



UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

DANILO SOUZA DE ASSUNÇÃO

**Classificação automática de espécies de abelhas por meio da morfologia de
suas asas capturadas em imagens digitais utilizando rede neural
convolucional**

São Paulo

2021

DANILO SOUZA DE ASSUNÇÃO

**Classificação automática de espécies de abelhas por meio da morfologia de
suas asas capturadas em imagens digitais utilizando rede neural
convolucional**

Texto de qualificação apresentado à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo como etapa do mestrado no Programa de Pós-graduação em Sistemas de Informação.

Área de concentração: Metodologia e Técnicas da Computação

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em xx de xxxxxxxxxxxxxxxx de xxxx. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: Prof. Dr. Helton Hideraldo Biscaro

Coorientador: Prof. Dr. Luciano Antonio Digiampietri

São Paulo

2021

Dissertação de autoria de Danilo Souza de Assunção, sob o título **“Classificação automática de espécies de abelhas por meio da morfologia de suas asas capturadas em imagens digitais utilizando rede neural convolucional”**, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação, aprovada em ____ de _____ de _____ pela comissão julgadora constituída pelos doutores:

Prof. Dr.
Instituição
Presidente

Prof. Dr.
Instituição

Prof. Dr.
Instituição

Prof. Dr.
Instituição

Prof. Dr.
Instituição

Resumo

ASSUNÇÃO, Danilo. **Classificação automática de espécies de abelhas por meio da morfologia de suas asas capturadas em imagens digitais utilizando rede neural convolucional**. 2021. 64 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2022.

O processo de classificação de espécies de abelhas é uma atividade importante para a preservação das abelhas, pois possibilita o estabelecimento de estratégias mais precisas de conservação ao obter informações detalhadas de uma determinada espécie e também para uma localidade específica. A realização automática da análise visual da morfologia das abelhas se convém devido às despesas necessárias nos processos de identificação manual, e, para esta análise morfológica, os conjuntos de características extraídos a partir das asas têm se mostrado uma eficiente maneira para a identificação das espécies utilizando métodos estatísticos ou computacionais. O aprendizado profundo tem sido amplamente aplicado em atividades relacionadas à visão computacional. Diferente dos métodos tradicionais de aprendizado de máquina, o modelo de aprendizado profundo pode aprender recursos automaticamente a partir de uma grande quantidade de amostras de dados, e não requer a assistência de um especialista no domínio para a extração destas características. No aprendizado profundo, as Redes Neurais Convolucionais (CNN) são bem conhecidas por seu sucesso em muitas tarefas de visão computacional. Neste trabalho será desenvolvido um método utilizando CNN para classificação de espécies de abelhas através da morfologia de suas asas presente em imagens digitais.

Palavras-chaves: Redes Neurais Convolucionais. CNN. Aumento de dados. Aprendizado por transferência. Reconhecimento de espécies. Reconhecimento de espécies de abelhas. Morfologia das asas. Reconhecimento por imagens digitais.

Abstract

ASSUNÇÃO, Danilo. **Classificação automática de espécies de abelhas por meio da morfologia de suas asas capturadas em imagens digitais utilizando rede neural convolucional**. 2021. 64 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2022.

Bees' species classification process turns out to be an important activity in bee preservation, as it makes it possible to establish more precise conservation strategies when obtaining detailed information about a specific species and also for a specific location. The automatic realization of the visual analysis of the morphology of the bees is convenient due to the necessary expenses in the processes of manual identification, and, for this morphological analysis, the sets of characteristics extracted from the wings have shown to be an efficient way for the identification of the species using statistical methods. Deep learning has been widely applied in activities related to computer vision. Unlike traditional machine learning methods, the deep learning model can automatically learn resources from a large number of data samples and does not require the assistance of an expert in the field to extract these characteristics. In deep learning, Convolutional Neural Networks (CNN) are well known for their success in many computer vision tasks. In this work, a method will be developed using CNN to classify bee species through the morphology of their wings present in digital images.

Keywords: Convolutional Neural Networks. CNN. Data augmentation. Transfer learning. Species recognition. Recognition of bee species. Wing morphology. Digital image recognition.

Lista de figuras

Figura 1 – Representação gráfica de subajuste e sobreajuste	20
Figura 2 – Arquitetura de uma rede neural artificial	23
Figura 3 – Constituição macro de uma CNN	24
Figura 4 – Processo da convolução em dados de entrada	25
Figura 5 – Stride 1 no processo convolucional	26
Figura 6 – funções de ativação não lineares	27
Figura 7 – funções de ativação não lineares	28
Figura 8 – funções de ativação não lineares	28
Figura 9 – Método de transferência de características	30
Figura 10 – Método de ajuste fino	31
Figura 11 – Diagrama do processo de revisão sistemática	38
Figura 12 – Ano de publicação dos artigos com base em palavras chaves	45
Figura 13 – Ano de publicação dos artigos selecionados	46
Figura 14 – País de afiliação dos pesquisadores	47
Figura 15 – Técnicas de pré-processamento de imagens aplicadas	48
Figura 16 – Técnicas de Aprendizado por transferência	49
Figura 17 – Técnicas de Aumento de dados	51
Figura 18 – Arquitetura de Redes Neurais Convolucionais	52
Figura 19 – Tipos agrupados de arquiteturas de Redes Neurais Convolucionais	53
Figura 20 – Método de classificação utilizado	54
Figura 21 – Imagem das nervuras destacadas de uma asa de abelha	56

Lista de tabelas

Tabela 1 – Sumário de todos os artigos analisados	41
Tabela 2 – Cronograma	59

Lista de abreviaturas e siglas

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
GAN	Generative Adversarial Networks
MNIST	Modified National Institute of Standards and Technology
RBF	Radius Basis Function
ReLU	Rectified Linear Unit
RGB	Red, Green and Blue
ResNet	Residual Neural Network
SVM	Support Vector Machine
VGG	Visual Geometry Group

Lista de símbolos

A_c	Acurácia
F_1	Medida F
F_n	Falso negativo
F_r	Falso positivo
P_r	Precisão
R_v	Revocação
V_n	Verdadeiro negativo
V_p	Verdadeiro positivo

Sumário

1	Introdução	11
1.1	<i>Hipótese</i>	12
1.2	<i>Objetivos</i>	12
1.3	<i>Justificativa</i>	12
1.4	<i>Estrutura do documento</i>	13
2	Conceitos fundamentais	14
2.1	<i>Visão computacional</i>	14
2.2	<i>Aprendizado de máquina</i>	15
2.2.1	Tarefas padrões em aprendizado de máquina	16
2.2.2	Terminologias	17
2.2.3	Cenários de aprendizagem	18
2.2.4	Sobreajuste e subajuste	19
2.2.5	Avaliação de desempenho de classificadores	20
2.3	<i>Aprendizado profundo</i>	22
2.3.1	Terminologias	22
2.4	<i>Redes neurais artificiais</i>	22
2.5	<i>Redes neurais convolucionais</i>	23
2.5.1	Camada de convolução	24
2.5.2	Função de ativação não linear	26
2.5.3	Camada de agrupamento	27
2.5.4	Camada totalmente conectada	29
2.6	<i>Técnica de aprendizado por transferência</i>	29
2.6.1	Transferência de características	29
2.6.2	Ajuste fino	30
2.7	<i>Técnica de aumento de dados</i>	31
2.7.1	Transformações geométricas	31
2.7.2	Transformações fotométricas	33
2.7.3	Aplicação de máscara	33
2.7.4	Mistura de imagens	34

2.7.5	Apagamento aleatório	34
2.8	<i>Conclusão</i>	34
3	Revisão sistemática	36
3.1	<i>Objetivos da revisão</i>	36
3.2	<i>Protocolo</i>	37
3.3	<i>Resultado da revisão sistemática</i>	40
3.3.1	Ano de publicação	45
3.3.2	País de afiliação dos autores	46
3.3.3	Técnicas de pré-processamento de imagens	47
3.3.4	Técnica de aprendizado por transferência	49
3.3.5	Técnicas de aumento de dados	50
3.3.6	Modelo de arquitetura da CNN	51
3.3.7	Modelo do classificador	53
3.4	<i>Conclusão</i>	54
4	Materiais e métodos	56
4.1	<i>Ferramenta de classificação</i>	57
4.2	<i>Avaliação</i>	57
5	Proposta de pesquisa	58
5.1	<i>Plano de trabalho e cronograma</i>	59
6	Considerações finais	60
	REFERÊNCIAS	61

1 Introdução

As plantações alimentícias possuem uma elevada dependência em relação as abelhas, cuja a responsabilidade de polinização se encontra por volta de 70% (DRAUSCHKE *et al.*, 2007), ou seja, as abelhas economizam um serviço no qual custariam cerca de 65 bilhões de dólares anualmente (PIMENTEL *et al.*, 1997). Sendo elas também responsáveis por um papel de extrema valia para a natureza, onde a preservação de ecossistemas terrestres em conjunto de seu relacionamento com as abelhas (LAWTON, 1998).

Diante disso, o processo de classificação de espécies de abelhas acaba por ser um importante sistema para a sua preservação, por obter informações detalhadas e possibilitando estabelecer estratégias mais precisas de conservação de uma determinada espécie e também para uma localidade específica (GOULSON *et al.*, 2015).

A realização automática da análise visual da morfologia das abelhas se convém devido as despesas necessárias nos processos de identificação manual, e para esta análise morfológica, os conjuntos de características extraídos a partir das asas têm se mostrado uma maneira eficiente para a identificação das espécies utilizando métodos estatísticos (FRANCOY *et al.*, 2008).

O aprendizado profundo tem sido amplamente utilizado em atividades relacionadas à visão computacional. É um ramo do aprendizado de máquina, baseado na aplicação de uma grande quantidade de dados para treinamento de um modelo. Diferente dos métodos tradicionais de aprendizado de máquina, o modelo de aprendizado profundo pode aprender recursos automaticamente a partir de uma grande quantidade de amostras de dados, e não requer a assistência de um especialista no domínio para extração destas características (LIU *et al.*, 2020b). No aprendizado profundo, as redes neurais convolucionais (CNN) são bem conhecidas por seu sucesso em muitas tarefas de visão computacional. (LE *et al.*, 2020).

Devido a inevitável presença de uma grande quantidade de dados para o treinamento deste tipo de rede, é essencial que, para conjuntos de dados menores, se tenha o auxílio de técnicas que permitam atingir bons resultados. Baseado nisso, serão adicionados a este trabalho, mecanismos de aumento de dados ou aprendizado por transferência, afim de se obter uma melhor convergência do modelo por meio da CNN com relação ao conjunto de dados das asas de abelhas disponibilizados para identificação de suas respectivas espécies.

1.1 Hipótese

O uso de CNN para automatização do processo de extração de características e classificação, podem trazer resultados promissores ao problema em questão, devido a sua capacidade de identificar e extrair características com mais qualidade das imagens.

Esta hipótese será testada levando em consideração as medidas obtidas a partir de métricas tradicionais de aprendizado de máquina do modelo proposto, e relacionando com o modelos de *baseline* adquiridos como o apresentado por [Liu et al. \(2020b\)](#).

Espera-se que o modelo proposto apresente resultados superiores aos modelos de *baseline* para as métricas sugeridas.

1.2 Objetivos

Este projeto de pesquisa tem como objetivo o desenvolvimento de uma ferramenta de classificação de espécies de abelhas por meio da morfologia de suas asas, utilizando uma CNN como mecanismo de extração de características e classificação destas imagens, junto à utilização de técnicas de aumento de dados ou aprendizado por transferência para otimização dos resultados da rede, afim de se obter melhores resultados que os modelos de *baseline* definidos. Desta forma, este trabalho tem como objetivos específicos as seguintes entregas:

1. Conjunto de técnicas para aplicação de aumento de dados do conjunto atual disponibilizado;
2. Extrator automático de características das imagens das asas das abelhas;
3. Mecanismo de classificação e validação das espécies de abelhas.

1.3 Justificativa

Uma vez que o conjunto demonstre cenários de sucesso com base nas avaliações aplicadas, tornará-se evidente que a aplicação de uma CNN poderá atingir um modelo mais próximo do estado da arte para o problema abordado.

1.4 Estrutura do documento

O restante deste documento está organizado da seguinte forma. O capítulo 2 descreve os conceitos fundamentais para o entendimento do trabalho. O capítulo 3 apresenta a revisão de literatura realizada. O capítulo 4 descreve o conjunto de dados e as técnicas a serem utilizadas. O capítulo 5 aborda a proposta de pesquisa. Por fim, o capítulo 6 contém as considerações finais acerca do projeto.

2 Conceitos fundamentais

Neste capítulo são apontados alguns tópicos elementares para a compreensão deste projeto. Dessa forma, primeiro é realizada uma curta introdução ao tema de visão computacional, logo depois são apresentados as definições referentes à aprendizado de máquina, aprendizado profundo, redes neurais artificiais, redes neurais convolucionais (CNN), técnicas de aprendizado por transferência e aumento de dados.

2.1 Visão computacional

A visão computacional é a ciência responsável pela visão de uma máquina, diante da forma como um computador visualiza o ambiente à sua volta, extraindo informações significativas a partir de imagens capturadas por dispositivos eletrônicos como câmeras digitais.

Neste contexto, uma imagem pode ser definida como um conjunto de pontos, chamados *pixels*, que são a menor parte da imagem. Cada *pixel* possui uma cor e uma coordenada na imagem. A resolução da imagem corresponde a quantidade de pixel sem cada uma de suas dimensões. Ou seja, uma imagem com resolução de 1024 por 768, comumente representada como “1024 x 768”, tem 1024 *pixels* no eixo horizontal e 768 na vertical.

Com o uso da visão computacional, é possível reconhecer, manipular e refletir sobre os componentes que compõem uma imagem. Dessa forma, a visão computacional fornece ao computador uma infinidade de informações precisas a partir de imagens e vídeos, de forma que o computador consiga executar determinadas tarefas ao simular o comportamento visual do ser humano. Para um modelo de visão computacional, tem-se a composição de determinadas etapas, sendo elas, aquisição de imagem, pré-processamento, segmentação, extração de características e classificação:

- Aquisição de imagem: trata-se do processo de obtenção de uma ou várias imagens a partir de dispositivos eletrônicos como câmeras digitais ou escâneres, onde os *pixels* em cada imagem obtida (bidimensional, tridimensional ou uma sequência de imagens) indica suas características de propriedades físicas capturadas.

- **Pré-processamento:** é um processo realizado para ajuste ou normalização de imagens, além de preparar tais imagens para a utilização em algoritmos de aprendizado de máquina, tendo como exemplo, destaques de contornos, bordas, destaques de figuras geométricas, remoção de ruídos, normalização de tom da imagem, redução dos canais de cores para o tom de cinza, alteração do brilho e contraste e dentre outras técnicas.
- **Segmentação:** é um mecanismo que permite dividir em segmentos uma imagem, destacando objetos e alterando a representação da imagem, para que o processo de reconhecimento de padrões seja realizado com mais acurácia ou até mesmo para que este processo de reconhecimento de padrões seja possível.
- **Extração de características:** é um processo que constitui a extração das composições de uma imagem, obtendo matematicamente tais características como textura, bordas ou formatos.
- **Classificação:** é o método que inclui a validação da satisfação dos dados obtidos, a estimativa de parâmetros sobre a imagem e a classificação dos objetos obtidos em diferentes categorias.

2.2 *Aprendizado de máquina*

Dentre as definições presentes na literatura, [Mohri, Rostamizadeh e Talwalkar \(2018\)](#) diz que, o aprendizado de máquina pode ser amplamente definido como métodos computacionais, usando a experiência para melhorar o desempenho ou realizar previsões precisas. A experiência se refere às informações anteriores disponíveis para o modelo de aprendizagem de máquina, que normalmente assumem a forma de dados eletrônicos coletados e disponibilizados para análise. Esses dados podem estar na forma de conjuntos de treinamento com identificação humana ou a diferentes tipos de informações obtidas por meio da interação com o ambiente. Em todos os casos, sua qualidade e tamanho são cruciais para o sucesso das previsões feitas pelo modelo.

O aprendizado de máquina consiste em projetar algoritmos de previsão eficientes e precisos ([MOHRI; ROSTAMIZADEH; TALWALKAR, 2018](#)). Como em outras áreas da ciência da computação, algumas medidas críticas sobre a qualidade desses algoritmos são a sua complexidade de tempo e espaço ([MOHRI; ROSTAMIZADEH; TALWALKAR, 2018](#)).

Como o sucesso de um algoritmo de aprendizado depende dos dados usados, o aprendizado de máquina está inerentemente relacionado à análise de dados e estatísticas (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). De maneira mais geral, as técnicas de aprendizagem são métodos orientados por dados que combinam conceitos fundamentais em ciência da computação com ideias de estatística, probabilidade e otimização (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018).

2.2.1 Tarefas padrões em aprendizado de máquina

A seguir estão algumas tarefas comumente aplicadas no campo de aprendizado de máquina:

- Classificação (*Classification*): trata-se do problema de atribuir uma categoria a cada item (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). Por exemplo, a classificação de imagens consiste em atribuir a cada imagem uma categoria como carro, navio ou avião. O número de categorias em tais tarefas costuma ser menor do que algumas centenas, mas pode ser muito maior em algumas tarefas difíceis, e até mesmo ilimitado, como em classificação de texto ou reconhecimento de fala (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018).
- Regressão (*Regression*): é a atividade de solucionar problemas já previstos a partir de dados históricos de cada item (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). Exemplos de regressão incluem a previsão de valores de estoque ou variações de variáveis econômicas. Na regressão, a penalidade para uma previsão incorreta depende da magnitude da diferença entre os valores verdadeiros e previstos, visto que, em contraste com o problema de classificação, onde normalmente não há noção de proximidade entre as várias categorias (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018).
- Ranqueamento (*Ranking*): proporciona à resolução do problema de aprender a ordenar os itens de acordo com algum critério (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). A pesquisa na web, por exemplo, ao retornar páginas da web relevantes para uma consulta de pesquisa, é o exemplo de classificação canônica. Muitos outros problemas de classificação semelhantes, surgem no contexto do projeto de

extração de informações ou sistemas de processamento de linguagem natural (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018).

- Clusterização (*Clustering*): é uma forma de solucionar o problema de particionar um conjunto de itens em subconjuntos homogêneos (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). A clusterização é frequentemente utilizada para analisar conjuntos de dados muito grandes (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). Por exemplo, no contexto da análise de rede social, os algoritmos de agrupamento tentam identificar comunidades naturais dentro de grandes grupos de pessoas (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018).
- Redução da dimensionalidade (*Dimensionality reduction*): consiste em transformar uma representação inicial de itens em uma representação de dimensão inferior, preservando algumas propriedades da representação inicial (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). Um exemplo comum envolve o pré-processamento de imagens digitais em tarefas de visão computacional.

2.2.2 Terminologias

A seguir está uma lista de definições e terminologias comumente usados no escopo de aprendizado de máquina:

- Amostras (*Examples*): itens ou instâncias de dados usados para aprendizagem ou avaliação em um modelo de aprendizado de máquina (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018).
- Características (*Features*): O conjunto de atributos, frequentemente representado como um vetor, associado a uma amostra (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). No caso de uma imagem, alguns recursos relevantes podem incluir a forma do objeto presente nesta imagem, cor e dentre vários outros aspectos.
- Rótulo ou Classe (*Label*): valores ou categorias atribuídos a exemplos (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). Em problemas de classificação, as amostras são atribuídas a categorias específicas. Por exemplo, dado imagens de objetos de veículos, se tem categorias de carro, avião ou navio, num cenário onde é tido um problema de classificação com múltiplas classes.

- Hiperparâmetros (*Hyperparameters*): parâmetros livres que não são determinados pelo algoritmo de aprendizagem, mas sim especificados como entrada para o algoritmo de aprendizagem (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018).
- Amostra de treinamento (*Training sample*): são amostras utilizadas para treinar um algoritmo de aprendizagem (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). Em um problema envolvendo reconhecimento de padrões por meio de imagens digitais, a amostra de treinamento consiste em um conjunto de exemplos de imagens junto a suas classes associadas.
- Amostra de validação (*Validation sample*): são exemplos usados para ajustar os parâmetros de um algoritmo de aprendizagem ao trabalhar com dados rotulados (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). A amostra de validação é usada para selecionar valores apropriados para os parâmetros livres do algoritmo de aprendizagem (hiperparâmetros) (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018).
- Amostra de teste (*Test sample*): são exemplos usados para avaliar o desempenho de um algoritmo de aprendizagem (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). A amostra de teste é separada dos dados de treinamento e validação, não sendo disponibilizada na fase de aprendizado (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018). Por exemplo, em um cenário de reconhecimento facial, a amostra de teste consiste em uma coleção de exemplos de imagens de rostos, para os quais o algoritmo de aprendizagem deve prever suas determinadas classes com base em recursos, sendo então comparadas com os rótulos da amostra de teste para medir o desempenho do algoritmo.
- Função de perda (*Loss function*): é um tipo de função que mede a diferença, ou perda, entre uma classe prevista e uma classe verdadeira de uma determinada amostra (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018).

2.2.3 Cenários de aprendizagem

Nesta subseção, será descrito alguns cenários de aprendizado de máquina que podem ser interpretados como tipos de treinamento. Esses cenários diferem nos tipos de dados de treinamento disponíveis para o modelo de aprendizagem, na ordem e no método pelo

qual os dados de treinamento são recebidos, e nos dados de teste usados para avaliar o algoritmo de aprendizagem.

- Aprendizagem supervisionada (*Supervised learning*): aprendizagem supervisionada é a aprendizagem por meio de entradas pré-rotuladas, atuando como alvos, e, tem como objetivo, a utilização de treinamentos para a redução de erros de classificação dos modelos, fazendo uso do cálculo correto do valor de saída do treinamento (O'SHEA; NASH, 2015). Haverá um conjunto de valores de entrada (vetores) para cada modelo de treinamento, e um ou mais valores de saída associados (O'SHEA; NASH, 2015). O objetivo desta forma de treinamento é reduzir o erro geral de classificação dos modelos, por meio do cálculo correto do valor de saída do treinamento (O'SHEA; NASH, 2015).
- Aprendizagem não supervisionada (*Unsupervised learning*): a aprendizagem não supervisionada diverge do modelo supervisionado devido ao conjunto de treinamento não incluir nenhuma classe (O'SHEA; NASH, 2015). A capacidade da rede em reduzir ou aumentar uma função de custo associada são as determinantes para o sucesso (O'SHEA; NASH, 2015). No entanto, nota-se que a grande parte das tarefas de reconhecimentos de padrões com foco em imagens, geralmente dependem de classificação usando aprendizado supervisionado (O'SHEA; NASH, 2015).

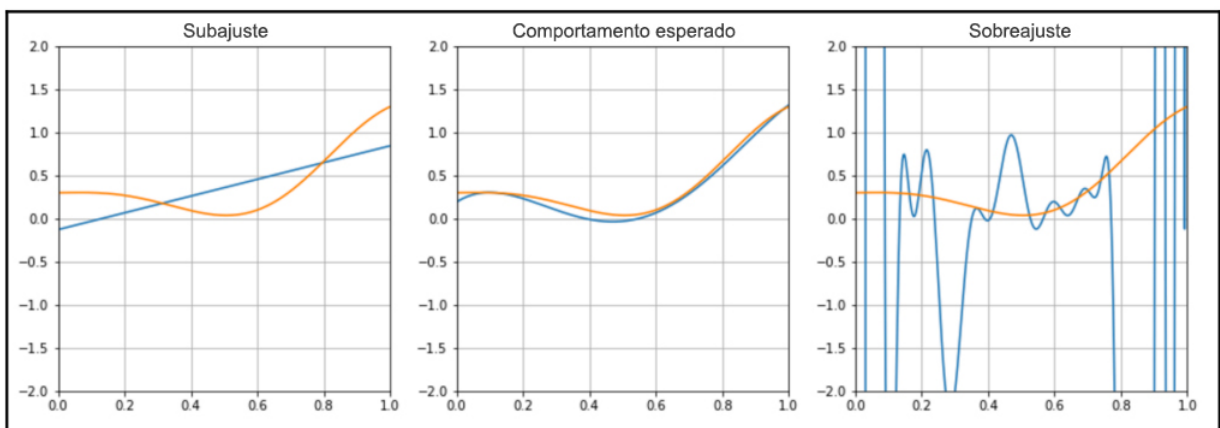
2.2.4 Sobreajuste e subajuste

O objetivo de um modelo de aprendizado de máquina é aproximar uma função desconhecida que associe elementos de entrada aos de saída (para um classificador, os chamamos de classes) (BONACCORSO, 2017). No entanto, um conjunto de treinamento é normalmente uma representação de uma distribuição global, sendo que, esta distribuição não pode conter todos os elementos possíveis, caso contrário, o problema poderia ser resolvido com uma associação de um para um (BONACCORSO, 2017). Portanto, ao treinar o modelo, é necessário pensar em ajustá-lo, mas mantendo-o livre para generalizar quando uma entrada desconhecida é apresentada. Infelizmente, essa condição ideal nem sempre é fácil de se alcançar, e este comportamento ocorre devido à duas características eminentes que são:

- Sobreajuste (*Overfitting*): o modelo não é mais capaz de generalizar, considerando a dinâmica original fornecida pelo conjunto de treinamento (BONACCORSO, 2017). Ele pode associar quase perfeitamente todas as amostras conhecidas aos valores de saída correspondentes, mas quando uma entrada desconhecida é apresentada, o erro de predição correspondente pode ser muito alto (BONACCORSO, 2017).
- Subajuste (*Underfitting*): isso significa que o modelo não é capaz de capturar a dinâmica mostrada pelo mesmo conjunto de treinamento (provavelmente porque sua capacidade é muito limitada) (BONACCORSO, 2017).

Na Figura 1 é possível observar os efeitos negativos causados pelo subajuste e sobreajuste. No gráfico da esquerda o efeito de subajuste mostra a incapacidade do modelo em capturar a dinâmica presente no modelo de treinamento, obtendo resultados ruins já neste procedimento. No gráfico localizado ao meio, temos um cenário adequado onde o modelo foi capaz de generalizar totalmente a distribuição apresentada. Por fim, no gráfico à direita, temos um cenário de sobreajuste, onde é possível verificar que, em determinados momentos ele é capaz de se adaptar, indicando assim os valores do conjunto de treinamento fornecido, mas para outros cenários do quadro de distribuição, ele não possui a mesma capacidade de identificação.

Figura 1 – Representação gráfica de subajuste e sobreajuste



Fonte – Bonaccorso (2017)

2.2.5 Avaliação de desempenho de classificadores

Ao categorizar uma amostra baseada em um algoritmo de classificação, pode-se determinar que tal amostra pertence a um dos quatro itens seguintes:

- Verdadeiro positivo (V_p)
- Verdadeiro negativo (V_n)
- Falso positivo (F_p)
- Falso negativo (F_n)

Tendo como exemplo, um Verdadeiro positivo entraria para uma classe determinada pelo classificador, identificando-a como pertencente à classe C. Em sequência, Verdadeiro negativo será o resultado do processo inverso, onde classificador identificaria a amostra como sendo não pertencente à classe C. O falso positivo se dá quando ocorre um erro pelo classificador, onde a identificação de uma amostra é determinada como sendo da classe C, sendo que ela não é. Em suma, o falso negativo são amostras nas quais o classificador não pode identificar, ou então as classificou como pertencente à outra classe.

Para fins de medição de qualidade do modelo de aprendizado de máquina, é utilizado funções de avaliação de classificadores. As funções mais usuais entre elas são:

- Precisão (P_r)
- Revocação (R_v)
- Medida F (F_1)
- Acurácia (A_c)

Precisão (P_r) é a parcela de amostras nas quais obtiveram uma identificação e classificação corretamente. Podendo ser descrita a partir de seu cálculo pela Fórmula 1.

$$P_r = \frac{V_p}{V_p + F_p} \quad (1)$$

Revocação (R_v) é a parcela de amostras com identificação correta, onde pertencem a uma determinada classe entre todas as amostra desta mesma classe. Seu cálculo pode ser descrito a partir da Fórmula 2.

$$R_v = \frac{V_p}{V_p + F_n} \quad (2)$$

Medida F (F_1) é a busca da média harmônica entre a precisão e a revocação. O valor da medição torna-se ideal quanto mais próximo de 0,5, enquanto que, quanto mais próximo de 0, expressa um resultado ruim. Sua equação pode ser descrita pela Fórmula 3.

$$F_1 = 2 \frac{P_r R_v}{P_r + R_v} \quad (3)$$

Acurácia (A_c) é a classificação correta sobre as parcelas de amostras, se elas pertencem ou não a uma classe, diante de todas as amostras testadas. Denominando assim a porcentagem de classificação correta das amostras. Sua equação pode ser descrita pela Fórmula 4.

$$A_c = \frac{V_p + V_n}{V_p + V_n + F_p + F_n} \quad (4)$$

2.3 Aprendizado profundo

Dentre as diversas definições presentes, no artigo de [LeCun, Bengio e Hinton \(2015\)](#), o aprendizado profundo é dito como um conjunto de algoritmos que modelam abstrações de alto nível de dados usando uma hierarquia com várias camadas de processamento, sendo esta uma forma de aprendizado de máquina. Como o computador reúne conhecimento com base na experiência, não há necessidade de um operador humano especificar todo o conhecimento necessário ao computador.

2.3.1 Terminologias

A seguir está uma lista de definições e terminologias utilizadas no escopo de aprendizado profundo:

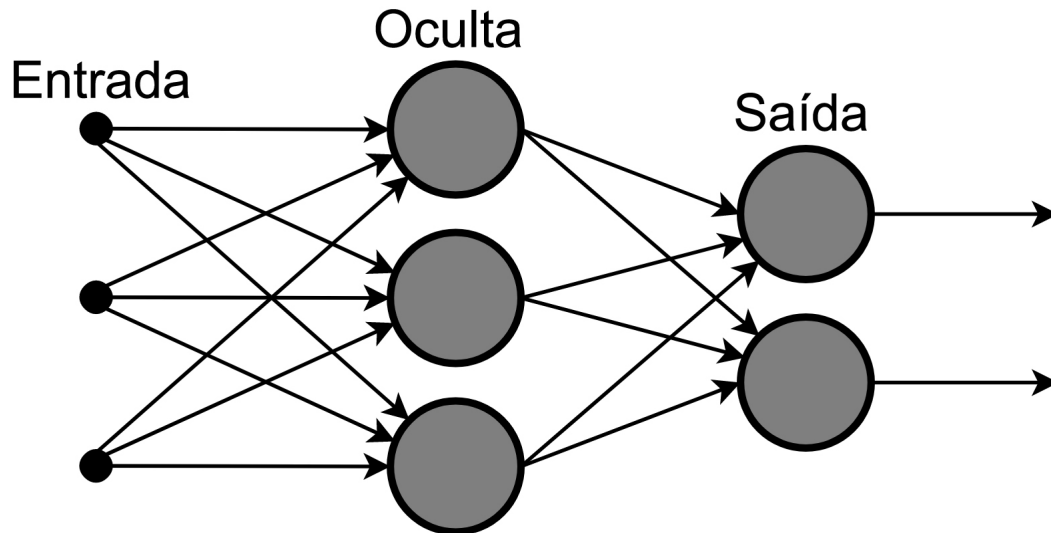
- Tensor: Um *tensor* é uma forma de representar os dados no aprendizado profundo. Este pode ser um vetor (*array*) unidimensional, bidimensional, tridimensional e dentre outras dimensões ([KOLDA; BADER, 2009](#)).

2.4 Redes neurais artificiais

Dentre as diversas definições correntes na literatura, [O'Shea e Nash \(2015\)](#) afirma que, as redes neurais artificiais (*artificial neural networks* - ANNs) são sistemas de processamento computacional que são fortemente inspirados pelo modo como os sistemas nervosos biológicos (como o cérebro humano) operam. As ANNs são compostas principalmente por um grande número de nós computacionais interconectados (referidos como neurônios), dos quais trabalham entrelaçados de forma distribuída para aprender coletivamente a partir

da entrada, a fim de otimizar sua saída final. A estrutura básica de uma ANN pode ser modelada como mostrado na Figura 2.

Figura 2 – Arquitetura de uma rede neural artificial



Fonte – https://en.wikipedia.org/wiki/Artificial_neural_network

Na Figura 2, é introduzido uma rede neural *feedforward* simples de três camadas, composta por uma camada de entrada (*input*), uma camada oculta (*hidden*) e uma camada de saída (*output*). Essa estrutura é a base de uma série de arquiteturas comuns de ANN (O'SHEA; NASH, 2015).

A entrada é carregada geralmente na forma de um vetor multidimensional que será distribuída para as camadas ocultas da rede (O'SHEA; NASH, 2015). As camadas ocultas tomarão decisões a partir da camada anterior e avaliarão como uma mudança estocástica em si mesma avaliando se o resultado prejudica ou melhora o valor final obtido, e isso é conhecido como o processo de aprendizagem (O'SHEA; NASH, 2015).

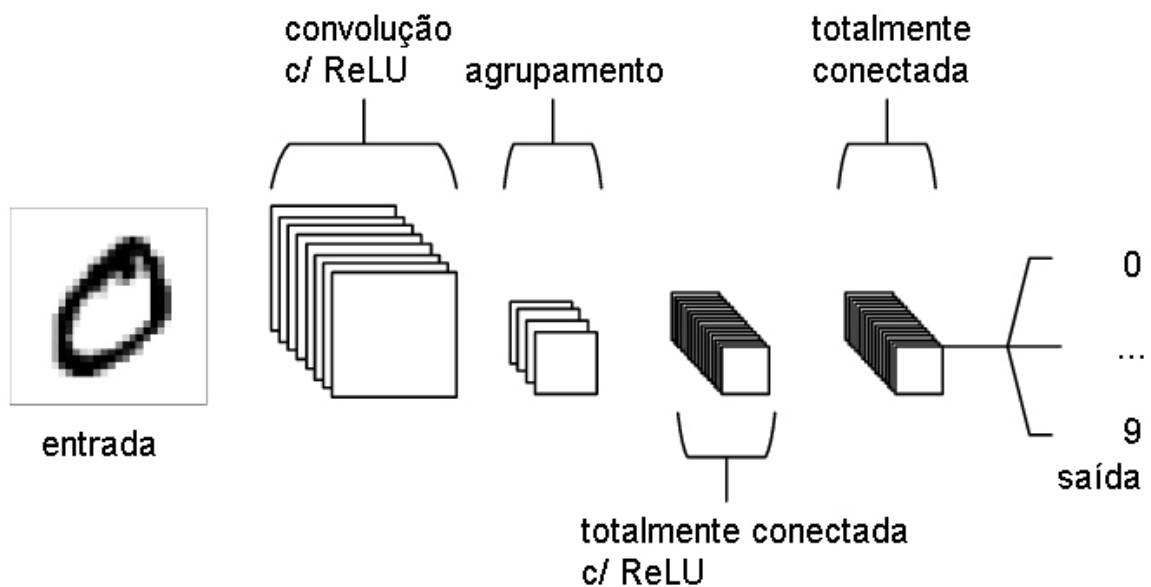
Existem diversas ramificações para determinação de um modelo de classificação para as redes neurais artificiais, normalmente os dois mais utilizados são aprendizado supervisionado e aprendizado não supervisionado.

2.5 Redes neurais convolucionais

Segundo (YAMASHITA *et al.*, 2018), as redes neurais convolucionais (CNN) são um tipo de modelo de aprendizado profundo para processamento de dados, inspiradas na

organização do córtex visual animal e projetada para aprender de forma automática e adaptativa, hierarquias espaciais de recursos disponíveis em imagens digitais de padrões de baixo a alto nível. A CNN normalmente é composta de três tipos de camadas, convolução (*convolutional layer*), agrupamento (*pooling layer*) e camadas totalmente conectadas (*fully-connected layer*) conforme o exemplo apresentado na Figura 3.

Figura 3 – Constituição macro de uma CNN



Fonte – O'Shea e Nash (2015)

As camadas de convolução e agrupamento executam a extração de recursos, enquanto a camada totalmente conectada mapeia os recursos extraídos na saída final, como a classificação.

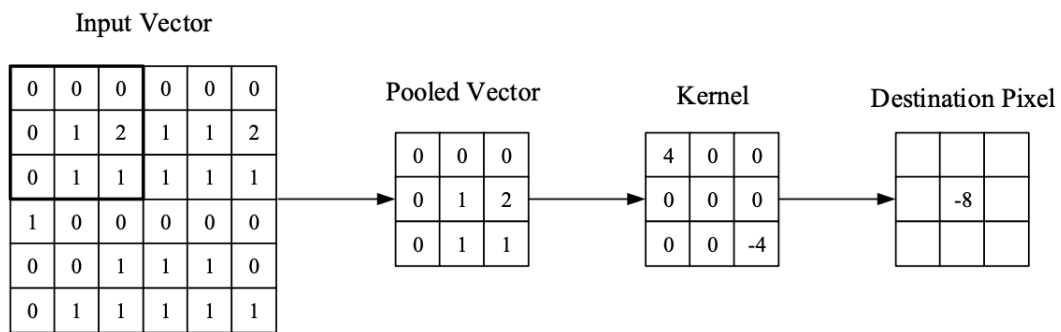
2.5.1 Camada de convolução

A camada de convolução (*convolutional layer*) é um componente fundamental da arquitetura de uma CNN, tendo como objetivo a extração de recursos de uma imagem digital, que consiste normalmente em uma combinação de operações lineares e não lineares, ou seja, operação de convolução e função de ativação (YAMASHITA *et al.*, 2018).

Convolução é um tipo especializado de operação linear, usado para o processo de extração de recursos onde um filtro (*kernel*) é aplicada na entrada que é uma matriz

de números, chamada de *tensor* (O'SHEA; NASH, 2015). Estes filtros são geralmente pequenos em dimensionalidade espacial, mas se espalham ao longo de toda a profundidade da entrada (O'SHEA; NASH, 2015). Quando os dados atingem uma camada convolucional, a camada convolve cada filtro na dimensionalidade espacial da entrada, para assim produzir um mapa de ativação 2D (bidimensional) (O'SHEA; NASH, 2015). Conforme navegamos pela entrada, o produto escalar é calculado para cada valor naquele filtro, como é mostrado na Figura 4. A partir disso, a rede aprenderá os filtros que são excitados quando veem um recurso específico em uma determinada posição espacial da entrada. Esses são comumente conhecidos como ativações (O'SHEA; NASH, 2015).

Figura 4 – Processo da convolução em dados de entrada



Fonte – O'Shea e Nash (2015)

A operação de convolução descrita acima, não permite que o centro de cada filtro se sobreponha ao elemento mais externo do *tensor* de entrada, e reduz a altura e largura do mapa de recursos de saída em comparação com o *tensor* de entrada (YAMASHITA *et al.*, 2018).

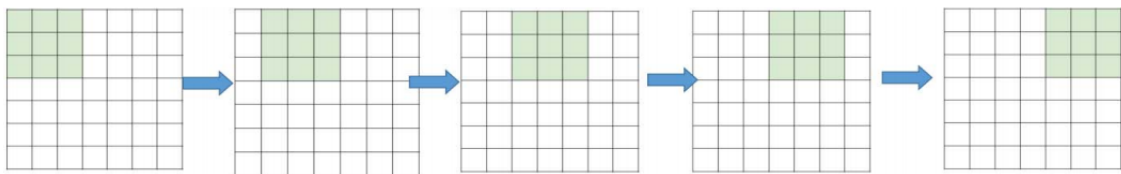
Camadas convolucionais também são capazes de reduzir significativamente a complexidade do modelo por meio da otimização de sua saída. Eles são otimizados por meio de três hiperparâmetros, a profundidade (*depth*), a passada (*stride*) e a configuração de preenchimento (*padding*) (O'SHEA; NASH, 2015).

A profundidade (*depth*) do volume de saída produzido pelas camadas convolucionais pode ser ajustada manualmente através do número de neurônios dentro da camada para a mesma região da entrada (O'SHEA; NASH, 2015). Isso pode ser visto com outras formas de ANNs, onde todos os neurônios na camada oculta são conectados diretamente a cada neurônio de antemão (O'SHEA; NASH, 2015). A redução desse hiperparâmetro pode minimizar significativamente o número total de neurônios da rede, mas também pode

reduzir significativamente os recursos de reconhecimento de padrão do modelo (O'SHEA; NASH, 2015).

A passada (*stride*) serve para definir a profundidade em torno da dimensionalidade espacial da entrada, e posicionar o campo receptivo (O'SHEA; NASH, 2015). Ou seja, se definíssemos uma passada como 1, conforme exibido na Figura 5, teríamos um forte campo receptivo sobreposto, produzindo ativações extremamente grandes (O'SHEA; NASH, 2015). Alternativamente, definir a passada para um número maior, reduzirá a quantidade de sobreposição e produzirá uma saída de dimensões espaciais menores (O'SHEA; NASH, 2015).

Figura 5 – Stride 1 no processo convolucional



Fonte – Albawi, Mohammed e Al-Zawi (2017)

Preenchimento (*Padding*), usualmente preenchimento de zero (*zero padding*), é uma técnica onde linhas e colunas de zeros são adicionadas em cada lado do tensor de entrada, de modo a ajustar o centro de um filtro no elemento mais externo, e mantendo a mesma dimensão no plano através da operação de convolução (YAMASHITA *et al.*, 2018). As arquiteturas modernas de CNN geralmente empregam preenchimento zero, para assim reter as dimensões no plano e aplicar mais camadas (YAMASHITA *et al.*, 2018). Sem o preenchimento de zero, cada mapa de recurso sucessivo ficaria menor após a operação de convolução (YAMASHITA *et al.*, 2018).

2.5.2 Função de ativação não linear

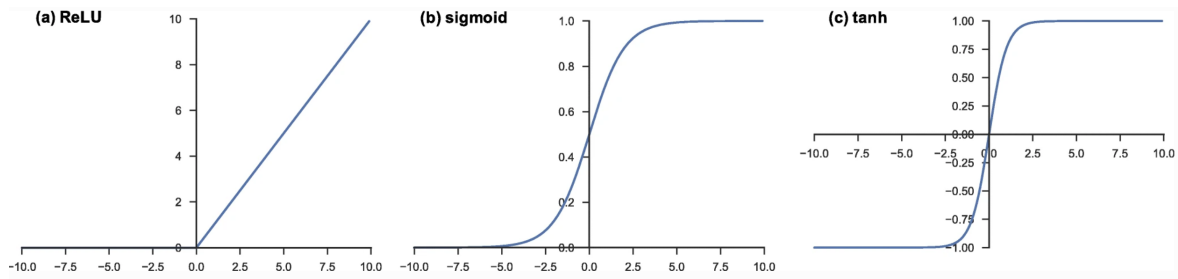
As saídas de uma operação linear, como a convolução, são passadas por uma função de ativação não linear (ALBAWI; MOHAMMED; AL-ZAWI, 2017). Embora funções não lineares, como a função sigmoide ou tangente hiperbólica (*tanh*), foram usadas anteriormente por serem representações matemáticas de um comportamento de neurônio biológico, a função de ativação mais comum usada atualmente é a unidade linear retificada

(ReLU), devido sua facilidade em computar sua derivadas. A ReLU é expressada pela seguinte Fórmula 5:

$$ReLU(x) = \max(0, x) \quad (5)$$

Ambas as funções não lineares citadas podem ser melhor observadas a partir da Figura 6, onde é possível observar suas composições do ponto de vista gráfico.

Figura 6 – funções de ativação não lineares



Fonte – Yamashita *et al.* (2018)

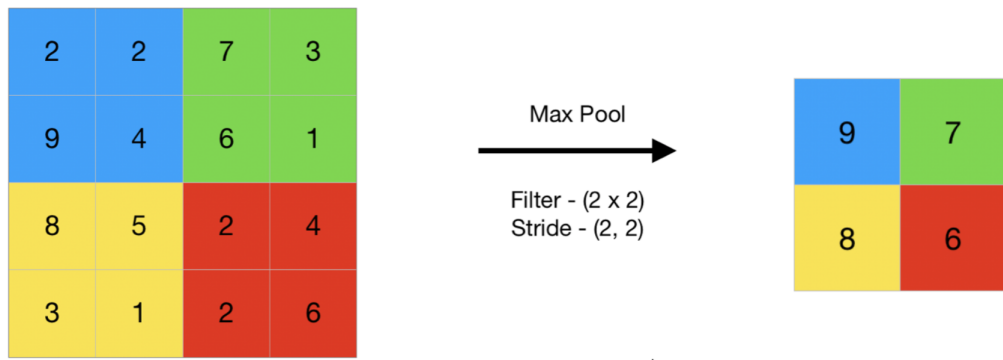
2.5.3 Camada de agrupamento

O objetivo da camada de agrupamento (*pooling layer*) é transformar a representação de característica conjunta em uma mais utilizável, que preserve informações importantes enquanto descarta detalhes irrelevantes (YU *et al.*, 2014). Ou seja, a camada de agrupamento visa reduzir gradualmente a dimensionalidade da representação e, assim, reduzir ainda mais o número de parâmetros e complexidade computacional do modelo (O'SHEA; NASH, 2015). A operação oferecida pelo agrupamento é tipicamente conhecido como *downsampling*. Devido a natureza presente no processo de agrupamento, um filtro com tamanho acima de 3 geralmente reduzirá muito o desempenho do modelo, portanto, na maioria das CNNs, filtros são configurados com dimensionalidade de 2 x 2 com passada de 2 na camada de agrupamento, o que permitirá que a camada se estenda por toda a dimensionalidade espacial da entrada recebida, reduzindo o mapa de ativação para 25% do tamanho original enquanto é mantido o volume de profundidade em seu tamanho padrão (YAMASHITA *et al.*, 2018). É importante notar que não há nenhum parâmetro aprendível em nenhuma das camadas de agrupamento, enquanto o tamanho do filtro, a passada (*stride*) e o preenchimento (*padding*) são também hiperparâmetros em operações de agrupamento, semelhantes às operações de convolução (YAMASHITA *et al.*, 2018). A

camada de agrupamento ajusta sua dimensionalidade usando uma função de agrupamento, podendo ser funções como a *Max Pooling* ou *Average Pooling*.

Max Pooling é a forma mais popular de operação de agrupamento, que extrai *patches* dos mapas de recursos de entrada, produz o valor máximo em cada *patch* e descarta todos os outros valores (YAMASHITA *et al.*, 2018). O *Max Pooling* com um filtro de tamanho 2×2 com uma passada de 2 é comumente usado na prática. Isso diminui a resolução da dimensão no plano dos mapas de recursos por um fator de 2, e, ao contrário da altura e da largura, a dimensão de profundidade dos mapas de recursos permanece inalterada (YAMASHITA *et al.*, 2018). Este método pode ser observado pela Figura 7.

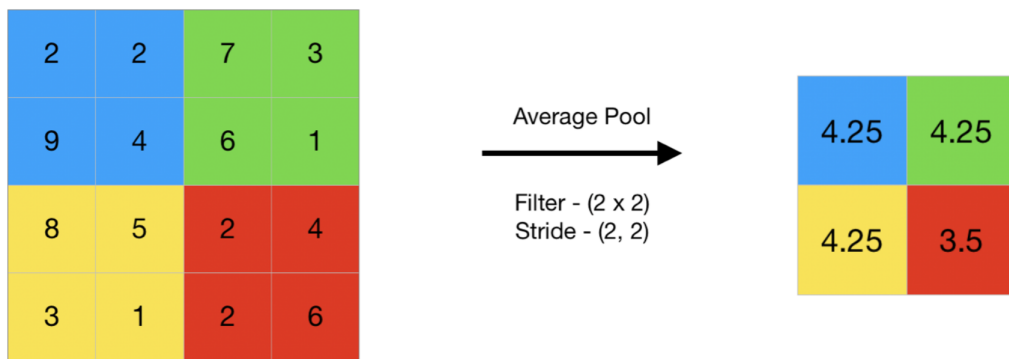
Figura 7 – funções de ativação não lineares



Fonte – Danilo Assunção, 2021

O *Average Pooling* extrai *patches* dos mapas de recursos de entrada e calcula a média destes elementos. Assim, enquanto o *Max Pooling* fornece o recurso mais proeminente em um *patch* específico do mapa de recursos, o *Average Pooling* fornece a média de recursos presentes em um *patch*, conforme a Figura 8.

Figura 8 – funções de ativação não lineares



Fonte – Danilo Assunção, 2021

2.5.4 Camada totalmente conectada

Os mapas de recursos de saída da convolução final ou camada de agrupamento, são normalmente achatados (*flattened*), ou seja, transformados em uma matriz unidimensional de números (ou vetor) e conectados a uma ou mais camadas totalmente conectadas, também conhecidas como camadas densas, em que cada entrada é conectada a cada saída por um peso aprendível (YAMASHITA *et al.*, 2018). Ao passar pelas camadas de convolução e agrupamento, o conteúdo resultando deste processo anterior é mapeado por um subconjunto de camadas totalmente conectadas para as saídas finais da rede (YAMASHITA *et al.*, 2018). A camada final totalmente conectada normalmente tem o mesmo número de nós de saída que o número de classes. Cada camada totalmente conectada é seguida por uma função não linear (YAMASHITA *et al.*, 2018).

2.6 Técnica de aprendizado por transferência

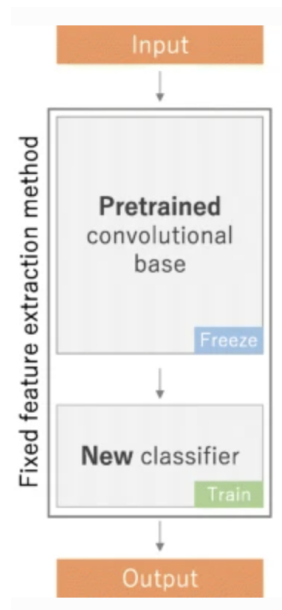
Segundo Shao, Zhu e Li (2014), a técnica de aprendizado por transferência (*transfer learning*) é um paradigma para evitar sobreajuste. O aprendizado por transferência funciona treinando uma rede em um elevado conjunto de dados, como *ImageNet*, para então usar tais pesos como os pesos iniciais em uma nova tarefa de classificação. Normalmente, apenas os pesos nas camadas convolucionais da CNN pré-treinada são copiados, em vez de toda a rede. Sua eficácia é notada pois muitos conjuntos de dados de imagem compartilham características espaciais de baixo nível, que são melhor aprendidas com grandes quantidades de dados. O aprendizado por transferência possui dois métodos aplicáveis disponíveis na literatura, sendo eles, a transferência de características (*feature learning*) e o ajuste fino (*fine-tuning*).

2.6.1 Transferência de características

O método de transferência de características (*feature learning*) é um processo de remoção das camadas totalmente conectadas de uma rede pré-treinada e ao mesmo tempo manter o restante da rede, cuja composição deste restante se dá em uma série de camadas de convolução e agrupamento, denominada base convolucional, funcionando como um

extrator de recursos fixo (LI; HOIEM, 2017). Nesse cenário, qualquer classificador de aprendizado de máquina podem ser adicionados no topo do extrator de recursos fixo, como exemplo as máquinas de vetores de suporte (SVM), florestas aleatórias (*random forest*) e bem como as camadas totalmente conectadas usuais em CNNs, resultando em um processo de treinamento focado no classificador adicionado em um determinado conjunto de dados de interesse (LI; HOIEM, 2017). Este mecanismo pode ser visto na Figura 9.

Figura 9 – Método de transferência de características

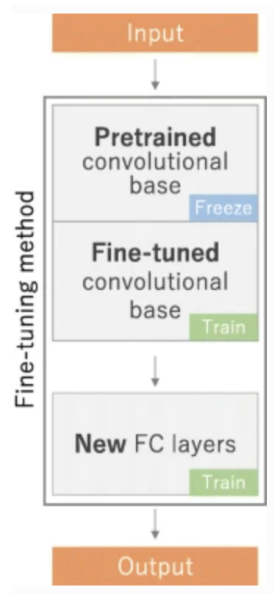


Fonte – Danilo Assunção, 2021

2.6.2 Ajuste fino

O método de ajuste fino substitui as camadas totalmente conectadas do modelo pré-treinado por um novo conjunto de camadas totalmente conectadas para retreinar em um determinado conjunto de dados, aplicando também o ajuste fino como um todo ou apenas partes dos componentes na base convolucional pré-treinada (LI; HOIEM, 2017). Todas as camadas na base convolucional podem ser ajustadas ou, alternativamente, algumas camadas anteriores podem ser corrigidas durante o ajuste fino do resto das camadas mais profundas (LI; HOIEM, 2017). Isso é motivado pela observação de que os recursos da camada inicial terem um fator maior de generalização, enquanto os recursos posteriores tornam-se progressivamente mais específicos para um determinado conjunto de dados (LI; HOIEM, 2017). Este mecanismo pode ser visto na Figura 10.

Figura 10 – Método de ajuste fino



Fonte – Danilo Assunção, 2021

2.7 Técnica de aumento de dados

Segundo [Shorten e Khoshgoftaar \(2019\)](#), a técnica de aumento de dados (*data augmentation*) abrange um conjunto de métodos que aumentam o tamanho dos conjuntos de dados de treinamento, nessa conformidade, melhores modelos de aprendizado profundo podem ser construídos com eles devido a sua capacidade em minimizar o sobreajuste. Os algoritmos de aumento de dados incluem transformações geométricas (*geometric transformations*), modificação de espaço de cores (*color space augmentations*), aplicação de máscara (*kernel filters*), misturas de imagens (*mixing images*), apagamento aleatório (*random erasing*).

2.7.1 Transformações geométricas

Esta subseção descreve as diferentes metodologias com base em transformações geométricas aplicadas em imagens. Os tipos descritos abaixo podem ser caracterizados por sua facilidade de implementação ([SHORTEN; KHOSHGOFTAAR, 2019](#)). Estes mecanismos foram divididos em 6 partes principais:

1. Inversão (*Flipping*): o método de inversão consiste em inverter o eixo horizontal ou vertical de uma imagem, onde a inversão do eixo horizontal costuma ser mais

- frequente do que a inversão do eixo vertical (SHORTEN; KHOSHGOFTAAR, 2019). Esta técnica é uma das mais simples de se implementar, e provou ser útil em determinados conjuntos de dados de imagens (SHORTEN; KHOSHGOFTAAR, 2019).
2. Espaço dos canais de cores (*Color space*): os dados de uma imagem digital são geralmente codificados como um tensor tridimensional (altura \times largura \times canais de cor), viabilizando a implementação de técnicas de aumento de dados no espaço de canais de cores (SHORTEN; KHOSHGOFTAAR, 2019). O método de espaço de cores inclui o isolamento de um simples canal de cor, tal como R, G ou B, sendo que, a imagem pode ter sua representação de cores rapidamente convertida ao isolar um canal de cor, e adicionando duas matrizes de zero nos demais canais (SHORTEN; KHOSHGOFTAAR, 2019). Adicionalmente, os valores nos canais de cores da imagem (RGB) podem ser manipulados com simples operações de matriz, ao aumentar ou diminuir o brilho da imagem.
 3. Corte (*Cropping*): cortar imagens é uma prática de processamento onde áreas relevantes da imagem são destacadas e redimensionadas (SHORTEN; KHOSHGOFTAAR, 2019). Por exemplo, uma imagem de dimensão 256 x 256 onde, após o corte, ficou com dimensão 224 x 224. Dependendo do limite de redução escolhido para o corte, pode ser uma transformação que não preserve a classe que representa a imagem.
 4. Rotação (*Rotation*): é a realização de rotação da imagem para a direita ou esquerda, em um eixo entre 1° e 359° (SHORTEN; KHOSHGOFTAAR, 2019). A segurança deste método é fortemente determinada pelo grau de rotação. Leves rotações entre 1 e 20 ou -1 a -20, podem ser úteis em tarefas de reconhecimento de dígitos, sendo que, à medida na qual o grau de rotação aumenta, a classe que representa os dados pode não mais ser preservada após aplicado tal transformação (SHORTEN; KHOSHGOFTAAR, 2019).
 5. Deslocamento (*shifting*): é um método de deslocação de imagens para a esquerda, direita, cima ou baixo, podendo ser uma transformação muito útil para evitar um viés no modelo e preservando as dimensões da imagem (SHORTEN; KHOSHGOFTAAR, 2019). Por exemplo, se todas as imagens em um conjunto de dados estiverem centralizadas, o que é comum em conjuntos de dados de reconhecimento de rosto, exigiria que o modelo também fosse testado em imagens perfeitamente centradas. Como a imagem original foi deslocada para um direção, o espaço restante pode ser

preenchido com um valor constante, no caso de um modelo na escala RGB, seria aplicada a cor preta (0, 0, 0) ou a cor branca (255, 255, 255), ou então com ruído aleatório ou gaussiano (SHORTEN; KHOSHGOFTAAR, 2019).

6. Injeção de ruído (*Noise injection*): A injeção de ruído consiste em injetar uma matriz de valores aleatórios, normalmente extraídos de uma distribuição gaussiana. Adicionar ruído às imagens pode ajudar uma CNN a aprender recursos mais robustos (SHORTEN; KHOSHGOFTAAR, 2019).

2.7.2 Transformações fotométricas

O canal de cores de uma imagem é composta por três matrizes bidimensionais empilhadas. Essas matrizes representam a cor RGB de um *pixel*. O enviesamento de iluminação está entre os desafios que ocorrem com mais frequência aos problemas de reconhecimento de imagem (SHORTEN; KHOSHGOFTAAR, 2019). Uma solução para imagens excessivamente claras ou escuras é percorrer as imagens e diminuir ou aumentar os valores da matriz de cores dos *pixels* aplicando uma determinada constante (SHORTEN; KHOSHGOFTAAR, 2019).

2.7.3 Aplicação de máscara

A aplicação de máscara, muito popular no processamento de imagens, é uma técnica que torna imagens mais nítidas ou desfocadas. Essas máscaras funcionam deslizando uma matriz bidimensional através de uma imagem com um filtro de desfoque Gaussiano, resultando em uma imagem mais desfocada ou um filtro de borda vertical ou horizontal de alto contraste, que resultará em uma imagem mais nítida ao longo das bordas (SHORTEN; KHOSHGOFTAAR, 2019). Intuitivamente, desfocar imagens para aumento de dados pode levar a uma maior resistência na classificação da imagem pelo modelo de aprendizado de máquina. Além disso, aumentar a nitidez de imagens para aumento de dados pode resultar no ocultamento de mais detalhes sobre regiões de interesse da imagem (SHORTEN; KHOSHGOFTAAR, 2019).

2.7.4 Mistura de imagens

Misturar imagens aplicando a média de seus valores de *pixel* é uma abordagem muito contra-intuitiva para o aumento de dados. As imagens produzidas com isso não parecerão uma transformação útil para um observador humano, sendo que, o emparelhamento de amostras pode ser desenvolvido em uma estratégia de aumento eficaz (SHORTEN; KHOSHGOFTAAR, 2019). Neste experimento, duas imagens são cortadas aleatoriamente de 256×256 a 224×224 , e invertidas aleatoriamente na horizontal. Essas imagens são então misturadas calculando a média dos valores de *pixel* para cada um dos canais RGB (SHORTEN; KHOSHGOFTAAR, 2019). Isso resulta em uma imagem mista que é usada para treinar um modelo de classificação. O rótulo atribuído à nova imagem é o mesmo da primeira imagem selecionada aleatoriamente (SHORTEN; KHOSHGOFTAAR, 2019).

2.7.5 Apagamento aleatório

O apagamento aleatório é inspirada nos mecanismos de regularização de *dropout* (SHORTEN; KHOSHGOFTAAR, 2019). *Dropout* é uma técnica de regularização para reduzir o sobreajuste. Esta técnica foi projetada especificamente para combater os desafios de reconhecimento de imagem devido à oclusão (SHORTEN; KHOSHGOFTAAR, 2019). Oclusão (*Occlusion*) se refere a algumas partes do objeto em uma imagem que não estão claras. O apagamento aleatório interromperá isso, forçando o modelo a aprender mais recursos descritivos sobre uma imagem, e evitando que ele se ajuste a um determinado recurso visual da imagem (SHORTEN; KHOSHGOFTAAR, 2019).

2.8 Conclusão

Neste capítulo foi introduzido uma série de conceitos importantes que estão amplamente relacionados a este trabalho de pesquisa, que envolve a aplicação de Redes Neurais Convolucionais (CNN) na resolução de cenários de reconhecimento de espécies. Contudo, a explicação para os conceitos de aprendizado de máquina e suas ramificações de mais baixo nível, como aprendizado profundo e CNN foram todas especificadas nesta seção, afim de

constituir uma maior compreensão de todos os conceitos que estão sendo abrangidos nos próximos capítulos.

3 Revisão sistemática

A revisão sistemática possibilita uma forma de análise e instrução a partir das informações dispostas pelos artigos disponíveis para uma pesquisa em específico, assunto de tópico ou tema de interesse, sendo que deve ser utilizada uma sistemática de confiança, auditável e íntegra. (HOLGADO *et al.*, 2020). O desenvolvimento da revisão sistemática se baseia em um protocolo devidamente definido para assim trazer resultados válidos e detalhados, trazendo assim a possibilidade de análise e contradição destes resultados a partir do uso do protocolo determinado.

3.1 *Objetivos da revisão*

Na área de visão computacional há a aplicação de técnicas com objetivo de identificação de insetos, onde a técnica de processamento baseado em imagens e algoritmos que reconhecem padrões e classificam assim, automaticamente as espécies de insetos. Tal processamento vem modificando o tradicional modelo manual e descritivo de características morfológicas já fornecido por estudos taxonômicos para a identificação destes insetos. Sendo que, apenas especialistas como taxonomistas e técnicos qualificados podem identificar tais insetos com precisão, pois isso requer conhecimento especializado e obtido por meio de uma vasta experiência (LIM; KIM; KIM, 2017). O aprendizado profundo tem sido amplamente utilizado em processamento de imagens, visão computacional e reconhecimento de padrões. Diferente dos métodos tradicionais de aprendizado de máquina, o modelo de aprendizado profundo pode aprender recursos automaticamente a partir de uma grande quantidade de imagens, e não requer um especialista com domínio na construção do extrator de características. (LIU *et al.*, 2020b).

As características extraídas das imagens são os fatores que mais influenciam na acurácia do classificador. Embora existam várias técnicas e ferramentas para extração de características automaticamente, ainda não se encontram trabalhos na literatura que avaliem estas técnicas de aprendizado profundo e nem ferramentas que as entreguem de forma sistemática.

Portanto, o objetivo desta revisão sistemática é apresentar e discutir o estado da arte das técnicas de reconhecimento automático de espécies por meio de suas imagens e

morfologias diversas, empregando técnicas de aprendizado profundo. Visando tal objetivo, foram analisados os artigos com foco na aplicação de Redes Neurais Convolucionais (CNN).

3.2 Protocolo

Foi coordenada em quatro fases para a realização desta revisão sistemática, sendo elas:

1. Planejamento
2. Condução
3. Extração
4. Consolidação

O planejamento consiste na criação do protocolo no qual as diretrizes da pesquisa serão baseadas. A condução é a execução da pesquisa e seleção de trabalhos de interesse, de acordo com os critérios de inclusão e exclusão definidos do protocolo. A extração é a fase na qual os artigos que foram selecionados são estudados e os dados importantes são separados. Em suma, temos a consolidação, na qual os dados são cruzados para gerar conhecimento e , assim, o artigo de revisão sistemática é escrito.

A fase de planejamento é a determinação dos critérios de pesquisa para a formulação do protocolo. Na fase de condução é realizada a pesquisa de trabalhos, seguido da inclusão e exclusão destes trabalhos a partir dos itens baseados no protocolo. Já na fase de extração ocorre o estudo e a separação de dados importantes extraídos dos trabalhos selecionados. Por fim a fase de consolidação, onde é realizada a comparação entre os dados obtidos para o ganho de ciência, e seguidamente criar um artigo de revisão sistemática.

Estes processos estão representados na Figura 11, na qual cada etapa do processo estão identificados com uma cor diferente.

Figura 11 – Diagrama do processo de revisão sistemática



Fonte – Danilo Assunção, 2021

Na etapa de planejamento foi levantada junto ao protocolo a distinta questão: “Quais são os métodos que sejam o estado da arte, para identificação de espécies utilizando Redes Neurais Convolucionais por meio de imagens digitais?”. A conclusão desta pergunta veio a partir de bases de dados junto às seguintes *strings* de busca:

- **IEEE Xplore** - (“convolutional neural network” OR “cnn” OR “convolution”) AND “feature extraction” AND (“insect” OR “pest”) AND (“image” OR “image segmentation” OR “image recognition”)
- **SCOPUS** - (“convolutional neural network” OR “cnn” OR “convolution”) AND “feature extraction” AND (“insect” OR “pest”) AND (“image” OR “image segmentation” OR “image recognition”) AND (EXCLUDE (DOCTYPE, “cp”) OR

EXCLUDE (DOCTYPE, “re”) OR EXCLUDE (DOCTYPE, “bk”) OR EXCLUDE (DOCTYPE, “ch”))

A palavra chave *pest* é incluída nas *strings* de busca por se tratar de um objeto de estudo muito recorrente nos meios de reconhecimento de espécies através de imagens, utilizando de técnicas similares às que são usadas para reconhecimento de insetos por meio de métodos utilizando CNN.

Este trabalho foi conduzido de dezembro de 2020 até fevereiro de 2021. Inicialmente 335 trabalhos foram encontrados nas bases de dados especificadas acima, dos quais 302 estavam na base de dados SCOPUS e 33 na base de dados IEEE Xplore.

Após aplicar os critérios de inclusão e exclusão nos trabalhos coletados, 31 artigos foram selecionados e analisados por completo, sendo 28 da base SCOPUS e 3 da base IEEE Xplore.

Os critérios utilizados para incluir ou excluir um trabalho nesta revisão foram:

- Inclusão
 - Artigos que utilizam Redes Neurais Convolucionais como ferramenta para discriminação de espécies por meio de imagens digitais;
 - Artigos que contém técnicas computacionais de segmentação e extração de características de imagens digitais de diversas espécies biológicas;
 - Artigos devem ter data de publicação maior ou igual a 2015.
- Exclusão
 - Artigos que não são relacionados com conceitos envolvendo técnicas de segmentação e extração de características por meio de imagens digitais;
 - Artigos que não utilizam Redes Neurais Convolucionais para classificação automática de espécies por meio de imagens digitais;
 - Artigos que não foram escritos com o idioma inglês ou português;
 - Artigos com data de publicação menor que 2015;
 - Documentos que não estejam categorizados como artigo em seu tipo de documento.

De cada trabalho selecionado foram extraídos tais atributos como, Ano de publicação, País de afiliação dos autores, Técnicas de pré-processamento de imagens, Técnica de

aprendizado por transferência, Técnicas de aumento de dados, Modelo de arquitetura da CNN e Modelo do classificador.

Com a finalidade de poder encontrar possíveis lapsos nas pesquisas utilizadas, seguidamente da extração de dados, foi exercida uma averiguação sobre os principais obstáculos e propensões praticadas.

3.3 Resultado da revisão sistemática

Os 31 trabalhos selecionados foram estudados e os seguintes atributos foram extraídos: técnica de extração de características, tipos de características extraídas, principal país de afiliação dos pesquisadores e ano de publicação. Um breve sumário dos trabalhos com os respectivos dados extraídos encontram-se no Tabela 1.

Tabela 1 – Sumário de todos os artigos analisados

Artigo	Ano de publicação	País de afiliação dos autores	Téc. de pré-processamento de imagens	Téc. de aprendizado por transferência	Téc. de aumento de dados	Modelo de arquitetura da CNN	Modelo do classificador
(LIU <i>et al.</i> , 2020b)	2020	Estados Unidos	Não aplicado	Fine-tuning	Skewing, Elastic distortion, Rotation	LeNet-5, AlexNet, VGGNet-16, VGGNet-19, ResNet-50, Inception v3, InceptionResNet v2	Camada totalmente conectada
(LE <i>et al.</i> , 2020)	2020	França	Resize	Fine-tuning	Constant to each channel, Split the channels of image	EB-Net	Camada totalmente conectada
(ZHANG; CHEN,)	2020	China	Não especificado	Não especificado	Não especificado	DenseNet-121	Não especificado
(LIU <i>et al.</i> , 2020a)	2020	China	Não aplicado	Não aplicado	High brightness, Low brightness, High contrast, Low contrast, High sharpness, Low sharpness, Rotation, Vertical symmetry, Horizontal symmetry, Gaussian noise, PCA Jittering	DICNN	Camada totalmente conectada
(LU; TUNG; KUO, 2020)	2020	Taiwan	Resize	Fine-tuning	Horizontal flipping, Vertical flipping, Width shifting, Height shifting, Rotation, Shearing, Zooming	VGGNet-16	Camada totalmente conectada
(BANAN; NA-SIRI; TAHERI-GARAVAND, 2020)	2020	Irã	Resize	Fine-tuning	Rotation, Height shifting, Width shifting	VGGNet-16	Camada totalmente conectada

Continua na próxima página

Tabela 1 – continuação da página anterior

Artigo	Ano de publicação	País de afiliação dos autores	Téc. de pré-processamento de imagens	Téc. de aprendizado por transferência	Téc. de aumento de dados	Modelo de arquitetura da CNN	Modelo do classificador
(TIWARI, 2020)	2020	India	Não especificado	Não aplicado	Não aplicado	CNN	Camada totalmente conectada
(LI <i>et al.</i> , 2020)	2020	China	Cropping, Resize	Não aplicado	Rotation, Clipping	VGGNet-16, Inception v3	Camada totalmente conectada
(CHEN <i>et al.</i> , 2020)	2020	China	Resize	Fine-tuning	Rotation, Blurred, Occluded	ResNeXt-101	Camada totalmente conectada
(HONG <i>et al.</i> , 2020)	2020	China	Resize, Cropping	Fine-tuning	Gaussian filter, Contrast enhancement	AlexNet	Camada totalmente conectada
(RAUF <i>et al.</i> , 2019)	2019	Paquistão, Estados Unidos	Resize, Transparent background	Não aplicado	Não aplicado	32-Layer CNN	Camada totalmente conectada
(GÓMEZ-RÍOS <i>et al.</i> , 2019)	2019	Espanha, Dinamarca	Não especificado	Fine-tuning	Não especificado	Inception v3, ResNet-50, ResNet-152, DenseNet-121, DenseNet-161	Camada totalmente conectada
(CIBUK <i>et al.</i> , 2019)	2019	Turquia, Estados Unidos	Não aplicado	Feature transfer	Não aplicado	VGGNet-16, AlexNet	SVM (Kernel RBF)
(LI <i>et al.</i> , 2019)	2019	China	Resize	Não aplicado	Multi-scale, Rotation	ResNet-50	Camada totalmente conectada
(KHALIFA <i>et al.</i> , 2019)	2019	Egito, Luxemburgo	Não especificado	Não aplicado	Reflection, Zooming, Gaussian noise, Poisson noise, Salt and pepper noise, Spectral noise	CNN	Camada totalmente conectada
(HSIANG <i>et al.</i> , 2019)	2019	Suécia, Reino Unido	Resize	Fine-tuning	Não especificado	VGGNet-16, DenseNet-121, Inception v3	Camada totalmente conectada

Continua na próxima página

Tabela 1 – continuação da página anterior

Artigo	Ano de publicação	País de afiliação dos autores	Téc. de pré-processamento de imagens	Téc. de aprendizado por transferência	Téc. de aumento de dados	Modelo de arquitetura da CNN	Modelo do classificador
(REN; LIU; WU, 2019)	2019	Japão	Resize, Cropping	Não aplicado	Resize, Cropping, Mean and standard deviation normalization	FR-ResNet	Camada totalmente conectada
(VALAN <i>et al.</i> , 2019)	2019	Suécia	Resize, Cropping	Feature transfer	Não aplicado	VGGNet-16	SVM (Kernel Linear)
(GYIRES-TÓTH <i>et al.</i> , 2019)	2019	Hungria	Cropping, Resize	Fine-tuning	Não aplicado	AlexNet, Inception V3	Camada totalmente conectada
(THENMOZHI; REDDY, 2019)	2019	India	Gray scale, Canny Edge Detection, Cropping, Resize	Fine-tuning	Reflection, Scaling, Rotation, Translation	VGGNet-16, VGGNet-19	Camada totalmente conectada
(TETILA <i>et al.</i> , 2019)	2019	Brasil	Não especificado	Fine-tuning, Feature transfer	Rotation, Scaling, Scrolling, Zooming	DenseNet-201, InceptionResNet v2, ResNet-50	Camada totalmente conectada
(TAN <i>et al.</i> , 2018)	2018	Malásia	Resize, Padding	Não aplicado	Não aplicado	D-Leaf	Camada totalmente conectada
(ZHU; ZHU; REN, 2018)	2018	China	Não especificado	Não aplicado	Não aplicado	Inception v2	Camada totalmente conectada
(ZHU <i>et al.</i> , 2018)	2018	China	Resize	Não especificado	Não aplicado	CNN	SVM (Kernel Linear)
(VILLON <i>et al.</i> , 2018)	2018	França	Não especificado	Não aplicado	Não aplicado	GoogLeNet	Camada totalmente conectada
(LEE <i>et al.</i> , 2017)	2017	Malásia, Reino Unido	Resize	Fine-tuning	Rotation	CNN	Camada totalmente conectada

Continua na próxima página

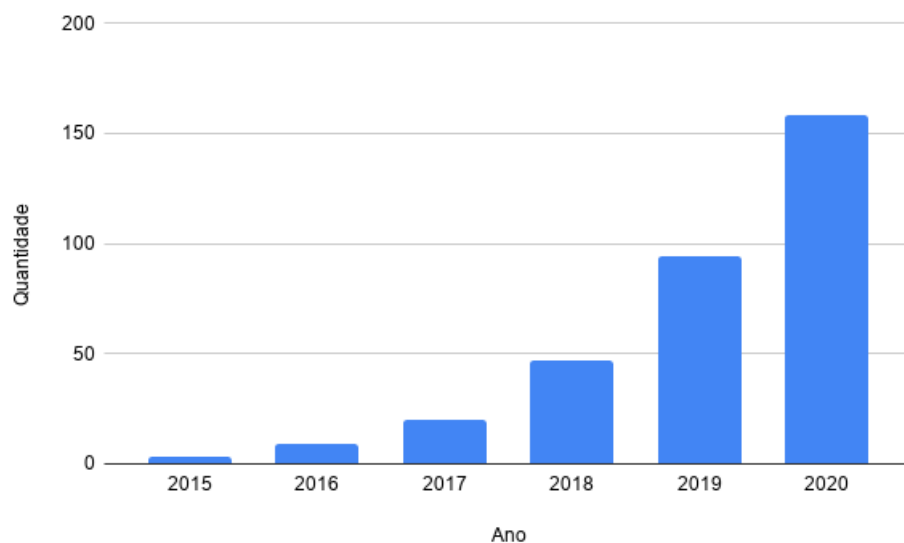
3.3.1 Ano de publicação

Nesta revisão sistemática foi coletado o ano de publicação dos artigos, dessa maneira, fez-se uma análise de tendência referente aos assuntos que abordam técnicas de reconhecimento de espécies utilizando redes neurais convolucionais, em domínios que envolvem assuntos relacionados à reconhecimento de espécies.

Visando uma melhor compreensão sobre a quantidade de publicações com base em assuntos que envolvem reconhecimento automático de espécies utilizando técnicas de aprendizado profundo, foi elaborado um gráfico com todos os artigos encontrados na base de dados pela consulta feita nas bases do SCOPUS e IEEE Xplore.

Na Figura 12 é apresentado uma linha de tendência aos assuntos envolvendo artigos que possuem as palavras chaves, aprendizado profundo, redes neurais convolucionais e reconhecimento de espécies.

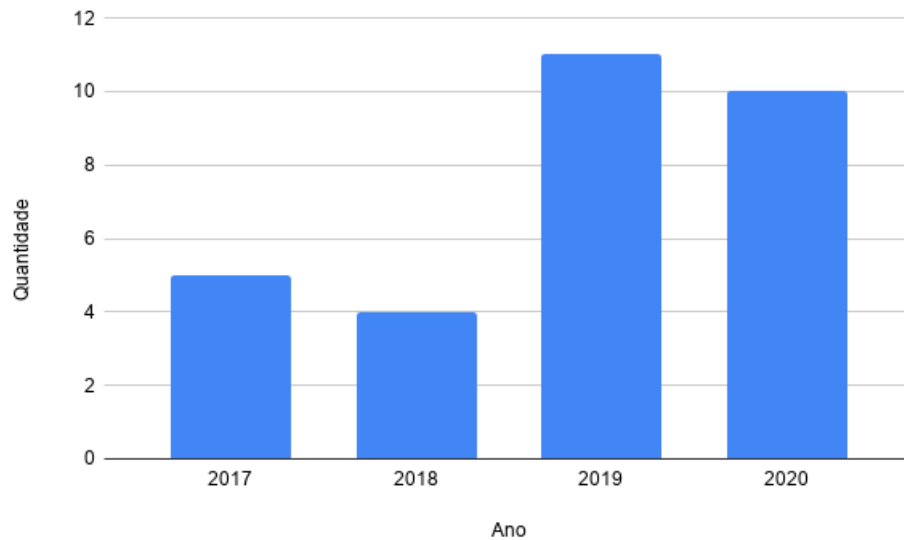
Figura 12 – Ano de publicação dos artigos com base em palavras chaves



Fonte – Danilo Assunção, 2021

Na Figura 13 é demonstrado uma relação de quantidade versus ano, para o total de artigos que foram filtrados e selecionados a partir do conjunto obtido nas consultas executadas nas bases do SCOPUS e IEEE Xplore para esta revisão sistemática.

Figura 13 – Ano de publicação dos artigos selecionados



Fonte – Danilo Assunção, 2021

Nota-se que a maior quantidade dos artigos selecionados foram publicados em 2019, representando um total de 36,67% (11). Já em 2020, é tido 33,33% (10) dos artigos. Em 2018 a quantia de artigos foi de 13,33% (4) e em 2017, 16,67% (5).

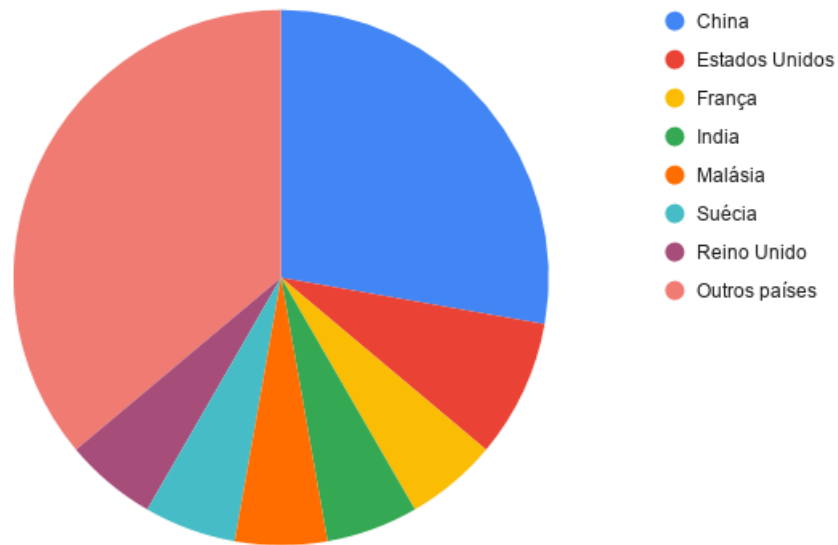
É possível perceber que, desde 2017 até 2020, houve um aumento significativo de artigos de interesse publicados com relação à aplicação de redes neurais convolucionais que almejam a resolução de problemas envolvendo reconhecimento de espécies em imagens digitais. Essa representatividade demonstra a veracidade de uma tendência de alta, com relação à utilização de técnicas envolvendo redes neurais convolucionais neste tipo de aplicação (LIU *et al.*, 2020b).

Os artigos de interesse obtidos possuem data de publicação a partir do ano de 2017, os artigos de 2015 a 2016 não se mostraram muito relevantes para os assuntos relacionados à esta revisão sistemática, sendo levado em consideração os critérios de exclusão definidos.

3.3.2 País de afiliação dos autores

Foi feito a extração de dados sobre os países de cada pesquisador contribuinte em determinado artigo e esta visão pode ser na partir da Figura 14.

Figura 14 – País de afiliação dos pesquisadores



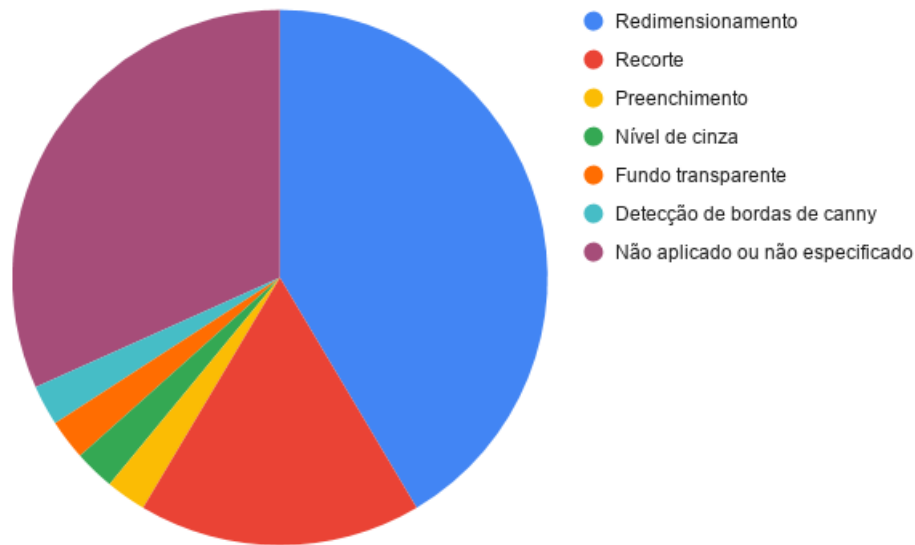
Fonte – Danilo Assunção, 2021

Pode ser visto que, com base nos artigos selecionados, a China possui a liderança com relação à produção de artigos científicos relacionados a identificação e classificação de espécies por meio do uso de Redes Neurais Convolucionais, possuindo 27,78% (10 artigos) em relação aos pesquisadores de outros países. Os Estados Unidos, em segundo colocado, possuindo 8,33% (3 artigos). A França, Índia, Malásia, Suécia e o Reino Unido possuem ambos 5,56% (2 artigos) artigos publicados. Já na categoria “Outros países”, que compõe os 36,11% dos artigos publicados restantes, são as demais regiões que possuem em cada um deles somente apenas um artigo científico publicado.

3.3.3 Técnicas de pré-processamento de imagens

Nos métodos abordados nos artigos selecionados é possível identificar a utilização de técnicas de pré-processamento de imagens em alguns cenários distintos como é demonstrado na Figura 15.

Figura 15 – Técnicas de pré-processamento de imagens aplicadas



Fonte – Danilo Assunção, 2021

O desempenho da tarefa de classificação é melhorado ao aplicar as técnicas de pré-processamento de imagem para extrair automaticamente o resultado da imagem introduzida antes de ser inserida nos modelos de aprendizagem profunda (THENMOZHI; REDDY, 2019). A técnica de Redimensionamento (*resize*) vem liderando com 41,46% de uso com relação as outras estratégias abordadas, e isso se deve aos diversos fatores que auxiliam no aumento de performance das redes neurais convolucionais, como a quantidade reduzida de parâmetros a serem introduzidos na rede. Em segundo com 17,07%, o Recorte (*cropping*) que pode vir a ser utilizado para destacar a região de interesse da imagem e reduzindo também o tamanho da imagem (THENMOZHI; REDDY, 2019). Preenchimento (*padding*), Nível de cinza (*gray scale*), Fundo transparente (*transparent background*) e Deteção de bordas de canny (*canny edge detection*) foram algumas das abordagens utilizadas, e ocupam apenas 2,44% de uso com relação aos outros métodos. Ocupando os 31,71% restantes, foram artigos que não aplicaram ou não especificaram diretamente o uso de técnicas de pré-processamento de imagens, isto pode ser ocasionado pelo fato do conjunto de dados utilizado já estar previamente pré-processado em sua disponibilização, ou de não haver necessidade direta de uma preparação antes de introduzi-las à rede.

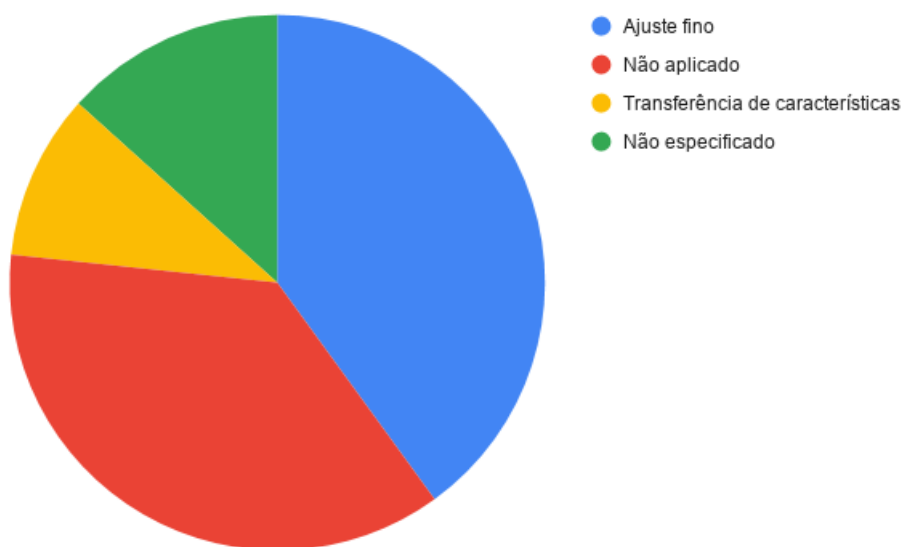
3.3.4 Técnica de aprendizado por transferência

Aprendizado por transferência (*Transfer learning*) é uma técnica que permite a utilização de uma rede pré-treinada, reduzindo a carga computacional e beneficiando-se do poder de uma rede neural convolucional mais sofisticada (VALAN *et al.*, 2019).

Dois métodos foram comumente utilizados como técnicas de aprendizado por transferência, sendo elas, o ajuste fino e a transferência de características. O ajuste fino (*fine-tuning*) modifica os parâmetros de uma Rede Neural Convolucional pré-treinada para o treinamento de uma nova tarefa, onde a camada de saída é estendida com pesos inicializados aleatoriamente para a nova tarefa, e uma pequena taxa de aprendizagem é usada para ajustar os parâmetros de seus valores originais para minimizar a perda no novo modelo (LI; HOIEM, 2017). A transferência de características (*feature transfer*) faz uso de uma rede neural convolucional pré-treinada para calcular recursos para uma imagem (LI; HOIEM, 2017). A extração de recursos não modifica a rede original e permite que novas tarefas se beneficiem de recursos complexos aprendidos em tarefas anteriores (LI; HOIEM, 2017).

Na Figura 16 será apresentado os dados coletados por meio dos artigos analisados que utilizam de técnicas de aprendizado por transferência, na resolução de cenários que abordam o uso de Redes Neurais Convolucionais.

Figura 16 – Técnicas de Aprendizado por transferência



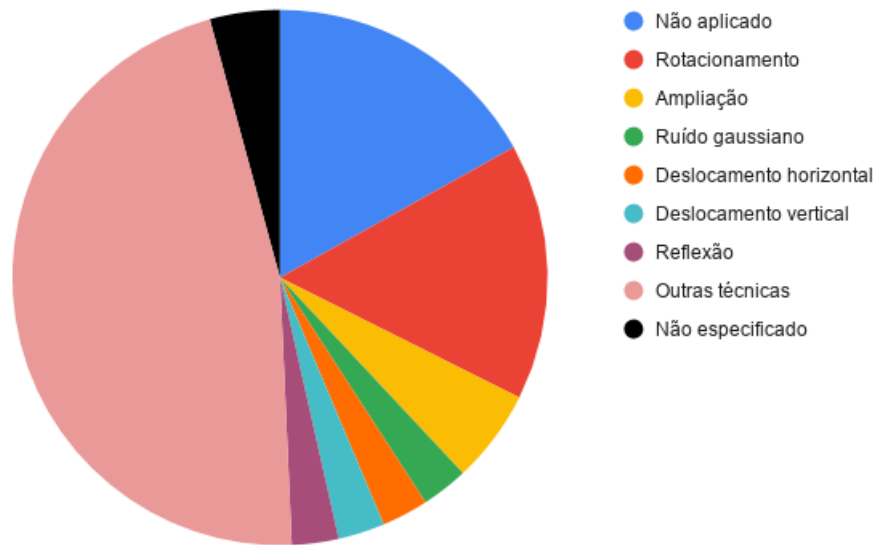
Fonte – Danilo Assunção, 2021

O ajuste fino foi o método mais utilizado com 40,00% (12) de uso em relação a todas as outras técnicas encontradas nos artigos analisados. Podendo ser devido ao ajuste fino funcionar bem quando o modelo especializado é semelhante à nova tarefa proposta (VALAN *et al.*, 2019), e este mecanismo foi utilizado para diversas tarefas com cenários semelhantes de reconhecimento de padrões. Em segundo, é tido em todos os artigos que não aplicaram nenhuma técnica de Aprendizado por transferência ocupando 36,67% (11), devendo-se ao fato de que nem todo cenário pode vir a necessitar do uso de técnica de reaproveitamento deste tipo, e também, este mecanismo é suscetível a sobreajuste na tarefa especializada quando os conjuntos de dados são pequenos, pois pode associar incorretamente uma categoria rara a um recurso irrelevante, como um tipo especial de fundo, que por acaso está presente nas poucas imagens dessa categoria e no conjunto de treinamento (VALAN *et al.*, 2019). O método de transferência de características teve um uso de 10,00% (3) com relação às outras técnicas abordadas. O restante mapeado, com 13,33% (4) não especificou diretamente o tipo de mecanismo de Aprendizado por transferência utilizado.

3.3.5 Técnicas de aumento de dados

Métodos de Aumento de dados (*Data augmentation*) permite aumentar artificialmente o tamanho do conjunto de treinamento. O aumento do conjunto de treinamento é feito ao aplicar várias distorções às imagens originais, como zoom, virando-as horizontalmente ou verticalmente, girando-as, deslocando-as e dentre outros (GÓMEZ-RÍOS *et al.*, 2019). Na Figura 17 é demonstrado graficamente as técnicas utilizadas para o Aumento de dados nos artigos abordados.

Figura 17 – Técnicas de Aumento de dados



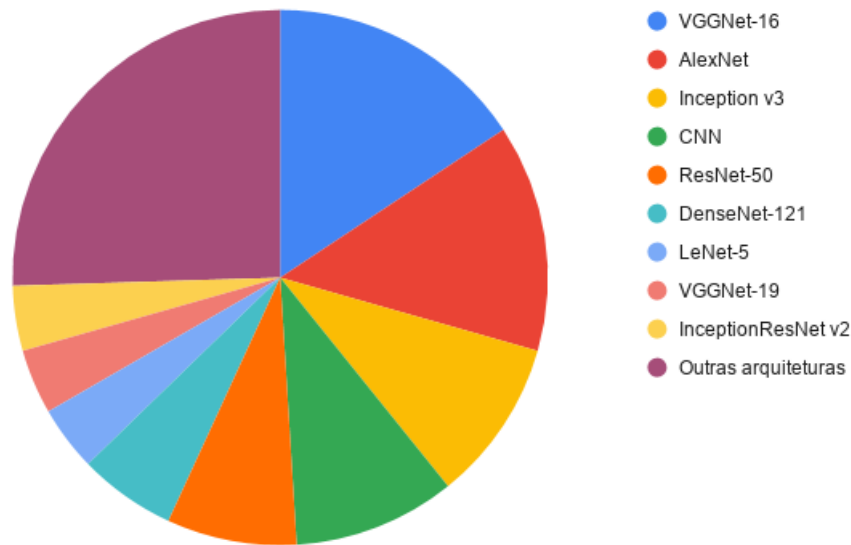
Fonte – Danilo Assunção, 2021

Em maioria, os artigos não aplicaram técnicas de Aumento de dados, com uma representação de 63,90% (12) de todos os exemplares analisados, isto pode ser ocasionado pelo fato dos conjuntos de dados utilizados pelos autores já serem grandes o suficiente ou devido ao uso de outras técnicas para supressão de sobreajuste. Em seguida, é possível observar que as técnicas de Aumento de dados utilizando Rotacionamento (*Rotation*) são bem prioritárias quando este mecanismo é aplicado, com um total de uso de 15,49% (11) com relação aos outros meios. Ampliação (*Zooming*), sendo a terceira metodologia mais aplicada com 5,63% (4). Ruído gaussiano (*Gaussian noise*), Deslocamento horizontal (*Width shifting*), Deslocamento vertical (*Height shifting*) e Reflexão (*Reflection*) representando 2,82% (2) nos demais cenários. As Outras técnicas agrupadas são métodos que foram utilizados somente uma única vez e o conjunto não especificado são aos artigos que não esclareceram de maneira explícita o tipo de técnica de Aumento de dados utilizada.

3.3.6 Modelo de arquitetura da CNN

É possível identificar na literatura uma grande quantidade de arquiteturas de Redes Neurais Convolucionais sendo utilizadas em diversas soluções de problemas que envolvem a classificação morfológica de espécies, por meio de suas imagens digitais. Na Figura 18 é possível observar graficamente estes dados.

Figura 18 – Arquitetura de Redes Neurais Convolucionais

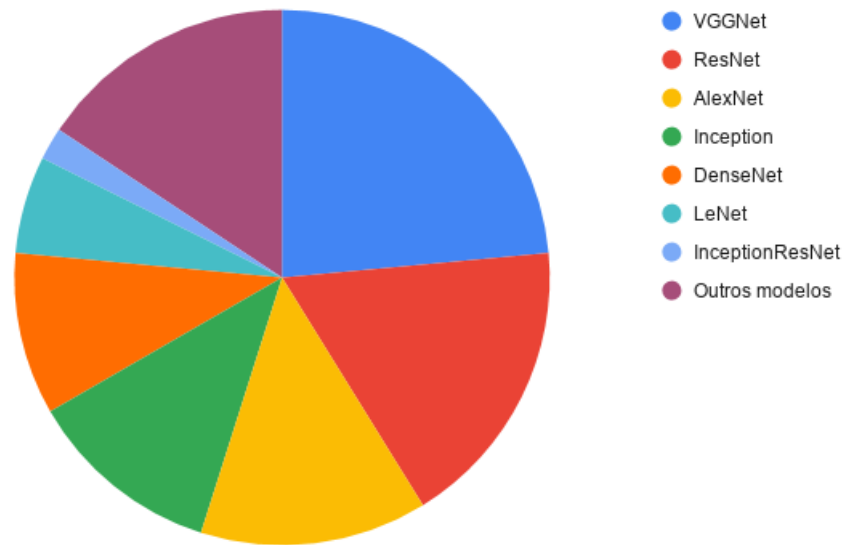


Fonte – Danilo Assunção, 2021

A VGGNet-16 obteve uma utilização de 14,29% (8), encontrando-se em primeiro com relação as outras arquiteturas destacadas. Em comparação a AlexNet posicionada em segundo com 12,50% (7), as arquiteturas de VGGNet aumentam a profundidade da rede e reduz o tamanho do *kernel* de convolução, podendo assim reduzir os parâmetros a serem computados, além disso, o desempenho de generalização de arquiteturas VGGNet costumam ter bons resultados (LI *et al.*, 2020). Os modelos de Inception v3 e CNN, onde a categoria CNN representa os modelos implementados pelos autores em seus artigos ocupam 7,81% (5). ResNet-50 com 6,25% (4) e DenseNet-121 com 4,69% (3). LeNet-5, VGGNet-19 e InceptionResNet v2 que ocupam 3,13% (2) em aplicabilidade. Já as “Outras arquiteturas” que compõe 20,31% do restante das arquiteturas utilizadas, possuem uma única utilização em seus respectivos artigos.

Para uma diferente perspectiva de visualização de usabilidade das arquiteturas de Redes Neurais Convolucionais extraídas nos projetos em questão, foi feito um agrupamento por tipo de arquitetura para a identificação dos exemplares mais utilizados, como é demonstrado na Figura 19.

Figura 19 – Tipos agrupados de arquiteturas de Redes Neurais Convolucionais



Fonte – Danilo Assunção, 2021

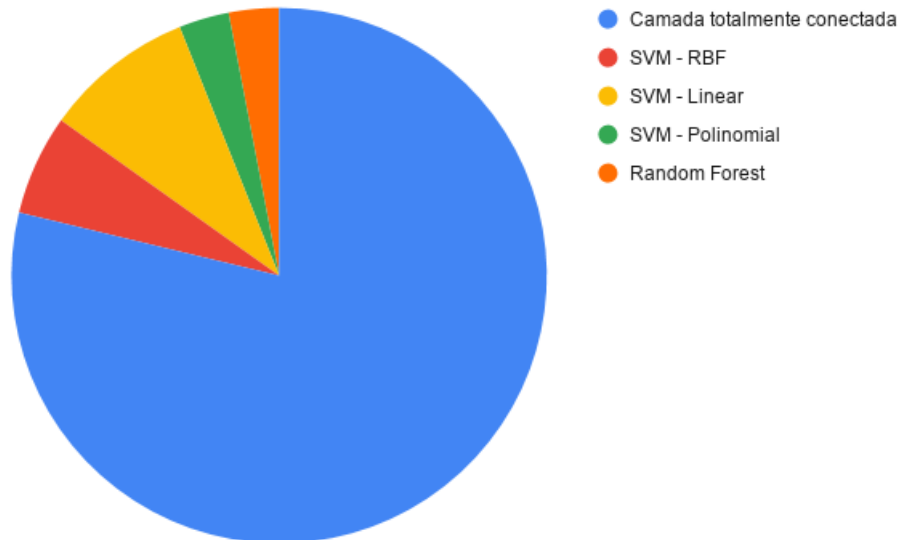
Nota-se que modelos de VGGNet com uso de 23,53% (12) ainda dominam em termos de arquiteturas aplicadas na resolução de problemas envolvendo a identificação de espécies. Arquiteturas do tipo ResNet ocupam o segundo lugar com 17,65% (9), isso devido ao seu comportamento de aumentar a profundidade da rede, sendo assim, as ResNets oferecem melhores recursos de aproximação de função (convergência ao resultado) à medida em que ganham mais parâmetros, contribuindo com sucesso na resolução de problemas como os de dissipação de gradiente, onde alguns modelos mais tradicionais acabam não conseguindo solucioná-lo, tendo resultados inferiores ao desejado (SUN *et al.*, 2017). Tendo a AlexNet em terceiro com 13,73% (7), sendo este tipo de arquitetura um modelo pré-treinado de rede neural convolucional que permite a aplicação de Aprendizado por Transferência para que atinja melhores resultados (HONG *et al.*, 2020). Modelos de Inception 11,76% (6), DenseNet 9,80% (5). LeNet 5,88% (3) e InceptionResNet 1,96% (1) possuem a menor quantidade de uso para os cenários avaliados. Por último, os “Outros modelos” que são arquiteturas de CNN desenvolvidas pelos autores dos artigos com especificidade para solucionar seus respectivos problemas abordados.

3.3.7 Modelo do classificador

Os métodos de classificação são aplicados nas últimas camadas das Redes Neurais Convolucionais, utilizando as características extraídas como entrada para o processo de

classificação dos dados e predição do resultado. Na Figura 20 é demonstrado os exemplares aplicados nos artigos selecionados.

Figura 20 – Método de classificação utilizado



Fonte – Danilo Assunção, 2021

A partir do processo de convolução da Rede Neural Convolucional, é seguido o agrupamento e divisão das imagens em recursos. Em sequência desse processo, tais recursos são utilizados para alimentar a estrutura de rede neural totalmente conectada, pode direcionar a decisão de classificação final (RAUF *et al.*, 2019).

Decorrendo em 78,79% (26) de utilização apresentada nos artigos e representada como “Camada totalmente conectada”. Nos demais cenários onde a rede neural convolucional é utilizada somente para extração de características, é visto a aplicação conjunta com algoritmos de SVM (Máquina de vetores de suporte) com *kernel* Linear 9,09% (3), RBF (Função de Base Radial) 6,06% (2) e Polinomial 3,03% (1). Por fim, houve apenas uma implementação do algoritmo *Random Forest* ocupando 3,03% (1).

3.4 Conclusão

Após o término da revisão sistemática é perceptível que o uso de técnicas de aprendizado profundo, como redes neurais convolucionais, vem sendo empregadas em diversos cenários onde há a necessidade de otimização no processo de extração de características das imagens, para assim, solucionar problemas que envolvem reconhecimento de espécies.

Um dos atributos mais importantes é a sua capacidade de aprender as características de mais alto nível de uma imagem, e fornecer prontamente uma quantidade de informações relevantes que permitam que os algoritmos de classificação possam identificar com mais eficácia o objeto visualizado. Entretanto, é importante ressaltar que o uso deste mecanismo exige uma quantidade muito grande de dados, justamente para que seu resultado venha a convergir para um modelo que tenha a capacidade de efetivar tal atividade, e para isto foi observado na literatura o uso de técnicas como transferência de aprendizado e aumento de dados que permitem a utilização de redes neurais convolucionais, mesmo com um conjunto pequeno de dados disponíveis e demonstrando também que o uso de pré-processamento ainda pode vir a ser extremamente útil em cenários diversos envolvendo processamento de imagens digitais.

4 Materiais e métodos

Visando a atingir o objetivo desta pesquisa, serão conduzidos experimentos com base em um conjunto de dados de treinamento e de validação, unindo técnicas de aprendizado por transferência ou aumento de dados aplicados a uma rede neural convolucional, contendo uma estrutura para a extração de características e classificação supervisionada das imagens das asas de maneira automática. No conjunto de treinamento e validação, serão utilizadas as imagens de asas de abelhas obtidas em um ângulo que mostre a face da asa, de tal maneira a expor claramente suas nervuras, como pode ser visto na Figura 21. As imagens contidas nos conjuntos de dados já foram previamente rotuladas com a espécie da abelha por um especialista da área.

Figura 21 – Imagem das nervuras destacadas de uma asa de abelha



Fonte – Danilo Assunção, 2021

Para levantar o estado da arte das técnicas de aprendizado por transferência, aumento de dados e CNN, foi elaborada uma revisão sistemática utilizando os conceitos aplicados em [Liu *et al.* \(2020b\)](#).

Após o levantamento do estado da arte como base, pretende-se construir uma CNN como ferramenta de classificação, empregando as melhores técnicas obtidas e, em caso pertinente, uma combinação do uso de técnicas de aprendizado por transferência ou aumento de dados, devem ser introduzidas, a fim de melhorar o estado da arte do método aplicado.

4.1 *Ferramenta de classificação*

O processo definido para a elaboração da ferramenta de classificação, utilizando imagens das asas de abelhas, está dividido em quatro partes principais:

1. Pré-processamento;
2. Carregamento;
3. Extração das características;
4. Classificação.

No pré-processamento, serão feitas as aplicações necessárias de processamento de imagem, nas imagens do conjunto de dados, para tratamentos gerais como, a dimensão da imagem, recorte e a possível aplicação de técnicas de aumento de dados, para a geração de um conjunto maior de dados. No passo de carregamento, a imagem pré-processada será transformada em uma representação computacional, a qual a CNN consiga computar e aceitar como valor de entrada. No processo de extração de características, os dados passarão pela arquitetura da CNN, onde será feito todo o processo convolucional junto ao redimensionamento dos dados dessa imagem, até a extração das melhores características destas imagens. Por fim, o processo de classificação será responsável pelo mecanismo de discriminação das espécies, por meio das características adquiridas no passo anterior.

A etapa de classificação contará com dois passos essenciais, o treinamento e classificação. No treinamento será utilizado um conjunto misto de dados, intitulados como conjunto de treinamento e validação para treinamento da camada de classificação da CNN, e um conjunto de teste deverá ser aplicado em cima do resultado obtido pela rede, para a visualização e confirmação do valores deferidos na saída deste procedimento.

4.2 *Avaliação*

A avaliação é executada ao comparar os resultados produzidos automaticamente pela CNN, observando a aplicação de métricas de aprendizado de máquina supervisionado (revocação, precisão, acurácia), para assim mensurar os resultados obtidos no processo de discriminação das classes.

5 Proposta de pesquisa

Na literatura foi possível observar que, o processo de extração manual de características das imagens é algo fundamental para um melhor desempenho na classificação, mas é tida como uma tarefa especializada e trabalhosa. Muitos desses recursos de mais alto nível acabam por não poder ser extraídos manualmente ao utilizar apenas de métodos de segmentação. Sendo assim, um procedimento que permita determinar automaticamente os recursos adequados para um deliberado problema, que envolva a identificação por meio de imagens vem sendo muito requisitado (WÄLDCHEN; MÄDER, 2018). No campo da ecologia, uma das tarefas é a identificação de espécies de abelhas a partir de imagens de suas asas (LIU *et al.*, 2020b), visto que a combinação de técnicas que envolvem o mecanismo de aprendizado profundo é raramente explorado (LIU *et al.*, 2020b). Entretanto, pesquisadores têm demonstrado, nos últimos anos, um aumento no interesse pela área de aprendizado de máquina, na qual a construção de sistemas de classificação de espécies mais eficientes, que utilizam técnicas de aprendizagem de máquina, acaba por ser de grande interesse e aplicabilidade no ramo científico (LIU *et al.*, 2020b). Para a aplicação de uma CNN na resolução de problemas envolvendo imagens e reconhecimento de espécies, é necessário uma quantidade maior de dados para que a rede consiga convergir para um bom quadro de hipóteses, permitindo alcançar melhores resultados no processo de reconhecimento. Portanto, dois problemas ocorrem no treinamento destes modelos em conjunto de dados ecológicos, sendo eles o conjunto de dados de imagens limitado e os números de amostras desequilibrados entre as diferentes classes, dessa maneira, as precisões de teste para os modelos de CNN existentes no conjunto de dados ecológicos podem ser baixas (LIU *et al.*, 2020b). Visando a resolver os problemas mencionados, neste trabalho será utilizada uma CNN para, assim, realizar a extração automática das características contidas na morfologia das asas das abelhas presentes em imagens digitais. A partir destas características, identificar as respectivas espécies, em conjunto à técnicas de aumento de dados na série de imagens disponibilizadas, com o intuito de ampliar em diversas vezes o seu tamanho, visando também a elevar a eficiência da velocidade de convergência da rede a partir de técnicas de aprendizado por transferência. Almeja-se, assim, um melhor desempenho nos resultados desejados para o processo de classificação das espécies de abelhas.

5.1 Plano de trabalho e cronograma

A seguir são apresentados um cronograma representado pela Tabela 2 e todas as atividades que englobam este projeto de pesquisa:

1. Revisão bibliográfica: uma revisão sistemática da literatura foi realizada a fim de identificar o estado da arte do uso de CNN, com técnicas que proveem um mecanismo que permita o uso deste tipo de rede em diversos cenários para a identificação de espécies;
2. Teste de proficiência em inglês: realização da prova de proficiência em inglês solicitada pelo programa;
3. Qualificação: realização do exame de qualificação;
4. Preparação dos dados: levantamento de todo o conjunto de dados das abelhas disponível contendo imagens de suas asas e aplicação de possíveis técnicas de aumento de dados para ampliação deste conteúdo.
5. Preparação do modelo de CNN: definição e desenvolvimento de um algoritmo de CNN para o modelo;
6. Extração de características: aplicação da CNN desenvolvida para extração de características, podendo ser utilizadas técnicas de aprendizado por transferência;
7. Classificação: compilado do resultado obtido no passo de extração em um algoritmo de classificação supervisionado estabelecido na CNN;
8. Avaliação: testes utilizando os algoritmos de classificação propostos para identificação das imagens;
9. Escrita da dissertação;
10. Defesa.

Tabela 2 – Cronograma

Atividade		2020					2021												2022			
#	Descrição	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4
1	Revisão bibliográfica	x	x	x	x	x	x	x	x	x												
2	Teste de proficiência em inglês										x											
3	Qualificação									x												
4	Preparação dos dados										x	x	x	x	x							
5	Preparação do modelo CNN												x	x	x							
6	Extração de características														x	x	x					
7	Classificação														x	x	x					
8	Avaliação																x	x				
9	Escrita da dissertação												x	x	x	x	x	x	x	x	x	
10	Defesa																					x

Fonte – Danilo Assunção, 2021

6 Considerações finais

Este trabalho propõe a utilização de visão computacional para a construção de um modelo de CNN, o qual realize o reconhecimento de espécies de abelhas a partir da morfologia de suas asas presentes em imagens digitais. Pretende-se utilizar conjunto a este mecanismo, técnicas de aumento de dados ou aprendizado por transferência. Inicialmente, foi realizada uma revisão sistemática da literatura, que identificou as técnicas de aumento de dados e aprendizado por transferência utilizando CNN para tarefas de reconhecimento de espécies. Esta revisão permitiu o entendimento das diversas possibilidades disponíveis para a aplicação de técnicas de aprendizado profundo, mesmo quando se tem poucos dados disponíveis, o que permite uma maior autonomia e flexibilidade no colhimento das características das imagens, junto a uma automatização como um todo no processo de reconhecimento. Com este método, espera-se obter bons resultados na classificação das espécies de abelhas, utilizando um modelo de CNN como automatizador de todo o processo de extração e classificação deste domínio.

Referências

- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: IEEE. *2017 International Conference on Engineering and Technology (ICET)*. [S.l.], 2017. p. 1–6. Citado na página 26.
- BANAN, A.; NASIRI, A.; TAHERI-GARAVAND, A. Deep learning-based appearance features extraction for automated carp species identification. *Aquacultural Engineering*, Elsevier, v. 89, p. 102053, 2020. Citado na página 41.
- BONACCORSO, G. *Machine learning algorithms*. [S.l.]: Packt Publishing Ltd, 2017. Citado 2 vezes nas páginas 19 e 20.
- CHEN, Y.; ZHANG, X.; CHEN, W.; LI, Y.; WANG, J. Research on recognition of fly species based on improved retinanet and cbam. *IEEE Access*, IEEE, v. 8, p. 102907–102919, 2020. Citado na página 42.
- CIBUK, M.; BUDAK, U.; GUO, Y.; INCE, M. C.; SENGUR, A. Efficient deep features selections and classification for flower species recognition. *Measurement*, Elsevier, v. 137, p. 7–13, 2019. Citado na página 42.
- DRAUSCHKE, M.; STEINHAGE, V.; VEGA, A. P. de la; MÜLLER, S.; FRANCOY, T. M.; WITTMANN, D. Reliable biometrical analysis in biodiversity information systems. In: *PRIS*. [S.l.: s.n.], 2007. p. 27–36. Citado na página 11.
- FRANCOY, T. M.; WITTMANN, D.; DRAUSCHKE, M.; MÜLLER, S.; STEINHAGE, V.; BEZERRA-LAURE, M. A.; JONG, D. D.; GONÇALVES, L. S. Identification of africanized honey bees through wing morphometrics: two fast and efficient procedures. *Apidologie*, EDP Sciences, v. 39, n. 5, p. 488–494, 2008. Citado na página 11.
- GÓMEZ-RÍOS, A.; TABIK, S.; LUENGO, J.; SHIHAVUDDIN, A.; HERRERA, F. Coral species identification with texture or structure images using a two-level classifier based on convolutional neural networks. *Knowledge-Based Systems*, Elsevier, v. 184, p. 104891, 2019. Citado 2 vezes nas páginas 42 e 50.
- GOULSON, D.; NICHOLLS, E.; BOTÍAS, C.; ROTHERAY, E. L. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science*, American Association for the Advancement of Science, v. 347, n. 6229, p. 1255957, 2015. Citado na página 11.
- GYIRES-TÓTH, B. P.; OSVÁTH, M.; PAPP, D.; SZÜCS, G. Deep learning for plant classification and content-based image retrieval. *Cybernetics and Information Technologies*, Sciendo, v. 19, n. 1, p. 88–100, 2019. Citado na página 43.
- HOLGADO, A. G.; PABLOS, S. M.; PEÑALVO, F. J. G. *et al.* Guidelines for performing systematic research projects reviews. *International Journal of Interactive Multimedia and Artificial Intelligence*, v. 6, n. 2, p. 9, 2020. Citado na página 36.
- HONG, S.; MEHDI, S. R.; HUANG, H.; SHAHANI, K.; ZHANG, Y.; JUNAIDULLAH, K. R.; KHAN, M. A. Classification of freshwater zooplankton by pre-trained convolutional neural network in underwater microscopy. 2020. Citado 2 vezes nas páginas 42 e 53.

- HSIANG, A. Y.; BROMBACHER, A.; RILLO, M. C.; MLENECK-VAUTRAVERS, M. J.; CONN, S.; LORDSMITH, S.; JENTZEN, A.; HENEHAN, M. J.; METCALFE, B.; FENTON, I. S. *et al.* Endless forams: 34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks. *Paleoceanography and Paleoclimatology*, Wiley Online Library, v. 34, n. 7, p. 1157–1177, 2019. Citado na página 42.
- KHALIFA, N. E. M.; TAHA, M. H. N.; HASSANIEN, A. E.; HEMEDAN, A. A. Deep bacteria: robust deep learning data augmentation design for limited bacterial colony dataset. *International Journal of Reasoning-based Intelligent Systems*, Inderscience Publishers (IEL), v. 11, n. 3, p. 256–264, 2019. Citado na página 42.
- KOLDA, T. G.; BADER, B. W. Tensor decompositions and applications. *SIAM review*, SIAM, v. 51, n. 3, p. 455–500, 2009. Citado na página 22.
- LAWTON, J. *Daily, GC (Ed.). 1997. Nature's services. Societal dependence on natural ecosystems*. [S.l.]: Island Press, Washington, DC, 1998. Citado na página 11.
- LE, V.-L.; BEURTON-AIMAR, M.; ZEMMARI, A.; MARIE, A.; PARISEY, N. Automated landmarking for insects morphometric analysis using deep neural networks. *Ecological Informatics*, Elsevier, v. 60, p. 101175, 2020. Citado 2 vezes nas páginas 11 e 41.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015. Citado na página 22.
- LEE, S. H.; CHAN, C. S.; MAYO, S. J.; REMAGNINO, P. How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition*, Elsevier, v. 71, p. 1–13, 2017. Citado na página 43.
- LI, K.; LIN, J.; LIU, J.; ZHAO, Y. Using deep learning for image-based different degrees of ginkgo leaf disease classification. *Information*, Multidisciplinary Digital Publishing Institute, v. 11, n. 2, p. 95, 2020. Citado 2 vezes nas páginas 42 e 52.
- LI, R.; WANG, R.; ZHANG, J.; XIE, C.; LIU, L.; WANG, F.; CHEN, H.; CHEN, T.; HU, H.; JIA, X. *et al.* An effective data augmentation strategy for cnn-based pest localization and recognition in the field. *IEEE Access*, IEEE, v. 7, p. 160274–160283, 2019. Citado na página 42.
- LI, Z.; HOIEM, D. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 40, n. 12, p. 2935–2947, 2017. Citado 2 vezes nas páginas 30 e 49.
- LIM, S.; KIM, S.; KIM, D. Performance effect analysis for insect classification using convolutional neural network. In: IEEE. *2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*. [S.l.], 2017. p. 210–215. Citado 2 vezes nas páginas 36 e 44.
- LIU, B.; DING, Z.; TIAN, L.; HE, D.; LI, S.; WANG, H. Grape leaf disease identification using improved deep convolutional neural networks. *Frontiers in Plant Science*, Frontiers, v. 11, p. 1082, 2020. Citado na página 41.

- LIU, S.; SHIH, F. Y.; RUSSELL, G.; RUSSELL, K.; PHAN, H. Classification of ecological data by deep learning. *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific, v. 34, n. 13, p. 2052010, 2020. Citado 7 vezes nas páginas 11, 12, 36, 41, 46, 56 e 58.
- LU, Y.-C.; TUNG, C.; KUO, Y.-F. Identifying the species of harvested tuna and billfish using deep convolutional neural networks. *ICES Journal of Marine Science*, Oxford University Press, v. 77, n. 4, p. 1318–1329, 2020. Citado na página 41.
- MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of machine learning*. [S.l.]: MIT press, 2018. Citado 4 vezes nas páginas 15, 16, 17 e 18.
- O'SHEA, K.; NASH, R. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015. Citado 7 vezes nas páginas 19, 22, 23, 24, 25, 26 e 27.
- PIMENTEL, D.; WILSON, C.; MCCULLUM, C.; HUANG, R.; DWEN, P.; FLACK, J.; TRAN, Q.; SALTMAN, T.; CLIFF, B. Economic and environmental benefits of biodiversity. *BioScience*, JSTOR, v. 47, n. 11, p. 747–757, 1997. Citado na página 11.
- RAUF, H. T.; LALI, M. I. U.; ZAHOR, S.; SHAH, S. Z. H.; REHMAN, A. U.; BUKHARI, S. A. C. Visual features based automated identification of fish species using deep convolutional neural networks. *Computers and Electronics in Agriculture*, Elsevier, v. 167, p. 105075, 2019. Citado 2 vezes nas páginas 42 e 54.
- REN, F.; LIU, W.; WU, G. Feature reuse residual networks for insect pest recognition. *IEEE Access*, IEEE, v. 7, p. 122758–122768, 2019. Citado na página 43.
- SHAO, L.; ZHU, F.; LI, X. Transfer learning for visual categorization: A survey. *IEEE transactions on neural networks and learning systems*, IEEE, v. 26, n. 5, p. 1019–1034, 2014. Citado na página 29.
- SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, Springer, v. 6, n. 1, p. 1–48, 2019. Citado 4 vezes nas páginas 31, 32, 33 e 34.
- SUN, Y.; LIU, Y.; WANG, G.; ZHANG, H. Deep learning for plant identification in natural environment. *Computational intelligence and neuroscience*, Hindawi, v. 2017, 2017. Citado 2 vezes nas páginas 44 e 53.
- TAN, J. wei; CHANG, S.-W.; ABDUL-KAREEM, S.; YAP, H. J.; YONG, K.-T. Deep learning for plant species classification using leaf vein morphometric. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, IEEE, v. 17, n. 1, p. 82–90, 2018. Citado na página 43.
- TETILA, E. C.; MACHADO, B. B.; MENEZES, G. V.; BELETE, N. A. de S.; ASTOLFI, G.; PISTORI, H. A deep-learning approach for automatic counting of soybean insect pests. *IEEE Geoscience and Remote Sensing Letters*, IEEE, v. 17, n. 10, p. 1837–1841, 2019. Citado na página 43.
- THENMOZHI, K.; REDDY, U. S. Crop pest classification based on deep convolutional neural network and transfer learning. *Computers and Electronics in Agriculture*, Elsevier, v. 164, p. 104906, 2019. Citado 2 vezes nas páginas 43 e 48.

- TIWARI, S. A comparative study of deep learning models with handcraft features and non-handcraft features for automatic plant species identification. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, IGI Global, v. 11, n. 2, p. 44–57, 2020. Citado na página [42](#).
- VALAN, M.; MAKONYI, K.; MAKI, A.; VONDRÁČEK, D.; RONQUIST, F. Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. *Systematic Biology*, Oxford University Press, v. 68, n. 6, p. 876–895, 2019. Citado 3 vezes nas páginas [43](#), [49](#) e [50](#).
- VILLON, S.; MOUILLOT, D.; CHAUMONT, M.; DARLING, E. S.; SUBSOL, G.; CLAVERIE, T.; VILLÉGER, S. A deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecological informatics*, Elsevier, v. 48, p. 238–244, 2018. Citado na página [43](#).
- WÄLDCHEN, J.; MÄDER, P. Machine learning for image based species identification. *Methods in Ecology and Evolution*, Wiley Online Library, v. 9, n. 11, p. 2216–2225, 2018. Citado na página [58](#).
- WANG, R.; ZHANG, J.; DONG, W.; YU, J.; XIE, C. J.; LI, R.; CHEN, T.; CHEN, H. A crop pests image classification algorithm based on deep convolutional neural network. *Telkomnika*, v. 15, n. 3, p. 1239–1246, 2017. Citado na página [44](#).
- YAMASHITA, R.; NISHIO, M.; DO, R. K. G.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, Springer, v. 9, n. 4, p. 611–629, 2018. Citado 7 vezes nas páginas [23](#), [24](#), [25](#), [26](#), [27](#), [28](#) e [29](#).
- YU, D.; WANG, H.; CHEN, P.; WEI, Z. Mixed pooling for convolutional neural networks. In: SPRINGER. *International conference on rough sets and knowledge technology*. [S.l.], 2014. p. 364–375. Citado na página [27](#).
- ZHANG, X.; CHEN, G. An automatic insect recognition algorithm in complex background based on convolution neural network an automatic insect recognition algorithm in complex background based on convolution neural network. Citado na página [41](#).
- ZHU, H.; LIU, Q.; QI, Y.; HUANG, X.; JIANG, F.; ZHANG, S. Plant identification based on very deep convolutional neural networks. *Multimedia Tools and Applications*, Springer, v. 77, n. 22, p. 29779–29797, 2018. Citado na página [43](#).
- ZHU, X.; ZHU, M.; REN, H. Method of plant leaf recognition based on improved deep convolutional neural network. *Cognitive Systems Research*, Elsevier, v. 52, p. 223–233, 2018. Citado na página [43](#).
- ZIELIŃSKI, B.; PLICHTA, A.; MISZTAL, K.; SPUREK, P.; BRZYCHCZY-WŁOCH, M.; OCHONSKA, D. Deep learning approach to bacterial colony classification. *PloS one*, Public Library of Science San Francisco, CA USA, v. 12, n. 9, p. e0184554, 2017. Citado na página [44](#).