

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO
MESTRADO EM TECNOLOGIAS COMPUTACIONAIS PARA O AGRONEGÓCIO

LEANDRO DE OLIVEIRA

**CLASSIFICAÇÃO DO TIPO DE TOMATE POR MEIO DE IMAGEM: UM ESTUDO
DE CASO UTILIZANDO REDES NEURAS CONVOLUCIONAIS**

DISSERTAÇÃO

MEDIANEIRA

2024

LEANDRO DE OLIVEIRA

**CLASSIFICAÇÃO DO TIPO DE TOMATE POR MEIO DE IMAGEM: UM ESTUDO
DE CASO UTILIZANDO REDES NEURAS CONVOLUCIONAIS**

**Tomato type classification through images: a case study using convolutional
neural networks**

Dissertação apresentada como requisito para obtenção do título de Mestre em Tecnologias Computacionais para o Agronegócio da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador: Prof. Dr. José Ailton Azevedo dos Santos.

Coorientador: Prof. Dr. Evandro Andre Konopatzki

MEDIANEIRA

2024



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite remixe, adaptação e criação a partir do trabalho, para fins não comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



**Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Medianeira**



LEANDRO DE OLIVEIRA

CLASSIFICAÇÃO DO TIPO DE TOMATE POR MEIO DE IMAGEM: UM ESTUDO DE CASO UTILIZANDO REDES NEURAIAS CONVOLUCIONAIS

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Tecnologias Computacionais Para O Agronegócio da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Tecnologias Computacionais Aplicadas À Produção Agrícola E Agroindústria.

Data de aprovação: 07 de Março de 2024

Dr. Jose Airton Azevedo Dos Santos, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Carlos Aparecido Fernandes, Doutorado - Universidade Tecnológica Federal do Paraná

Dra. Luciana Del Castanhe Peron Da Silva, Doutorado - Universidade Estadual do Oeste do Paraná (Unioeste)

Dr. Paulo Cesar Tonin, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 07/03/2024.

Dedico este trabalho a toda minha família.

AGRADECIMENTOS

Primeiramente gostaria de agradecer a toda a minha família pelo suporte ao longo de mais uma etapa importante de minha vida. Sem eles nada disso seria possível.

Um agradecimento em especial ao meu orientador, Prof. Dr. José Aírton Azevedo dos Santos, por ter aceitado a orientação e por sua paciência e dedicação.

Agradeço aos meus amigos pelas palavras de incentivo.

Agradeço à psicóloga Gislaine pela paciência em me escutar e me direcionar em cada sessão.

Agradeço a cada pessoa que contribuiu direta ou indiretamente para que o caminho trilhado por mim fizesse chegar a este momento.

“Agricultura sem ciência é uma fazenda
de porteiras fechadas.”
(FERNANDES, 2024)

RESUMO

DE OLIVEIRA, Leandro. Classificação do tipo de tomate por meio de imagens: um estudo de caso utilizando redes neurais convolucionais. 2024. Dissertação de Mestrado – Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio, Universidade Tecnológica Federal do Paraná. Medianeira, 2024.

Este trabalho tem como objetivo classificar, por meio de rede neural artificial, diferentes tipos de tomates (variedade ou cultivar), dispostos à venda em estabelecimentos da Região Oeste Paranaense. O processo de classificação ocorreu por meio da utilização de modelos baseados em redes neurais convolucionais, e compreendeu a aquisição e processamento de imagens, treinamento, validação e teste dos modelos de classificação. Utilizou-se, para construção da base de dados de imagens de tomates, uma câmera de smartphone. Os modelos escolhidos para classificação foram: ResNet50V2, MobileNet, MobileNetV2, Xception, EfficientNetV2B0, EfficientNetV2B1, EfficientNetV2B2 e EfficientNetV2B3, os quais foram implementados na linguagem Python, utilizando a API Keras do *framework* TensorFlow. Os resultados obtidos, dos oito modelos, foram comparados por meio da acurácia (*accuracy*) e do erro (*loss*). Verificou-se, para um conjunto de teste de 12 imagens, significativamente diferente do conjunto de treinamento, que o modelo MobileNet apresentou o melhor desempenho, com uma acurácia de teste de 83,33% e um erro de 0,622.

Palavras-chave: inteligência artificial; visão computacional; python.

ABSTRACT

DE OLIVEIRA, Leandro. Tomato type classification through images: a case study using convolutional neural networks. 2024. Master's thesis – Postgraduate Program in Computer Technologies for Agribusiness, Federal University of Technology – Paraná. Medianeira, 2024.

The aim of this work is to use artificial neural networks to classify different types of tomatoes (variety or cultivar) on sale in establishments in the western region of Paraná. The classification process was carried out using models based on convolutional neural networks, and included image acquisition and processing, training, validation and testing of the classification models. A smartphone camera was used to build the tomato image database. The models chosen for classification were: ResNet50V2, MobileNet, MobileNetV2, Xception, EfficientNetV2B0, EfficientNetV2B1, EfficientNetV2B2 and EfficientNetV2B3, which were implemented in the Python language using the Keras API of the TensorFlow framework. The results obtained from the eight models were compared using accuracy and loss. For a test set of 12 images, significantly different from the training set, the MobileNet model performed best, with a test accuracy of 83.33% and an error of 0.622.

Keywords: artificial intelligence; computer vision; python.

LISTA DE ABREVIATURAS E SIGLAS

ANN	Artificial Neural Network
API	Application Programming Interface
CNN	Convolutional Neural Network
CPU	Central Processing Unit
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FPS	Frames Per Second
GPU	Graphics Processing Unit
LSPA	Levantamento Sistemático da Produção Agrícola
RAM	Random Access Memory
RNA	Rede Neural Artificial
RNC	Rede Neural Convolutacional

LISTA DE FIGURAS

Figura 1 – Modelo do perceptron: (a) com detalhes; (b) simplificado.....	17
Figura 2 – Estrutura de uma rede neural multicamadas com uma camada oculta.....	17
Figura 3 – Camadas de inserção da rede <i>Neocognitron</i>	18
Figura 4 – À esquerda um tomateiro, à direita formas e tamanhos variados de tomates.....	19
Figura 5 – Top 10 dos produtores de tomate no mundo em 2021.....	20
Figura 6 – Tomate do grupo Cereja.....	22
Figura 7 – Tomate do grupo Italiano.....	22
Figura 8 – Tomate do grupo Salada.....	23
Figura 9 – Tomate do grupo Santa Cruz.....	24
Figura 10 – Procedimentos metodológicos.....	26
Figura 11 – Representação de uma rede convolucional.....	29
Figura 12 – Conjunto de imagens de tomates (classes).....	30
Figura 13 – Conjunto de imagens de tomates com características similares...	30
Figura 14 – Pontos de vista de captura das imagens das amostras.....	31
Figura 15 – Configuração da validação cruzada.....	32
Figura 16 – Amostras dos tipos de tomates contidos no banco de imagens <i>Fruits 360</i>	35
Figura 17 – <i>Overfitting</i>	36
Figura 18 – Etapas do trabalho.....	37
Figura 19 – Resultados do erro e da acurácia para 100 épocas – Conjunto de Validação.....	41
Figura 20 – Nova predição com as imagens de teste – modelo MobileNet.....	42

LISTA DE TABELAS

Tabela 1 – Produção de tomate no Brasil em 2022.....	20
Tabela 2 – Alguns híbridos pertencentes ao grupo Cereja.....	22
Tabela 3 – Alguns híbridos pertencentes ao grupo Italiano.....	23
Tabela 4 – Alguns híbridos pertencentes ao grupo Salada.....	24
Tabela 5 – Alguns híbridos pertencentes ao grupo Santa Cruz.....	25
Tabela 6 – Desempenho obtido no treinamento e teste dos modelos FCN8s, U-Net e BR-Net.....	28
Tabela 7 – Modelos de Redes Neurais Convolucionais.....	34
Tabela 8 – Característica do banco de imagens de tomates.....	38
Tabela 9 – Média dos resultados de treino.....	39
Tabela 10 – Média dos resultados de validação.....	39
Tabela 11 – Média dos resultados de validação para um treinamento com 100 épocas.....	40
Tabela 12 – Parâmetros e Métricas – Conjunto de teste.....	42

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos	15
1.1.1	Objetivo geral.....	15
1.1.2	Objetivos específicos.....	15
2	REVISÃO DE LITERATURA	16
2.1	Redes neurais artificiais	16
2.1.1	Perceptron.....	16
2.1.2	Redes neurais multicamadas.....	17
2.1.3	Rede neural convolucional.....	18
2.2	A cultura do tomate no mundo e no Brasil	18
2.2.1	O tomate.....	18
2.2.2	Perspectiva global da produção do tomate.....	19
2.2.3	Perspectiva nacional da produção de tomate.....	20
2.2.4	Variedades e cultivares.....	21
2.3	Aprendizado de máquina	25
3	MATERIAIS E MÉTODOS	26
3.1	Abordagens metodológicas	26
3.2	Recurso computacional	27
3.2.1	<i>Hardware</i>	28
3.2.2	<i>Software</i>	28
3.3	Arquitetura da rede neural convolucional	28
3.4	Criação do banco de imagens de tomates	29
3.4.1	Tipos de tomates escolhidos.....	29
3.4.2	Captura das imagens.....	30
3.4.3	Pré-processamento das imagens.....	31
3.5	Modelos de rede convolucional	33
3.6	Métricas de desempenho	34
3.7	<i>Fruits-360</i>	34
3.8	<i>Overfitting</i>	35
3.9	Etapas do trabalho	36
4	RESULTADOS E DISCUSSÕES	38

4.1	Características da base de dados.....	38
4.2	Treinamento e validação.....	38
4.3	Teste.....	41
5	CONCLUSÃO.....	44
	REFERÊNCIAS.....	45

1 INTRODUÇÃO

O tomate, conhecido pelo nome científico *Lycopersicon esculentum*, é um dos vegetais mais versáteis do mundo, pode ser consumido *in natura* ou processado. É rico em nutrientes como: carotenoides, flavonoides, ácidos ascórbicos, vitaminas E, etc. O tomate é nativo dos Andes na América do Sul, sendo mais tarde levado a Europa pelos espanhóis (OLIVEIRA, 2015).

O Brasil, não é só um grande consumidor de tomates, mas um produtor eficiente desta hortaliça, é o 9º produtor mundial. Sendo destaque mundial, não somente pelo volume produzido, mas também pelas tecnologias adotadas em sua produção (AGRISHOW DIGITAL, 2023).

No Brasil, a cultura do tomate é economicamente importante, devido a exportação brasileira anual de aproximadamente quatro milhões de toneladas. Observa-se também que o tomate é uma das hortaliças mais consumidas no mundo, ficando atrás somente das batatas. Destaca-se pela relevância social e contingente de mão de obra que emprega (OLIVEIRA, 2015).

Várias atividades necessitam da interação humana nos processos decisórios, dentre essas atividades está a classificação dos tomates. Realiza-se, na classificação dos tomates, a separação deste produto em diferentes classes. No entanto, esse processo de classificação é de caráter manual e gera custos de mão de obra (SÁNCHEZ, 2023).

Em um mundo cada vez mais globalizado, novas tecnologias surgem constantemente, sendo uma delas a inteligência artificial, cuja origem se deu nos anos 1960 através do surgimento dos primeiros computadores, que permitiu pesquisadores desenvolverem o *perceptron*, passo primordial para a evolução das Redes Neurais Artificiais (RNAs) como conhecemos atualmente (CAMACHO OLMEDO *et al.*, 2018; GRAVES, 2014; HAYKIN, 2001; SANTOS, 2021).

Baseada em conexões neurais do cérebro humano, uma Rede Neural Artificial é composta por vários neurônios artificiais interconectados. A complexidade de uma rede neural está diretamente ligada ao número de neurônios e como eles se conectam. Dessa forma, pesquisadores ao longo do tempo investiram na criação de diferentes arquiteturas (BISHOP; NASRABADI, 2006).

Apesar do avanço de implementação de arquiteturas mais complexas, necessitava-se, em conjunto, um acompanhamento da capacidade tecnológica

computacional, para garantir que os modelos pudessem ser postos em prática. Somente a partir de 2009 houve uma revolução na indústria de placas gráficas, a qual permitiu o início do suprimento do poder de processamento exigido (ELHASSOUNY; SMARANDACHE, 2019).

Com modelos mais robustos e maior capacidade computacional disponível, surgiram então as aplicações. Uma das mais conhecidas atualmente é a detecção de diferentes tipos de objetos através de imagens utilizando Rede Neural Convolutiva (RNC), que dentro desse contexto, o presente trabalho visa classificar diferentes tipos de tomates.

1.1 Objetivos

1.1.1 Objetivo geral

Este trabalho tem como objetivo classificar, por meio de redes neurais convolucionais, diferentes tipos de tomates (variedade ou cultivar) colocados à venda em estabelecimentos da região oeste paranaense.

1.1.2 Objetivos específicos

- a) Criar um banco de imagens com diferentes tipos de tomates por meio de um smartphone;
- b) Realizar o pré-processamento das imagens para eliminar inconsistências;
- c) Treinar, validar e testar modelos de redes neurais convolucionais, utilizando o banco de imagens construído.

2 REVISÃO DE LITERATURA

2.1 Redes neurais artificiais

Redes Neurais Artificiais, do inglês *Artificial Neural Networks* (ANNs), recebem este nome, pois foram inspiradas em neurônios e suas conexões no cérebro humano. Os nós de uma Rede Neural remetem-se aos neurônios, onde, os mesmos são interconectados entre as camadas, de modo a maximizar o acerto na predição da saída (O'SHEA, NASH, 2015).

2.1.1 Perceptron

O perceptron é conhecido como a menor unidade em uma rede neural artificial, além disso, se a rede possui somente um perceptron, pode-se dizer que ela se encontra em sua forma mais simples (Camacho Olmedo *et al.*, 2018). Foi inicialmente proposto por Frank Rosenblatt em 1962 através de seu trabalho "*Principles of Neuro-dynamics: Perceptrons and the Theory of Brain Mechanisms*" – Princípios da Neurodinâmica: Perceptrons e a Teoria dos Mecanismos do Cérebro (BISHOP; NASRABADI, 2006).

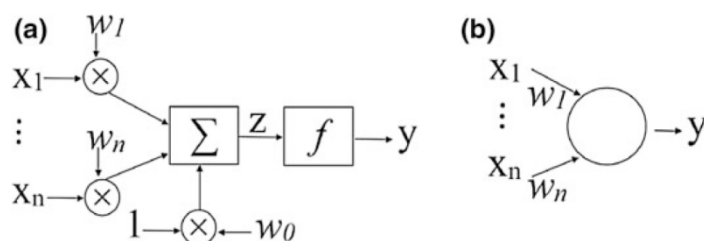
Um perceptron pode ser representado através da equação matemática (Equação 1) e por diagramas (Figura 1) (CAMACHO OLMEDO *et al.*, 2018).

$$y = f(z) \quad \text{onde,} \quad z = \sum_{i=0}^n \omega_i \cdot x_i = \omega_0 + \sum_{i=1}^n \omega_i \cdot x_i \quad (1)$$

Onde:

- x_i : entradas (*inputs*);
- ω_i : pesos (*weights*);
- ω_0 : limiar (*bias*);
- $f(z)$: função de ativação (*activation function*);
- y : saída (*output*).

Figura 1 – Modelo do perceptron: (a) com detalhes; (b) simplificado



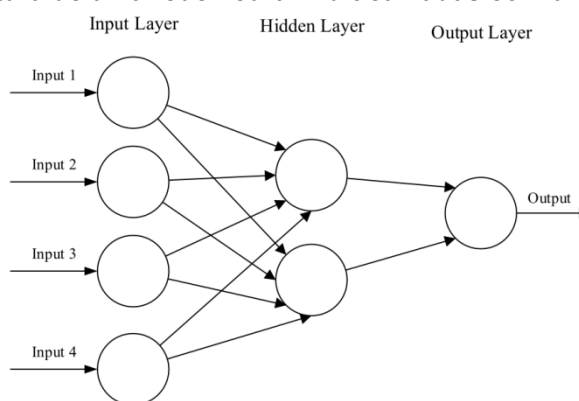
Fonte: Camacho *et al.* (2018, p. 452)

2.1.2 Redes neurais multicamadas

Segundo O'Shead e Nash (2015), uma rede neural é alimentada através da camada de entrada por um vetor multidimensional, que por sua vez é distribuído para as camadas ocultas, responsáveis pela tomada de decisão do processamento da informação recebida da camada anterior. Ao receber as informações, as camadas então as repassam até a última, que é responsável pela predição, para que a rede possa então analisar seu erro e aprender com o mesmo, realizando então a atualização dos pesos.

Uma Rede Neural Multicamadas é geralmente conhecida como Rede Neural Profunda, do inglês *Deep Neural Network*. Na Figura 2 é possível observar uma rede com apenas uma camada oculta, entretanto, ela pode conter quantas forem necessárias para satisfazer a predição de saída.

Figura 2 – Estrutura de uma rede neural multicamadas com uma camada oculta



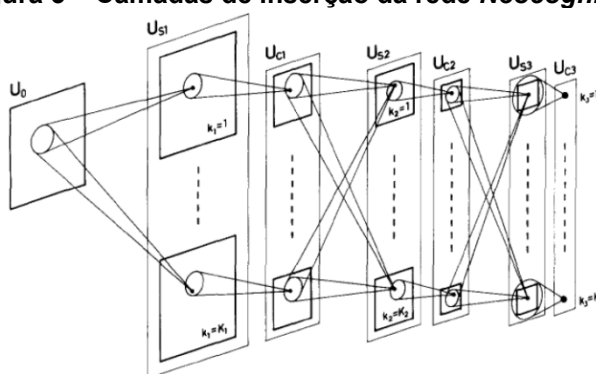
Fonte: O'Shea e Nash (2015)

2.1.3 Rede neural convolucional

A arquitetura de uma Rede Neural Convolucional (*Convolutional Neural Network* – CNN) é composta por camadas do tipo: *convolutional* (convolucional), *activation* (ativação), *pooling* (“agrupamento”), etc. A arquitetura de uma CNN não é a mesma para todos os tipos de situações, para isso modelos específicos vem sendo implementados desde 1960.

A primeira arquitetura de uma rede convolucional implementada foi a *Neocognitron* (Figura 3), responsável pelos mecanismos de reconhecimento de padrões. Até 2009, as arquiteturas das redes neurais (no geral) eram limitadas, devido às tecnologias computacionais ainda não atenderem ao desempenho necessário (armazenamento, processamento, etc.).

Figura 3 – Camadas de inserção da rede Neocognitron



Fonte: Elhassouny e Smarandache (2019)

A partir de 2009 surgem arquiteturas mais robustas, devido ao aumento do poder computacional disponível no mercado. Isso possibilitou a criação de modelos de redes convolucionais, como: AlexNet (2012), ZFNet (2013), VGGNet (2014), GoogLeNet (2014), dentre várias outras (ELHASSOUNY; SMARANDACHE, 2019).

2.2 A cultura do tomate no mundo e no Brasil

2.2.1 O tomate

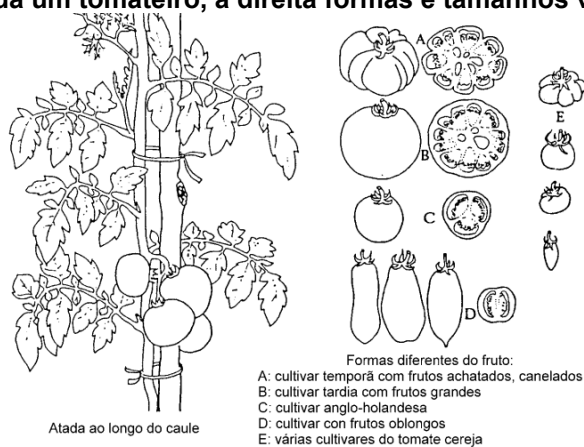
Lycopersicon esculentum Mill., hortaliça popularmente conhecida como tomate, é uma das mais consumidas no mundo. Pertence à família das solanáceas, assim como o pimentão, a pimenta, a berinjela, a batata, entre outras. Sua origem

histórica deu-se no México, América Central e América do Sul. Foi levada à Europa no século XVI pelos espanhóis, onde mais tarde durante século XIX seu consumo tomaria maiores proporções.

O tomate possui propriedades importantes como: vitaminas B e C, ferro e fósforo, propriedade essas, importantes para a manutenção da saúde do nosso organismo. Pode ser consumido fresco, em saladas, cozido, em sopas, em molhos e combinados com muitos pratos (Dusi *et al.*, 1993). Além, de seu potencial na saúde, possui grande impacto na economia, já que é uma hortaliça de fácil cultivo e pode ser armazenada de diversas formas, em latas e desidratada, prolongando sua vida útil.

O tomateiro (Figura 4) é uma planta anual, porém na América do Sul é possível realizar a colheita a partir da mesma planta, mais de uma vez. Existem tomates de diversas formas e tamanhos, até mesmo de várias cores, podendo sua coloração variar do amarelo ao vermelho (SHANKARA *et al.*, 2005).

Figura 4 – À esquerda um tomateiro, à direita formas e tamanhos variados de tomates

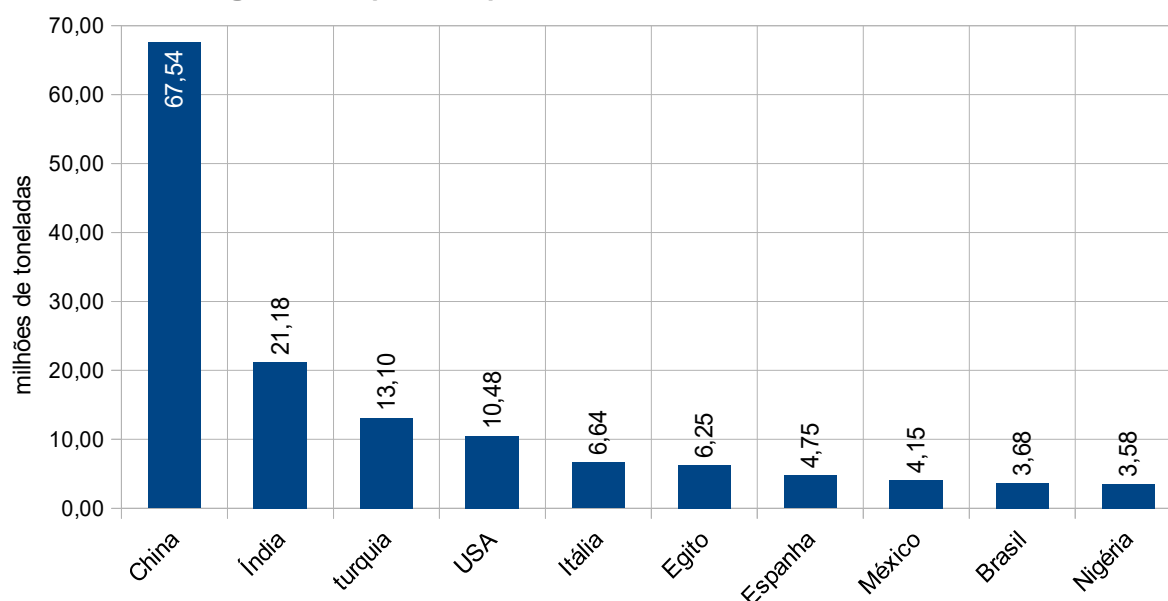


Fonte: Shankara et al. (2005)

2.2.2 Perspectiva global da produção do tomate

Através do gráfico da Figura 5 pode-se observar os 10 maiores produtores de tomate no mundo em 2021. A China Continental foi a maior produtora com 67,54 milhões de toneladas, em seguida Índia (21,18 milhões de toneladas) e logo após turquia com 13,1 milhões. O Brasil ocupou o 9º lugar do *ranking*, com uma produção de 3,68 milhões de toneladas.

Figura 5 – Top 10 dos produtores de tomate no mundo em 2021



Fonte: Adaptado de FAOSTAT (2023)

2.2.3 Perspectiva nacional da produção de tomate

Segundo o Levantamento Sistemático da Produção Agrícola (LSPA) de dezembro de 2022, o Brasil produziu um total de 3.856.430 toneladas de tomate, sendo o rendimento médio por hectare de aproximadamente 71,14 toneladas e uma produção per capita de 18,55 quilogramas por habitante (SIDRA-IBGE, 2023).

O estado que mais produziu foi Goiás, com 1.147.440 toneladas, sendo responsável por quase 1/3 da produção nacional (29,75%), em seguida o estado de São Paulo com 892.100 toneladas (23,13%) (Tabela 1). O estado do Acre, Amazonas, Amapá, Tocantins e Sergipe, não possuem dados para a análise do período.

Tabela 1 – Produção de tomate no Brasil em 2022

UF	Produção [ton/ano]	Área plantada [ha]	Rendimento médio [ton/ha]	Rendimento per capita [kg/hab]	Produção [%]
GO	1.147.440,00	12.583,00	91,19	165,08	29,75%
SP	892.100,00	11.400,00	78,25	19,38	23,13%
MG	578.116,00	7.720,00	74,89	27,88	14,99%
PR	255.000,00	4.000,00	63,75	21,55	6,61%
BA	178.004,00	4.240,00	41,98	12,14	4,62%
CE	170.104,00	2.335,00	72,85	19,03	4,41%
ES	151.636,00	2.364,00	64,14	36,91	3,93%
RJ	145.481,00	2.186,00	66,55	8,76	3,77%
SC	118.537,00	2.062,00	59,81	15,27	3,07%
RS	90.731,00	1.972,00	46,13	8,18	2,35%

UF	Produção [ton/ano]	Área plantada [ha]	Rendimento médio [ton/ha]	Rendimento per capita [kg/hab]	Produção [%]
PE	44.540,00	1.040,00	42,95	4,92	1,15%
DF	27.844,00	350,00	79,55	9,52	0,72%
PB	19.059,00	670,00	30,01	4,73	0,49%
AL	9.303,00	215,00	44,51	2,98	0,24%
RN	6.120,00	195,00	31,39	1,85	0,16%
RR	4.689,00	264,00	17,76	7,39	0,12%
PI	4.343,00	175,00	24,82	1,33	0,11%
PA	4.068,00	165,00	24,66	0,48	0,11%
MT	3.573,00	176,00	20,30	0,94	0,09%
MA	3.125,00	154,00	20,29	0,46	0,08%
MS	1.617,00	51,00	33,00	0,57	0,04%
RO	1.000,00	26,00	38,46	0,62	0,03%
AC	-	-	-	-	-
AM	-	-	-	-	-
AP	-	-	-	-	-
TO	-	-	-	-	-
SE	-	-	-	-	-
Brasil	3.856.430,00	54.343,00	71,14	18,55	100,00%

Fonte: SIDRA-IBGE (2023)

2.2.4 Variedades e cultivares

Segundo a EMBRAPA (2023), as primeiras espécies de tomates existentes foram disseminadas ao longo do tempo por agricultores. A expressão “variedade” está atrelada à propagação da planta ao longo do tempo sem a interferência seletiva humana. Já “cultivar” se remete a preferência do cultivo de frutos que possuam características que sejam desejáveis para o homem (Jorge, 2023). Visando somente as características dos frutos, o termo “tipo” será usado para se referir a um fruto em particular ou um grupo com qualidades em comum.

Basicamente, as cultivares de tomate consumidas *in natura* estão divididas em quatro grandes grupos: Cereja, Italiano, Salada e Santa Cruz. Dentre os quatro, o grupo Cereja (Figura 6) possui os menores frutos, podendo seus cachos produzirem entre 12 e 18 tomates, essa cultivar é muito usada na decoração de pratos por conta de seu tamanho. Além de sua coloração avermelhada, o fruto também pode ter variação de cor amarelada.

Figura 6 – Tomate do grupo Cereja

Fonte: EMBRAPA (2023)

Na Tabela 2 apresentam-se híbridos pertencentes ao grupo cereja. Pode-se observar, por meio desta tabela, que a massa de um híbrido pertencente ao grupo, em estágio de maturação, pode variar de 15 g (Zamir) até 40 g (Red Petit).

Tabela 2 – Alguns híbridos pertencentes ao grupo Cereja

Híbrido	Empresa	Massa
Zamir	Embrapa e Agrocinco	15 g
Renata	Sakama	18 g
Red sugar	Sakama	20 g
Sindy	Sakama	15 a 20 g
Sweet Gold	Sakata	15 a 25 g
Samambaia	TopSeed	20 a 30 g
Red Petit	Sakama	40 g

Fonte: EMBRAPA (2023)

Os tomates do tipo Italiano possuem uma coloração vermelha intensa, uma polpa densa e firme. Possuem formato longo, medindo de 7 a 10 cm de comprimento, podendo alguns frutos terem uma extremidade pontiaguda. Uma demonstração do fruto pertencente a esse grupo pode ser observada na Figura 7.

Figura 7 – Tomate do grupo Italiano

Fonte: EMBRAPA (2023)

Alguns híbridos pertencentes ao grupo Italiano podem ser vistos na Tabela 3. Pode-se analisar que a massa de um fruto mais leve já é, pelo menos, duas vezes maior que a massa do fruto do tipo Cereja, ou seja, temos frutos partindo de 95 gramas (San Vito) indo até 200 gramas (BRS Nagai e Giuliana).

Tabela 3 – Alguns híbridos pertencentes ao grupo Italiano

Híbrido	Empresa	Massa
San Vito	Embrapa	95 a 105 g
IPA 6	Hortec	100 a 150 g
Grande HT	Hortec	110 a 130 g
Andrea	Sakata	120 a 140 g
Júpiter	Hortec	130 a 150 g
Andrea Vistory	Sakata	140 a 160 g
Tyna	Sakata	160 a 180 g
BRS Nagai / Giuliana	Agrocinco / Sakata	180 a 200 g

Fonte: EMBRAPA (2023)

A cultivar Salada vista na Figura 8, possui coloração rosada ou avermelhada. Seu formato é globular com “protuberâncias” alternadas, de tamanho grande com híbridos que podem chegar até 500 g de massa.

Figura 8 – Tomate do grupo Salada



Fonte: EMBRAPA (2023)

Na Tabela 4 apresenta-se informações sobre alguns tipos híbridos pertencentes ao grupo da cultivar Salada, com frutos que podem variar entre 180 e

350 gramas.

Tabela 4 – Alguns híbridos pertencentes ao grupo Salada

Híbrido	Empresa	Massa
Lumi	Sakata	180 a 200 g
BRS Portinari	Agrocinco	180 a 240 g
Aliança	Hortec	200 a 220 g
Sheila	Sakata	200 a 240 g
Tyler	Sakata	200 a 250 g
Ivete	Sakata	210 a 220 g
Pleno F1	Hortec	220 a 270 g
Gisele	Sakata	280 a 350 g

Fonte: EMBRAPA (2023)

O tomate do tipo Santa Cruz (Figura 9) foi inicialmente selecionado por agricultores, mais tarde, empresas por meio de pesquisas utilizaram técnicas de melhoramento genético para criar frutos maiores e mais resistentes. Possui híbridos que variam sua massa, indo de 80 até 220 gramas. Detém um sabor mais ácido, tem um preço mais em conta e é muito procurado em supermercados.

Figura 9 – Tomate do grupo Santa Cruz



Fonte: EMBRAPA (2023)

Através da Tabela 5 é possível identificar vários híbridos do grupo da cultivar Santa Cruz, com frutos que possuem massa de 130 a 140 gramas, como é o caso da Débora Plus, e frutos até 200 g como o híbrido BRS Kiara.

Tabela 5 – Alguns híbridos pertencentes ao grupo Santa Cruz

Híbrido	Empresa	Massa
Débora Plus	Sakata	130 a 140 g
Santa Cruz Kada	Hortec	130 a 150 g
Delta / Débora Max	Hortec / Sakata	140 a 160 g
Santa Clara / Santa Clara VF 5600	Hortec / Sakata	150 a 180 g
Débora Victory / Kombat	Sakata / Hortec	160 a 180 g
Carina	Sakata	180 a 200 g
BRS Kiara	Agrocinco	Até 200 g

Fonte: EMBRAPA (2023)

2.3 Aprendizado de máquina

O aprendizado é a capacidade de se adaptar, modificar e aprimorar as respostas e comportamentos, além de ser um atributo fundamental dos seres chamados de inteligentes, sejam eles humanos ou não (Osório; Bittencourt 2000). O aprendizado de máquina é o uso de métodos computacionais para coletar dados, aprender padrões com eles, e então fazer uma determinação ou previsão utilizando esses padrões (ROSA, 2019).

Segundo Rosa (2019) e Braga *et al.* (2000), as tarefas de treinamento de aprendizado de máquina podem ser divididas em três campos, de acordo com o tipo de *feedback* utilizado no treinamento. São eles:

- **Aprendizado Supervisionado:** No treinamento supervisionado existe um supervisor externo que fornece a rede tanto os valores de entrada disponíveis quanto os valores de saída esperados;
- **Aprendizado não Supervisionado:** Considera-se um aprendizado não supervisionado, como o próprio nome evidencia, quando não há um “supervisor externo”, ou seja, somente os padrões de entrada estão disponíveis à rede;
- **Aprendizado de Reforço:** no aprendizado de reforço a máquina interage com o ambiente de forma dinâmica e recebe um *feedback* baseado em uma função de avaliação, para ajustar seus pesos e melhorar a cada iteração.

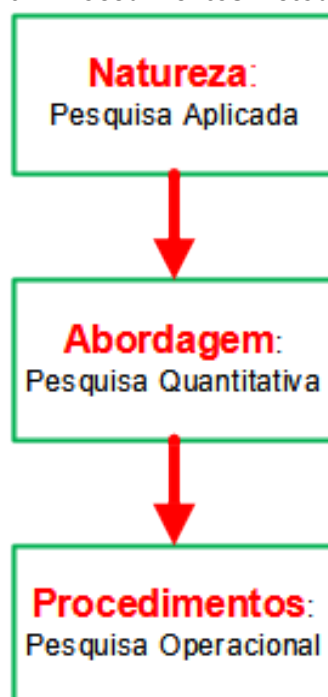
3 MATERIAIS E MÉTODOS

Nesse capítulo abordam-se a metodologia e as tecnologias utilizadas na pesquisa e no desenvolvimento do trabalho.

3.1 Abordagens metodológicas

Os procedimentos metodológicos, desenvolvidos nesta pesquisa, são apresentados no fluxograma da Figura 10 (Gil, 2010). Neste estudo, a metodologia utilizada pode ser considerada quanto:

Figura 10 – Procedimentos metodológicos



Fonte: Autoria própria (2023)

- **Natureza:** como Pesquisa Aplicada: porque visa gerar conhecimento para uma aplicação prática;
- **Abordagem:** como Quantitativa: porque tem como objetivo quantificar um problema e entender a dimensão dele;
- **Procedimentos:** como Operacional: porque busca solução para um determinado problema por meio de modelos matemáticos ou estatístico.

3.2 Recurso computacional

Os recursos computacionais são primordiais para um bom desempenho durante a execução de uma rede neural, o *hardware* e o *software* devem garantir uma boa eficiência no tempo de execução, visto que sejam necessários, no decorrer do processo de construção da arquitetura, vários ajustes de parâmetros.

O *hardware* para redes neurais deve atender os requisitos básicos atrelados aos fatores: quantidade de dados de entrada (memória RAM), número de iterações e operações matemáticas em baixa escala (CPU) e operações matemáticas em grande escala (GPU). O *software* é escrito em uma determinada linguagem de programação, o qual é a ponte que interligará os dados ao *hardware*.

A escolha dos mesmos impacta o fluxo de trabalho, onde há a necessidade de produtividade e desempenho. A produtividade está atrelada ao nível de abstração do *software* em que se implementa a rede neural (facilidade de escrita e interpretação do problema). Já o desempenho depende da eficácia da linguagem de programação e do *hardware*, para que trabalhem em conjunto para resolver a tarefa demandada.

Como exemplo se pode citar uma comparação, feita por (WU et al., 2018), onde foram utilizados os seguintes recursos de *hardware* e *software*: processador (CPU) baseado na arquitetura de 64 bits, placa de vídeo (GPU) NVIDIA® GeForce® GTX™ 1070, 8 Giga Bytes de memória RAM, sistema operacional Linux Ubuntu de 64 bits. Os recursos foram responsáveis por treinar 3 arquiteturas de redes neurais diferentes, a FCN8s, U-Net e BR-Net, sendo todas implementadas através da biblioteca PyTorch disponível para a linguagem de programação Python.

Para realizar o procedimento de comparação os autores utilizaram os seguintes parâmetros e hiperparâmetros: otimizador estocástico Adam, taxa de aprendizagem de 0,0002, pesos e limiar variando entre 0,9 e 0,999, *batch size* (tamanho do lote de amostras) de 24, e por fim, 10.000 iterações. Observando a Tabela 6 nota-se que o desempenho foi medido em FPS (*Frames Per Second*) tanto para o treino quanto para o teste. Durante o treinamento o FPS é menor, uma vez que são realizadas operações mais complexas, já para a fase de teste essa complexidade não é necessária, resultando em um desempenho 3 a 4 vezes maior.

Tabela 6 – Desempenho obtido no treinamento e teste dos modelos FCN8s, U-Net e BR-Net

Stage	FCN8s	U-Net	BR-Net (–BN/ReLU)	BR-Net (+BN/ReLU)	BR-Net (–BN/LeakyReLU)	BR-Net (+BN/LeakyReLU)
Training (FPS)	29.3	91.7	88.1	80.2	86.6	78.9
Testing (FPS)	130.2	280.6	276.5	252.5	274.1	249.9

Fonte: WU et al. (2018)

3.2.1 Hardware

Para Treinar Os Modelos De Redes Neurais Convolucionais foram utilizados os seguintes requisitos:

- **CPU:** Intel Core I5-1135g7;
- **GPU:** Intel Iris Xe Graphics Tiger Lake-Up3 GT2 (Integrada);
- **Memória RAM:** 20 Gb DDR-4 SDRAM.

3.2.2 Software

Para o treinamento das redes neurais convolucionais foram necessários alguns elementos base, como: linguagem de programação, *framework* de *machine learning* e bibliotecas complementares para o tratamento de dados e imagem. Assim, tem-se os seguintes elementos de *software* utilizados:

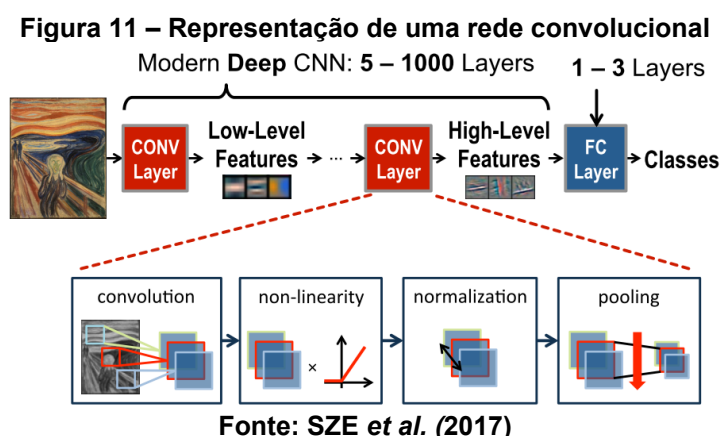
- **Linguagem de programação:** Python versão 3.10 em ambiente Conda;
- **Framework de machine learning:** TensorFlow DirectML Plugin (versão 0.4.0.dev230202) e Keras 2.10;
- **Bibliotecas complementares:** Numpy 1.24.2, OpenCV 4.8.0.76, Sklearn 0.0.post9 e Matplotlib 3.8.

3.3 Arquitetura da rede neural convolucional

Uma arquitetura de rede neural é composta por parâmetros e hiperparâmetros, basicamente os hiperparâmetros são elementos definidos antes do início do treinamento da rede, exemplo: taxa de aprendizagem (*learning rate*), número de iterações ou épocas (*epochs*), número de camadas escondidas (*hidden layers*), número e neurônios por camada escondida, função de ativação (*activation*

function), entre outros. Já os parâmetros são atualizados durante o treinamento, ou seja, são os elementos que aprendem, são eles: pesos e limiares (*weights* e *biases*).

Para uma rede convolucional, além dos parâmetros e hiperparâmetros citados anteriormente, há as camadas de convolução (que dá nome a rede neural) e *pooling* para compor a arquitetura. A Figura 11 mostra a composição de uma rede neural convolucional moderna, cuja quantidade de camadas pode ser consideravelmente grande, dependendo das necessidades da aplicação.



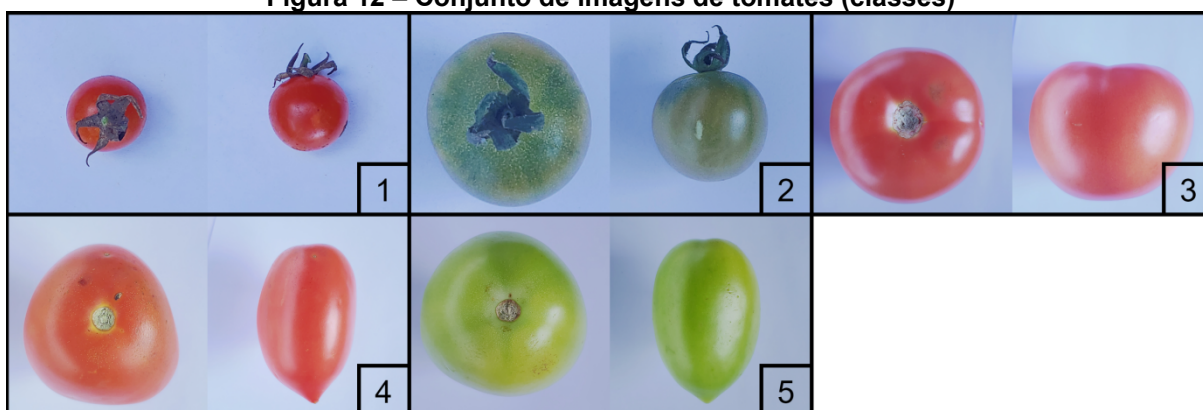
3.4 Criação do banco de imagens de tomates

3.4.1 Tipos de tomates escolhidos

Uma das etapas primordiais para a execução do trabalho foi a escolha dos tipos de tomates a serem utilizados para o treinamento das redes neurais convolucionais. Optou-se ao menos, por três tipos, dos quais, pelo menos dois deles possuísem características semelhantes entre si. Isso se dá pelo fato de que a investigação buscou colocar em teste o desempenho dos modelos utilizados neste trabalho.

Pode-se observar, de acordo com a Figura 12, os seguintes tipos escolhidos: tomate cereja maduro – 1 (grupo cereja); tomate cereja verde – 2; tomate longa vida maduro – 3 (grupo salada); tomate rasteiro maduro – 4 (grupo italiano) e tomate rasteiro verde – 5. Os respectivos nomes são os apresentados para venda nos estabelecimentos em que foram adquiridos.

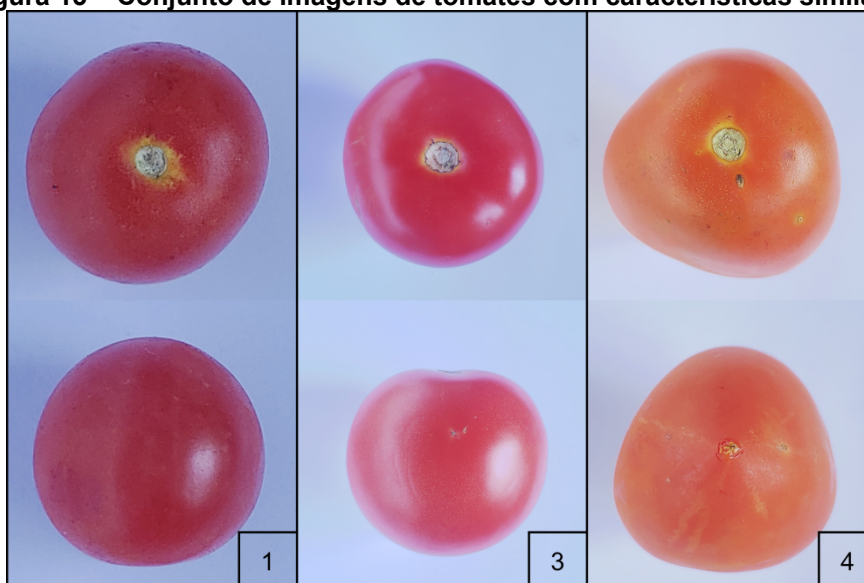
Figura 12 – Conjunto de imagens de tomates (classes)



Fonte: Autoria própria (2023)

Observa-se, na figura anterior, uma certa semelhança de cor entre os tipos 1 e 3, e de características em relação aos tomates 3 e 4. Na Figura 13, para reforçar ainda mais as semelhanças, pode-se observar outras amostras. Nota-se que há uma semelhança ainda maior entre os tomates do tipo cereja e do tipo longa vida (1 e 3), sendo somente a cor do tomate 4 ligeiramente diferente.

Figura 13 – Conjunto de imagens de tomates com características similares



Fonte: Autoria própria (2023)

3.4.2 Captura das imagens

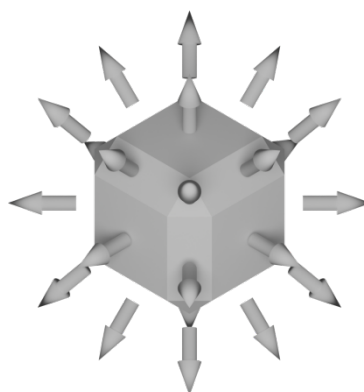
Para capturar as imagens foi acoplado um *smartphone* a um tripé apontado para o elemento de interesse. O elemento em questão foi colocado em frente a uma folha de papel comum de cor branca para padronizar o fundo de todas as imagens e

não gerar vieses com elementos que pudessem alterar de forma significativa o foco do contexto.

Além do fundo das imagens, a questão da iluminação também foi um elemento importante. As imagens foram capturadas em ambiente com luz natural, pois foi necessário para ressaltar as características dos tomates.

A Figura 14 representa quais foram os pontos de vista em que foram realizadas as capturas das imagens dos tomates. No total foram ao menos 26 imagens para cada amostra, indicado a cada flecha. Estes pontos de vista são importantes para adicionar variedade e qualidade nos dados, influenciando em uma melhor taxa de acerto de predição dos modelos.

Figura 14 – Pontos de vista de captura das imagens das amostras



Fonte: Autoria própria (2023)

A câmera utilizada foi uma câmera do *smartphone* Galaxy S9® da fabricante Samsung™. Para capturar as imagens ainda foi necessário utilizar o aplicativo Open Camera para o sistema Android em uma configuração de resolução de 2976x2976 pixels. O aplicativo em questão foi escolhido pois uma configuração importante como a exposição da luz, por exemplo, pode ser travada de forma fácil, impedindo o ajuste automático.

3.4.3 Pré-processamento das imagens

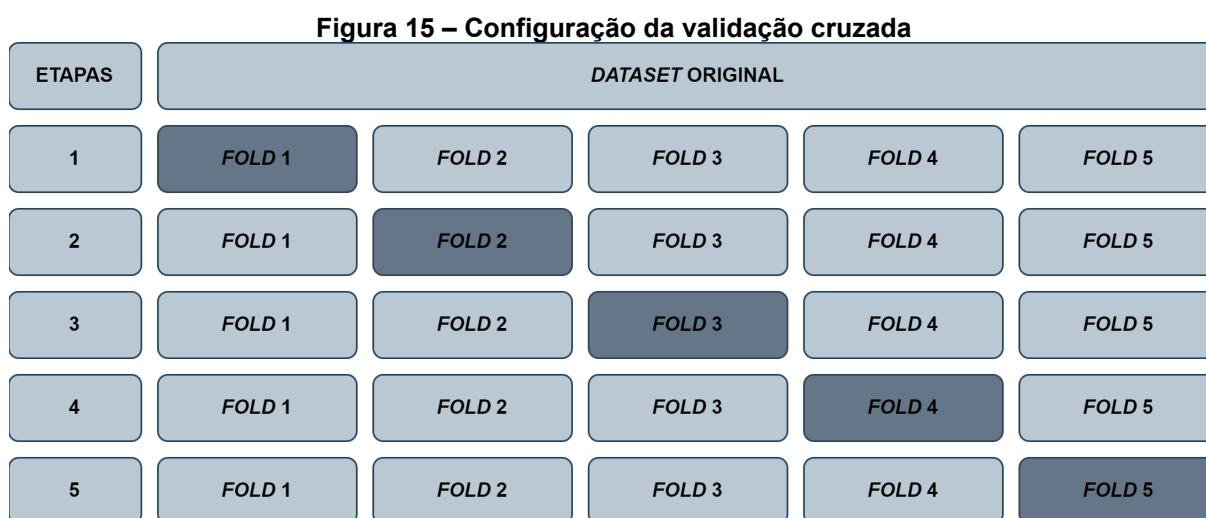
Realizou-se, após obter as imagens das amostras, uma inspeção manual para identificar possíveis elementos indesejados nas mesmas. Observa-se que não houve a necessidade de descarte de nenhuma imagem. Em seguida, para uma melhor organização, foi realizada a separação de cada tipo de amostra em pastas

separadas (ver Figura 12).

Com as pastas organizadas, foi então realizada a etapa de separação dos dados em *folds* (dobras ou partes). Esta etapa foi importante para a validação cruzada dos dados (*cross-validation*), ou seja, uma verificação de quais partes do banco de imagem o modelo de rede neural convolucional melhor generalizou em relação a predição das classes na fase de validação.

Inicialmente, realizou-se o embaralhamento das imagens para gerar uma melhor distribuição das mesmas. Na Figura 15 pode-se observar a forma como cada etapa da validação cruzada foi realizada. A primeira linha se trata da dimensão do *dataset* original, utilizado como referência.

Cada linha representa uma das 5 formas em que os dados de treinamento e validação foram separados. Cada *fold* destacado em fundo cinza escuro representa 20% da linha e foi utilizado na etapa da validação. Os outros quatro *folds* (80%) representam os dados que foram utilizados durante a fase de treinamento.



Fonte: Autoria própria (2023)

Após a separação, de cada etapa da validação cruzada, fez-se a configuração para a geração do aumento de dados (*data augmentation*), mas antes disso foi realizado a normalização dos dados (*rescale*), ou seja, como os dados de entrada da rede neural são os valores das cores dos *pixels* das imagens que variam de 0 até 255, esses valores foram então normalizados para uma faixa de 0 a 1.

A técnica de *data augmentation* visa expandir e gerar diversidade no tamanho do banco de dados, seja rotacionando, escalonando, distorcendo, modificando cores, etc. (TENÓRIO, 2019).

A obtenção de imagens, no mundo real, pode ter sido realizada a partir de um conjunto limitado de condições. Contudo o modelo treinado precisa classificar em uma variedade de condições, tais como brilho, localização, escala, entre outras. Por isso, utiliza-se a técnica *data augmentation* para treinar a rede neural com dados adicionais, modificados sinteticamente (ROSA, 2019).

As técnicas de *data augmentation*, utilizadas inicialmente neste trabalho, foram:

- **Faixa de cisalhamento (*shear range*):** visa distorcer a imagem de modo que seu formato fique semelhante a um paralelogramo. O valor escolhido foi de 20%;
- **Faixa de rotação (*rotation range*):** gira as amostras de imagens a cada 30°;
- **Faixa de zoom (*zoom range*):** amplia ou diminui a imagem. O valor escolhido foi de 50% para mais e 50% para menos;
- **Giro horizontal (*horizontal flip*):** espelha a imagem em relação ao eixo y;
- **Faixa de brilho (*brightness range*):** visa alterar o valor do brilho da imagem para simular cenários com pouca ou muita iluminação. Foram escolhidos valores de 50% menos brilho e 50% mais brilho.

Observa-se que as amostras de treino e validação são as amostras que passaram pelas etapas de pré-processamento e *data augmentation*.

3.5 Modelos de rede convolucional

Na Tabela 7 apresentam-se os modelos de redes neurais convolucionais utilizados neste trabalho. Os modelos possuem características distintas, como, por exemplo, o número de parâmetros. Observa-se que devido as restrições, de hardware e quantidade de classes (tipos de tomates – Figura 12) a serem treinadas, optou-se por modelos com quantidades de parâmetros de 3,5 milhões a 25,5 milhões.

Tabela 7 – Modelos de Redes Neurais Convolucionais

Modelo	Autor(es)	Ano de publicação	Parâmetros (milhões)
ResNet50V2	HE, Kaiming et al.	2016	25,60
MobileNet	François Chollet	2017	4,30
MobileNetV2	François Chollet	2017	3,50
Xception	François Chollet	2017	22,90
EfficientNetV2B0	Mingxing Tan, Quoc V. Le	2021	7,20
EfficientNetV2B1	Mingxing Tan, Quoc V. Le	2021	8,20
EfficientNetV2B2	Mingxing Tan, Quoc V. Le	2021	10,20
EfficientNetV2B3	Mingxing Tan, Quoc V. Le	2021	14,50

Fonte: TEAM (2023)

3.6 Métricas de desempenho

Para avaliar o desempenho das redes convolucionais foram utilizadas as métricas de acurácia (*accuracy*) e erro (*loss*). São métodos utilizados para medir o desempenho de um modelo de classificação. A acurácia (*accuracy*) (Equação 2) avalia simplesmente o percentual de acertos, ou seja, ela pode ser obtida pela razão entre a quantidade de acertos e o total de entradas (MARIANO, 2023):

$$Acurácia = \frac{Total\ de\ Acertos}{Total\ de\ Itens} \quad (2)$$

Já o erro (*loss*) é uma função de perda, também conhecida como função de custo, leva em consideração as probabilidades ou incertezas de uma previsão com base no quanto a previsão varia do valor real. Isso nos dá uma visão mais detalhada do desempenho do modelo.

Ao contrário da acurácia, a perda não é uma porcentagem, mas uma soma dos erros cometidos para cada amostra nos conjuntos de treinamento ou validação. A perda é frequentemente usada no processo de treinamento para encontrar os “melhores” valores de parâmetros para o modelo, por exemplo, os pesos da rede neural (AI WIKI, 2023).

3.7 Fruits-360

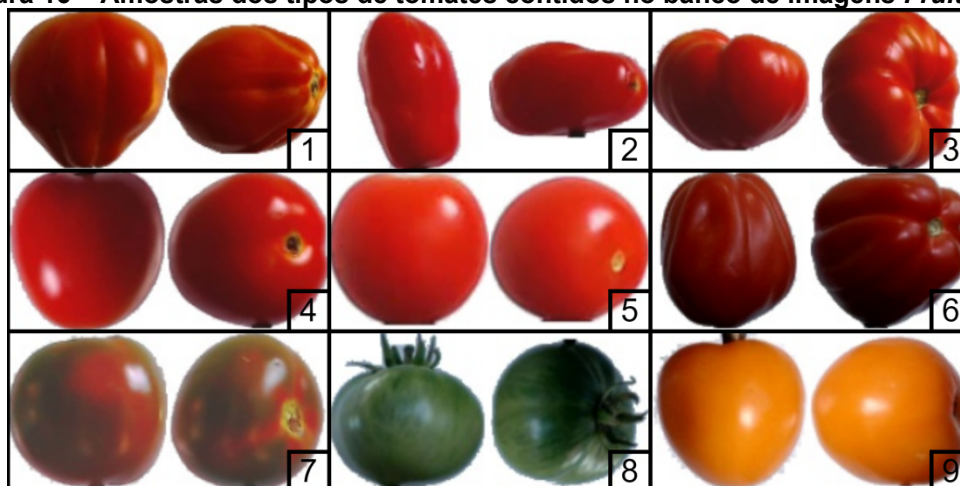
Para o presente trabalho, serão utilizados dois bancos de imagens. O primeiro se refere a um banco chamado *Fruits 360* obtido em repositório na internet

(Mureşan; Oltean, 2018). O banco contém vários tipos de vegetais, porém só foram aproveitadas as imagens de tomates.

Por se tratar de imagens que já passaram pela etapa de pré-processamento, são ideais para iniciar já na fase de construção, treinamento e testes da arquitetura da rede convolucional, gerenciando o tempo enquanto o segundo banco esteja sendo criado.

Amostras do banco de imagens de tomates pertencentes ao conjunto *Fruits 360* podem ser vistas na Figura 16. O autor identificou os frutos pelos seguintes nomes: 1 – *Tomato 1*; 2 – *Tomato 2*; 3 – *Tomato 3*; 4 – *Tomato 4*; 5 – *Tomato Cherry Red*; 6 – *Tomato Heart*; 7 – *Tomato Maroon*; 8 – *Tomato not Ripened*; 9 – *Tomato Yellow*. Esses nomes serão importantes para identificação da imagem durante a etapa de treinamento da rede neural convolucional.

Figura 16 – Amostras dos tipos de tomates contidos no banco de imagens *Fruits 360*



Fonte: Mureşan e Oltean (2023)

3.8 Overfitting

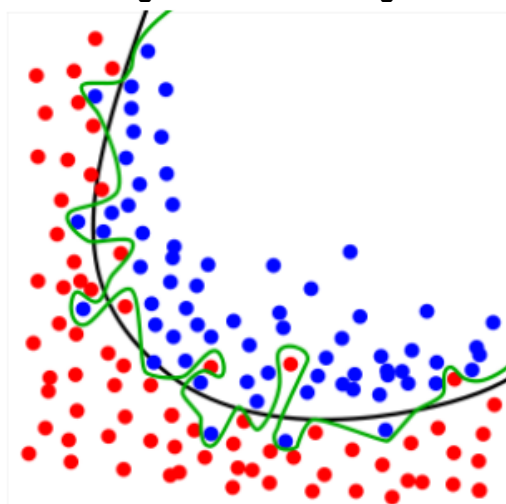
Overfitting (sobreajuste) é um fenômeno que pode ocorrer durante o treinamento de modelos de aprendizado de máquina. Este fenômeno acontece quando o modelo treinado se ajusta muito bem ao conjunto de dados de treinamento, porém é ineficaz ao tentar classificar novos dados. Apresenta alta acurácia com os dados de treinamento, porém não é uma representação confiável da realidade (ROSA, 2019).

Para minimizar o impacto do overfitting, utilizou-se, neste trabalho, os seguintes métodos:

- **Cross-validation (validação cruzada):** método que tem como objetivo o intercâmbio gradual entre amostras do banco de imagens do treino e de validação com intuito de melhorar a generalização da rede;
- **Dropout (“abandonar”):** este método remove neurônios desnecessários da rede com base no desempenho do modelo.

Na Figura 17 apresenta-se um modelo de predição em que a linha verde representa um modelo com *overfitting* e a linha preta um modelo bem ajustado (“*just right*”).

Figura 17 – *Overfitting*



Fonte: Rosa (2019)

3.9 Etapas do trabalho

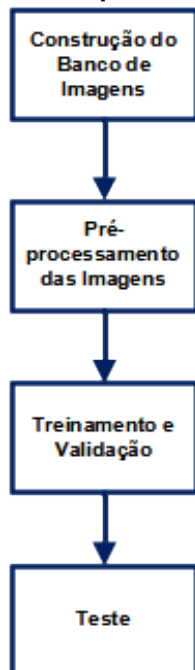
Neste trabalho, para realizar a classificação das imagens de tomate, foram realizadas as etapas apresentadas a seguir (Figura 18):

- **Banco de Imagens:** Inicialmente, na primeira etapa, foi criado um banco de imagens com diferentes tipos de tomates, e realizou-se o pré-processamento das imagens para eliminar inconsistências;
- **Treinamento e Validação:** Na segunda etapa foram realizados o treinamento e a validação. Foram selecionados então, por meio das métricas, os melhores modelos dentre: MobileNet, MobileNetV2, Xception, ResNet50V2, EfficientNetV2B0, EfficientNetV2B1, EfficientNetV2B2,

EfficientNetV2B3;

- **Teste:** Nessa última etapa, os melhores modelos, encontrados na etapa anterior, foram então submetidos aos dados do conjunto de imagens de teste (conjunto de imagens que não participaram da etapa de treinamento e validação).

Figura 18 – Etapas do trabalho



Fonte: Autoria própria (2023)

4 RESULTADOS E DISCUSSÕES

4.1 Características da base de dados

Na Tabela 8 apresentam-se as características obtidas, da base de dados, depois da etapa de pré-processamento das imagens. Nesta tabela apresentam-se: o número de classes (ver Figura 12), o tamanho das amostras de treino, validação e teste, a dimensão das amostras (largura x altura da imagem) e a quantidade de canais das amostras (valores em pixels que representam as cores RGB – *Red*, *Green* e *Blue*, com valores que variam de 0 a 255, que posteriormente foram normalizados entre 0 e 1).

Tabela 8 – Característica do banco de imagens de tomates

Características	Valor
Classes	5
Amostras de treino	521
Amostras de validação	131
Amostras de teste	12
Dimensão das amostras (pixels)	400x400
Canais por amostra	3

Fonte: Autoria própria (2023)

As dimensões das amostras escolhidas, largura e altura em pixels das imagens, tiveram como objetivo manter um certo grau das características das imagens originais e mesmo assim ganhar performance na hora de treinar os modelos.

Observa-se que treinar padrões de imagens com uma qualidade muito superior pode ser custoso e treinar com qualidade muito inferior pode ocasionar perda de características importantes. Decidiu-se então, neste trabalho, por uma dimensão de camada de entrada da rede neural de 400 x 400 x 3 resultando em 480 mil *pixels*.

4.2 Treinamento e validação

Optou-se, no processo de treinamento e validação, por iniciar os parâmetros das redes com os parâmetros previamente treinados do banco de imagens ImageNet, disponibilizado pelo próprio *framework* de *machine learning* Keras. Essa

técnica, denominada de *transfer learning*, utiliza modelos pré-treinados como ponto de partida no processo de classificação de imagens (TENÓRIO, 2019).

Inicialmente, para gerar resultados prévios mais rapidamente, treinou-se os modelos com uma quantidade de épocas reduzida (20 épocas), por conta do tempo que os modelos levaram para treinar uma época, que foi entre 13 e 60 minutos (do menor ao maior modelo). Os resultados obtidos na etapa inicial de treinamento e validação foram avaliados pelo erro (*loss*) e pela acurácia (*accuracy*).

As Tabelas 9 e 10 apresentam as médias, dos resultados das métricas de treinamento e validação inicial, para cada um dos modelos apresentados na Tabela 7. As últimas 10 épocas foram escolhidas para se obter a média, pois é quando os resultados das métricas de treinamento e validação estão em fase de relativa estabilidade, sem muito efeito da variação. Os valores em negrito, nas Tabelas 9 e 10 indicam o melhor desempenho das métricas entre os arranjos da validação cruzada.

Tabela 9 – Média dos resultados de treino

Método	Modelo	Batches	Epochs	Média de cada fold				
				1	2	3	4	5
Erro (<i>loss</i>)	MobileNet	8	20	0,18	0,25	0,28	0,27	0,26
	MobileNetV2	8	20	0,22	0,46	0,45	0,41	0,46
	Xception	8	20	0,22	0,44	0,38	0,49	0,39
	ResNet50V2	8	20	0,50	1,46	1,50	1,07	1,34
	EfficientNetV2B0	8	20	0,29	0,47	0,45	0,55	0,56
	EfficientNetV2B1	8	20	0,34	0,75	0,68	0,54	0,60
	EfficientNetV2B2	8	20	0,32	0,70	0,71	0,73	0,68
	EfficientNetV2B3	8	20	0,37	0,74	0,72	0,67	0,72
Acurácia (<i>accuracy</i>)	MobileNet	8	20	0,94	0,91	0,90	0,90	0,90
	MobileNetV2	8	20	0,92	0,83	0,83	0,85	0,83
	Xception	8	20	0,92	0,84	0,86	0,81	0,86
	ResNet50V2	8	20	0,81	0,34	0,35	0,52	0,35
	EfficientNetV2B0	8	20	0,89	0,81	0,83	0,78	0,76
	EfficientNetV2B1	8	20	0,87	0,67	0,70	0,78	0,76
	EfficientNetV2B2	8	20	0,88	0,71	0,69	0,68	0,72
	EfficientNetV2B3	8	20	0,87	0,69	0,71	0,73	0,68

Fonte: Autoria própria (2023)

Tabela 10 – Média dos resultados de validação

Método	Modelo	Batches	Epochs	Média de cada fold				
				1	2	3	4	5
	MobileNet	8	20	0,13	0,18	0,17	0,14	0,12
	MobileNetV2	8	20	13,82	14,95	31,92	8,52	30,54
	Xception	8	20	2,01	0,48	0,78	0,53	0,43

Método	Modelo	Batches	Epochs	Média de cada <i>fold</i>				
				1	2	3	4	5
Erro (<i>loss</i>)	ResNet50V2	8	20	106,50	605,56	158,33	2,01	285,12
	EfficientNetV2B0	8	20	2,21	6,49	5,32	8,61	6,99
	EfficientNetV2B1	8	20	3,73	3,74	8,22	5,31	4,61
	EfficientNetV2B2	8	20	4,25	3,43	5,55	8,56	5,28
	EfficientNetV2B3	8	20	2,05	4,11	4,70	2,59	4,37
Acurácia (<i>accuracy</i>)	MobileNet	8	20	0,95	0,93	0,93	0,95	0,95
	MobileNetV2	8	20	0,35	0,26	0,26	0,36	0,26
	Xception	8	20	0,83	0,84	0,82	0,85	0,88
	ResNet50V2	8	20	0,60	0,35	0,33	0,53	0,34
	EfficientNetV2B0	8	20	0,39	0,29	0,27	0,38	0,22
	EfficientNetV2B1	8	20	0,22	0,27	0,15	0,26	0,26
	EfficientNetV2B2	8	20	0,19	0,21	0,23	0,22	0,26
	EfficientNetV2B3	8	20	0,42	0,26	0,22	0,38	0,30

Fonte: Autoria própria (2023)

Observou-se, dos resultados apresentados nas Tabelas 9 e 10, que houve, tanto para o treinamento quanto para a validação, um melhor desempenho da primeira etapa da validação cruzada (*fold* 1).

Na sequência, com o objetivo de encontrar os modelos com melhor desempenho (Acurácia maior que 80%), realizou-se a etapa de treinamento e validação agora para 100 épocas, utilizando somente a primeira etapa da validação cruzada (*fold* 1). Na Tabela 11 apresentam-se os resultados das médias das últimas 10 épocas para o conjunto de imagens da validação.

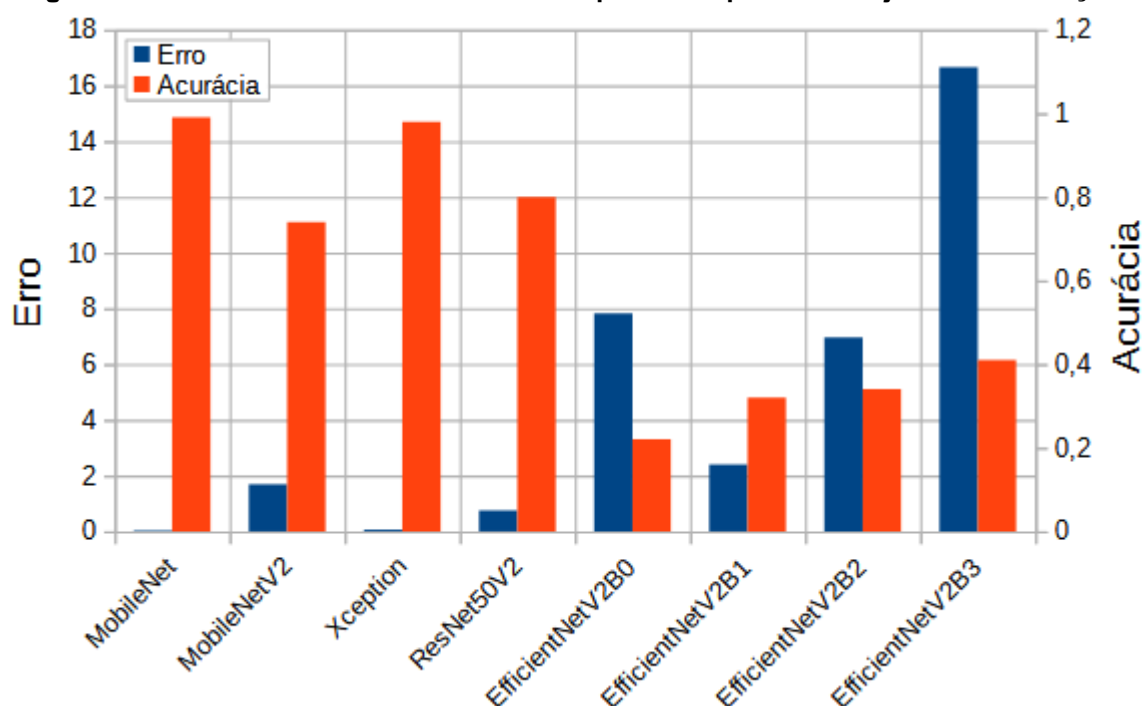
Tabela 11 – Média dos resultados de validação para um treinamento com 100 épocas

Modelo	Erro	Acurácia
MobileNet	0,03	0,99
MobileNetV2	1,69	0,74
Xception	0,06	0,98
ResNet50V2	0,75	0,80
EfficientNetV2B0	7,82	0,22
EfficientNetV2B1	2,40	0,32
EfficientNetV2B2	6,96	0,34
EfficientNetV2B3	16,66	0,41

Fonte: Autoria própria (2023)

Na Figura 19 apresentam-se, para o conjunto de validação, os resultados, na forma gráfica, dos erros e das acurácias dos oito modelos.

Figura 19 – Resultados do erro e da acurácia para 100 épocas – Conjunto de Validação



Fonte: Autoria própria (2023)

Observa-se, dos resultados apresentados na Tabela 11 e Figura 19, que os modelos MobileNet e Xception apresentaram os melhores desempenhos, acurácias de 98,7% e 97,63%, respectivamente.

4.3 Teste

Para avaliar a generalização, dos modelo MobileNet e Xception, foi construído uma base de dados de imagens (conjunto de teste – 12 imagens) que não participaram da etapa de treinamento e validação. A etapa de teste é importante porque em situações reais os cenários e condições das amostras podem sofrer mudanças significativas, em muitos casos, imprevisíveis.

Para o conjunto de teste foram ajustados alguns métodos de geração de dados (*data augmentation*) como a retirada do *zoom range* e *shear range*. Preferiu-se não adicionar distorções às imagens, uma vez que as mesmas podem confundir o modelo, gerando padrões semelhantes aos padrões de outra classe.

Foram também adicionados, para gerar diversidade nas amostras, fundos as imagens (além do fundo branco original), que se assemelham a folhas de um tomateiro. Na Tabela 12 apresentam-se os parâmetros utilizados pelos modelos e os resultados das métricas (*Loss* e *Accuracy*) para o conjunto de teste.

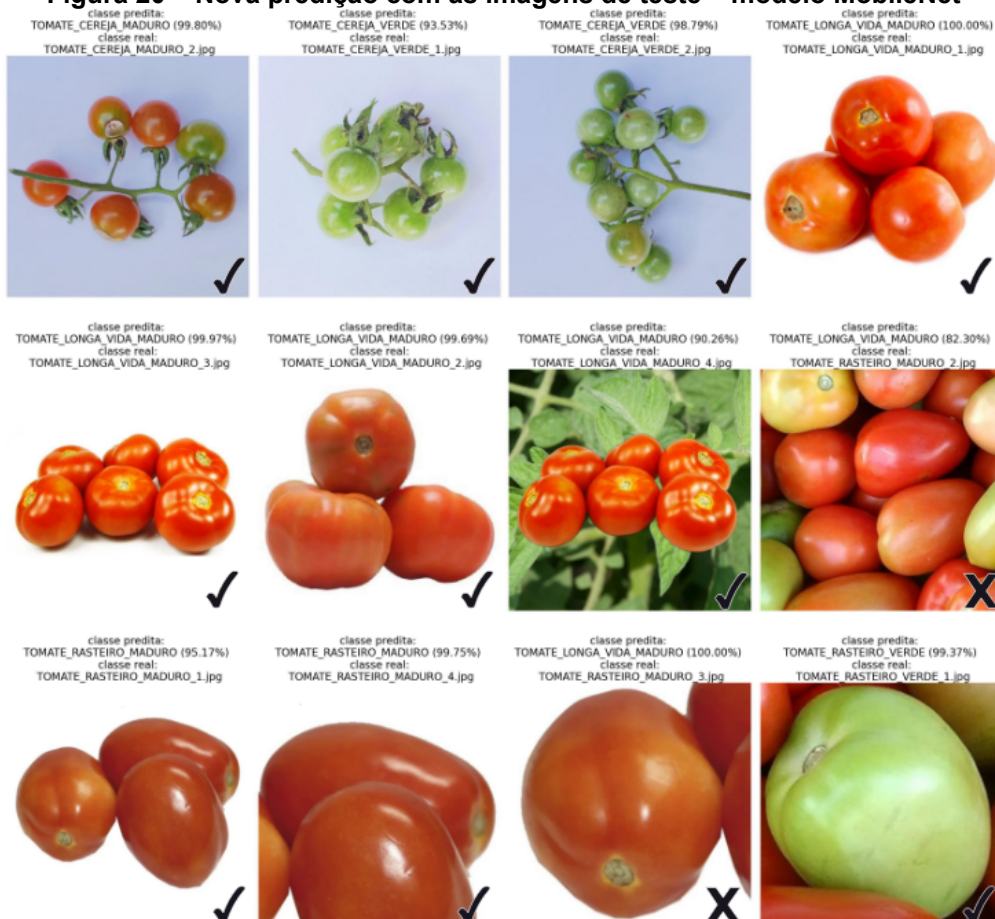
Tabela 12 – Parâmetros e Métricas – Conjunto de teste

Modelo	Epochs	Batch	Learning Rate	Loss	Accuracy	Trainable Parameters
MobileNet	29	16	0,004	0,622	83,33%	3.212.101
Xception	7	16	0,004	5,388	75,00%	20.817.197

Fonte: Autoria própria (2023)

Observa-se, dos resultados dos erros e acurácias, apresentados na tabela anterior, que o modelo MobileNet, em relação ao modelo Xception, obteve o melhor desempenho, com um erro de 0,622 e uma acurácia de 83,33%, ou seja, 10 das 12 imagens foram preditas corretamente. Para uma melhor compreensão da performance do modelo MobileNet, fez-se uma análise visual (Figura 20). O símbolo “✓” significa que o modelo realizou a predição corretamente e o “X” significa que o modelo não obteve êxito.

Figura 20 – Nova predição com as imagens de teste – modelo MobileNet



Fonte: Autoria própria; Horta do horto; Grupo sinos; Valdir; Cenourão (2023)

Vale destacar que nos conjuntos de treino e validação originais não foram adicionadas imagens ao qual continha mais de uma amostra por imagem. Isso

significa que o banco de teste foi utilizado, neste trabalho, com o papel de testar realmente o desempenho dos modelos.

5 CONCLUSÃO

Neste trabalho apresentou-se uma aplicação, de modelos de redes neurais convolucionais, para classificação do tipo de tomate (variedade ou cultivar). Para tal proposta foram utilizados os modelos ResNet50V2, MobileNet, MobileNetV2, Xception, EfficientNetV2B0, EfficientNetV2B1, EfficientNetV2B2 e EfficientNetV2B3.

Inicialmente, por meio da acurácia (*accuracy*) e do erro (*loss*), selecionaram-se, na etapa de treinamento e validação, os melhores modelos, ResNet50V2, MobileNet, MobileNetV2, Xception, EfficientNetV2B0, EfficientNetV2B1, EfficientNetV2B2 e EfficientNetV2B3. Os modelos MobileNet e Xception obtiveram, no conjunto de validação, os melhores desempenhos, acurácias de 98,7% e 97,63%, respectivamente.

Na sequência, classificaram-se, por meio dos modelos MobileNet e Xception as imagens do conjunto de teste, imagens que não participaram dos processos de treinamento e validação. Observou-se, considerando o erro (*loss*) e a acurácia (*accuracy*), que o modelo MobileNet obteve, em relação ao modelo Xception, um melhor desempenho no conjunto de teste (0,622 e 83,33%), demonstrando uma boa capacidade de generalização do modelo MobileNet.

Embora o modelo MobileNet tenha apresentado bons resultados, quando comparado com o modelo Xception, sugere-se então, para trabalhos futuros, a aplicação de outras técnicas de *machine learning* além das Redes Neurais Convolucionais utilizadas neste trabalho. Espera-se, com a evolução da Inteligência Artificial, que outros modelos mais precisos e/ou mais eficientes sejam implementados pela comunidade científica.

REFERÊNCIAS

AGRISHOW DIGITAL. **Cenário do tomate no Brasil: tendências e dificuldades de cultivo**. Disponível em: <https://digital.agrishow.com.br/hortifruti/cenario-do-tomate-no-brasil-tendencias-e-dificuldades-de-cultivo>. Acesso em: 5 fev. 2023.

AI WIKI. **Accuracy and loss**. Disponível em: <https://machine-learning.paperspace.com/wiki/accuracy-and-loss>. Acesso em: 5 fev. 2023.

BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. [s.l.] Springer, 2006. v. 4

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 1. ed. Rio de Janeiro: Editora LTC, 2000

CAMACHO OLMEDO, M. T. *et al.* Geomatic approaches for modeling land change scenarios. An introduction. Em: **Geomatic Approaches for Modeling Land Change Scenarios**. [s.l.] Springer, 2018. p. 1–8.

CENOURÃO. **Tomate rasteiro**. Disponível em: <http://www.cenourao.com.br/produto/82471-tomate-rasteiro-700g>. Acesso em: 24 nov. 2023.

DUSI, A. N. *et al.* **A cultura do tomateiro (para mesa)**. EMBRAPA, 1993.

ELHASSOUNY, A.; SMARANDACHE, F. **Trends in deep convolutional neural networks architectures: a review**. 2019 International Conference of Computer Science and Renewable Energies (ICCSRE). Morocco, 2019.

EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária). **Cultivares**. EMBRAPA Hortaliças. Disponível em: <https://www.embrapa.br/hortalicas/tomate-de-mesa/cultivares2>. Acesso em: 30 jan. 2023.

FAOSTAT. **Production quantities of tomatoes by country**. Disponível em: <https://www.fao.org/faostat/en/#data/QCL/visualize>. Acesso em: 26 jan. 2023.

FERNANDES, I. **Pensador**. Disponível em: https://www.pensador.com/autor/iago_fernandes/. Acesso em: 27 mar. 2024.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2010.

GRAVES, A. **Towards end-to-end speech recognition with recurrent neural networks**. Proceedings [...]. 31st International Conference on Machine Learning (ICML-14), Beijing, China, 2014.

GRUPO SINOS. **Os diferentes tipos de tomate - Gourmet**. Disponível em: http://gourmet.gruposinos.com.br/_conteudo/2015/07/home/194926-os-diferentes-tipos-de-tomates.html. Acesso em: 24 nov. 2023.

HAYKIN, S. **Neural networks: a comprehensive foundation**. New Delhi: Pearson Prentice Hall, 2001.

HORTA DO HORTO. **Tomate longa vida**. Disponível em: <https://www.hortadohorto.com.br/legumes/Tomate-Longa-Vida>. Acesso em: 24 nov. 2023.

JORGE. **What is the difference between genus, species, variety and cultivar?** Disponível em: <https://www.primrose.co.uk/blog/guest-posts/what-is-the-difference-between-genus-species-variety-and-cultivar/>. Acesso em: 9 fev. 2023.

MARIANO, D. **Métricas de avaliação em machine learning**. Disponível em: <https://diegomariano.com/metricas-de-avaliacao-em-machine-learning/>. 2021. Acesso em: 27 nov. 2023.

MUREȘAN, H.; OLTEAN, M. Fruit recognition from images using deep learning. **Acta Universitatis Sapientiae, Informatica**, v. 10, p. 26–42, 1 jun. 2018.

MUREȘAN, H.; OLTEAN, M. **Fruits-360**: A dataset of images containing fruits and vegetables. 7 fev. 2023. Disponível em: <https://github.com/Horea94/Fruit-Images-Dataset>. Acesso em: 7 fev. 2023

O'SHEA, K.; NASH, R. **An introduction to convolutional neural networks**. 2015. Disponível em: <https://arxiv.org/abs/1511.08458>. Acesso em: 7 fev. 2023.

OLIVEIRA, C. M. de. **Efeito do gene Mi e dos altos teores foliares de acilacúcares e de zingibereno na resistência do tomateiro a artrópodes-praga**. Tese, Universidade Federal de Lavras. Lavras, 2015.

ONU. **Mundo com 8 bilhões oferece infinitas possibilidades**. Disponível em: <https://news.un.org/pt/story/2023/04/1813112>. Acesso em: 27 nov. 2023.

OSÓRIO, F. S.; BITTENCOURT, J. R. **Sistemas inteligentes baseados em redes neurais artificiais aplicados ao processamento de imagens**. I Workshop de inteligência artificial. 2000.

ROSA, L. A. **Classificação de imagens de frutas utilizando aprendizado de máquina**. [Trabalho de Conclusão de Curso]. Universidade Federal de Santa Catarina, 2019.

SÁNCHEZ, D. A. R. S. **Comparación de diferentes métodos de visión por computadora para la clasificación del tomate chonto por tamaño y estado de maduración**. Proyecto fin de Carrera. Universidade de los Andes, 2023.

SANTOS, J. A. A. Aplicação de redes neurais artificiais na previsão do preço do milho no estado do Paraná. **Revista de Engenharia e Tecnologia**, v. 13, p. 124-134, 2021.

SHANKARA, N. J. Lidt de J, Marja de G, Martin H, Barbara VD. **Cultivation of**

tomato production, processing and marketing, Wageningen: Agromisa Foundation and CTA (Agrodok 17), 2005.

SIDRA-IBGE. **Série histórica da estimativa anual da área plantada, área colhida, produção e rendimento médio dos produtos das lavouras**. Disponível em: <https://sidra.ibge.gov.br/tabela/6588>. Acesso em: 5 fev. 2023.

SZE, V. *et al.* **Hardware for machine learning: challenges and opportunities**. 2017. Disponível em: <https://arxiv.org/abs/1612.07625>. Acesso em: 5 fev. 2023.

TEAM, K. **Keras documentation: Keras Applications**. Disponível em: <https://keras.io/api/applications/>. Acesso em: 10 out. 2023.

TENÓRIO, G. L. **Applications of deep learning for crop monitoring: classification of crop type, health and maturity**. [Tese]. PUC-Rio, 2019.

VALDIR. **Tomate rasteiro**. Disponível em: <https://www.facebook.com/sacolaodovaldir/photos/promo-%C3%A7%C3%A3otomate-rasteiro-r-299-kgaproveite-que-%C3%A9-s%C3%B3-nessa-final-de-semana/1864951363750583/>. Acesso em: 24 nov. 2023.

WU, G. *et al.* A boundary regulated network for accurate roof segmentation and outline extraction. **Remote Sensing**, v. 10, n. 8, p. 1195, 2018.