

**UNIVERSIDADE DE ITAÚNA**  
**FACULDADE DE ENGENHARIA**  
**CIÊNCIA DA COMPUTAÇÃO**

**Gustavo Soares Mendonça**

**Luiz Paulo Bittencourt Guimarães**

**Análise de comparação de algoritmos de redes neurais  
convulsionais para classificação de imagens de plantio  
agrícolas**

**Itaúna - MG**

**2023**

**UNIVERSIDADE DE ITAÚNA**  
**FACULDADE DE ENGENHARIA**  
**CIÊNCIA DA COMPUTAÇÃO**

**Gustavo Soares Mendonça**

**Luiz Paulo Bittencourt Guimarães**

**Análise de comparação de algoritmos de redes neurais  
convulsionais para classificação de imagens de plantio  
agrícolas**

Trabalho de Curso apresentado como requisito parcial para obtenção do título de Ciência da Computação no curso de Ciência da Computação da Faculdade de Engenharia da Universidade de Itaúna.

Orientador: Adriano Benigno Moreira

**Itaúna - MG**

2023

**UNIVERSIDADE DE ITAÚNA**  
**FACULDADE DE ENGENHARIA**  
**CIÊNCIA DA COMPUTAÇÃO**

**Gustavo Soares Mendonça**  
**Luiz Paulo Bittencourt Guimarães**

**Análise de comparação de algoritmos de redes neurais  
convulsionais para classificação de imagens de plantio  
agrícolas**

Trabalho de Curso apresentado ao Curso de Ciência da Computação como requisito parcial para obtenção do grau de Ciência da Computação.

Este trabalho foi julgado adequado para obtenção da aprovação na disciplina Trabalho de Curso do Curso de Ciência da Computação da Universidade de Itaúna.

**Banca**

---

**Professor orientador e membro da banca: Adriano Benigno Moreira**

---

**Professor examinador: Wallison Willian Guimarães**

**Aluno(s)**

---

**Gustavo Soares Mendonça**

---

**Luiz Paulo Bittencourt Guimarães**

Data\_\_\_\_/\_\_\_\_/\_\_\_\_

## TERMO DE RESPONSABILIDADE DE AUTORIA

Curso: Ciência da Computação Turno: Noturno Período: 8º

Declaramos que estamos cientes de que, nos termos da Lei de Direitos Autorais 9.610/98, reproduzir integralmente um texto, mesmo indicando a fonte, mas sem a autorização do autor, pode constituir crime de violação de direitos autorais. Da mesma forma é considerado “utilização indevida” e/ou “plágio”, os seguintes casos:

- a) inclusão de texto cuja autoria de terceiros não esteja claramente identificada;
- b) texto supostamente produzido pelo aluno, mas que se trata de texto adaptado em parte ou totalmente;
- c) texto produzido por terceiros sob encomenda do aluno mediante pagamento (ou não) de honorários profissionais, que não cite a autoria e não tenham sido autorizados em casos especiais pelo orientador;
- d) texto já previamente preparado sem que tenha havido participação do professor orientador na sua produção durante o processo ou que não tenha sido levado ao conhecimento do mesmo;
- e) texto supostamente produzido pelo aluno sem que ele consiga responder perguntas acerca do tema, ou sem que ele consiga elucidar seu conteúdo de forma sistemática, seja em parte ou na sua totalidade.

Dessa forma, declaramos ser de nossa inteira responsabilidade a autoria do texto referente ao Trabalho de Curso e trabalho acadêmico realizado na disciplina Trabalho de Curso de Engenharia de Produção da Universidade de Itaúna.

Itaúna,                /                /

Ass.

Nome do aluno: Gustavo Soares Mendonça

CIU 80709

Ass.

Nome do aluno: Luiz Paulo Bittencourt Guimarães

CIU 77865

Data\_\_\_\_/\_\_\_\_/\_\_\_\_

## Resumo

Neste trabalho, propõe-se uma aplicação de tecnologias de visão computacional e aprendizado de máquina na agricultura, visando obter informações detalhadas sobre os cultivos agrícolas. Para isso, utilizou-se uma base de imagens abrangente de diferentes tipos de plantações e comparou-se o desempenho de diversas arquiteturas de redes neurais convolucionais na tarefa de análise e classificação dessas imagens. Realizou-se uma análise comparativa de diversos algoritmos de redes neurais convolucionais para a classificação de imagens de cultivos agrícolas. Utilizou-se um conjunto de dados, que continham imagens de 30 variedades de culturas agrícolas. Testaram-se várias arquiteturas de redes neurais, incluindo uma sem transferência de aprendizado e cinco com transferência de aprendizado, que incluem MobileNet, MobileNetV2, MobileNetV3Small, VGG16 e VGG19, para avaliar o desempenho na tarefa de classificação. Os resultados revelaram que as redes neurais com transferência de aprendizado superaram significativamente aquelas sem transferência de aprendizado em termos de acurácia na classificação. O objetivo principal deste estudo foi propor um modelo de rede neural convolucional capaz de detectar e classificar diferentes tipos de plantações com base em imagens de um amplo banco de dados agrícolas. Esse trabalho desempenha um papel no avanço da visão computacional aplicada à agricultura, fornecendo tanto uma metodologia sólida quanto uma avaliação prática das capacidades dos algoritmos de aprendizado profundo nesse contexto.

**Palavras-chave:** Visão computacional. Aprendizado profundo, Plantios Agrícolas, Redes neurais convolucionais. Transferência de Aprendizado.

## **Abstract**

In this article, the use of technology in agriculture, particularly through computer vision systems and machine learning, can be a valuable tool for obtaining detailed information about agricultural crops. This study specifically focuses on the application of these technologies in analyzing images of various types of crops. In this work, the authors conducted a comparative analysis of various convolutional neural network algorithms for the classification of agricultural crop images. They used a dataset containing images of 30 different crop types. Several neural network architectures were tested, including one without transfer learning and five with transfer learning, namely MobileNet, MobileNetV2, MobileNetV3Small, VGG16, and VGG19, to assess their performance in the classification task. The results revealed that neural networks with transfer learning significantly outperformed those without transfer learning in terms of classification accuracy. The primary objective of this study was to propose a convolutional neural network model capable of detecting and classifying different types of crops based on images from a large agricultural database. This work plays a significant role in advancing computer vision applied to agriculture, providing both a robust methodology and a practical evaluation of deep learning algorithms in this context.

**Key-words:** Key-words: Computer vision. Deep learning. agricultural crops. Convolutional neural networks. transfer learning.

## Lista de Ilustrações

FIGURA 1 - Intensidade de cores usando RGB	13
FIGURA 2 - Intensidade de cores usando na roda de cores HSV	13
FIGURA 3 - Artificial intelligence, machine learning, and deep learning relation	16
FIGURA 4 - Diagrama de rede neural	17
FIGURA 5: Caminho de processo das convoluções	18
FIGURA 6: Exemplo de uma convolução de filtro 3x3	20
FIGURA 7: Maxpooling com 2x2 de área	20
FIGURA 8: Matriz de Confusão	21
FIGURA 9 - Fórmula Acurácia	22
FIGURA 10 - Função de perda	23
FIGURA 11 - Ilustração da MobileNet	24
FIGURA 13 - Cpu do Colab.	34
FIGURA 14 - Diagrama de embolsando a metodologia.	35
FIGURA 15 - Rede Neural sem aprendizagem por transferência.	41
FIGURA 16 - Rede Neural Convolucional MobileNet.	42
FIGURA 17 - Rede Neural Convolucional MobileNetV2.	43
FIGURA 18 - Rede Neural Convolucional MobileNetV3.	44
FIGURA 19 - Rede Neural Convolucional VGG16.	45
FIGURA 20 - Rede Neural Convolucional VGG19.	46
FIGURA 21 - Quadro de comparação entre Rede Neurais acurácia e função de perda.	47
FIGURA 22 - Quadro de comparação entre Rede Neurais pelo tempo.	48

### **Lista de abreviaturas e siglas**

RGB - Red green blue(Vermelho verde azul)

HSV - hue saturation e value(matiz saturação valor)

ConvNet - Convolutional Neural Network(Rede Neural Convolucional)

CNN - Convolutional Neural Network(Rede Neural Convolucional)

TCC - Trabalho de Conclusão de Curso

AUTOML - Aprendizado de Máquina Automatizado

VGG - Visual Geometry Group(Grupo de Geometria Visual)

TP - True Positive(Verdadeiro Positivo)

TN - True Negative(Verdadeiro Negativo)

FP - False Positive(Falso Positivo)

FN - Falso Negative(Falso Negativo)

GPU - Graphics Processing Unit(Unidade de Processamento Gráfico)

DBN - Deep Belief Networks(Redes de Crenças Profundas)



## SUMÁRIO

1 INTRODUÇÃO	10
1.1 TRABALHOS RELACIONADOS	11
2 REVISÃO BIBLIOGRÁFICA	12
2.1 Visão Computacional	12
2.1 Inteligência Artificial	14
2.2 Aprendizagem de máquina	14
2.3 Aprendizado Supervisionado e Não Supervisionado	15
2.4 Aprendizagem profundo	15
2.5 Transferência de aprendizado	17
2.6 Rede Neural Convolucional	17
2.6.1 Principais Camadas da Rede Neural Convolucional	18
2.7 Métricas de Desempenho	21
2.7.1 Acurácia	21
2.7.2 Função de Perda	22
2.8 Arquiteturas de Redes Convolucionais	23
2.8.1 Mobilenet	23
2.8.2 MobilenetV2 e MobilenetV3	25
2.8.3 VGG16 e VGG19	26
2.9 Overfitting	27
3 PROBLEMA, PROPOSTA, OBJETIVOS E JUSTIFICATIVA	29
3.1 Problema	29
3.2 Proposta	29
3.3 Objetivos	29
3.3.1 Objetivo Principal	29
3.3.2 Objetivo Secundário	30
3.1 Justificativa	30
4 METODOLOGIA	31
4.1 Base de imagens	31
4.2 Materiais	34
4.3 Método proposto	35
4.4 Tratamento de Imagem	35
4.5 Supervisão de treinamento	36
4.6 Análise e comparação dos resultados	39
5 DESENVOLVIMENTO	40
5.1 Introdução	40
5.2 Contextualização	40
5.2 Análise de resultados	41
5.4 Comparação de resultados	47
6 CONCLUSÃO	49
Referências	51

## 1 INTRODUÇÃO

O avanço tecnológico possibilita uma revolução na agricultura, com a integração de sistemas de automação e tecnologias de monitoramento nas práticas de plantio e cultivo, com uso de máquinas agrícolas autônomas.

Segundo (INAMASU, 2021), há um grande aumento de máquinas agrícolas autônomas no setor, que deve crescer 10% ao ano em todo o mundo até 2031. Nesse contexto, a análise de imagens é fundamental para reconhecer e classificar as plantações agrícolas de forma eficiente e precisa. Este trabalho visa investigar e avaliar o desempenho de diversos algoritmos de aprendizado profundo nesse domínio específico. Um exemplo é o trabalho de (NETO et al., 2019), que faz a detecção de laranjas no campo usando redes neurais.

A classificação precisa das imagens agrícolas é essencial para a tomada de decisões efetivas pelos agricultores e profissionais do campo, impactando diretamente na produtividade, na gestão de recursos e na redução de perdas. Para alcançar seu objetivo, este estudo utiliza uma base de imagens (AZAM, 2022) abrangente de plantio agrícolas, que inclui uma variedade de culturas e cenários. É feita uma análise comparativa das arquiteturas de redes neurais convolucionais na área de visão computacional, incluindo MobileNet, MobileNetV2, MobileNetV3Small, VGG16 e VGG19. Essa análise considera não apenas o desempenho das redes quando treinadas do zero, mas também quando se emprega a técnica de transferência de aprendizado a partir de modelos pré-treinados, visando avaliar a eficácia dessa estratégia no contexto da classificação de imagens agrícolas.

Além disso, a crescente demanda cria uma limitação por métodos eficazes na agricultura moderna, e isso impulsiona o desenvolvimento e a aplicação de tecnologias de inteligência artificial, tornando este estudo ainda mais relevante. Conhecer e comparar o desempenho de diferentes algoritmos de redes neurais convolucionais em um contexto agrícola é fundamental para aprimorar a eficiência das operações agrícolas e, consequentemente, contribuir para a segurança alimentar global. A demanda crescente é um desafio e a automação das plantações com redes neurais convolucionais é um elemento fundamental para uma solução viável.

O trabalho está estruturado da seguinte maneira: após esta introdução, no próximo, a revisão bibliográfica. Nesta seção, são revisados os trabalhos e informações relevantes existentes

sobre o tema, oferecendo um panorama do estado atual do conhecimento na área, o Tema, Problema, Objetivos e Justificativa. Este tópico aborda o problema de pesquisa, a proposta de solução, os objetivos do estudo e a justificativa para a realização da pesquisa. A metodologia descreve o plano e os métodos usados para conduzir a pesquisa. Desenvolvimento, nessa seção, é apresentada a parte principal do estudo, onde os resultados, análises e discussões são abordados em detalhes e pode incluir experimentos, descobertas, dados coletados e outras informações relevantes, a conclusão resume os principais resultados e descobertas do estudo, enfatizando como eles se relacionam com os objetivos definidos anteriormente.

## 1.1 TRABALHOS RELACIONADOS

Nesta seção tem o objetivo de mostrar trabalhos que usaram técnicas de aprendizado de máquina em outro contexto, semelhantes ao que este trabalho pretende se inspirar.

Neste artigo (PATRÍCIO; RIEDER, 2018) apresenta uma revisão sistemática que visa identificar a aplicabilidade da visão computacional na agricultura de precisão para a produção dos cinco grãos mais produzidos no mundo: milho, arroz, trigo, soja e cevada. O artigo analisa 25 trabalhos selecionados nos últimos cinco anos com diferentes abordagens para tratar aspectos relacionados à detecção de doenças, qualidade dos grãos e fenotipagem. A partir dos resultados da revisão sistemática, é possível identificar grandes oportunidades, como a exploração de GPU (Graphics Processing Unit) e técnicas avançadas de inteligência artificial, como DBN (Deep Belief Networks) na construção de métodos robustos de visão computacional aplicados à agricultura de precisão.

Neste artigo (OLIVEIRA,2020) aborda a aplicação de Deep Learning para a criação de modelos que detectam frutos de acerola em imagens digitais de aceroleiras e classificam esses frutos de acordo com sua cor, foi proposta uma metodologia para classificação de frutos de acerola em imagens digitais, de acordo com suas cores, verde ou vermelha. Para tal, utilizou-se a arquitetura SDD para geração dos modelos, ela uma arquitetura de RNC que detecta objetos em imagens usando uma única rede neural profunda. Os principais sistemas de detecção de objetos por muito tempo seguiram a abordagem de criar hipóteses de caixas delimitadoras de objetos e aplicação de classificadores nos possíveis objetos detectados.

Neste artigo (DHANYA et al., 2022) faz uma revisão dos avanços da visão computacional baseada em aprendizado profundo para resolver diversos problemas agrícolas, como análise de qualidade de sementes, análise de solo, manejo de irrigação, análise de saúde de plantas, manejo de ervas daninhas, manejo de gado e estimativa de rendimento. O artigo conclui que as redes neurais convolucionais são a base das abordagens modernas de visão computacional e que seu sucesso depende da construção do modelo em um conjunto de dados de qualidade e da oferta de soluções em tempo real.

Neste artigo (SOUSA, 2018) Foi realizado o diagnóstico em doenças através de radiografia do tórax o trabalho apresentou a técnica de como acelerar este diagnóstico de pneumonia

usando técnicas aprendizam profunda Na literatura apresentando arquiteturas possíveis rede neurais convolucionais Os modelos que em contratos por meio da otimização de arquiteturas de redes neurais convolucionais através do uso da biblioteca HyperOpt.

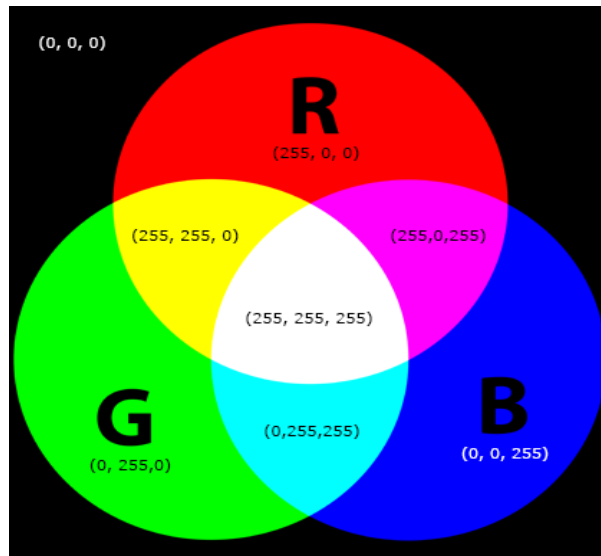
## **2 REVISÃO BIBLIOGRÁFICA**

Neste capítulo, realizaremos uma revisão bibliográfica abordando o tema proposto e sua literatura acadêmica relacionada. Essa revisão fornecerá uma base sólida para a construção do conhecimento e o aprofundamento do entendimento do objetivo específico deste estudo, ao mesmo tempo, em que oferecerá uma visão geral das principais pesquisas e teorias relacionadas ao tema proposto. Neste projeto, foi aplicado conhecimentos da área de aprendizado de máquina, com ênfase nas redes neurais convolucionais. Também foi explorado alguns conceitos de visão computacional para melhorar a compreensão deste trabalho.

### **2.1 Visão Computacional**

A visão computacional é uma área da computação que se dedica a imitar a capacidade de visão humana por meio de algoritmos e tecnologias. Seu objetivo é fazer com que os computadores possam analisar imagens e compreender o mundo ao seu redor, extraindo informações úteis de imagens e vídeos. Embora o conceito seja simples, sua execução é complexa devido à capacidade de distinguir objetos em meio a interferências, condições de luz e semelhanças entre objetos. As máquinas “veem” em pixels de acordo com (ROSA et al., 2019) os pixels, são representados por 3 cores em uma escala de 0 a 255, sendo vermelho, verde, azul que é o padrão chamado RGB(do inglês:red, green e blue) em uma escala de 0 a 255 como pode observar na Figura 1.

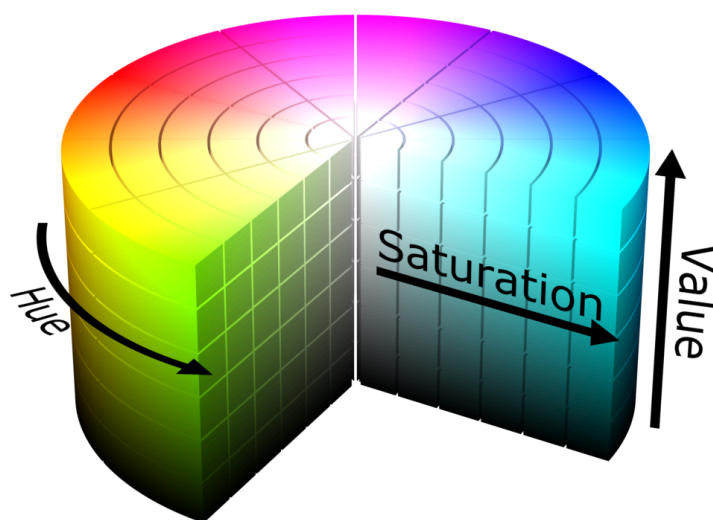
FIGURA 1 - Intensidade de cores usando RGB



Fonte: (LAUBE, 2013)

Outro sistema mencionado por (ROSA et al., 2019) para representar imagens é o HSV (do inglês: hue, saturation e value). HSV modela a forma como as tintas de diferentes cores se misturam, com a dimensão da saturação, que se assemelha a vários tons de tinta de cores vivas e a dimensão de valor que se assemelha à mistura dessas pinturas com diferentes quantidades de tinta preta ou branca, como mostra a Figura 2.

FIGURA 2 - Intensidade de cores usando na roda de cores HSV



Fonte: (FAIZUNNABI, 2020)

Após ler a imagem corretamente, podem-se aplicar várias transformações de visão computacional para obter informações das características da imagem. Esse processo de obtenção envolve reduzir os dados de entrada ao descrever a imagem usando suas características, como, por exemplo, usar métodos para identificar partes ou formas importantes em uma imagem e assim diminuir a quantidade de processamento.

## **2.1 Inteligência Artificial**

Inteligência Artificial permite imitar alguns sentidos humanos, com o propósito de desenvolver abordagens abrangentes para a simulação de atividades perceptivas, cognitivas e manipulativas por meio de sistemas computacionais. A Inteligência Artificial abrange uma dimensão de pesquisa fundamental, acompanhada por experimentação, e uma dimensão tecnológica. A união dessas duas vertentes é responsável por uma revolução da forma automatização de software, sendo essa a automatização das capacidades mentais por meio da modelização em computador como diz (PEREIRA, 1988).

## **2.2 Aprendizagem de máquina**

A aprendizagem de máquina (ou “*machine learning*” em inglês) é uma sub-área da inteligência artificial que se concentra em desenvolver algoritmos e modelos computacionais capazes de aprender a partir de dados e realizar tarefas específicas sem serem explicitamente programados para isso. De acordo com (VELLINO, 1986), uma das possibilidades da Inteligência Artificial seria que “as máquinas repliquem o pensamento humano”. Não é absurda e nem óbvia. Ao invés de seguir um conjunto de regras rígidas, os algoritmos de aprendizagem de máquina são projetados para aprender a partir de exemplos e padrões em dados, permitindo que eles se adaptem a novas situações e melhorem seu desempenho ao longo do tempo. Informalmente, o termo “inteligência artificial” é usado quando uma máquina consegue desempenhar funções que os humanos associam a mentes humanas, como “aprendizado” e “resolução de problemas”. O aprendizado é um aspecto vital das máquinas. Portanto, a aprendizagem de máquina é um subcampo da inteligência artificial. Inteligência artificial é o conceito mais amplo de permitir que uma máquina ou um sistema detecte e performe como um ser humano. Aprendizagem de máquina é uma aplicação de IA que

permite às máquinas extrair conhecimento dos dados e aprender com eles de maneira autônoma.

### 2.3 Aprendizado Supervisionado e Não Supervisionado

Segundo (PAIXÃO et al., 2022) os métodos de aprendizado de máquina podem ser classificados como supervisionado, não-supervisionado:

- No aprendizado supervisionado, o modelo possui uma referência daquilo que está certo e daquilo que está errado e seus parâmetros vão sendo ajustados iterativamente de modo a melhorar a capacidade de classificação do modelo. O treinamento é baseado na comparação entre o resultado obtido do modelo e o rótulo previamente classificados;
- Na aprendizagem não supervisionada, o modelo é treinado em um conjunto de dados, extraíndo as características dos dados e construindo uma representação sem o conhecimento prévio dos rótulos de cada dado. O objetivo é encontrar estruturas ou padrões nos dados sem orientação externa;

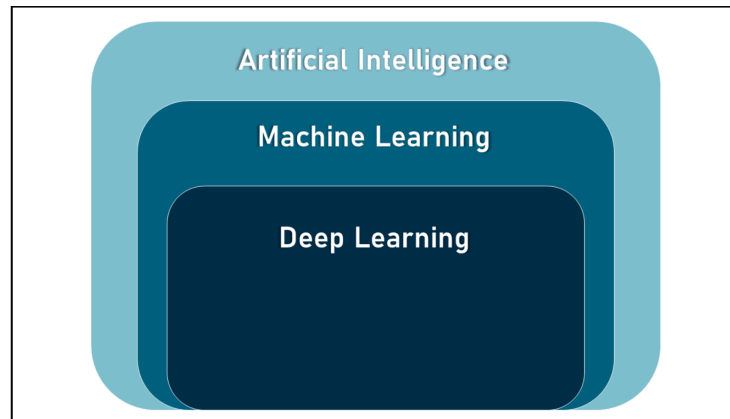
### 2.4 Aprendizagem profundo

Devido aos elevados esforços dos cientistas para evolução do aprendizado de máquina, o aprendizado profundo ou *deep learning* é uma tentativa que fizeram e tiveram ao qual foi produto dessas pesquisas. Ele é um subconjunto do aprendizado de máquina. Conforme o trabalho em aprendizado avança em diversas áreas, sempre tem alguma aplicabilidade nova em várias áreas mais recentes e é sempre uma tarefa da comunidade de pesquisa descobrir novas coisas como diz (SHINDE; SHAH, 2018).

Na Figura 3 mostra a hierarquia e subconjuntos da inteligência artificial.



FIGURA 3 - Artificial intelligence, machine learning, and deep learning relation

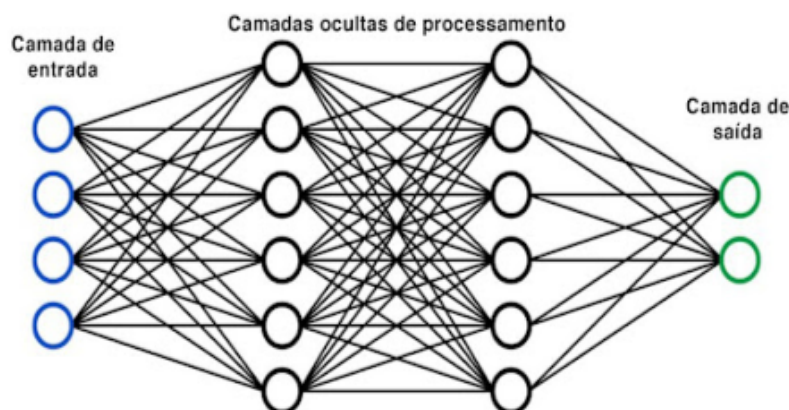


Fonte: (RAAD ALI KHALIL; BOUZOUANE, 2023)

Segundo (SHINDE; SHAH, 2018) o aprendizado profundo utiliza uma cascata de múltiplas camadas de unidades de processamento não lineares para extração e transformação de características. As camadas inferiores próximas à entrada de dados aprendem características simples, enquanto camadas superiores aprendem características mais complexas derivadas das características das camadas inferiores. A arquitetura forma uma representação hierárquica e poderosa de características. Isso significa que o aprendizado profundo é adequado para analisar e extrair conhecimento útil tanto de grandes quantidades de dados quanto de dados coletados de diferentes fontes.

Na Figura 4 é uma representação em rede neural de 2 camadas do aprendizado profundo onde o input é feito pelo camada de entrada e assim nas camadas ocultas cada colunas de nós seria uma representação de um característica específica ao qual está procurando por consequência filtrando até a camada de saída.

FIGURA 4 - Diagrama de rede neural



Fonte: (ÓTISON, 2022)

## 2.5 Transferência de aprendizado

De acordo (PAN; YANG, 2009) o aprendizado por transferência é impulsionado pelo fato de que as pessoas podem usar de maneira inteligente o conhecimento adquirido anteriormente por modelos com pesos treinados para resolver novos problemas semelhantes de forma mais rápida ou com soluções mais eficazes. Como dizia (WEISS; KHOSHGOFTAAR; WANG, 2016) o aprendizado por transferência surge quando há uma limitação de dados para treinamento. Isso pode ocorrer devido à raridade dos dados, ao alto custo de coleta e rotulação dos dados, ou à inacessibilidade dos mesmos. Com o aumento da disponibilidade de repositórios de Big data, o uso de conjuntos de dados existentes relacionados, mas nem sempre a Base de Dados tem alguma relação ao objetivo de interesse, e torna as soluções de aprendizado por transferência uma abordagem atraente.

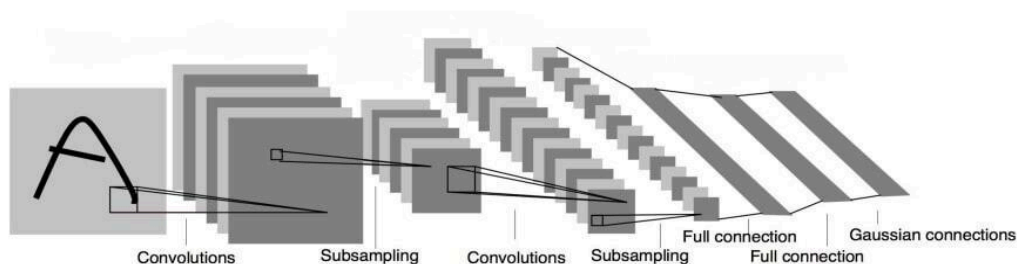
## 2.6 Rede Neural Convolucional

Uma rede neural é um sistema criado para imitar a maneira como o cérebro executa uma tarefa específica. Geralmente, essa imitação é realizada por meio de componentes eletrônicos ou através de simulação em um computador digital. Para atingir um alto nível de desempenho, as redes neurais utilizam uma extensa rede de células computacionais simples, chamadas de "neurônios" ou unidades de processamento, fortemente interconectadas (FLECK et al., 2016). Uma ConvNet (Rede Neural Convolucional) é um tipo de algoritmo

de Aprendizado Profundo que tem a capacidade de compreender uma imagem de entrada, atribuindo significado (através de pesos ajustáveis e valores de viés que são ajustados durante o treinamento) a diferentes características ou objetos presentes na imagem. Isso permite que a rede possa discernir e distinguir entre esses elementos. Segundo (RASCHKA; MIRJALILI, 2017) na camada convolucional, os neurônios atuam como filtros aplicados aos dados de entrada, sendo que, neste contexto, os dados de entrada são imagens. Cada filtro é composto por uma matriz de pesos.

A principal finalidade da convolução em uma rede neural convolucional (CNN) é capturar características distintas da imagem de entrada. Como dizia (CUNHA, 2020) comumente, as redes convolucionais são estruturadas em camadas, e entre essas camadas, encontram-se as camadas de convolução. Sua função é a detecção de características na imagem, e podem ser visualizadas como filtros que exploram a imagem, cada um procurando por características específicas representando na Figura 5. Tais características como bordas nítidas, curvaturas particulares e cores básicas.

FIGURA 5: Caminho de processo das convoluções



Fonte: (ALVES, 2023)

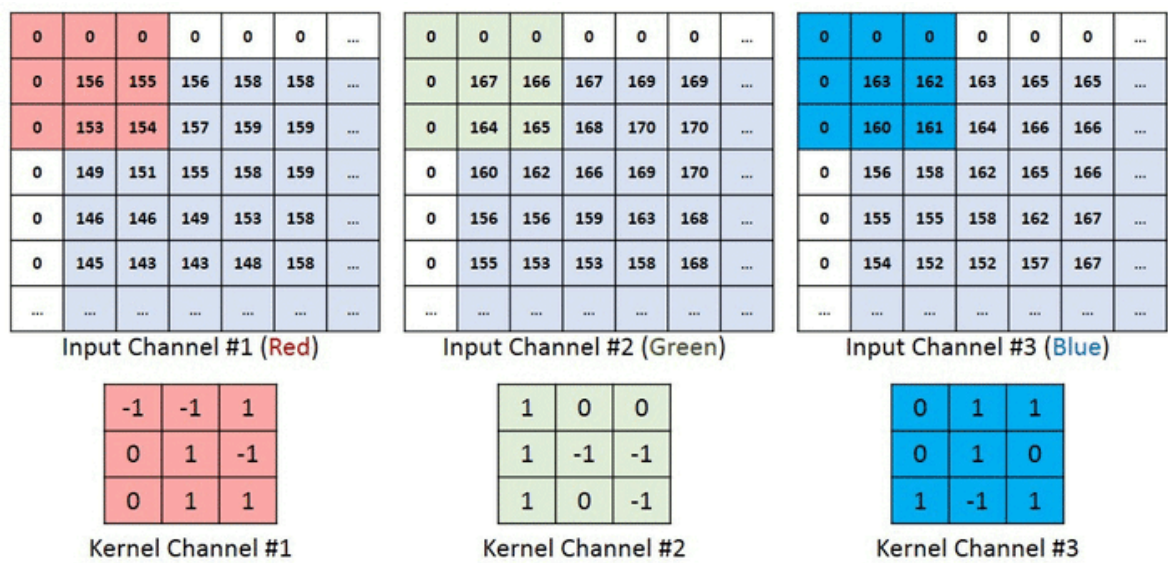
### 2.6.1 Principais Camadas da Rede Neural Convolucional

- Segundo (RODRIGUES, 2018) Camada de Convolução, ocorre a extração de características da entrada por meio do uso de filtros convolucionais de pequenas dimensões. Durante o treinamento da rede, os filtros são ajustados para identificar características comuns, tais como bordas e cores representada na figura 6. À medida que o treinamento avança, esses filtros se aprimoram na capacidade de reconhecer estruturas cada vez mais complexas. Basicamente, a camada convolução consiste em operações entre matrizes para extrair recursos de imagem e a partir disso, gerar uma

nova imagem. "Pooling" se refere a uma etapa comum em redes neurais convolucionais (CNNs) usadas em tarefas de processamento de imagens representado na figura 7. Essa camada tem uma de suas funções reduzir a dimensionalidade dos mapas de características gerados pelas camadas de convolução, mantendo as características mais importantes, as camadas de pooling ajudam a reduzir a quantidade de parâmetros na rede, tornando-a mais eficiente e evitando o *overfitting* como fala (KRIZHEVSKY; SUTSKEVER; HINTON, 2012);

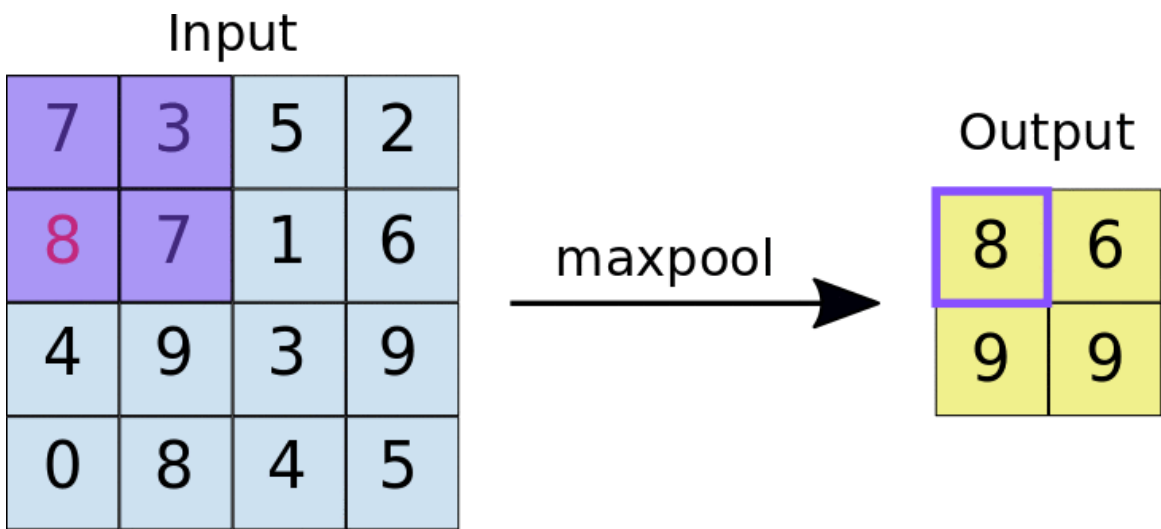
- Segundo (ALMEIDA; SILVA, 2021) a camada de ativação em uma rede neural convolucional (CNN) é aplicada após a camada de convolução e é responsável por determinar a relevância de cada neurônio na rede para a tarefa de classificação. Essas funções determinam se um neurônio deve ser ativado com base no cálculo da relevância desse neurônio para a classificação. O resultado da função de ativação é então transmitido para o próximo neurônio na rede, permitindo a aprendizagem de representações complexas dos dados;
- Segundo (YAMASHITA et al., 2018) a camada totalmente conectada em redes neurais desempenha um papel crítico e segue geralmente as camadas de convolução e pooling em arquiteturas de redes convolucionais. Sua função principal é realizar a classificação e tomar decisões finais com base nas características extraídas pelas camadas anteriores. A saída gerada na etapa de extração de características (por meio de camadas de convolução e pooling) é tipicamente transformada em um vetor unidimensional de números e conectada a uma ou mais camadas totalmente conectadas.

FIGURA 6: Exemplo de uma convolução de filtro 3x3



Fonte: (ALVES, 2023)

FIGURA 7: Maxpooling com 2x2 de área



Fonte: (ALVES, 2023)

## 2.7 Métricas de Desempenho

Métricas de desempenho são usadas para avaliar o quão bem um modelo de aprendizado de máquina ou algoritmo está funcionando em uma determinada tarefa. A escolha das métricas depende do tipo de problema que você está resolvendo.

### 2.7.1 Acurácia

Uma matriz de confusão é uma ferramenta comumente usada para avaliar o desempenho de um modelo de classificação, como um modelo de aprendizado de máquina. Ela fornece uma visão detalhada de como um modelo está se saindo na tarefa de classificação, comparando as previsões do modelo com os valores reais dos dados. Para quantificar os acertos e erros do modelo são utilizadas métricas como a acurácia como diz (MAGIRI, 2023).

A Figura 8 mostra uma Matriz de Confusão como os valores podem ser classificados como verdadeiros e falsos em relação ao contexto de cada dado.

FIGURA 8: Matriz de Confusão

		Valor Predito	
		Sim	Não
Imagem Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Fonte: (MAGIRI, 2023)

- **Verdadeiro Positivo (TP):** Representa os casos em que o modelo classificou corretamente uma instância como positiva (classe verdadeira) quando ela realmente era positiva;

- Falso Positivo (FP): Representa os casos em que o modelo classificou incorretamente uma instância como positiva quando ela era, na verdade, negativa (um erro de classificação positiva incorreta);
- Verdadeiro Negativo (TN): Representa os casos em que o modelo classificou corretamente uma instância como negativa (classe verdadeira) quando ela realmente era negativa;
- Falso Negativo (FN): Representa os casos em que o modelo classificou incorretamente uma instância como negativa quando ela era, na verdade, positiva (um erro de classificação negativa incorreta);

A acurácia seria proporção de classificações corretas em relação ao total de classificações. É uma métrica geralmente usada, mas pode ser enganosa em conjuntos de dados desequilibrados. A Figura 9 mostra a fórmula matemática da acurácia.

FIGURA 9 - Fórmula Acurácia

$$Acurácia = \frac{TP + TN}{P + N}$$

Fonte: (MAGIRI, 2023)

### 2.7.2 Função de Perda

De acordo com (HO; WOOKEY, 2019) a função de perda Cross-Entropy, também conhecida como entropia cruzada, é uma métrica amplamente usada em aprendizado de máquina e redes neurais, especialmente em tarefas de classificação. Ela é usada para quantificar a discrepância entre as previsões de um modelo e os rótulos reais dos dados de treinamento. A entropia cruzada é particularmente eficaz em problemas de classificação, onde a saída da rede é uma distribuição de probabilidade sobre as classes de interesse.

FIGURA 10 - Função de perda

$$H(p, q) = - \sum_x p(x) \log q(x).$$

Fonte: Imagem do próprio Autor

Na figura 10 mostra fórmula matemática a Função de perda, ela mede a discrepância entre as probabilidades reais  $p(x)$  que fica em uma classificação Básica que normalmente está em *hot code encoding* e as probabilidades previstas  $q(x)$  que é elevado a um logaritmo natural e multiplicado por  $p(x)$  assim filtrando pelo resultado escolhido. O objetivo durante o treinamento de uma rede neural é minimizar essa função de perda, ajustando os parâmetros da rede para que as probabilidades previstas se aproximem ao máximo das probabilidades reais, resultando em uma entropia cruzada próxima de zero. Quanto menor o valor da entropia cruzada, melhor o desempenho da rede na tarefa de classificação.

## 2.8 Arquiteturas de Redes Convolucionais

Essa subseção ira apresentar uma visão geral das diferentes arquiteturas de CNNs que serão empregadas em nosso estudo. Cada uma dessas arquiteturas possui características distintas, como profundidade, complexidade e capacidade de generalização, e são amplamente reconhecidas na comunidade de aprendizado de máquina, e visa aprofundar a compreensão dessas redes neurais e avaliar como elas se comportam na classificação.

### 2.8.1 Mobilenet

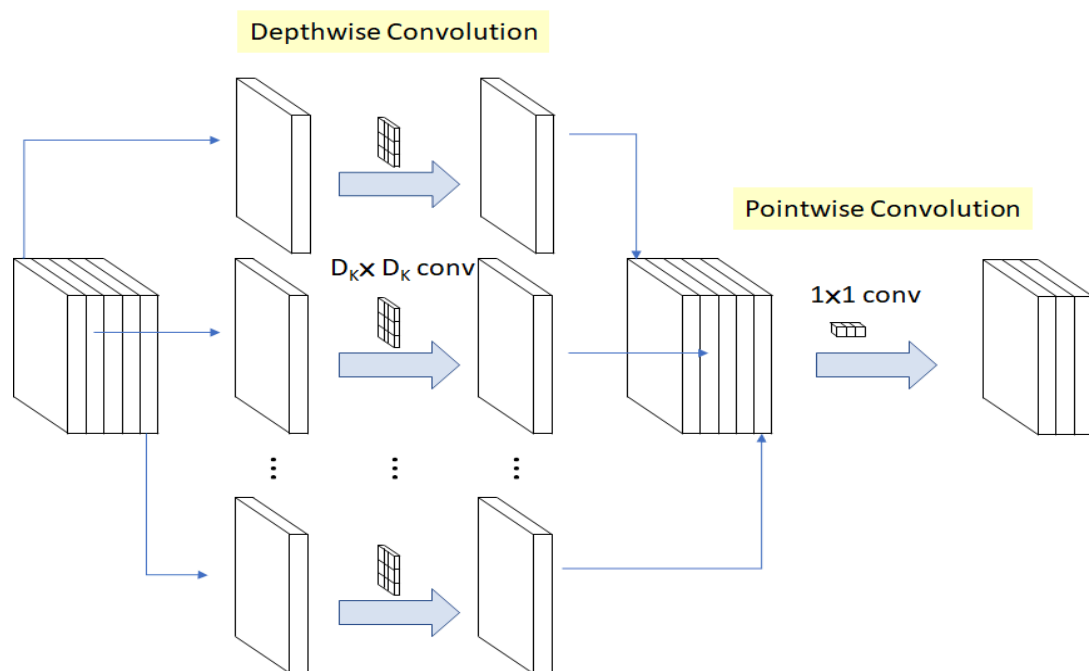
A MobileNet é uma arquitetura de rede neural convolucional (CNN) que foi projetada para ser eficiente e leve, ideal para dispositivos com recursos limitados, como smartphones e sistemas embarcados. No geral, a MobileNet é uma escolha popular para aplicações que exigem eficiência computacional e boa capacidade de classificação de imagens, especialmente em dispositivos com recursos limitados. Ela oferece um equilíbrio entre desempenho e eficiência que a torna uma arquitetura valiosa em uma variedade de cenários.



A arquitetura descrita no trabalho de (HOWARD et al., 2017) envolve uma rede que usa um tipo específico de operação chamada "convolução separável em profundidade". Essa operação é, na verdade, uma maneira de desmembrar uma convolução tradicional em duas partes distintas: uma convolução em profundidade e uma convolução pontual. O conceito é semelhante a decompor a convolução em duas etapas: primeiro, uma convolução em profundidade que examina as características em toda a imagem, seguida por uma convolução pontual que combina essas características em um resultado final. Isso é feito para otimizar o processamento e reduzir a quantidade de cálculos necessários. Em termos mais simples, imagine que é como se você estivesse quebrando uma tarefa em duas partes menores, uma que lida com as características gerais e outra que ajusta essas características para obter o resultado desejado. Isso melhora a eficiência do processo.

A Figura 11 no artigo ilustra esse processo, mostrando como a convolução separável em profundidade funciona visualmente.

FIGURA 11 - Ilustração da MobileNet



Fonte: (HOWARD et al., 2017)

Algumas características dessa rede são:

- **Arquitetura Profunda:** Apesar de ser uma rede leve, a MobileNet é profunda. Ela usa camadas de convolução profunda para extrair recursos relevantes das imagens, permitindo um bom desempenho de classificação;
- **Convoluções Separáveis:** Uma característica distintiva da MobileNet é o uso de convoluções separáveis em vez de convoluções tradicionais. Essas convoluções separáveis dividem o processo de convolução em duas etapas: uma convolução espacial seguida por uma convolução em profundidade ( $1 \times 1$ ). Isso reduz significativamente a quantidade de cálculos necessários, tornando a rede mais eficiente;
- **Flexibilidade de Tamanho:** A MobileNet pode ser dimensionada em termos de tamanho e complexidade. Isso significa que você pode ajustar a profundidade e a largura da rede para atender às necessidades específicas do seu aplicativo, equilibrando o desempenho e os recursos computacionais disponíveis;
- **Aprendizado por transferência:** Devido à sua eficiência e desempenho razoável, a MobileNet é frequentemente usada como uma base para aprendizado por transferência em tarefas de classificação de imagens. Você pode usar uma MobileNet pré-treinada em grandes conjuntos de dados e ajustá-la para tarefas específicas com menos dados de treinamento;

### **2.8.2 MobilenetV2 e MobilenetV3**

Como diz (SANDLER et al., 2018) A MobileNetV2 se destaca por apresentar melhor desempenho em comparação com a MobileNetV1, oferecendo maior precisão na classificação de imagens. Além disso, sua eficiência computacional é notável, tornando-a mais eficaz em termos de uso de recursos computacionais. A inclusão de camadas residuais em sua arquitetura é um avanço significativo, permitindo o treinamento de redes mais profundas sem comprometer o desempenho. Isso simplifica o processo de treinamento, mesmo quando visa reduzir ainda mais o número de parâmetros e a latência da rede.

A MobilenetV3, conforme diz (HOWARD et al., 2019) que em contraste com as estruturas anteriores, que foram desenvolvidas de acordo com a visão específica de seus pesquisadores, a arquitetura atual foi gerada por meio de AutoML (Aprendizado de Máquina

Automatizado). Essa abordagem utiliza técnicas automatizadas para pesquisar e identificar arquiteturas de redes neurais que sejam bem adequadas para tarefas de reconhecimento de imagens em sistemas com recursos computacionais limitados. Em outras palavras, em vez de projetar manualmente a arquitetura da rede, o AutoML realiza uma busca automatizada em uma ampla gama de possíveis arquiteturas para encontrar aquelas que funcionam melhor para a tarefa em questão. Isso permite desenvolver modelos de aprendizado profundo e eficiente em termos de computação, os quais são particularmente úteis em dispositivos com recursos de processamento restritos.

E a MobilenetV3 pode ser até 27% mais rápida que a MobilenetV2 e ainda obter resultados similares em relação à precisão em tarefas de classificação, segmentação ou detecção de objetos. MobileNet tornou-se uma das redes mais populares no desenvolvimento de tarefas de visão computacional usando redes neurais em dispositivos com poder de processamento limitado.

### 2.8.3 VGG16 e VGG19

A VGG é uma arquitetura de rede neural convolucional (CNN) desenvolvida por Karen Simonyan e Andrew Zisserman em 2014, que atingiu um desempenho de última geração no conjunto de dados ImageNet (KRICHEN, 2023). A arquitetura VGG consiste em várias camadas de pequenos filtros convolucionais  $3 \times 3$ , seguidos por camadas Max Pooling. As camadas totalmente conectadas utilizam ativação ReLU e regularização de dropout para evitar o *overfitting*.

O nome VGG é uma abreviação para "*Visual Geometry Group*" VGG é um grupo de pesquisa em visão computacional e processamento de imagem da Universidade de Oxford, localizado no Departamento de Engenharia de Ciência da Computação, portanto, VGG não é uma abreviação para uma expressão específica, mas sim o nome de um grupo de pesquisa e o nome da rede neural convolucional que eles desenvolveram.

A arquitetura VGG pode ser representada da seguinte forma:

- Camada Convolucional com 64 filtros de tamanho  $3 \times 3$ , seguida por uma camada Max Pooling de tamanho  $2 \times 2$ ;

- Camada Convolucional com 128 filtros de tamanho  $3 \times 3$ , seguida por uma camada Max Pooling de tamanho  $2 \times 2$ ;
- Camada Convolucional com 256 filtros de tamanho  $3 \times 3$ , seguida por uma camada Max Pooling de tamanho  $2 \times 2$ ;
- Camada Convolucional com 512 filtros de tamanho  $3 \times 3$ , seguida por uma camada Max Pooling de tamanho  $2 \times 2$ ;
- Camada Convolucional com 512 filtros de tamanho  $3 \times 3$ , seguida por uma camada Max Pooling de tamanho  $2 \times 2$ ;
- Camada totalmente conectada com 4096 unidades e função de ativação ReLU, seguida por regularização de dropout;
- Camada de Saída com 1000 unidades e função de ativação softmax;

Além disso, a VGG se destaca em termos de desempenho. Ela obteve resultados de alto nível em competições de visão computacional, incluindo o prestigioso Desafio de Classificação de Imagens do ImageNet. A profundidade da arquitetura permite que ela capture com precisão características complexas em imagens, sendo fundamental para a classificação precisa.

No entanto, vale ressaltar que, embora a VGG seja uma arquitetura sólida, ela não é a mais eficiente em termos de recursos computacionais e não é tão adequada para aplicações em dispositivos com recursos limitados, como smartphones e sistemas embarcados. Nesses casos, arquiteturas mais recentes, como MobileNets, podem ser mais apropriadas.

Por último, mas não menos importante, a VGG demonstra uma notável capacidade de generalização. Isso significa que ela pode ser aplicada com êxito em uma ampla gama de tarefas de visão computacional sem a necessidade de ajustes significativos. Essa versatilidade a torna uma escolha sólida em diversas aplicações.

## 2.9 Overfitting

O *overfitting*, conforme (YING, 2019), é um problema comum no aprendizado de máquina supervisionado. Ele ocorre quando um modelo se ajusta muito bem aos dados de treinamento, a ponto de memorizá-los, mas tem dificuldade em generalizar esse

conhecimento para dados novos. Isso resulta em um desempenho excelente no conjunto de treinamento, mas um desempenho ruim para novos dados, porque o modelo se tornou excessivamente especializado nos dados de treinamento, incluindo ruído e detalhes insignificantes que não são relevantes para a tarefa em questão.

O *overfitting* é causado por várias razões, incluindo:

- Ruído nos dados de treinamento: Quando os dados de treinamento contêm erros ou informações irrelevantes, o modelo pode aprender esses detalhes irrelevantes e super ajustar-se a eles;
- Complexidade do modelo: Modelos excessivamente complexos, com muitos parâmetros ou hipóteses, têm maior probabilidade de super ajustar-se por terem mais flexibilidade para se adaptar aos dados de treinamento;
- Procedimentos de comparação múltipla: Algoritmos que envolvem comparação de múltiplas hipóteses ou modelos, como algoritmos de aprendizado de máquina, podem ser propensos ao *overfitting*, especialmente se houver muitos parâmetros ou recursos envolvidos;

A revisão bibliográfica apresentada neste capítulo fornece uma base sólida para a compreensão do tema proposto e sua literatura acadêmica relacionada. Exploramos conceitos fundamentais de visão computacional, inteligência artificial e aprendizado de máquina, destacando a importância das redes neurais convolucionais (CNNs) como uma abordagem poderosa para a análise de imagens. Além disso, discutimos métricas de desempenho comuns, como acurácia e funções de perda, cruciais para avaliar modelos de aprendizado de máquina. No próximo capítulo problema, proposta, objetivo e justificativa, terão uma compreensão consistente dos princípios essenciais necessários para o desenvolvimento e avaliação de modelos de propostos anteriormente, com foco em aprimorar a classificação de imagens e a utilização eficaz de transferência de aprendizado.

### 3 PROBLEMA, PROPOSTA, OBJETIVOS E JUSTIFICATIVA

Neste capítulo, se estabelece o alicerce fundamental do trabalho, abordando o problema central que motiva essa pesquisa, delineando a proposta de solução. Além disso, definem-se os objetivos específicos que orientam nosso estudo e fornecem uma visão clara do que pretende alcançar. Finalmente, discute a justificativa por trás da escolha desse problema, demonstrando sua relevância no contexto atual.

#### 3.1 Problema

A tarefa de classificação de imagens de plantio agrícola é de grande importância no campo da agricultura de precisão. A identificação precisa das plantas é muito importante para auxiliar os agricultores a otimizar o uso de recursos, como água, fertilizantes e pesticidas, aumentando a eficiência da produção e reduzindo os impactos ambientais. O problema que este trabalho se propõe a abordar é a necessidade de desenvolver e identificar algoritmos de aprendizado de máquina, em particular, redes neurais convolucionais, que sejam altamente eficazes na tarefa de classificação de imagens de plantio agrícola.

#### 3.2 Proposta

Para abordar o desafio de identificação de cultivos agrícolas, propõe-se a aplicação de *deep learning* no âmbito das redes neurais convolucionais (CNN) com o propósito de classificar tipos de plantações agrícolas. Essa abordagem envolve a comparação e avaliação de diversos modelos de CNNs para determinar qual deles apresenta o desempenho mais eficaz na tarefa de classificação de cultivos agrícolas. Esse estudo tem o intuito de contribuir para o avanço das técnicas de visão computacional aplicadas à agricultura, permitindo a identificação precisa e eficiente de diferentes tipos de plantações, o que pode ser benéfico para o monitoramento e o gerenciamento de áreas agrícolas.

### 3.3 Objetivos

#### 3.3.1 Objetivo Principal

O Objetivo Principal visa a categorização de imagens de plantios agrícolas por meio da aplicação e comparação de diversos algoritmos de redes neurais convolucionais e comparar diferentes algoritmos, com o intuito de identificar quais modelos de redes neurais são mais eficazes para essa finalidade.

#### 3.3.2 Objetivo Secundário

Para atingir o objetivo primário deste trabalho é necessário atingir os seguintes objetivos específicos:

- Definir uma base de dados de imagem de plantio agrícolas que seja diversa;
- Recorrer à literatura para encontrar métodos computacionais a serem aplicados;
- Determinar arquitetura de detecção e classificação;
- Treinar a arquitetura construída;
- Implementar a solução de detecção com base no treinamento;
- Avaliar o desempenho do modelo de detecção e classificação;

### 3.1 Justificativa

O propósito é contribuir para o progresso das tecnologias no domínio da pesquisa em *deep learning* aplicado a produtos agrícolas, com ênfase na identificação dos modelos de redes neurais convolucionais que se destacam em termos de eficácia para essa finalidade. Com o aumento da população, a necessidade de alimentos cresceu, o que levou a um aumento na produção de alimentos agrícolas. O objetivo deste trabalho é aprimorar a classificação e identificação desses alimentos por meio de métodos de aprendizado de máquina, redes neurais e reconhecimento de imagens. Podendo ajudar a melhorar a eficiência da distribuição desses alimentos, permitindo que a distribuição seja mais eficiente.

Este capítulo, estabelece o alicerce fundamental do trabalho, abordando o problema central que motiva essa pesquisa, delineando a proposta de solução. Além disso, define os objetivos específicos que orientam o estudo e fornece uma visão clara do que se pretende alcançar. E discute a justificativa por trás da escolha desse problema, demonstrando sua relevância no contexto atual. No próximo capítulo será abordado como irá ser estruturado os métodos, como são tratadas as imagens, matérias e os treinamentos dos modelos, a relação entres esses dois capítulos ambos são complementares, pois sem um objetivo é impossível definir com quais modelos de CNN o problema se encaixa melhor.



## 4 METODOLOGIA

Este capítulo, apresentará em detalhes a metodologia empregada na realização deste trabalho, sendo descritas as etapas presentes no desenvolvimento utilizada por este projeto, detalhando a base escolhida e as ferramentas computacionais utilizadas para o desenvolvimento dos algoritmos.

### 4.1 Base de imagens

As imagens de referência empregadas nesta pesquisa foram primeiramente coletadas no site Kaggle e disponibilizadas por (AZAM, 2022). Essa base de dados ao todo possui 829 imagens, aos quais se dividem em 30 pastas rotuladas pelos alimentos escolhidos. Sendo na seguinte ordem *cherry* (cerejas), *coffe-plant* (café), *cucumber* (pepino), *fox nut* (noz de raposa), *lemon* (limão), *olive-tree* (oliveira), *pearl millet* (milheto pérola), *tobacco-plant* (tabaco), *almond* (amêndoa), *banana*, *cardamom* (cardamomo), *chilli* (pimenta), *clove* (cravo da índia), *coconut* (coco), *cotton* (algodão), *gram* (Grão-de-bico), *jowar* (sorgo), *jute* (juta), *maize* (milho), *mustard-oil* (mostarda), *papaya* (mamão), *pineapple* (abacaxi), *rice* (arroz), *soyabean* (Soja), *sugarcane* (cana-de-açúcar), *sunflower* (girassol), *tea* (chá), *tomato* (tomate), *vigna-radiati* (feijão-verde), *wheat* (trigo). A base de treinamento foi separada em duas partes: uma para treinamento e outra para validação. Após cada ciclo completo de treinamento, a base de validação foi usada para testar o modelo. Essas duas partes de dados tinham tamanhos e características iguais durante todo o experimento, o que garantiu uma comparação imparcial dos resultados entre diferentes arquiteturas de modelo em uma única sessão de testes. Essa abordagem proporciona uma avaliação justa do desempenho dos modelos.

Na parte do código, uma função é usada para criar conjuntos de treinamento e teste a partir de um conjunto de dados, garantindo que o modelo seja treinado e avaliado de maneira justa e representativa. Ela é amplamente usada no desenvolvimento e avaliação de modelos de aprendizado de máquina para garantir que os modelos sejam robustos e generalizar bem para dados não vistos. A base de dados foi dividida em dois conjuntos: um conjunto de

treinamento (com 70% dos dados) e um conjunto de validação (com 30% dos dados). A Figura 12 é um simples sumário da base de dados.

FIGURA 12 - Sumário da base dados

Alimentos	Numeros de imagens por pasta
Pearl_millet(bajra)	39
Tobacco-plant	33
cotton	32
Cherry	32
wheat	31
Cucumber	31
banana	31
maize	31
clove	30
soyabean	30
Olive-tree	30
jowar	30
Coffee-plant	29
rice	29
mustard-oil	28
Lemon	28
vigna-radiati(Mung)	27
tomato	26
gram	25
coconut	25
sugarcane	25
pineapple	25
sunflower	24
Fox_nut(Makhana)	23
jute	23
papaya	23
chilli	23
tea	23
cardamom	22
almond	21

Fonte: Imagem do próprio autor

## 4.2 Materiais

O Google Colab foi a plataforma escolhida, é uma plataforma baseada na web que fornece acesso a máquinas virtuais com recursos de CPU, GPU e memória. É uma escolha popular para a comunidade de aprendizado de máquina devido à sua facilidade de uso e, muitas vezes, oferece acesso a GPUs, o que pode acelerar o treinamento de modelos de *machine learning*.

No Google Colab, foi o ambiente de execução *Python 3* no tipo de ambiente de execução que se refere ao ambiente de execução onde pode-se escolher para criar e executar um código. Esses ambientes de execução determinam qual linguagem de programação pode-se usar para criar um notebook. O *Python 3* é o ambiente de execução padrão no Google Colab e permite que crie e execute notebooks usando a linguagem de programação *Python*, é uma escolha comum para a maioria dos usuários, pois o Colab oferece uma ampla variedade de bibliotecas e recursos *Python* prontos para uso.

Em relação a Hardware no ambiente virtualizado do Google Colab foi utilizado o CPU disponibilizado, e como observa-se na Figura 13 um detalhamento das especificações do processador, e 12.7 GB de memória ram e 107.7 GB de espaço em disco.

A linguagem de programação utilizado para o projeto foi *Python 3.10.12* com as bibliotecas *tensorflow* (versão 2.14.0), *Keras* (versão 2.14.0) utilizando e a biblioteca *os* (versão 3.10.12) organizar as imagens do banco de dados em forma que possa ser usada no algoritmos de redes neurais.

FIGURA 13 - Cpu do Colab.

```

Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Address sizes:          48 bits physical, 48 bits virtual
Byte Order:             Little Endian
CPU(s):                 2
  On-line CPU(s) list:  0,1
Vendor ID:              AuthenticAMD
Model name:             AMD EPYC 7B12
CPU family:             23
Model:                  49
Thread(s) per core:     2
Core(s) per socket:     1
Socket(s):              1
Stepping:               0
BogoMIPS:               4499.99
Flags:                  fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clf
                        lush mmx fxsr sse sse2 ht syscall nx mmxext fxsr_opt pdpe1gb rdtscp lm cons
                        tant_tsc rep_good nopl nonstop_tsc cpuid extd_apicid tsc_known_freq pni pcl
                        mulqdq ssse3 fma cx16 sse4_1 sse4_2 movbe popcnt aes xsave avx f16c rdrand
                        hypervisor lahf_lm cmp_legacy cr8_legacy abm sse4a misalignsse 3dnowprefetch
                        h osvw topoext ssbd ibrs ibpb stibp vmmcall fsgsbase tsc_adjust bmi1 avx2 s
                        mep bmi2 rdseed adx smap clflushopt clwb sha_ni xsaveopt xsavec xgetbv1 clz
                        ero xsaveerptr arat npt nrip_save umip rdpid

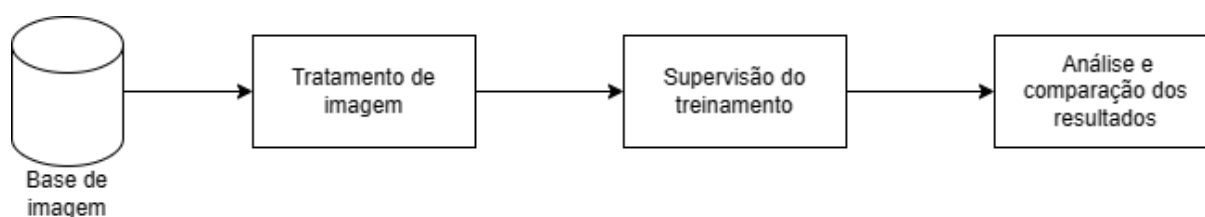
```

Fonte: Imagem do próprio Autor

### 4.3 Método proposto

Esta seção mostra passo a passo do método proposto do Objetivo que se pretende alcançar, buscamos abranger todas as nuances e etapas necessárias para atingir com sucesso esse objetivo. A Figura 14 mostra um diagrama que descreve o passo a passo para alcançar esse objetivo.

FIGURA 14 - Diagrama de embolsando a metodologia.



Fonte: Imagem do próprio autor

### 4.4 Tratamento de Imagem

Essa etapa tem a finalidade de preparação e tratamentos no conjunto de dados de imagens para treinar os modelos de redes neurais, as técnicas usadas nessa parte deste código têm o propósito de melhorar a precisão dos modelos de aprendizado de máquina. Elas fazem isso ao introduzir variações nas imagens de treinamento, como girar, ampliar, deslocar, espelhar,

mudar o brilho, distorcer e variar as cores. Essas variações ajudam o modelo a aprender a reconhecer objetos em diferentes situações, tornando-o mais eficaz na hora de fazer previsões em situações do mundo real, onde as imagens podem variar em termos de posição, orientação, iluminação e outras características. Em resumo, essas técnicas de aumento de dados aprimoram a capacidade do modelo de se adaptar a diferentes condições e, conseqüentemente, melhora sua precisão.

O código usa a classe *ImageDataGenerator* da biblioteca *Keras/TensorFlow* para aplicar uma série de tratamentos de imagem conhecidos como aumento de dados (data augmentation). Aqui estão os tratamentos de imagem que são realizados:

- Rotação (rotation\_range=10): As imagens podem ser giradas até 10 graus no sentido horário ou anti-horário. Isso ajuda o modelo a aprender a lidar com diferentes orientações das imagens;
- Zoom (zoom\_range=0.1): As imagens podem ser ampliadas ou reduzidas em até 10% do tamanho original. Isso ajuda a aumentar a diversidade das imagens e a tornar o modelo mais robusto;
- Deslocamento horizontal (width\_shift\_range=0.1) e vertical (height\_shift\_range=0.1): As imagens podem ser deslocadas horizontalmente ou verticalmente em até 10% do tamanho da imagem isso introduz variações na posição dos objetos na imagem;
- Espelhamento horizontal (horizontal\_flip = True) e vertical (vertical\_flip=True): As imagens podem ser espelhadas horizontalmente ou verticalmente. Isso ajuda o modelo a aprender a reconhecer objetos independentemente da orientação;
- Variação de brilho (brightness\_range=(0.8, 1.2)): O brilho das imagens pode variar entre 80% e 120% do brilho original. Isso ajuda a tornar o modelo mais resistente a variações de iluminação. ;
- Cisalhamento (shear\_range=0.1): As imagens podem ser distorcidas levemente em até 10 graus. Isso ajuda o modelo a lidar com deformações em objetos;

- Variação de canal (`channel_shift_range=0.1`): Pode ocorrer uma pequena variação nos valores dos canais de cor das imagens, dentro do intervalo especificado. Isso introduz variações de cor nas imagens;

#### 4.5 Supervisão de treinamento

Essa etapa tem a finalidade de supervisionar, ou seja, orientar os treinamentos dos Modelos, redes neurais com ou sem aprendizado de máquina com setar seus hiper parâmetros e como suas camadas são criadas e qualquer outra configuração que serão utilizados para treinamento dos modelos.

Os experimentos envolveram a utilização de cinco modelos de aprendizado de máquina pré-treinados, nomeadamente Mobilenet, MobilenetV2, MobilenetV3Small, VGG16 e VGG19 visando realizar comparação entre com e sem Transferência de aprendizado. A importância dos hiperparâmetros em uma rede neural recorrente (RNN) é uma etapa essencial no desenvolvimento de modelos de aprendizado de máquina. A escolha cuidadosa desses hiperparâmetros é crucial, uma vez que eles podem afetar a capacidade do modelo de aprender relações temporais em dados sequenciais, sua habilidade de generalização e a eficácia em tarefas específicas. As mesmas configurações de hiperparâmetros foram aplicados a todos modelos e também foi aplicada normalização. Os pesos desses modelos pré-treinados, foi a partir do banco de imagens da Imagenet. E para o treinamento dos modelos, foram selecionados os seguintes hiperparâmetros.

Hiperparâmetros do Modelo:

- `input_shape=(224, 224, 3)`: Especifica o tamanho e o número de canais das imagens de entrada;
- `alpha=1.0`: Fator de escala que afeta a arquitetura do modelo MobileNetV3Small;
- `include_top=False`: Indica se as camadas finais da rede (camadas densas) estão incluídas no modelo ou não;
- `weights="imagenet"`: Define a fonte dos pesos pré-treinados, neste caso, a base de dados ImageNet;
- `input_tensor=None`: Especifica o tensor de entrada (imagem);

- pooling=None: Tipo de camada de pooling aplicada à saída da rede;
- dropout\_rate=0.5: A taxa de dropout aplicada à camada densa;

Hiperparâmetros de Treinamento:

- Função de Perda: Categorical Cross Entropy. Esta é a perda padrão em casos de classificação;
- Número de épocas: 20;
- Optimizador: RMSprop para os algoritmos do Mobile Net e Adam para os algoritmos do VGG;
- batch\_size=4;

Em relação à construção da camada da rede neural que não faz uso de transferência de aprendizado, ela foi construída da seguinte forma:

Camada de Convolução 2D (Conv2D):

- 64 filtros;
- Tamanho do filtro: (3, 3);
- Função de ativação: ReLU;
- Tamanho da entrada: (150, 150, 3);

Camada de MaxPooling 2D:

- Tamanho do pool: (2, 2);

Outra Camada de Convolução 2D:

- 64 filtros;
- Tamanho do filtro: (3, 3);
- Função de ativação: ReLU;

Outra Camada de MaxPooling 2D:

- Tamanho do pool: (2, 2);



Camada de Flattening:

- Transforma os mapas de características 2D em um vetor 1D;

Camada de Dropout:

- Taxa de dropout: 0,5;

Camada Densa (Fully Connected Layer):

- 512 neurônios;
- Função de ativação: ReLU;

Outra Camada Densa:

- 30 neurônios;
- Função de ativação: Softmax;

Em relação à construção da camada da rede neural que fazem uso de transferência de aprendizado ela foi construída da seguinte forma

Camada Densa:

- 1024 neurônios;
- Função de ativação: ReLU;

Camada de Dropout:

- Taxa de dropout: 0,3;

Camada Densa:

- 30 neurônios;
- Função de ativação: softmax.

#### **4.6 Análise e comparação dos resultados**

Nessa etapa será apresentada uma análise abrangente e uma comparação detalhada dos resultados obtidos entre as diferentes arquiteturas para compreender quais modelos são mais adequados para a tarefa específica em questão. Essa etapa fornece uma visão abrangente do desempenho das CNNs e o motivo para a tomada de decisão da escolha dos melhores modelos em aplicações de aprendizado de máquina relacionadas à classificação de imagens de problema em questão. Os resultados são comparados, quando aplicável, para destacar as diferenças entre modelos. As descobertas são discutidas em relação aos objetivos da pesquisa, fornecendo uma visão clara do significado prático dos resultados.

A análise mencionada faz parte do próximo capítulo de desenvolvimento que chama a Análise de resultados, explica os procedimentos e técnicas usadas para realizar o estudo que é falado neste tópico. A seção de Análise de resultados dos resultados vem determinar a eficácia de cada modelo de CNN.

## **5 DESENVOLVIMENTO**

Neste capítulo, são apresentados as análises e resultados das arquiteturas de redes neurais convolucionais (CNN) obtidos por meio da aplicação do método proposto, juntamente com comparações feitas em relação a resultados disponíveis. Essa análise inclui a classificação das arquiteturas com base nas métricas previamente discutidas. A avaliação das métricas desempenha um papel fundamental na compreensão do desempenho das CNNs, permitindo uma análise criteriosa e comparativa que ajuda a identificar quais modelos se destacam na resolução do problema em questão.

### **5.1 Introdução**

Essa seção visa retomar e relembrar os objetivos que direcionam este trabalho, bem como os desafios e problemas que se propõem a resolver. Esse TCC teve como objetivo geral, e encontra o problema de achar e implementar algoritmos para classificar uma base de dados de plantio de agricultura e sua metodologia entrega para solução do problema ao qual seria usando redes neurais convolucionais faria uma comparação diferentes algoritmos para classificação de imagem para plantio de agricultura. Assim, esta introdução sinaliza o caminho que este capítulo seguirá, proporcionando um contexto claro para a análise e discussão que se desenrolará nas seções subsequentes.

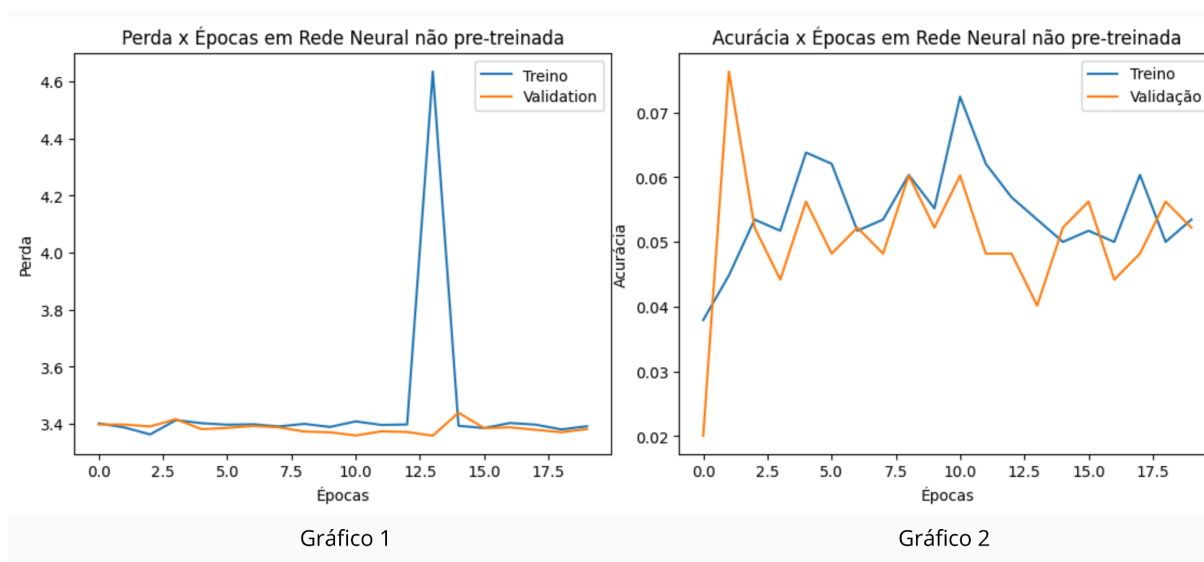
### **5.2 Contextualização**

Todos os resultados apresentados consideraram o modelo de CNN utilizando a arquitetura MobileNet, MobileNetV1, MobileNetV3Small e VGG16 e VGG19 e um modelo sem aprendizagem por transferência. As métricas que serão aplicadas para medir o desempenho será a acurácia e a função de perda. Dois gráficos serão plotado para cada rede neural sendo o primeiro gráfico o seu eixo X igual a épocas da rede e no seu eixo Y a função de perda, o segundo gráfico o seu eixo X igual a épocas da rede e no seu eixo Y a função de acurácia, e ambos os gráficos serão traçados: duas linhas uma marcando o treinamento e outra a validação durante o treinamento das redes neurais.

## 5.2 Análise de resultados

Esta seção é um componente crucial deste estudo, uma vez que permite visualizar com mais clareza as diferenças de cada arquitetura de redes neurais convolucionais. Durante esta análise, será examinado minuciosamente as métricas previamente definidas para determinar a eficácia de cada modelo de CNN. Além disso, identificar padrões e tendências que podem fornecer uma visão valiosa sobre como melhorar ainda mais a compreensão e aplicação da agricultura.

FIGURA 15 - Rede Neural sem aprendizagem por transferência.

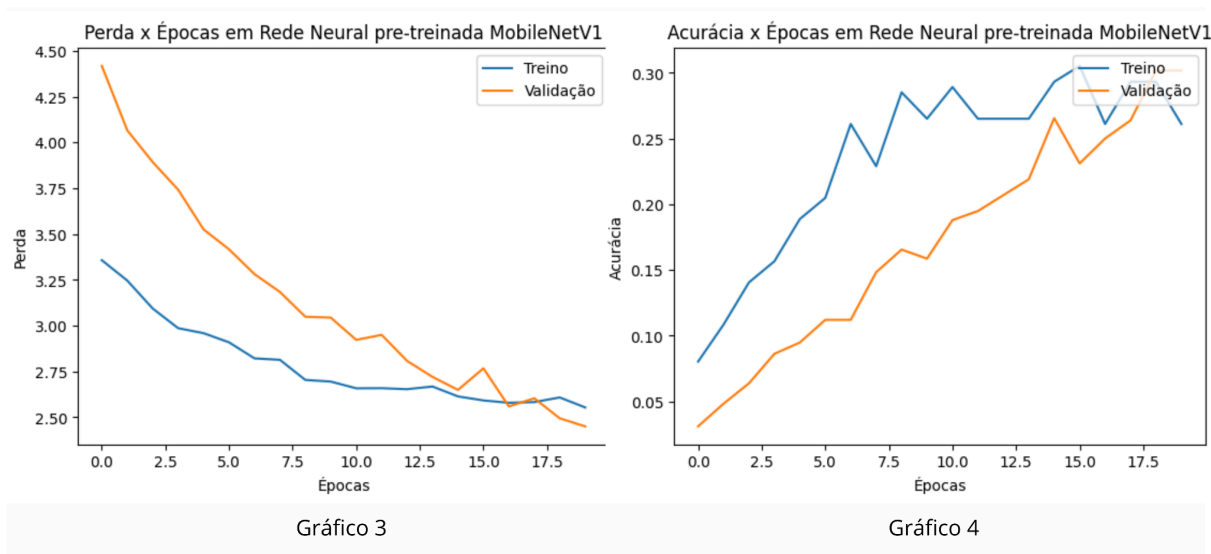


Fonte: Imagem do próprio autor

Na Figura 15 mostra uma rede neural que não usa aprendizagem por transferência, e no gráfico 1 tendo sua função de perda final de treinamento sendo 3,3817 e validação 3,3917, no gráfico 2 acurácia final de treinamento 0,0534 e validação 0,0522, com 20 épocas, e tempo de execução de 1560 segundos. Sua acurácia é a menor, 0,0534 no conjunto de treinamento, o que sugere que o modelo tem dificuldade em aprender com os dados. E a sua função de perda é a maior, tendo 3,3817, indicando que o modelo está tendo dificuldade em minimizar o erro durante o treinamento. No conjunto de validação, a acurácia e a função de perda são semelhantes, o desempenho geral é pior entre todos os modelos testados. Seus valores durante o treinamento ficaram estagnados, não tendo quase nada de variação durante seu aprendizado, e vale ressaltar que se a arquitetura da CNN for muito simples, com poucas

camadas ou poucos parâmetros, isso pode levar ao *underfitting*, já que o modelo não tem capacidade para aprender padrões complexos. Com valores muito abaixo demonstra que uma rede neural sem aprendizagem por transferência se mostrou o pior entre todos os modelos testados para se classificar uma base de dados com 30 possibilidades diferentes.

FIGURA 16 - Rede Neural Convolutacional MobileNet.

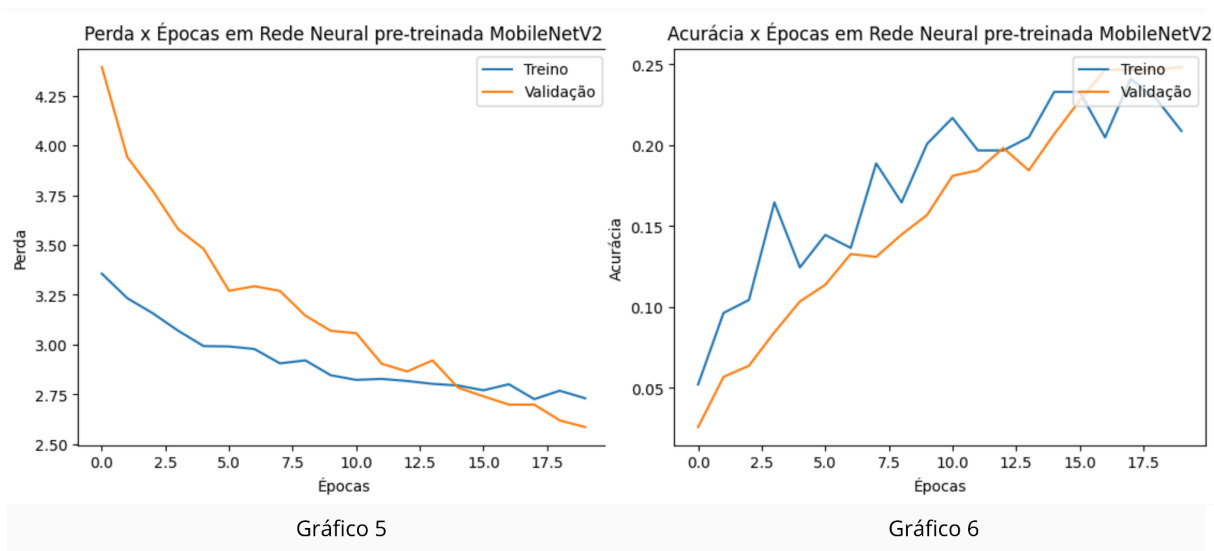


Fonte: Imagem do próprio autor

Na Figura 16 mostra um rede neural que usa aprendizagem por transferência MobileNet, tendo no gráfico 3 a função de perda de treinamento sendo 2,4504 e validação 2,5537 e no gráfico 4 sua acurácia final de treinamento 0,3017 e validação 0,2610, com 20 épocas, e tempo de execução de 1606 segundos. A acurácia é melhor em comparação com a rede não pré-treinada, mas ainda relativamente baixa 0,3017 no conjunto de treinamento. A função de perda no treinamento 2,4504 é menor do que a da rede não pré-treinada, indicando algum progresso na otimização. O modelo parece estar fazendo um trabalho melhor no conjunto de validação em comparação com o conjunto de treinamento (acurácia de validação de 0,2610 versus acurácia de treinamento de 0,3017). A rede MobileNet mostrou uma melhora em relação a uma rede sem aprendizagem por transferência, mas mesmo assim mostrou nível de acerto baixo em relação às redes neural MobileNetV2 e V3 e VGG. Com valores muito abaixo demonstra que uma rede neural MobileNet se mostrou pior que os modelos

MobileNetV2 e V3 e os modelos VGG para se classificar uma base de dados com 30 possibilidades diferentes.

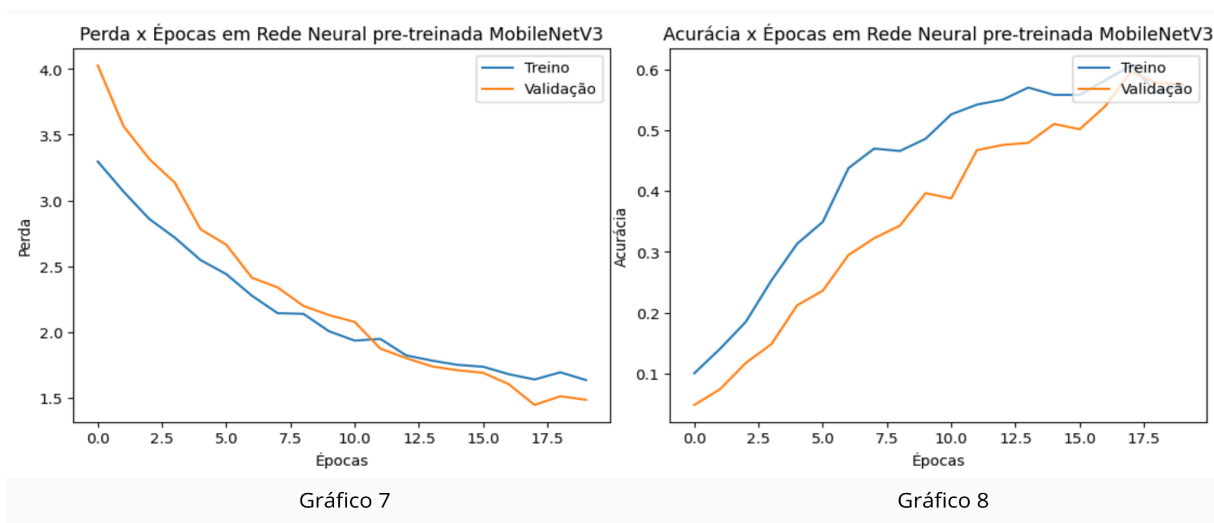
FIGURA 17 - Rede Neural Convolutacional MobileNetV2.



Fonte: Imagem do próprio autor

Na Figura 17 mostra um rede neural que usa aprendizagem por transferência MobileNetV2, no gráfico 5 tendo sua função de perda de treinamento sendo 2,5858 e validação 2,5537 e no gráfico 6 sua acurácia final de treinamento 0,3017 e validação 0,0522, com 20 épocas, e tempo de execução de 786 segundos. A acurácia (0,2483) e a função de perda (2,5858) no conjunto de treinamento são semelhantes aos resultados do MobileNet. O modelo tem dificuldade em generalizar o conjunto de validação, onde a acurácia é baixa 0,2088 em relação ao MobileNetV3 e os modelos VGG e a sua função de perda de validação é semelhante ao modelo MobileNet 2,7305. A rede MobileNetV2 não mostrou nenhuma mudança do modelo MobileNet e ainda teve uma leve piora em relação à acurácia ao MobileNet, e sua acurácia foi mais baixa do que o MobileNet. Com valores abaixo que uma rede MobileNet se mostrou pouco eficiente para se classificar uma base de dados com 30 possibilidades diferentes.

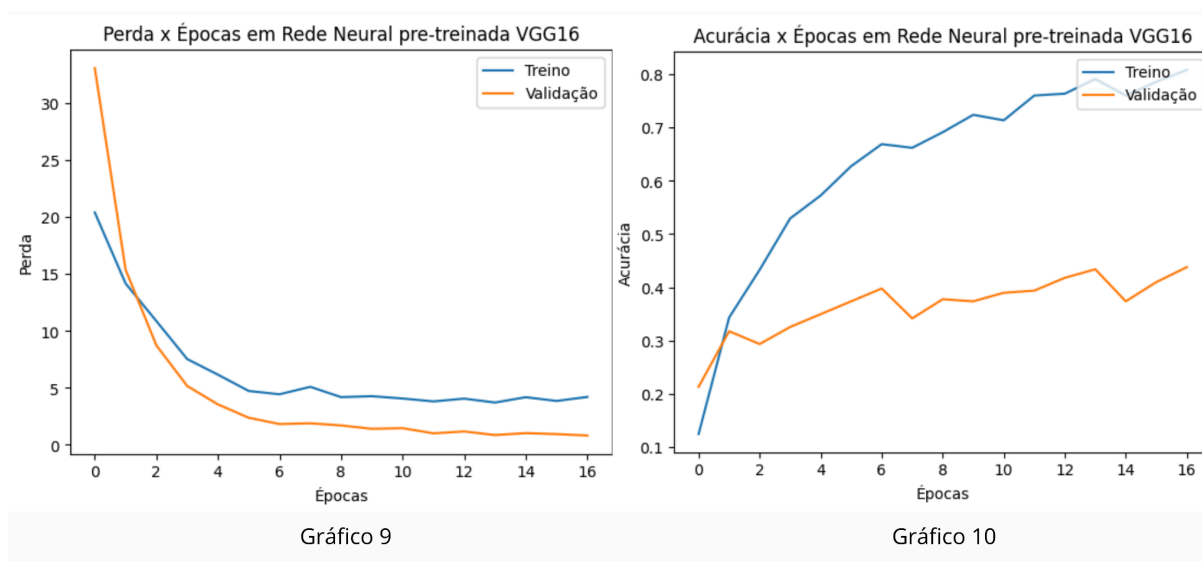
FIGURA 18 - Rede Neural Convolucional MobileNetV3.



Fonte: Imagem do próprio autor

Na Figura 18 mostra um rede neural que usa aprendizagem por transferência MobileNetV3, no gráfico 7 tendo sua função de perda de treinamento sendo 1,4861 e validação 1,6358 e no gráfico 8 sua acurácia final de treinamento 0,5759 e validação 0,5743, com 20 épocas, e tempo de execução de 724 segundos. A acurácia 0,5759 no conjunto de treinamento é significativamente melhor em comparação com os modelos MobileNetV2, MobileNet, e não treinado. A função de perda 1,4861 é menor de todos os modelos mostrados anteriormente, sugerindo um melhor ajuste dos dados. No entanto, o modelo ainda tem alguma dificuldade em generalizar para o conjunto de validação, onde a acurácia de validação 0,5743 é próxima da acurácia de treinamento, e a função de perda de validação 1,6358 é mais alta em relação os modelos VGG 16 e 19. A rede MobileNetV3 mostrou uma grande melhora em relação ao MobileNetV2 e MobileNet e não treinado, porque modelos maiores com mais camadas podem aprender representações mais complexas, o que pode ser benéfico para tarefas mais complexas. A MobileNetV3 demonstra que uma rede neural aprendizagem por transferência que se demonstrou melhor que as redes anteriores para se classificar uma base de dados com 30 possibilidades diferentes.

FIGURA 19 - Rede Neural Convolucional VGG16.

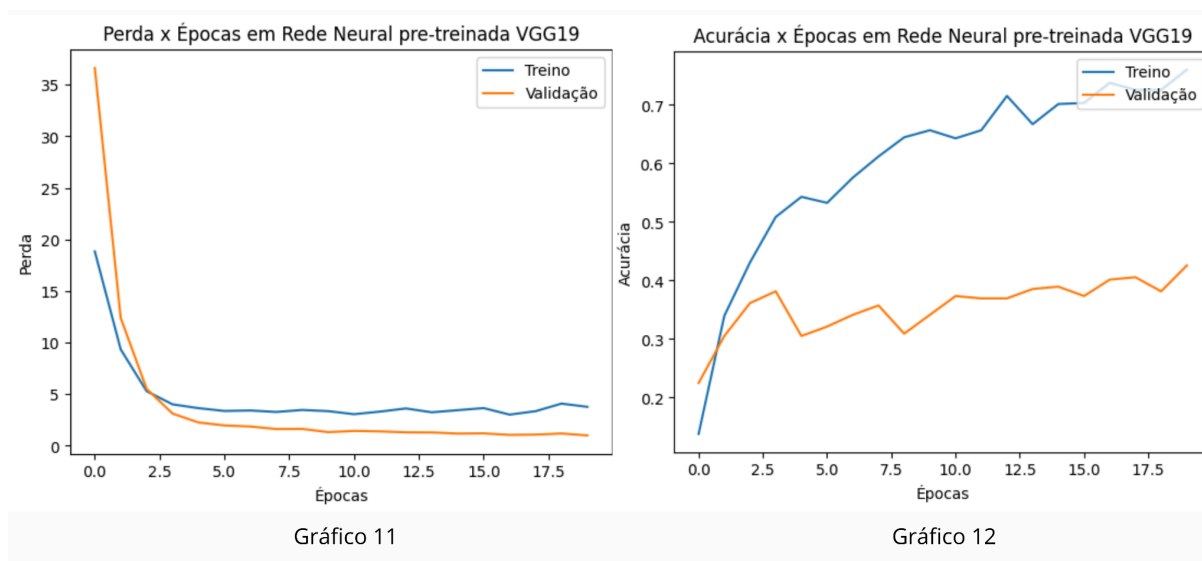


Fonte: Imagem do próprio autor

Na Figura 19 mostra uma rede neural que usa aprendizagem por transferência VGG16, no gráfico 9 tendo sua função de perda de treinamento sendo 0,8192 e validação 4,2094 e no gráfico 10 sua acurácia final de treinamento 0,8086 e validação 0,4378, com 20 épocas, e tempo de execução de 5138 segundos. A VGG16 demonstra um o melhor resultado no conjunto de treinamento, com uma alta acurácia 0,8086 e a função de perda mais baixa 0,8192. No entanto, o modelo parece sofrer de *overfitting*, pois a acurácia de validação 0,4378 é consideravelmente mais baixa em relação ao conjunto de treinamento, e a função de perda de validação 4,2094 é significativamente mais alta em relação a validação. A rede VGG16 mostrou uma grande melhora em relação ao algoritmos MobileNet, a rede VGG16 é uma rede com 16 camadas convolucionais e tende a oferecer alto desempenho em tarefas de visão computacional e é frequentemente usado quando a precisão é a principal prioridade, o que pode ser benéfico para tarefas mais complexas, e os algoritmos MobileNet são projetados para serem mais eficientes e leves. Ele mostrou o nível de acurácia mais alto dos modelos testados. O modelo VGG16 demonstra a maior acurácia entre todos os modelos para se classificar uma base de dados com 30 possibilidades diferentes.



FIGURA 20 - Rede Neural Convolucional VGG19.



Fonte: Imagem do próprio autor

Na Figura 20 mostra um rede neural que usa aprendizagem por transferência VGG19, no gráfico 11 tendo sua função de perda de treinamento sendo 0,9973 e validação 3,7664 e no gráfico 12 sua acurácia final de treinamento 0,7603 e validação 0,4257, com 20 épocas, e tempo de execução de 6444 segundos. A VGG19 tem um padrão semelhante ao da VGG16, com uma acurácia elevada no conjunto de treinamento 0,7603 e uma função de perda mais alta no conjunto de validação 0,9973. Mais uma vez, o modelo mostra sinais de overfitting, com a acurácia de validação 0,4257 sendo consideravelmente mais baixa que a acurácia de treinamento e a função de perda de validação 3,7664 mais alta que função de perda de treinamento. A rede VGG19 como mostrada anteriormente com rede a VGG19 também tem uma grande melhora em relação ao algoritmo MobileNet. A utilização da rede VGG19 em bases de dados pequenas apresenta um desafio significativo dada sua profundidade, pela demanda por um volume considerável de dados de treinamento para aprender representações robustas é uma limitação, uma vez que bases de dados pequenas e médias não fornecem informações suficientes para ser mais eficaz. Ela mostrou um nível de acurácia semelhante ao VGG 16. O VGG19 demonstra a segunda maior acurácia entre todos os modelos para se classificar uma base de dados com 30 possibilidades diferentes.

## 5.4 Comparação de resultados

Na seção a seguir, será realizada uma análise comparativa dos resultados objetivos das redes neurais convolucionais (CNNs) implementadas. O objetivo é avaliar o desempenho de cada modelo em relação às métricas definidas, tais como acurácia, e a função de perda. Isso permitirá extrair informação sobre quais CNNs se desempenham na tarefa de classificação de alimentos agrícolas.

FIGURA 21 - Quadro de comparação entre Rede Neurais acurácia e função de perda.

Modelo	Acuracia	Acuracia de Validação	Função de Perda	Função de Perda de Validação
Rede não pré-treinada	0,0534	0,0522	3,3817	3,3917
MobileNet	0,3017	0,2610	2,4504	2,5537
MobileNetV2	0,2483	0,2088	2,5858	2,7305
MobileNetV3Small	0,5759	0,5743	1,4861	1,6358
VGG16	0,8086	0,4378	0,8192	4,2094
VGG19	0,7603	0,4257	0,9973	3,7664

Fonte: Imagem do próprio autor

Com base nos resultados apresentados na Figura 21, é possível ver a avaliação dos modelos de classificação de imagens de plantio agrícola. Os resultados indicam variações significativas de desempenho entre os modelos, com diferenças notáveis entre as métricas de treinamento e validação.

A análise revela que modelos não pré-treinados apresentam um desempenho inferior a todos os outros modelos testados, caracterizado pela menor taxa de acurácia e o maior valor na função de perda, destacando a importância na transferência de aprendizado para melhores resultados. Por outro lado, modelos como MobileNet, MobileNetV2, MobileNetV3Small, VGG16 e VGG19 demonstram resultados variados. Alguns deles alcançam uma acurácia mais alta em relação às outras, mas com desafios em relação à generalização, conforme evidenciado pela diferença notável entre as métricas de treinamento e validação. Além disso, a análise ressalta a necessidade de lidar com problemas de overfitting, especialmente em modelos mais complexos.

FIGURA 22 - Quadro de comparação entre Rede Neurais pelo tempo.

Modelo	Tempo
Rede não pré-treinada	1560
MobileNet	1606
MobileNetV2	786
MobileNetV3Small	724
VGG16	5138
VGG19	6444

Fonte: Imagem do próprio autor

Na Figura 22, mostra o tempo gasto em cada modelo. Os resultados indicam variações significativas em tempo, com diferenças notáveis entre algoritmos VGG e MobileNet.

A análise revela que modelo não pré-treinado apresenta tempo maior que MobileNetV2 e MobileNetV3Small tendo sua acurácia menor que ambos, destacando a importância na transferência de aprendizado. Por outro lado, modelos como VGG16 e VGG19 demonstram os maiores tempos gasto em todos os modelos, mas tendo a maior acurácia. Além disso, a análise ressalta a necessidade de lidar com problemas de *overfitting*, especialmente em modelos mais complexos.

Esses resultados forneceram uma visão abrangente do desempenho das CNN na classificação de imagens de plantio agrícola e estabelecem as bases para futuras melhorias e otimizações. No próximo capítulo, serão abordadas as conclusões e implicações desses resultados, bem como sugestões para trabalhos futuros na área de aprendizado de máquina aplicado à agricultura.

## 6 CONCLUSÃO

O presente trabalho explorou a aplicação de redes neurais convolucionais (CNNs) na classificação de alimentos agrícolas, visando aprimorar a identificação precisa desses produtos por meio da visão computacional. Durante o estudo, foi implementado e avaliado diversos modelos de CNNs, incluindo o VGG, o MobileNet e rede sem aprendizado por transferência, e assim verificar o desempenho mais eficaz na tarefa de classificação de cultivos agrícolas.

Nesta análise comparativa dos modelos, pode-se observar uma variedade de desempenhos ao avaliar as métricas-chave: acurácia, acurácia de validação, função de perda e função de perda de validação. Os resultados revelam a importância da escolha da arquitetura da rede neural convolucional (CNN) e a influência da transferência de aprendizado com modelos pré-treinados.

Observa-se que a rede não pré-treinada apresentou uma acurácia muito baixa e uma função de perda alta em relação a todas as outras redes testadas, indicando a necessidade de um grande conjunto de dados de treinamento para alcançar resultados satisfatórios. Por outro lado, as arquiteturas MobileNet, MobileNetV2 e MobileNetV3Small demonstraram desempenho moderado em relação às demais, com a MobileNetV3Small obtendo a melhor acurácia e a menor função de perda entre elas.

Surpreendentemente, as redes VGG16 e VGG19, conhecidas por sua profundidade e capacidade de aprendizado, alcançaram as maiores acurácias, mas apresentaram uma grande discrepância entre a acurácia de treinamento e validação. Isso sugere um problema de overfitting, no qual essas redes se adaptaram demais aos dados de treinamento, resultando em desempenho inferior em dados não vistos.

Já o tempo de execução tem uma grande diferença, a VGG16 e a VGG19 precisaram de mais tempo de execução em relação às outras, sendo quase 4000 segundos a mais. A mais eficiente seria a MobileNetV3, sendo seu tempo 724 segundos o menor tempo entre os algoritmos, mostrando que nessa questão de desempenho ela se destaca.

Em resumo, as redes VGG16 e a VGG19 demonstram um desempenho superior no conjunto de treinamento em relação às demais redes, mas tendo vista que levam mais tempo, com dificuldade em generalizar para o conjunto de validação, sugerindo overfitting. O MobileNetV3Small, embora tenha um desempenho médio no treinamento comparado com

as demais, também parece ter alguma dificuldade em generalizar. O MobileNet e o MobileNetV2 inferior a MobileNetV3 e aos modelos VGG, com uma melhor generalização em relação aos modelos não pré-treinados.

Esse estudo fornece informações valiosas sobre a seleção de modelos da CNN e destaca a importância de técnicas de regularização e validação cruzada para garantir resultados robustos e generalização eficaz em tarefas de classificação de imagens. Com base nos resultados apresentados, uma limitação evidente é a diferença entre a acurácia de treinamento e a acurácia de validação em alguns modelos, como VGG16 e VGG19. Essa discrepância sugere um problema de *overfitting*, no qual esses modelos se ajustam demais aos dados de treinamento e, como resultado, seu desempenho em dados não vistos (validação) é menor.

Além disso, a MobileNet, embora tenha uma acurácia mais equilibrada em relação às demais, ainda mostra uma diferença significativa entre as métricas de treinamento e validação. Isso sugere que esse conjunto de dados com a transferência de aprendizado no algoritmo da MobileNet pode não ser ideal em termos de generalização. Para superar essas limitações e melhorar o desempenho geral, é necessário implementar técnicas de regularização, como *dropout*, e ajustar os hiperparâmetros adequadamente.

Além disso, o uso de um conjunto de dados de validação maior e mais diversificado pode ajudar a avaliar o desempenho dos modelos de forma mais precisa. Essas medidas são essenciais para garantir que os modelos sejam capazes de generalizar eficazmente para novos dados e evitar o *overfitting*. Alguns meios de reduzir o *overfitting* incluem:

- Adicionar dados de treinamento: quanto mais dados de treinamento o modelo tiver, menor será a chance de ele aprender padrões ruidosos. Se não for possível obter mais dados, pode-se usar técnicas de aumento de dados, que consistem em criar novas amostras a partir das existentes, aplicando transformações como rotação, translação, escala, ruído, etc.
- Regularização: é uma técnica que consiste em adicionar um termo à função de custo do modelo que penaliza os valores extremos dos parâmetros. Isso impede que o modelo se ajuste demais aos dados de treinamento e favorece soluções mais simples e robustas.

- Simplificar o modelo: se o modelo for muito complexo em relação aos dados disponíveis, ele pode capturar padrões que não são generalizáveis. Uma forma de evitar isso é simplificar o modelo, reduzindo o número de parâmetros, camadas, neurônios, etc. Isso pode ser feito por meio de técnicas como poda, compressão ou destilação de modelos.

Em suma, esse trabalho oferece uma base sólida para avançar na automação da classificação de alimentos agrícolas, e traz uma análise para otimizar processos que pretendem utilizar visão computacional na área de plantio agrícolas e contribuir para futuras pesquisas na área de classificação de plantio agrícolas.

Algumas sugestões de temas futuros, assuntos e continuação para futuras pesquisas:

- diversificar e aumentar o repertório de bases de plantio agrícolas;
- Investigar métodos para automatizar a seleção de hiperparâmetros, a fim de otimizar o desempenho do modelo;
- Ampliar a pesquisa para incluir a detecção de doenças e pragas em plantas, essencial para a saúde das colheitas;
- Explorar como incorporar informações de sensores, como dados meteorológicos e de solo, juntamente com imagens de plantio, para melhorar a precisão da classificação;

## Referências

- ALMEIDA, Leonardo de Borges, SILVA, Marcus Vinícius da. Estudo comparativo de redes neurais convolucionais para a classificação da qualidade de imagens de documentos de identidade. 2021.
- ALVES, Gisely. Entendendo Redes Convolucionais (CNNs). 2023. Acessado em 2023-11-26 Disponível em: <<https://dev.to/giselyalves13/entendendo-redes-convolucionais-cnns-42h9>>
- AZAM, Md Waquar. Agricultural crops image classification. kaggle, 2022. Disponível em: <https://www.kaggle.com/datasets/mdwaquarazam/agricultural-crops-image-classification> Acessado em 10/29/2023 às 18:05.
- OLIVEIRA, Patrick Ágton de. Deep learning na segmentação automática de imagens de satélite. 2020.
- CUNHA, Leonardo Cardoso da. Redes neurais convolucionais e segmentação de imagens: uma revisão bibliográfica. 2020.
- DHANYA, VG; SUBEESH, A; KUSHWAHA, NL; VISHWAKARMA, Dinesh Kumar; KUMAR, T Nagesh; RITIKA, G; SINGH, AN. Deep learning based computer vision approaches for smart agricultural applications. Artificial Intelligence in Agriculture, Elsevier, 2022
- FAIZUNNABI, Syed. roda de cores HSV. 2020. Disponível em: <https://buzzneers.com/wpcontent/uploads/2020/08/HSVcolorsolidcylinder-1024x768.png> Acessado em 29/10/2023.
- FLECK, Leandro; TAVARES, Maria Hermínia Ferreira; EYNG, Eduardo; HELMANN, Andrieli Cristina; ANDRADE, Minéia Aparecida de Moares. Redes neurais artificiais: Princípios básicos. [S.l.: s.n.], 2016. 47–57 p.
- HO, Yaoshiang; WOOKEY, Samuel. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. IEEE access, IEEE, v. 8, p. 4806–4813, 2019
- HOWARD, Andrew; SANDLER, Mark; CHU, Grace; CHEN, Liang-Chieh; CHEN, Bo; TAN, Mingxing; WANG, Weijun; ZHU, Yukun; PANG, Ruoming; VASUDEVAN, Vijay et al. Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. [S.l.: s.n.], 2019. p. 1314–1324.
- HOWARD, Andrew G; ZHU, Menglong; CHEN, Bo; KALENICHENKO, Dmitry; WANG, Weijun; WEYAND, Tobias; ANDREETTO, Marco; ADAM, Hartwig. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017

INAMASU, Ricardo. A era das máquinas inteligentes no campo. 2021. Disponível em: <https://summitagro.estadao.com.br/colunistas/a-era-dasmaquinas-inteligentes-no-campo/> Acessado em 25/10/2023.

KRICHEN, Moez. Convolutional neural networks: A survey. *Computers*, MDPI, v. 12, n. 8, p. 151, 2023.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. *Curran Associates, Inc.*, v. 25, 2012. Disponível em: <https://proceedings.neurips.cc/paperfiles/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>. Acessado em 25/10/2023.

LAUBE, Alini Lessa. Pixel e sistema rgb. 2013. Disponível em: <https://www.render.com.br/artigos/o-que-e-pixel>. Acessado em 05/10/2023.

MAGIRI, Danilo Sampaio. Classificação de imagens de satélite com redes neurais convolucionais. Universidade Federal de São Carlos, 2023.

NETO, J CAMARGO; TERNES, Sônia; SOUZA, Kleber Xavier Sampaio de; YANO, IH; QUEIROS, LR. Uso de redes neurais convolucionais para detecção de laranjas no campo. In: IN: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 12., 2019, INDAIATUBA. ANAIS . . . . [S.l.], 2019.

PAIXÃO, Gabriela Miana de Mattos; SANTOS, Bruno Campos; ARAUJO, Rodrigo Martins de; RIBEIRO, Manoel Horta; MORAES, Jermana Lopes de; RIBEIRO, Antonio L. Machine learning na medicina: Revisão e aplicabilidade. *Arquivos Brasileiros de Cardiologia*, SciELO Brasil, v. 118, n. 1, p. 95, 2022.

PAN, Sinno Jialin; YANG, Qiang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 22, n. 10, p. 1345–1359, 2009.

PATRÍCIO, Diego Inácio; RIEDER, Rafael. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and electronics in agriculture*, Elsevier, v. 153, p. 69–81, 2018.

PEREIRA, Luís Moniz. Inteligência artificial: mito e ciência. *Revista Colóquio-Ciências*, v. 3, p. 1–13, 1988.

RAAD ALI KHALIL; BOUZOUANE, Abdenour; Ibrahim Hussein; ADDA MEHDI; Moshawrab Mohammad Mahmoud;. Smart Wearables for Cardiovascular Disease Detection and Prediction: A Systematic Literature Review. 2023. Acessado em 10/10/2023. Disponível em: <https://www.mdpi.com/1424-8220/23/2/828> Acessado em 20/10/2023.

RASCHKA, Sebastian; MIRJALILI, Vahid. *Python Machine Learning*. S.l.: Packt Publishing Ltd, 2017.



- RODRIGUES, Diego Alves. Deep learning e redes neurais convolucionais: reconhecimento automático de caracteres em placas de licenciamento automotivo. Universidade Federal da Paraíba, 2018.
- ROSA, Ayrton Lima da et al. Classificação de imagens de frutas utilizando aprendizado de máquina. Florianópolis, SC, 2019.
- SANDLER, Mark; HOWARD, Andrew; ZHU, Menglong; ZHMOGINOV, Andrey; CHEN, Liang-Chieh. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2018. p. 4510–4520.
- SHINDE, Pramila P; SHAH, Seema. A review of machine learning and deep learning applications. In: IEEE. 2018 Fourth international conference on computing communication control and automation (ICCUBE). [S.l.], 2018. p. 1–6
- SOUSA, Gabriel Garcez Barros. Deep learning para a detecção e classificação de pneumonia por radiografias do tórax. Universidade Federal do Maranhão, 2018.
- VELLINO, Andre. Artificial intelligence: The very idea: J. haugeland, (mit press, cambridge, ma, 1985); 287 pp. Artificial Intelligence, v. 29, p. 349–353, 09 1986.
- WEISS, Karl; KHOSHGOFTAAR, Taghi M; WANG, DingDing. A survey of transfer learning. Journal of Big data, SpringerOpen, v. 3, n. 1, p. 1–40, 2016.
- YAMASHITA, Rikiya; NISHIO, Mizuho; DO, Richard Kinh Gian; TOGASHI, Kaori. Convolutional neural networks: an overview and application in radiology. Insights into imaging, Springer, v. 9, p. 611–629, 2018.
- YING, Xue. An overview of overfitting and its solutions. In: IOP PUBLISHING. Journal of physics: Conference series. [S.l.], 2019. v. 1168, p. 022022.
- ÓTISON. O que são redes neurais? importância e como funciona? 2022. Disponível em: <https://www.opencadd.com.br/blog/o-que-sao-redes-neurais>. Acessado em 05/10/2023.