Diante disso, seria factível congelar as camadas escondidas e treinar apenas a última camada. Congelar as camadas, nesse caso, corresponde em impedir que os pesos reutilizados sejam atualizados durante a etapa de retropropagação (*backpropagation*).

Em um cenário onde há um pouco mais de dados, mas ainda insuficientes para generalizar o aprendizado a partir de pesos inicializados aleatoriamente em tempo e acurácia aceitáveis, além de um pouco mais de recursos computacionais disponíveis para treinamento, é possível congelar menos camadas. Desse modo, a probabilidade de se obter melhores resultados é maior, dado que mais pesos estão sendo ajustados ao problema.

### 3 AVALIAÇÃO EXPERIMENTAL

Os experimentos conduzidos nesse trabalho abrangeram cinco principais etapas, que constituem em "aquisição de dados", "anotação dos dados", "pré-processamento", "treinamento dos modelos" e "avaliação do aprendizado" (Figura 3). Ao longo deste capítulo, serão abordadas as metodologias empregadas em cada uma delas.



#### 3.1 Aquisição de Dados

Três conjuntos distintos de dados foram utilizados para treinamento e avaliação dos modelos. Dois deles, "Lemon Quality Dataset" e "Mangas Embrapa", possuem imagens coletadas apenas na faixa visível do espectro eletromagnético (imagens RGB), e um terceiro, "Maçãs 1.0", é composto por imagens tanto coloridas quanto infravermelhas (imagens monocromáticas). Tais conjuntos correspondem a diferentes tipos de frutos em diferentes estágios de maturação separados entre as classes "boa qualidade" e "má qualidade".

##### 3.1.1 Lemon Quality Dataset

Trata-se de um conjunto de dados públicos disponibilizado no portal Kaggle (KOROGLU, 2022). Tal conjunto é composto por 2533 imagens comprimidas no formato JPEG de resolução 300x300 pixels. Cada imagem corresponde a um limão posicionado sobre uma superfície de concreto texturizada, fotografada em condições de iluminação solar indireta. As imagens apresentam variações de escala, rotação e nitidez e estão rotuladas entre duas classes distintas: "má qualidade", com 942 amostras e "boa qualidade" com 1125 amostras. Outras 466 imagens não contêm frutos e foram descartadas.

##### 3.1.2 Mangas Embrapa

A base de dados Mangas Embrapa (nome fantasia) (INSTRUMENTACAO, 2014) é composta por 51 imagens de mangas do tipo Palmer. Sua aquisição foi realizada no verão do ano de 2014 por pesquisadores da Embrapa Instrumentação, localizada na cidade de São Carlos - SP. Ao todo, foram utilizados 12 exemplares para compor o conjunto de dados, sendo que para cada exemplar, foram capturadas imagens de diferentes lados da fruta em diferentes estágios de maturação, além da polpa da fruta. Cada imagem possui tamanho de 1024x768 pixels e 3 camadas (RGB) em formato JPEG.

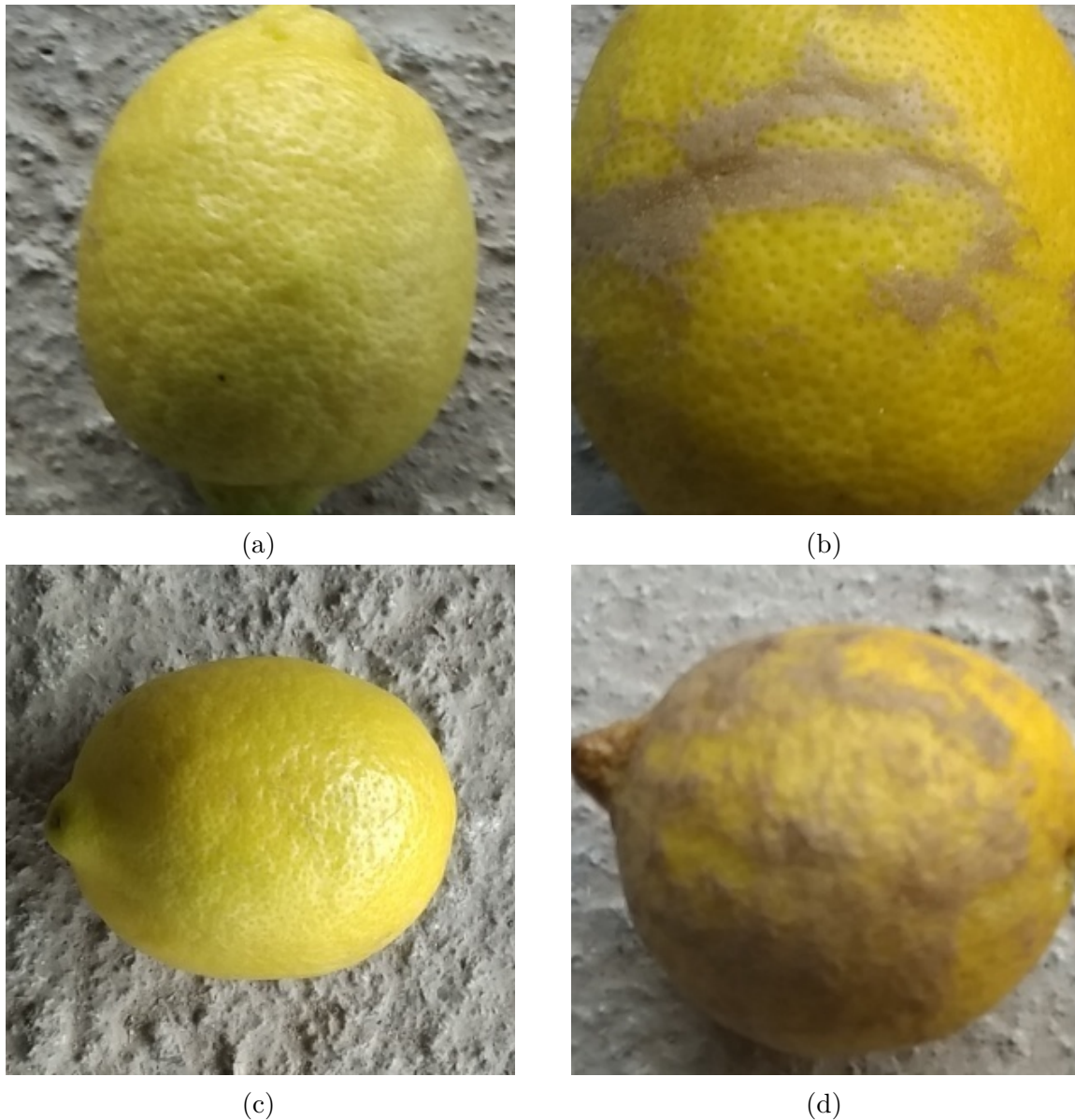


Figura 5 – Exemplo de amostras da base de dados "Lemon Quality Dataset". As imagens a) e c) correspondem a limões de boa qualidade, e as imagens b) e d), limões de má qualidade

### 3.1.3 Maçãs 1.0

A base de dados Maçãs 1.0 corresponde à um conjunto de imagens de maçãs adquiridas pelo autor deste trabalho no mês de abril de 2023. Para a coleta dos dados, foram utilizadas duas câmeras distintas: a High Quality Camera (HQ Cam) para Raspberry Pi, equipada com uma lente varifocal de 8 a 50 mm para capturar imagens na faixa de luz visível, e a câmera Mako G-125B com filtro BP735 para capturar imagens na faixa do infravermelho próximo. Foi utilizada uma light box da marca Puluz, com dimensões de 40x40x40 cm, para controlar a iluminação durante as capturas e a distância entre a câmera e o objeto foi de aproximadamente 1 metro. Ambas as câmeras foram posicionadas lado a lado e o acionamento de captura foi realizado por meio de software específico para cada



(a)



(b)



(c)



(d)

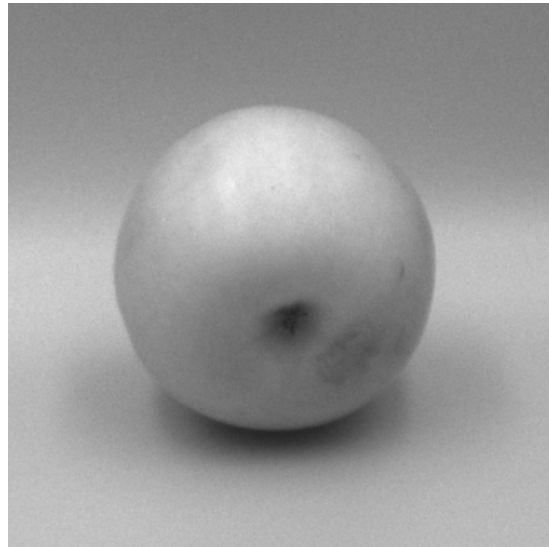
Figura 6 – Exemplo de amostras da base de dados "Mangas Embrapa". Em a), uma imagem inteira com identificação da fruta. Em b), c) e d), exemplares de mangas em diferentes estágios de maturação.

modelo. Para a HQ Cam, utilizou-se o software raspistill, enquanto para a Mako G-125B utilizou-se o software Vimba versão 6.0, disponível no site da fabricante Allied Vision.

Ao todo, foram capturadas 70 imagens RGB de resolução de 4056x3040 e 70 imagens monocromáticas de resolução de 1292x964. Para essa aquisição, foram utilizadas 6 maçãs, sendo 4 do tipo "nacional" e 2 do tipo "Argentina".



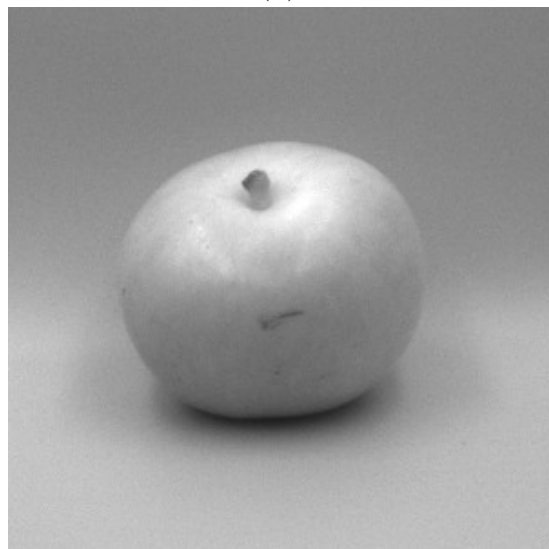
(a)



(b)



(c)



(d)

Figura 7 – Exemplo de amostras da base de dados "Maçãs 1.0". Na primeira coluna, formada pelas amostras a) e c), é possível visualizar duas maçãs capturadas com a câmera HQ Cam. Em b) e d) as mesmas frutas obtidas utilizando a câmera espectral Mako G-125B.

### 3.2 Anotação

O processo de anotação dos dados foi realizado utilizando a ferramenta CVAT versão 2.5.0 para Desktop. Para cada amostra, foi associado um valor dentre as categorias "alta qualidade" e "baixa qualidade" mediante uma análise visual das imagens. Haja visto que o anotador não apresentava conhecimento técnico qualificado para avaliar a qualidade

do fruto, o critério estabelecido baseou-se na presença de manchas na casca, de modo que frutos com manchas ou anomalias foram classificados como de "baixa qualidade", enquanto aqueles com cascas de textura uniforme foram categorizados como de "alta qualidade". Dos três conjuntos de dados utilizados no estudo, o "Mangas Embrapa" e "Maças 1.0" precisaram ser anotados. As anotações da base "Lemon Quality Dataset" foram disponibilizadas juntamente com as imagens.

A anotação dos dados se deu através da delimitação do fruto por meio de bounding boxes. Essa abordagem possibilitou, a partir de suas coordenadas, reduzir a quantidade de informação das imagens e filtrar padrões irrelevantes para o escopo do estudo que pudessem produzir algum viés de aprendizado do modelo, como a identificação numérica das mangas. Ao todo, 32 mangas foram classificadas como sendo de "boa qualidade" e 18 como sendo de "má qualidade". Já para a base de dados maçãs, 34 foram classificadas como de "boa qualidade" e 36 como de "má qualidade".

### 3.3 Pré-processamento

Três algoritmos foram utilizados para pré-processamento dos dados. A partir das imagens obtidas, foram extraídas as regiões de interesse com base nas coordenadas dos bounding boxes anotados. Para padronizar o tamanho das amostras, aplicou-se a técnica de zero padding, que corresponde em adicionar pixels de valor 0 nas bordas da imagem. Apenas as bordas do eixo de menor dimensão sofreram alterações, resultando, assim, em uma imagem de dimensões iguais. Por fim, todas as imagens foram redimensionadas para 300x300 pixels.

A divisão das bases de dados entre os subconjuntos de treinamento, validação e teste ocorreu de forma aleatória utilizando a semente *ia – bigdata – 2023*. Do total de imagens da base de dados Maças 1.0, 70% foram destinadas para o subconjunto de treinamento, 20% para o subconjunto de validação e 10% para o subconjunto de teste. Para as bases restantes, utilizadas nesse estudo como pré-treinamento, 70% das amostras foram utilizadas para treinamento e 30% para validação.

Para evitar overfitting dos modelos de classificação e direcionar o aprendizado para os padrões relevantes de cada fruto, alguns experimentos foram realizados utilizando técnicas de aumento de dados no subconjunto de treinamento. Foram selecionados cinco algoritmos da biblioteca Albumentations. Para cada imagem pertencente ao subconjunto de treinamento, foram aplicados, com probabilidade de 50% cada, algoritmos cujo parâmetros estão detalhados na Tabela 1. Destaca-se que alguns desses parâmetros correspondem a uma *range* de valores, e, portanto, referem-se apenas aos valores mínimos e máximos que podem assumir. Com isso, os valores reais são determinados aleatoriamente e em tempo de execução, processo tal qual é conhecido como aumento de dados online.

Algoritmo	Parâmetros
RandomRotate90	-
Downscale	scale_min=0.87, scale_max=0.99
GridDistortion	num_steps=1, distort_limit=(-0.12, 0.07), interpolation=0, border_mode=4, value=(0, 0, 0), mask_value=None, normalized=False
RandomBrightnessContrast	brightness_limit=(-0.16, 0.52), contrast_limit=(-0.3, 0.2), brightness_by_max=True
ShiftScaleRotate	shift_limit_x=(-0.01, -0.01), shift_limit_y=(-0.01, -0.01), scale_limit=(-0.007999999999999996, 0.08599999999999999), rotate_limit=(0, 0), interpolation=4, border_mode=4, value=(0, 0, 0), mask_value=None, rotate_method=largest_box

Tabela 1 – Algoritmos e parâmetros utilizados para aumento de dados durante a etapa de treinamento

### 3.4 Configuração Experimental

Os primeiros experimentos envolveram o treinamento do modelo ResNet50 utilizando pesos pré-treinados da base de dados ImageNet (DENG *et al.*, 2009), disponibilizados pela biblioteca TorchVision na versão 0.14.0. Durante esta fase, foi realizado um treinamento para cada conjunto de dados, utilizando lotes de 16 imagens e uma taxa de aprendizado de 0.001. Esses experimentos correspondem aos itens 1, 2, 3 e 4 na coluna “Treinamento” da Tabela 2. Os experimentos de 5 a 7 foram conduzidos utilizando os pesos obtidos no experimento 1 como ponto de partida. Da mesma forma, os experimentos de 8 a 9 utilizaram os pesos do experimento 5, o experimento 10 utilizou os pesos do experimento 8, e os experimentos 11 e 12 utilizaram os pesos dos experimentos 2 e 3, respectivamente. A partir do experimento 5, todos os treinamentos foram realizados com taxa de aprendizado de 0.0001.

Os treinamentos dos modelos foram conduzidos por meio de ciclos iterativos de ajuste de pesos, usando tanto o subconjunto de treinamento como o de validação. O número total de épocas foi estabelecido em 100 para cada experimento. Para a arquitetura do modelo e a manipulação de tensores, foi utilizada a biblioteca PyTorch na versão 1.13.0. A função Binary Cross-Entropy (BCE) foi adotada como função de perda, e para

otimização dos parâmetros dos modelos, empregou-se o otimizador Stochastic Gradient Descent (SGD). Os pesos salvos ao final de cada treinamento correspondem à época de menor valor de loss obtido no conjunto de validação.

Experimento	Base de dados	Pesos iniciais
1	Lemon Quality	ImageNet
2	Mangas Embrapa	ImageNet
3	Maçãs 1.0 - NIR	ImageNet
4	Maçãs 1.0 - RGB	ImageNet
5	Mangas Embrapa	ImageNet + Lemon Quality
6	Maçãs 1.0 - NIR	ImageNet + Lemon Quality
7	Maçãs 1.0 - RGB	ImageNet + Lemon Quality
8	Maçãs 1.0 - NIR	ImageNet + Lemon Quality + Mangas Embrapa
9	Maçãs 1.0 - RGB	ImageNet + Lemon Quality + Mangas Embrapa
10	Maçãs 1.0 - RGB	ImageNet + Lemon Quality + Mangas Embrapa + Maçãs 1.0 - NIR
11	Maçãs 1.0 - RGB	ImageNet + Mangas Embrapa
12	Maçãs 1.0 - RGB	ImageNet + Maçãs 1.0 - NIR

Tabela 2 – Treinamentos realizados. A coluna "Base de dados" indica a base de dados utilizada para treinar o classificador e a coluna "Pesos iniciais", o conjunto de pesos utilizado como ponto de partida.

### 3.5 Avaliação do aprendizado

A avaliação do desempenho do modelo foi realizada através de métricas quantitativas de precisão (Equação 3.1), recall (Equação 3.2), acurácia (Equação 3.3) e F1-Score (Equação 3.4). Além disso, foram geradas a matriz de confusão e a curva de precision-recall. Para a análise, considerou-se um verdadeiro positivo (VP) amostras classificadas corretamente como de “má qualidade”, verdadeiro negativo (VN), amostras corretamente classificadas como “boa qualidade”. Falso positivo (FP), amostras de “boa qualidade” identificadas como de “má qualidade”, e falso negativo (FN) amostras classificadas como de “boa qualidade” quando são de fato de “má qualidade”. Abaixo, são listadas as fórmulas para cálculo da precisão, recall, acurácia e F1-Score.

$$Precisao = \frac{VP}{VP + FP} \quad (3.1)$$

$$Recall = \frac{VP}{VP + FN} \quad (3.2)$$

$$Acuracia = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.3)$$

$$F1 - Score = 2 * \frac{Precisao * Recall}{Precisao + Recall} \quad (3.4)$$

### 3.6 Resultados e Discussões

Nesta seção, são apresentados os resultados dos experimentos organizados na Tabela 2. A Tabela 3 contém os menores valores de *loss* obtidos durante os treinamentos, de modo que as “colunas Loss BCE (Treina.)” e “Loss BCE (Valida.)” indicam, respectivamente, os valores de *loss* nos conjuntos de treinamento e validação. O Experimento 1, realizado apenas utilizando a base de dados Lemon Quality a partir de um treinamento do modelo ResNet50 com a base de dados ImageNet, apresentou os índices mais baixos em ambos os conjuntos. Esse desempenho destaca-se, em grande parte, devido à extensiva quantidade e distribuição equilibrada de amostras nesta base de dados.

Experimento	Base de dados	Pesos Iniciais	Loss BCE (Treina.)	Loss BCE (Valida.)
1	Lemon Quality	ImageNet	0.018	0.012
2	Mangas Embrapa	ImageNet	0.618	0.665
3	Maçãs 1.0 - NIR	ImageNet	0.675	0.631
4	Maçãs 1.0 - RGB	ImageNet	0.675	0.682
5	Mangas Embrapa	Experimento 1	0.441	0.662
6	Maçãs 1.0 - NIR	Experimento 1	0.494	0.253
7	Maçãs 1.0 - RGB	Experimento 1	0.494	0.467
8	Maçãs 1.0 - NIR	Experimento 5	0.396	0.314
9	Maçãs 1.0 - RGB	Experimento 5	0.514	0.411
10	Maçãs 1.0 - RGB	Experimento 8	0.471	0.452
11	Maçãs 1.0 - RGB	Experimento 2	0.686	0.682
12	Maçãs 1.0 - RGB	Experimento 3	0.676	0.680

Tabela 3 – Resultados dos experimentos de classificação de qualidade de frutas

Os Experimentos 2, 3 e 4, por conseguinte, exibiram valores de *loss* relativamente mais elevados, sugerindo desafios específicos associados às bases de dados Mangas Embrapa, Maçãs 1.0 - NIR e Maçãs 1.0 RGB. Adicionalmente, essas bases possuem uma quantidade reduzida de amostras em comparação com a Lemon Quality, um fator crucial para o aumento dos índices de *loss*.

Os Experimentos 5, 6 e 7 utilizaram os pesos do Experimento 1 como ponto de partida, e o Experimento 6 apresentou uma redução significativa no *loss* de validação em

comparação com o Experimento 3. Isso indica, à primeira vista, que o pré-treinamento realizado com a base de dados Lemon Quality contribuiu para melhorias no desempenho.

Os Experimentos 8 e 9 partiram dos pesos treinados no Experimento 5. Observa-se que a base de dados Mangas Embrapa contribuiu para otimizar o treinamento com a base de dados Maçãs 1.0 - NIR. No entanto, o mesmo não se aplicou à versão RGB dessa base de dados. Uma possível interpretação é que os defeitos presentes nas mangas têm altas frequências, assim como nas imagens da base Maçãs 1.0 - NIR. Contudo, isso não ocorre na versão RGB dessa base de dados.

Outro ponto a ser destacado é que o Experimento 10, treinado com base no Experimento 8, não apresentou avanços significativos em comparação com o Experimento 9. Isso sugere que tanto a base de dados Mangas Embrapa quanto a base Maçãs 1.0 - NIR, não contribuíram para melhorias no desempenho de treinamento. Os Experimentos 11 e 12 reforçam essa ideia, pois ambos apresentaram desempenhos semelhantes aos obtidos no Experimento 4, realizado a partir da base ImageNet. Portanto, o treinamento realizado utilizando a base de dados Lemon Quality foi a principal responsável pelas melhorias reportadas no desempenho do modelo ResNet50 com imagens de maçãs em RGB, mesmo sua versão em NIR sendo composta essencialmente pelas mesmas maçãs.

As avaliações realizadas no subconjunto de teste da base de dados Maçãs 1.0 - RGB estão organizadas na Tabela 4. A grande dificuldade em se extrair indicadores confiáveis foi o baixo número de amostras. Ao todo, foram utilizadas 5 amostras classificadas como “boa qualidade” e 2 amostras como “má qualidade”. Ainda assim, é importante ressaltar que esse cenário caracteriza-se como corriqueiro no dia-a-dia de um especialista em aprendizado de máquinas e que, em determinados casos, é necessário recorrer à outros recursos como por exemplo modelos mais robustos e técnicas mais avançadas para suprir a carência de dados, não cobertas nesse estudo.

Experimento	Base de dados	Pesos Iniciais	Acurácia	Precisão	Recall	F1-Score
4	Maçãs 1.0 - RGB	ImageNet	0.714	0.500	0.500	0.500
7	Maçãs 1.0 - RGB	Experimento 1	0.571	0.571	1.000	0.400
9	Maçãs 1.0 - RGB	Experimento 5	0.428	0.333	1.000	0.500
10	Maçãs 1.0 - RGB	Experimento 8	0.428	0.333	1.000	0.500
11	Maçãs 1.0 - RGB	Experimento 2	0.285	0.285	1.000	0.444
12	Maçãs 1.0 - RGB	Experimento 3	0.571	0.400	1.000	0.571

Tabela 4 – Resultados dos experimentos de classificação de qualidade de frutas

A princípio, nota-se que a maior parte dos experimentos resultou em Recall igual a 1. Em outras palavras, isso significa que as duas amostras de “má qualidade” foram corretamente classificadas como sendo de “má qualidade”. Em contrapartida, os valores

de Precisão se mostraram baixos, o que corresponde a dizer que o modelo apresentou dificuldades para classificar algumas das 5 amostras de “boa qualidade” como sendo, realmente, de “boa qualidade”. Embora os resultados obtidos não forneçam uma clara ideia da capacidade de generalização do modelo ao inferir novas amostras, devido ao conjunto de testes não ser representativo o suficiente, os valores apresentados na Tabela 3 fornecem fortes evidências de que a escolha da base de dados e a estratégia de transferência de aprendizado têm impacto direto na eficácia do modelo. Além disso, os resultados sugerem que foi mais vantajoso realizar uma tarefa de pré-treinamento em uma base de dados composta por uma grande quantidade de imagens de limões do que aprender em uma base de dados pequena composta pelas mesmas maçãs na banda NIR, onde os defeitos estão mais realçados. Por fim, é importante ressaltar que acumular o aprendizado em bases de dados diversas através de sucessivas transferências de aprendizado não necessariamente resulta em uma maior robustez do modelo.

## 4 CONCLUSÕES

Neste trabalho, exploramos o impacto da transferência de aprendizado no treinamento do modelo ResNet50 para classificação de maçãs em duas classes: “boa qualidade” e “má qualidade”. Para isso, criamos uma base de dados composta por 70 imagens de maçãs nas bandas RGB e infravermelho próximo (NIR). Além das imagens de maçãs na faixa NIR, outras bases de dados de diferentes frutos, como limões e mangas, fizeram parte do pré-treinamento. Os resultados obtidos sugerem que a escolha da base de dados e a estratégia de transferência de aprendizado têm impacto direto na eficácia do modelo. Além disso, os resultados sugerem que o aprendizado em bases de dados diversas através de sucessivas transferências de aprendizado não necessariamente resulta em uma maior robustez do modelo. Das bases estudadas, a que mais contribuiu para a melhoria do desempenho de treinamento para classificação de maçãs RGB foi a base composta por limões. Apesar de apresentar características físicas distintas das maçãs, como cor, textura e aspecto dos defeitos, ela possui mais imagens e isso pareceu contribuir significativamente para um melhor desempenho do classificador durante o treinamento.



## REFERÊNCIAS

- AGYEMAN, P. C. *et al.* Using spectral indices and terrain attribute datasets and their combination in the prediction of cadmium content in agricultural soil. **Computers and Electronics in Agriculture**, v. 198, p. 107077, 2022. ISSN 0168-1699. Available at: <<https://www.sciencedirect.com/science/article/pii/S0168169922003945>>.
- BACCA, J.; MARTINEZ, E.; ARGUELLO, H. Computational spectral imaging: A contemporary overview. **arXiv preprint arXiv:2303.04848**, 2023.
- BALLARD, W. **Hands-on deep learning for images with TensorFlow: build intelligent computer vision applications using TensorFlow and Keras**. [S.l.: s.n.]: Packt Publishing Ltd, 2018.
- CHOLLET, F. **Deep learning with Python**. [S.l.: s.n.]: Simon and Schuster, 2021.
- DENG, J. *et al.* Imagenet: A large-scale hierarchical image database. *In: IEEE. 2009 IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2009. p. 248–255.
- ELGENDY, M. **Deep learning for vision systems**. [S.l.: s.n.]: Simon and Schuster, 2020.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.: s.n.]: MIT press, 2016.
- HE, K. *et al.* Deep residual learning for image recognition. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.
- HERRERA, L. C. G. P. *et al.* Hyperspectral image segmentation: a preliminary study on the oral and dental spectral image database (odsi-db). **Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization**, Taylor & Francis, p. 1–9, 2023.
- IKEUCHI, K. *et al.* **Active lighting and its application for computer vision**. [S.l.: s.n.]: Springer, 2020.
- INSTRUMENTACAO, E. **Mangas Embrapa: banco de dados**. 2014. Acesso em 04-12-2014.
- KOROGLU, Y. E. **Lemon Quality Dataset**. 2022. Available at: <<https://www.kaggle.com/datasets/yusufemir/lemon-quality-dataset>>.
- LI, B.; ZHANG, D.; SHEN, Y. Study on terahertz spectrum analysis and recognition modeling of common agricultural diseases. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 243, p. 118820, 2020. ISSN 1386-1425. Available at: <<https://www.sciencedirect.com/science/article/pii/S138614252030799X>>.
- LIU, J. *et al.* Deep visual anomaly detection in industrial manufacturing: A survey. **arXiv preprint arXiv:2301.11514**, 2023.

- MICHELUCCI, U. **Advanced applied deep learning: convolutional neural networks and object detection**. [*S.l.: s.n.*]: Springer, 2019.
- MONAKHOVA, K. *et al.* Spectral diffusercam: lensless snapshot hyperspectral imaging with a spectral filter array. **Optica**, Optica Publishing Group, v. 7, n. 10, p. 1298–1307, 2020.
- PENG, X. *et al.* Rgb-nir image categorization with prior knowledge transfer. **EURASIP Journal on Image and Video Processing**, SpringerOpen, v. 2018, n. 1, p. 1–11, 2018.
- POPESCU, G. **Principles of Biophotonics, Volume 2: Light emission, detection, and statistics**. [*S.l.: s.n.*]: IOP Publishing, 2019.
- RAMO, K. **Hands-On Java Deep Learning for Computer Vision: Implement machine learning and neural network methodologies to perform computer vision-related tasks**. [*S.l.: s.n.*]: Packt Publishing Ltd, 2019.
- ROSEBROCK, A. **Deep learning for computer vision with python: Starter bundle**. [*S.l.: s.n.*]: PyImageSearch, 2017.
- SEIDLITZ, S. *et al.* Robust deep learning-based semantic organ segmentation in hyperspectral images. **Medical Image Analysis**, Elsevier, v. 80, p. 102488, 2022.
- SILVA, I. N. d.; SPATTI, D. H.; FLAUZINO, R. A. Redes neurais artificiais para engenharia e ciências aplicadas. 2010.
- STEVENS, E.; ANTIGA, L.; VIEHMANN, T. **Deep learning with PyTorch**. [*S.l.: s.n.*]: Manning Publications, 2020.
- STUMP, E. A. *et al.* Meta-learning for color-to-infrared cross-modal style transfer. **arXiv preprint arXiv:2212.12824**, 2022.
- THIELE, C. J. *et al.* Microplastics in european sea salts – an example of exposure through consumer choice and of interstudy methodological discrepancies. **Ecotoxicology and Environmental Safety**, v. 255, p. 114782, 2023. ISSN 0147-6513. Available at: <<https://www.sciencedirect.com/science/article/pii/S0147651323002865>>.
- TRASK, A. W. **Grokking deep learning**. [*S.l.: s.n.*]: Simon and Schuster, 2019.
- VIVONE, G. Multispectral and hyperspectral image fusion in remote sensing: A survey. **Information Fusion**, Elsevier, v. 89, p. 405–417, 2023.
- WU, G. *et al.* Early identification of strawberry leaves disease utilizing hyperspectral imaging combing with spectral features, multiple vegetation indices and textural features. **Computers and Electronics in Agriculture**, Elsevier, v. 204, p. 107553, 2023.
- XIANG, Y. *et al.* Deep learning and hyperspectral images based tomato soluble solids content and firmness estimation. **Frontiers in Plant Science**, Frontiers Media SA, v. 13, 2022.
- XISHUAI, P. *et al.* Multi-modal scene categorization using multi-tasks learning. *In*: **IEEE 2016 IEEE 13th International Conference on Signal Processing (ICSP)**. [*S.l.: s.n.*], 2016. p. 1106–1111.

XU, Z.; TANG, G.; PANG, B. An infrared thermal image few-shot learning method based on capnet and its application to induction motor fault diagnosis. **IEEE Sensors Journal**, IEEE, v. 22, n. 16, p. 16440–16450, 2022.

YAN, Y.; REN, J.; SUN, H. Nondestructive quality control in powder metallurgy using hyperspectral imaging. **arXiv preprint arXiv:2207.12966**, 2022.

ZHANG, A. *et al.* Dive into deep learning. **arXiv preprint arXiv:2106.11342**, 2021.