

UNIVERSIDADE FEDERAL DO PARANÁ

ALESSANDRA KARINE CARNEIRO

CLASSIFICAÇÃO DE IMAGENS DE CARAVELAS-PORTUGUESAS UTILIZANDO
REDES NEURAS CONVOLUCIONAIS

CURITIBA

2022

ALESSANDRA KARINE CARNEIRO

CLASSIFICAÇÃO DE IMAGENS DE CARAVELAS-PORTUGUESAS UTILIZANDO
REDES NEURAS CONVOLUCIONAIS

Trabalho de conclusão de curso apresentado ao curso de Graduação em Ciência da Computação, Setor de Informática, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Ciência da Computação.

Orientadoras: Prof^a. Dr^a. Aurora Trinidad Ramirez Pozo e Prof^a. Dr^a. Carmem Satie Hara

CURITIBA

2022

AGRADECIMENTOS

Agradeço imensamente a minha família por todo carinho e incentivo, em especial à minha mãe Maristela, pelo apoio incondicional e por ser a minha maior referência.

Ao Victor, meu companheiro e amigo, por todo incentivo e por tornar os momentos difíceis mais leves.

À Prof^a. Aurora Pozo e à Prof^a Carmem Hara, pela oportunidade, orientação e auxílio durante a elaboração deste trabalho.

A todos que de alguma forma contribuíram para a minha formação.

RESUMO

As caravelas-portuguesas são organismos gelatinosos detentores de longos tentáculos capazes de causar queimaduras graves em seres humanos, o que pode provocar impactos negativos em atividades como o turismo e a pesca. A utilização de métodos alternativos para coleta de dados desses organismos pode contribuir com o seu monitoramento, uma vez que existem diversas lacunas de informações sobre a dinâmicas espaço-temporais dessa espécie. Um método que apresenta resultados promissores é a utilização dos dados presentes na rede social Instagram, através de uma metodologia de coleta e validação de registros publicados pelos usuários. O presente trabalho tem como objetivo validar a utilização de redes neurais convolucionais para a classificação de imagens de caravelas-portuguesas, com a intenção de automatizar uma das etapas da metodologia citada, que é realizada de forma manual. Foi gerado um banco de dados adequado para os experimentos, e as redes neurais utilizadas foram a VGG-16, a ResNet50 e a InceptionV3, com e sem treinamento na base de dados ImageNet, analisadas através das métricas de acurácia, precisão, recall e F1 Score. Os melhores resultados foram apresentados pela rede ResNet50 pré-treinada com a ImageNet, obtendo 94% na métrica de acurácia e 95% nas métricas de precisão, recall e F1 Score. A partir disso, é possível concluir que a utilização de redes neurais convolucionais pode ser bastante efetiva na automatização da etapa de extração de evidências, presente na metodologia de coleta de registros de caravelas-portuguesas da rede social Instagram.

Palavras-chave: Caravela-portuguesa. Aprendizagem Profunda. Redes Neurais Convolucionais. Rede Social.

ABSTRACT

Portuguese man-of-war are gelatinous organisms with long tentacles capable of causing severe burns in humans, which can have negative impacts on human activities, such as tourism and fishing. The use of alternative methods for collecting data from these organisms can contribute to their monitoring, since there is a lack of information about the spatiotemporal dynamics of this species. A method that presents promising results is the use of data presents in the social media Instagram, through a methodology for collecting and validating records published by the users. The present work aims to validate the use of convolutional neural networks for the Portuguese man-of-war images classification, with the intention of automating one of the steps of the aforementioned manual methodology. A database suitable for the experiments was developed, and the neural networks used were VGG-16, ResNet50 and InceptionV3, with and without training in the ImageNet database, analyzed through the accuracy, precision, recall and F1 Score metrics. The best results were presented by the ResNet50 network pre-trained with ImageNet, obtaining 94% in the accuracy metric and 95% in the precision, recall and F1 Score metrics. Therefore, it is possible to conclude that the use of convolutional neural networks can be very effective in automating the evidence extraction stage, present in the methodology for collecting portuguese man-of-war records from the social media Instagram.

Keywords: Portuguese Man-of-war. Deep Learning. Convolutional Neural Networks. Social Media.

LISTA DE FIGURAS

FIGURA 1 – METODOLOGIA DE COLETA DE DADOS DE CARAVELAS-PORTUGUESAS DAS REDES SOCIAIS	13
FIGURA 2 – REPRESENTAÇÃO MATEMÁTICA DE UM MODELO DE NEURÔNIO ARTIFICIAL	17
FIGURA 3 – REDE NEURAL FEEDFORWARD COM N CAMADAS	18
FIGURA 4 – ARQUITETURA DE UMA REDE NEURAL CONVOLUCIONAL	21
FIGURA 5 – PROCESSO DE CONVOLUÇÃO	22
FIGURA 6 – EXEMPLO DE MAX POOLING E AVERAGE POOLING	23
FIGURA 7 – EXEMPLO DE FLATTENING EM UMA ENTRADA 2X2X3	24
FIGURA 8 – DEMONSTRAÇÃO DA TÉCNICA EARLY STOPPING	27
FIGURA 9 – EXEMPLO DE DATA AUGMENTATION	28
FIGURA 10 – ESTRUTURA DO RESIDUAL LEARNING	31
FIGURA 11 – MÓDULO INCEPTION	32
FIGURA 12 – MATRIZ DE CONFUSÃO DE UM CLASSIFICADOR BINÁRIO	33
FIGURA 13 – IMAGENS COMPARATIVAS DE PHYSALIA PHYSALIS E VELELLA VELELLA	37
FIGURA 14 – AMOSTRA DE IMAGENS DO DATASET UTILIZADO NO SISTEMA JELLYTORING	38
FIGURA 15 – AMOSTRA DE IMAGENS PRESENTES NA BASE DE TREINAMENTO	47
FIGURA 16 – MATRIZ DE CONFUSÃO OBTIDA NO TESTE DA REDE RESNET50 PRÉ-TREINADA	51
FIGURA 17 – AMOSTRA DE FALSOS NEGATIVOS DA CLASSE CARAVELA	51
FIGURA 18 – AMOSTRA DE FALSOS POSITIVOS DA CLASSE CARAVELA	52

LISTA DE TABELAS

TABELA 1 – CATEGORIZAÇÃO DE IMAGENS ENCONTRADAS NOS RESULTADOS DA BUSCA DA HASHTAG #CARAVELAPORTUGUESA NO INSTAGRAM	41
TABELA 2 – IMAGENS OBTIDAS ATRAVÉS DA PLATAFORMA BING	44
TABELA 3 – QUANTIDADE DE IMAGENS PRESENTES NAS BASES DE TREINAMENTO, VALIDAÇÃO E TESTE	46
TABELA 4 – ACURÁCIA OBTIDA DURANTE O TREINAMENTO DE CADA ARQUITETURA	49
TABELA 5 – MÉDIA DOS RESULTADOS OBTIDOS NOS TESTES DE CADA ARQUITETURA	49
TABELA 6 – RESULTADOS POR CLASSE OBTIDOS NOS TESTES DE CADA ARQUITETURA	49

SUMÁRIO

1	INTRODUÇÃO	10
1.1	JUSTIFICATIVA	11
1.2	METODOLOGIA	12
1.3	OBJETIVOS	14
1.3.1	Objetivo geral	14
1.3.2	Objetivos específicos	15
1.4	ORGANIZAÇÃO DO CONTEÚDO	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	REDES NEURAS PROFUNDAS	16
2.1.1	Aprendizagem supervisionada	19
2.2	REDES NEURAS CONVOLUCIONAIS	20
2.2.1	Entrada	21
2.2.2	Arquitetura	21
2.2.3	Camadas convolucionais	22
2.2.4	Pooling	23
2.2.5	Flattening	24
2.2.6	Fully-connected	25
2.2.7	Camada de saída	25
2.2.8	Dropout	25
2.2.9	Treinamento	26
2.2.9.1	Data augmentation	28
2.2.10	Pré-treinamento	29
2.2.11	Arquiteturas referências	30
2.2.11.1	VGG-16	30
2.2.11.1	ResNet50	31
2.2.11.3	InceptionV3	32
2.2.11.4	Considerações	33
2.3	MÉTRICAS DE AVALIAÇÃO	33
2.4	CONSIDERAÇÕES	35
3	TRABALHOS RELACIONADOS	36

3.1	CARACTERÍSTICAS DAS CARAVELAS-PORTUGUESAS	36
3.2	ÁGUAS-VIVAS E APRENDIZAGEM PROFUNDA	37
3.2.1	Considerações	39
4	CONSTRUÇÃO DA BASE DE DADOS	41
4.1	ANÁLISE DE DADOS OBTIDOS DO INSTAGRAM	41
4.2	COLETA DOS DADOS	43
4.3	CONSIDERAÇÕES	45
5	ANÁLISE EXPERIMENTAL	46
5.1	IMPLEMENTAÇÃO DAS REDES NEURAIS	46
5.2	APRESENTAÇÃO DOS RESULTADOS	48
5.3	ANÁLISE DOS RESULTADOS	50
6	CONCLUSÃO	54
	REFERÊNCIAS	55

1 INTRODUÇÃO

A caravela-portuguesa (*Physalia physalis*) é um cnidário com aparência semelhante a de uma água-viva, mas se trata de um sifonóforo, um organismo composto por pequenos animais que vivem em colônia (TOTTON; MACKIE, 1960). Seus longos tentáculos são detentores de toxinas que, em contato com seres humanos, podem causar ferimentos graves e até fatais (PRIETO et al., 2015), gerando impactos negativos em diversas atividades, como o turismo e a pesca (CANEPA et al., 2020).

Os dados sobre a dispersão da espécie são de suma importância para entender melhor as dinâmicas das populações de caravelas-portuguesas, pois viabilizam interpretar e prever possíveis proliferações (BADALAMENTI et al., 2021). No entanto, existem diversas lacunas de informações sobre registros da espécie e seus impactos ao longo do tempo, causadas parcialmente pela dificuldade em coletar amostras desses animais, devido a sua fragilidade (RASKOFF, 2002).

No trabalho apresentado por (NASCIMENTO, 2020), é analisada a viabilidade de se utilizar dados extraídos de redes sociais para complementar as informações sobre a distribuição e fenologia de caravelas-portuguesas na costa brasileira. Para isso, uma metodologia de coleta, filtragem e armazenamento de dados de redes sociais foi estabelecida. O método foi aplicado na rede social Instagram. Apesar dos resultados serem preliminares e requererem mais análises, foi possível observar que a distribuição das caravelas-portuguesas baseada nas publicações coletadas é bastante similar às encontradas na literatura.

O processo de filtragem consiste, de forma simplificada, em analisar as informações presentes na publicação da rede social para validar se trata-se realmente de um registro de uma caravela-portuguesa. Para isso, são buscadas evidências como fotos ou vídeos do animal, além de informações sobre o local e a data da ocorrência. Se todas essas informações estiverem presentes, o registro é considerado válido.

Apesar dos resultados positivos, a metodologia aplicada, quando feita de forma manual, sofre as consequências das limitações humanas, resultando em um

processo cansativo e demorado. Além disso, a continuidade e expansão da aplicação do método exige um esforço humano permanente, tornando-a pouco escalável.

Parte do processo de filtragem consiste em uma tarefa de classificação de imagens. A utilização de técnicas de aprendizagem profunda para otimizar tarefas de mesma espécie com dados oriundos de redes sociais pode ser vista em diversos trabalhos, como em (KAGAYA; AIZAWA, 2015), (JEONG; CHO, 2017) e (HU et al., 2021). A parte do processo que envolve o tratamento do texto das publicações também pode ser otimizado com técnicas de aprendizagem profunda, porém não será abordada no presente trabalho.

O tratamento das mídias fundamenta-se na verificação da presença da caravela-portuguesa em pelo menos uma das imagens ou vídeos da publicação a ser analisada. Logo, esse processo pode ser modelado como um problema de classificação binária de imagens entre uma classe positiva e uma classe negativa para a presença da caravela-portuguesa, onde a classe negativa poderá ser representada por qualquer outra imagem aleatória.

Por se tratar de uma tarefa complexa, uma abordagem adequada para a resolução seria a utilização de Redes Neurais Convolucionais (do inglês *Convolutional Neural Networks* (CNN)), um modelo de rede neural profunda que apresentou bons resultados anteriores em aplicações de classificação de imagens complexas (ALBAWI et al., 2017).

Com isso, o presente trabalho propõe implementar e analisar alguns modelos de CNN aplicados ao problema de classificação de imagens de caravelas-portuguesas, com a intenção de validar a utilização desse tipo de tecnologia para a metodologia definida por (NASCIMENTO, 2020).

1.1 JUSTIFICATIVA

As caravelas-portuguesas são responsáveis pelos acidentes mais graves cometidos por gelatinosos na costa brasileira (MIGOTTO et al., 2010), além de causarem outros impactos socioeconômicos, como na pesca e no turismo (FERRER;

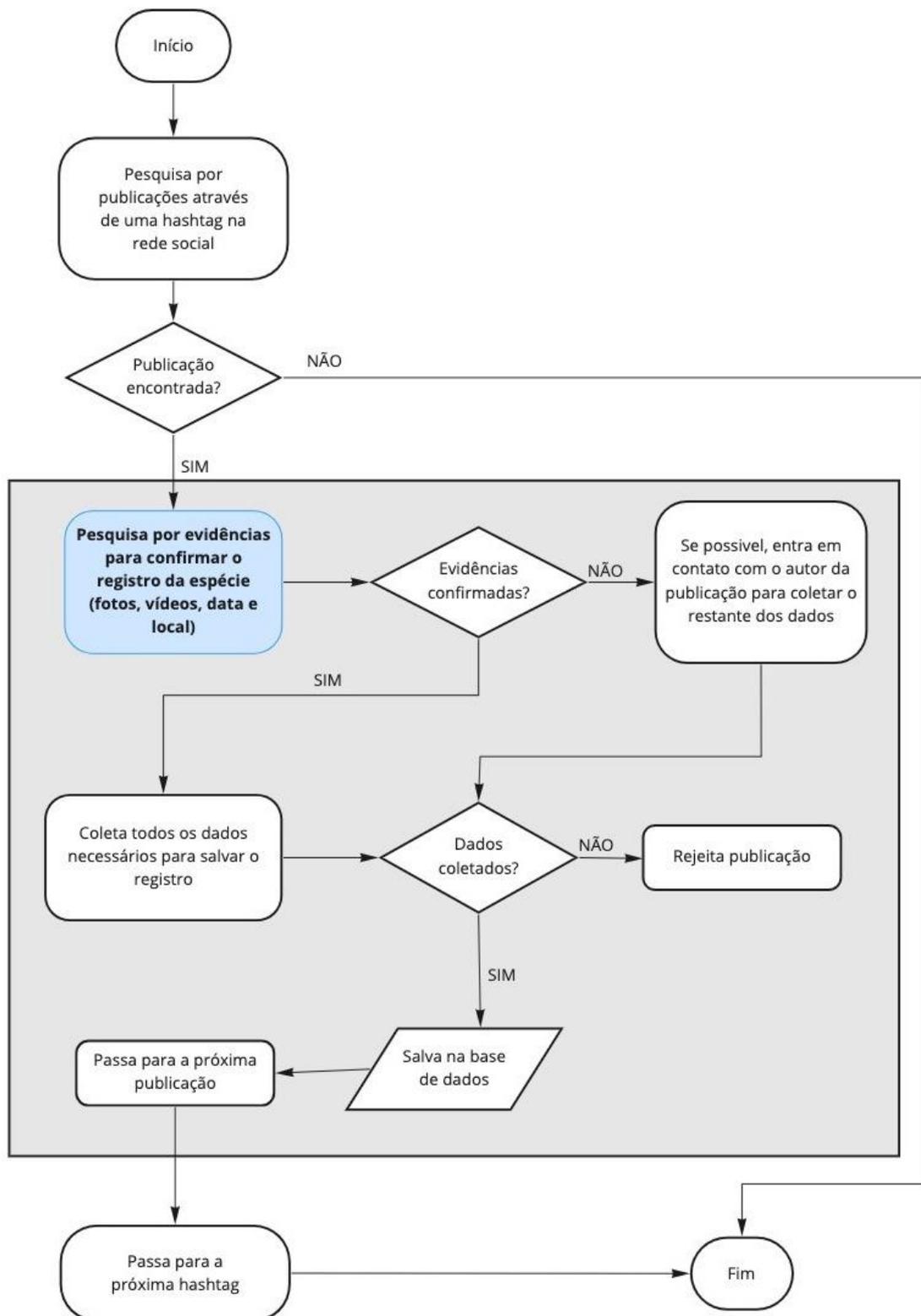
PASTOR, 2017). O entendimento das dinâmicas das populações da espécie é fundamental para a criação de medidas de prevenção e controle.

Por serem seres de difícil monitoramento, é possível se beneficiar da enorme quantidade de dados presentes nas redes sociais. Com um processo de filtragem eficiente, esses dados podem ser uma forma fácil e barata de se obter um complemento das informações já existentes da distribuição das caravelas-portuguesas (NASCIMENTO, 2020).

1.2 METODOLOGIA

Em (NASCIMENTO, 2020), é estabelecida uma metodologia para coleta manual, filtragem e armazenamento de dados de caravelas-portuguesas das redes sociais, mostrada de forma simplificada na Figura 1. O presente trabalho fundamenta-se na etapa de busca por evidências que confirmem o registro da caravela-portuguesa, destacada em azul na Figura 1, focando nos dados em formato de imagem.

FIGURA 1 – METODOLOGIA DE COLETA DE DADOS DE CARAVELAS-PORTUGUESAS DAS REDES SOCIAIS



FONTE: Adaptado de (NASCIMENTO, 2020).

O processo de validação das imagens coletadas como um registro válido para caravela-portuguesa foi modelado como um problema de classificação binária, onde uma imagem pode pertencer à classe Caravela, contendo imagens de caravelas-portuguesas, ou Não-Caravela, contendo imagens que não indicam um registro da espécie.

A escolha das CNNs como tecnologia base para resolução desse problema foi baseada no bom desempenho apresentado em trabalhos anteriores no tratamento de imagens complexas (ALBAWI et al., 2017), que consiste no tipo de imagem que compõe os dados de entrada.

Para a escolha dos modelos, foi utilizada como base a competição ImageNet Large Scale Visual Recognition Challenge (RUSSAKOVSKY et al., 2014), que tem como objetivo medir a eficiência de diversas arquiteturas para a classificação de imagens com o banco de imagens ImageNet. Os modelos selecionados foram a VGG-16 (SIMONYAN; ZISSERMAN, 2014), uma das arquiteturas destaques em 2014, a ResNet50 (HE et al., 2016), vencedora da competição em 2015, e a InceptionV3 (SZEGEDY et al., 2015), destaque também em 2015.

Para o treinamento e validação dos modelos, foi elaborada uma base de dados com uma classe Caravela e uma classe Não-Caravela. A classe Caravela é formada por imagens da caravela-portuguesa, enquanto a classe Não-Caravela é composta por imagens relacionadas às comumente encontradas na coleta dos dados estabelecida por (NASCIMENTO, 2020) que não se referem a uma caravela-portuguesa.

1.3 OBJETIVOS

1.3.1 Objetivo geral

Implementar e analisar métodos de classificação de imagens de caravelas-portuguesas com base em redes neurais convolucionais.

1.3.2 Objetivos específicos

- A. Construção da base de dados
- B. Implementação das arquiteturas VGG-16, ResNet50 e InceptionV3 que utilizem a base desenvolvida anteriormente para treinamento e validação
- C. Avaliação e comparação dos resultados obtidos pelos modelos implementados

1.4 ORGANIZAÇÃO DO CONTEÚDO

O trabalho está organizado da seguinte forma: no Capítulo 2 é apresentada a fundamentação teórica em que o presente estudo se baseia. No Capítulo 3 são apresentados os trabalhos relacionados ao tema. O Capítulo 4 contém a metodologia utilizada para a coleta dos dados que compõem a base de dados utilizada. O Capítulo 5 contém a análise experimental e apresentação dos resultados. Por fim, o Capítulo 6 apresenta as considerações finais do estudo.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados os principais assuntos que fundamentam esse trabalho, incluindo o embasamento teórico das técnicas de aprendizagem profunda utilizadas, arquiteturas e métricas escolhidas.

2.1 REDES NEURAIS PROFUNDAS

As técnicas de aprendizagem profunda, que utilizam como base modelos de redes neurais artificiais, têm como objetivo simular o comportamento do cérebro humano em tarefas como reconhecimento de imagens, reconhecimento de voz e processamento de linguagem natural (BEZERRA, 2016).

Como as redes neurais profundas são modelos matemáticos que se baseiam no funcionamento do cérebro humano, a sua estrutura básica é um neurônio artificial, que teve seu conceito apresentado pela primeira vez em (MCCULLOCH; PITTS, 1943). Este trabalho seminal serviu de base para a construção do perceptron em (ROSENBLATT, 1958).

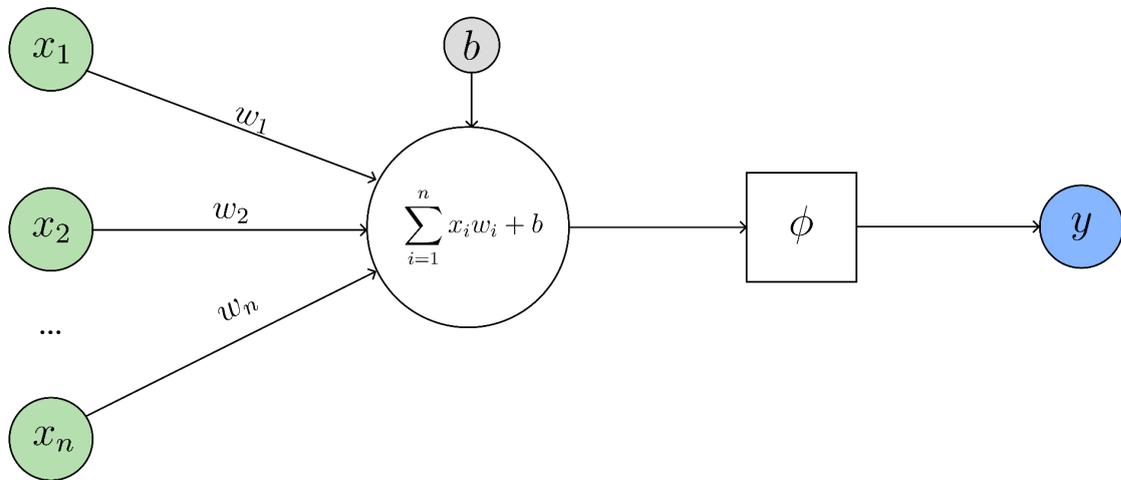
Um modelo genérico de neurônio recebe uma entrada vetorial $x \in \mathbb{R}^n$, e sua estrutura possui um viés b e um vetor de pesos $w \in \mathbb{R}^n$, onde $n \in \mathbb{N}$ e corresponde a dimensão da entrada (EMMERT-STREIB et al., 2020). Cada peso w_j fica associado a uma entrada x_j , representando a importância desta entrada (RAUBER, 2005). O somatório dos pesos multiplicados por cada entrada juntamente com o viés b resulta na saída z obtida pelo neurônio, como pode ser observado na fórmula abaixo (EMMERT-STREIB et al., 2020).

$$z = \sum_{i=1}^n x_i w_i + b$$

A saída do neurônio forma o argumento para uma função de ativação ϕ , que aplica uma transformação não-linear em z e resulta em y .

$$y = \phi(z)$$

FIGURA 2 – REPRESENTAÇÃO MATEMÁTICA DE UM MODELO DE NEURÔNIO ARTIFICIAL



FONTE: Adaptado de (EMMERT-STREIB et al., 2020).

O modelo de neurônio proposto por (MCCULLOCH; PITTS, 1943) utiliza a função de ativação Heaviside (função de escada), que pode ser definida como:

$$\phi(z) = \begin{cases} 1 & \text{se } z \geq 1 \\ 0 & \text{se } z < 0 \end{cases}$$

Já o modelo perceptron (ROSENBLATT, 1958) implementa como função de ativação a função sinal, definida por:

$$\phi(z) = \begin{cases} 1 & \text{se } z \geq 1 \\ -1 & \text{se } z < 0 \end{cases}$$

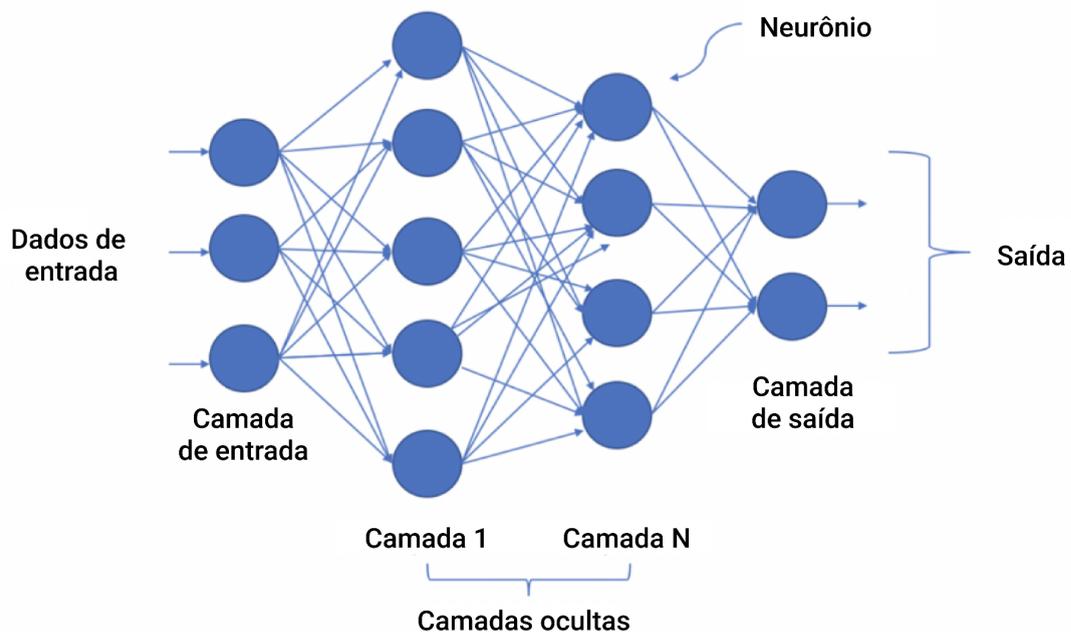
Existem diversos outros tipos de funções de ativação que podem ser utilizadas nesse processo, mas uma das mais populares atualmente é a ReLu (EMMERT-STREIB et al., 2020). A função ReLu aplica a seguinte operação na saída obtida pelo neurônio:

$$\phi(z) = \max\{0, z\}$$

O resultado dessa operação gera um resultado não-linear que possui determinados padrões evidenciados, importante para distinção de objetos em imagens complexas que podem conter diversos padrões irrelevantes para o reconhecimento do objeto (WU, 2017).

Uma rede neural é geralmente formada pela conexão de mais de um neurônio, e uma das estruturas mais simples nesse sentido é a rede feedforward, que pode ser visualizada na Figura 3. A informação de entrada da rede é passada em apenas uma direção através de todas as camadas, cada uma recebendo como entrada a saída da anterior. A primeira camada é denominada camada de entrada, já a última camada é denominada camada de saída, enquanto as camadas restantes são chamadas de camadas ocultas (DATA SCIENCE ACADEMY, 2021).

FIGURA 3 – REDE NEURAL FEEDFORWARD COM N CAMADAS



FONTE: Adaptado de (MOOLAYIL, 2019).

2.1.1 Aprendizagem supervisionada

As formas de aprendizagem de redes neurais podem ser divididas em duas categorias, a aprendizagem supervisionada e não-supervisionada. A aprendizagem supervisionada acontece quando o modelo é treinado com uma base de dados rotulada, e tem como objetivo prever um rótulo para uma entrada nova. Já a aprendizagem não-supervisionada é aplicada quando se tem uma base de dados sem rótulos e se deseja agrupar os dados semelhantes (BURKOV, 2019). No contexto do presente trabalho, será tratada apenas a aprendizagem supervisionada.

A aprendizagem supervisionada de redes neurais propõe utilizar o conjunto de treinamento para ajustar os valores dos pesos da rede, de forma a minimizar o erro de treinamento. Segundo (BEZERRA, 2016), o treinamento de uma rede neural dessa forma é convertido em um problema de otimização, cujo objetivo é minimizar o erro cometido pela rede, quando considerada toda a base de treinamento.

A comparação do resultado obtido pela rede com rótulo verdadeiro do objeto a ser classificado é feita através de uma função de perda. Existem diversas funções de perda que se adequam a cada objetivo, mas para problemas de classificação, uma das funções mais utilizadas é a entropia cruzada (do inglês *cross-entropy*) (WU, 2017), que pode ser definida por:

$$H(y_j, \bar{y}_j) = - \sum_{j=1}^M y_j \log \bar{y}_j$$

Onde M é a quantidade de objetos classificados, y_j é a probabilidade verdadeira e \bar{y}_j é a probabilidade obtida pela rede (DRAELOS, 2019).

Um problema bastante recorrente no treinamento de redes neurais, causado principalmente quando o conjunto de dados de treinamento não é grande o suficiente, é não conseguir obter uma boa capacidade de generalização, chamado de *overfitting*. Para evitar esse tipo de problema, é possível aplicar algumas técnicas conhecidas como técnicas de regularização, com a intenção de evitar o sobreajuste do modelo na base de treinamento (BEZERRA, 2016).

Um dos algoritmos mais utilizados como método de otimização de redes neurais é o gradiente descendente, também conhecido como GD. Seja J a função de perda aplicada em todo o conjunto de resultados obtidos por uma rede que possui um vetor de pesos θ . O GD considera a superfície multidimensional representada pela função J , e através de pequenas modificações nos pesos θ , tem a intenção de levar o vetor de pesos na direção da maior descida dessa superfície (BEZERRA, 2016).

Os métodos de otimização são comumente combinados com algoritmos de retropropagação de erros (do inglês *backpropagation*), fundamentais no treinamento de redes neurais (BEZERRA, 2016). O primeiro passo do algoritmo de retropropagação, chamado de propagação adiante (do inglês *forward step*), consiste em apresentar uma entrada para a rede e propagá-la até a camada de saída. O segundo passo, que embasa o nome do algoritmo, faz o caminho contrário pela rede, atualizando os valores dos pesos e vieses utilizados pelos neurônios, com a intenção de reduzir o erro cometido para aquela entrada (BEZERRA, 2016).

2.2 REDES NEURAIAS CONVOLUCIONAIS

Existem diversas arquiteturas de redes neurais profundas com as mais variadas aplicações. Segundo (BURKOV, 2019), uma das arquiteturas com resultados mais promissores na área de classificação de imagens é a Rede Neural Convolutiva (CNN). As CNNs ganharam popularidade em 2012 por sua utilização na popular competição de reconhecimento de objetos ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (RUSSAKOVSKY et al., 2014), onde obtiveram resultados bastante promissores em relação às outras tecnologias (BEZERRA, 2016).

Um dos principais fatores para o sucesso das CNNs é a sua capacidade de funcionar como extrator automático de características das imagens fornecidas, diferente de outras abordagens que geralmente exigem uma extração manual de características, como as máquinas de vetores de suporte (SVM) (BEZERRA, 2016).

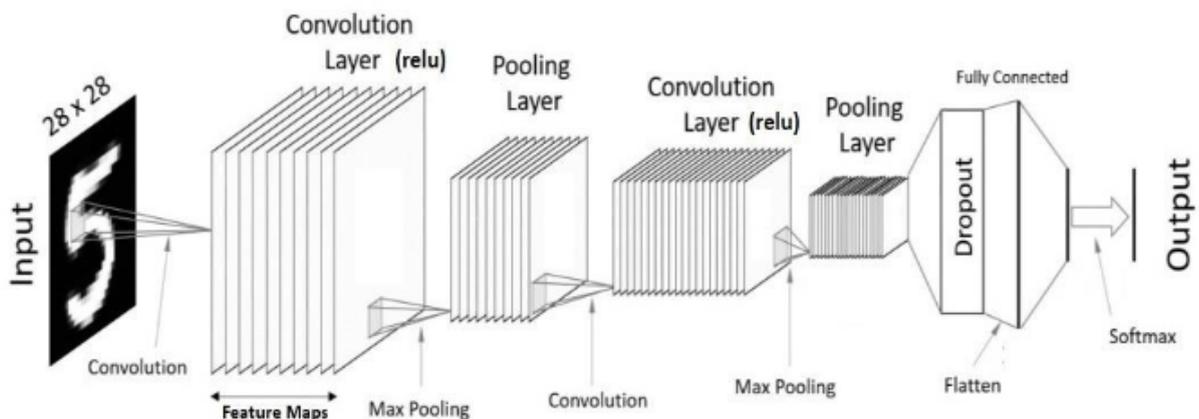
2.2.1 Entrada

Uma imagem de entrada de uma CNN é geralmente representada por uma matriz, que varia de tamanho conforme o número de linhas, colunas e canais da imagem. Se a imagem está salva em formato RGB, por exemplo, então ela possui 3 canais, M linhas e N colunas, e pode ser representada em uma matriz tridimensional de tamanho $M \times N \times 3$. Processar imagens coloridas é um processo mais custoso que processar imagens em nível de cinza, que são representadas por matrizes de tamanho $M \times N$. No entanto, os canais de cores podem conter informações extremamente relevantes em diversos problemas de reconhecimento de imagens (WU, 2017).

2.2.2 Arquitetura

A arquitetura de uma CNN pode possuir diversos formatos, dependendo de sua implementação. Como padrão, a entrada de uma CNN sempre passa por diversos processos sequencialmente, e cada etapa desses processos é chamada de camada. As camadas podem ser do tipo convolucional, *pooling*, normalização, totalmente conectada (do inglês *Fully Connected*), perda (do inglês *Loss*), entre outras (WU, 2017; ALOM et al., 2019). Uma arquitetura genérica de uma CNN pode ser observada na Figura 4.

FIGURA 4 – ARQUITETURA DE UMA REDE NEURAL CONVOLUCIONAL



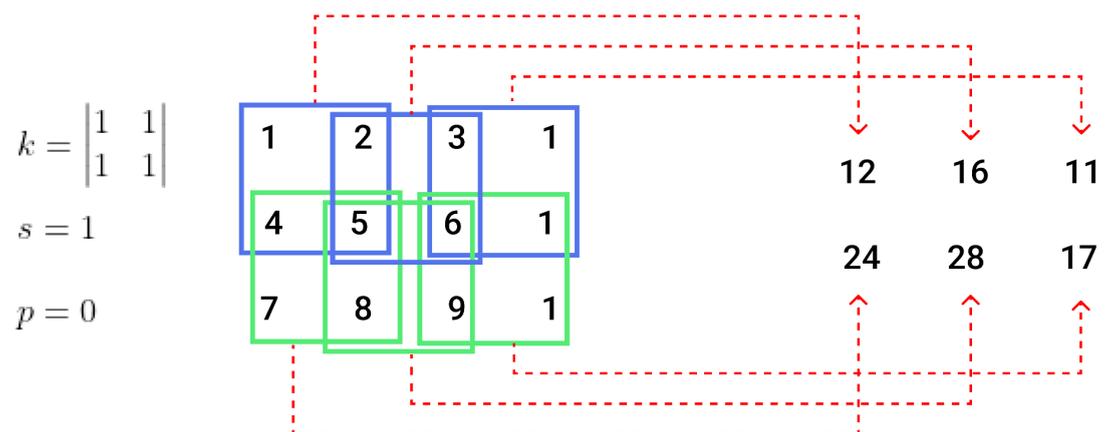
FONTE: (SHYAM, 2021).

2.2.3 Camadas convolucionais

As camadas convolucionais têm o objetivo de extrair características da sua entrada e formar um mapa de características, chamado de *feature map*. Cada entrada é convoluída com filtros que têm a capacidade de aprender e passa por uma função de ativação, formando um *feature map* de saída, que serve de entrada para a próxima camada da sequência (BOUVRIE, 2006).

A convolução é uma operação onde se aplica um filtro, também chamado de kernel (k), sobrepondo uma imagem e realizando operações de soma dos produtos dos elementos da imagem e do kernel que se encontram na mesma posição. O kernel se move por toda a imagem, da esquerda para direita e do topo para baixo, s pixels por vez, onde s é chamado de *stride* (WU, 2017). Também é definido um parâmetro p , chamado de *zero-padding*, que define a quantidade de zeros a serem preenchidos nas bordas da matriz de entrada, utilizado para manter a dimensão da entrada (EMMERT-STREIB et al., 2020).

FIGURA 5 – PROCESSO DE CONVOLUÇÃO



FONTE: Adaptado de (WU, 2017).

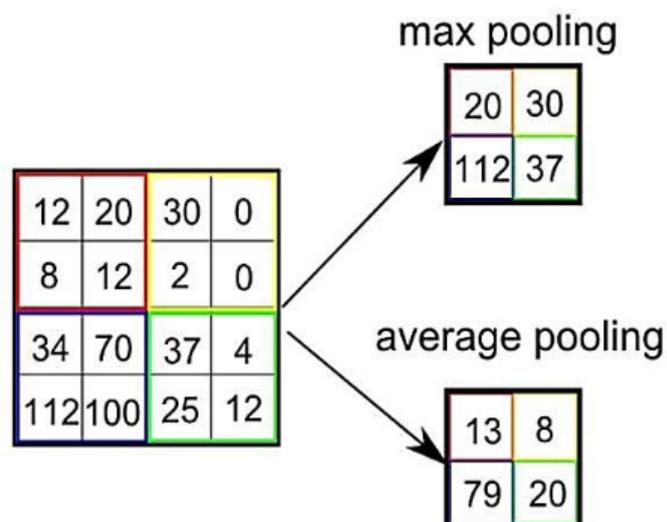
2.2.4 Pooling

As camadas de *pooling* são camadas sem parâmetros de aprendizagem, e atuam resumindo progressivamente a informação recebida pela camada convolucional anterior. Seu objetivo é reduzir a quantidade de características da entrada, fazendo com que a complexidade da rede seja menor e prevenindo a ocorrência de overfitting (SHYAM, 2021).

O *pooling* atua na entrada em uma área variável e pré-definida na construção da CNN, chamada de kernel, transformando tal área em apenas um número. O kernel se movimenta pela matriz de entrada s pixels por vez, onde s é chamado de *stride*, assim como os filtros das camadas convolucionais (SHYAM, 2021).

As funções pooling mais utilizadas são a *max pooling*, que mapeia uma região para o maior número do conjunto, e *average pooling*, que mapeia uma região para a média do conjunto (WU, 2017). Na Figura 6 é possível visualizar os resultados obtidos pela camada de pooling com a utilização de *max pooling* e *average pooling*, ambas com um kernel de tamanho 2×2 e $s = 2$.

FIGURA 6 – EXEMPLO DE MAX POOLING E AVERAGE POOLING

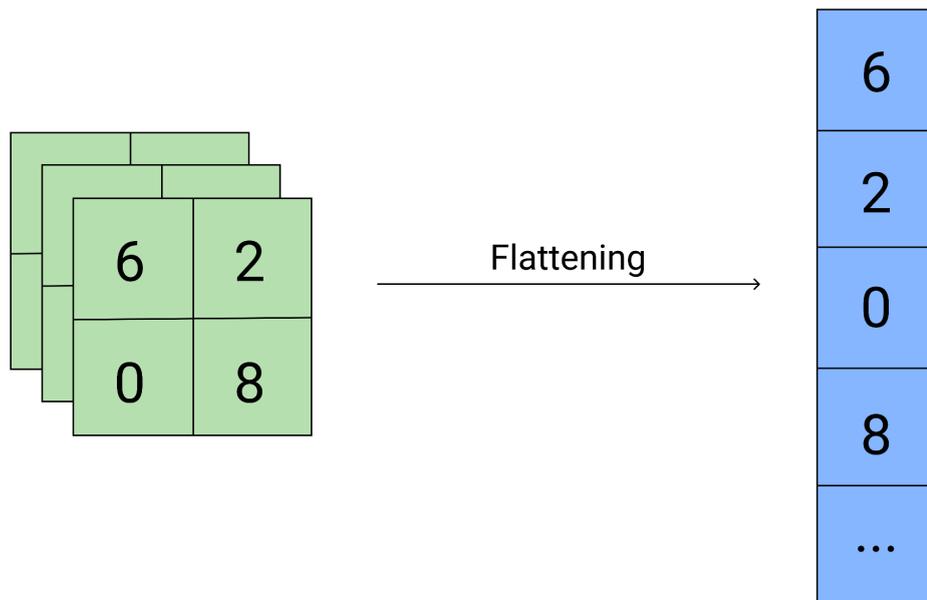


FONTE: (SHYAM, 2021).

2.2.5 Flattening

Após a entrada da rede passar por diversas camadas convolucionais e de pooling, ela servirá de entrada para as camadas *fully-connected*. Para isso, é importante transformar esses dados em dimensões adequadas, função exercida pela camada *flatten*. A saída das camadas anteriores geralmente está no formato matricial com dimensões $M \times N \times P$, e após passar pela camada *flatten*, se transforma em um vetor de dimensão \mathbb{R}^n , onde $n = M * N * P$ (SHYAM, 2021).

FIGURA 7 – EXEMPLO DE FLATTENING EM UMA ENTRADA 2X2X3



FONTE: Adaptado de (SHYAM, 2021).

2.2.6 Fully-connected

Após receber da camada *flatten* os mapas de características gerados pela camadas anteriores redimensionados em um vetor, a camada *fully-connected* insere esses dados como entrada de uma rede feedforward. A rede é modelada de acordo com o problema para retornar como saída um vetor de dimensão C , onde C é o número de classes apresentadas ao modelo.

As camadas *fully-connected* geralmente se encontram antes da camada de saída da CNN (BEZERRA, 2016), e são as responsáveis por determinar quais características estão mais relacionadas a cada uma das classes (SHYAM, 2021), através do formato de aprendizado descrito na seção 2.1.1.

2.2.7 Camada de saída

A camada de saída da CNN recebe, geralmente de uma camada *fully-connected*, um vetor x de dimensionalidade C , onde C corresponde ao número de classes apresentadas ao modelo. Seu retorno é um vetor também de dimensionalidade C , contendo a distribuição de probabilidade estimada de cada classe para uma determinada entrada da rede (WU, 2017).

A função mais utilizada na camada de saída é a softmax, que possui a seguinte definição:

$$\sigma(x) = \frac{e^{x_i}}{\sum_{j=1}^C e^{x_j}}, \text{ para cada } x_i \in x$$

Onde e representa a constante de Euler. Cada valor presente no vetor de saída de $\sigma(x)$ pertence ao intervalo $[0, 1]$ e a soma total dos valores é igual a 1 (BEZERRA, 2016).

2.2.8 Dropout

As camadas de Dropout são opcionais e podem ser inseridas como uma técnica de regularização no final da arquitetura da CNN. Têm o objetivo de promover a generalização dos resultados através da transformação de uma certa porcentagem de pesos da rede para 0, como uma forma de desconsiderá-los do resultado final. Esse processo é realizado com a intenção de prevenir o overfitting (WU, 2017).

2.2.9 Treinamento

O treinamento de uma CNN, assim como a maior parte dos modelos de aprendizado supervisionado, é geralmente realizado separando os dados disponíveis em três conjuntos: o conjunto de treinamento, de validação e de teste. O conjunto de treinamento é utilizado para construir o modelo e configurar os parâmetros da rede. Após cada treinamento, o modelo é exposto ao conjunto de validação, que serve para determinar o desempenho do modelo em dados desconhecidos, ajustando os seus pesos com a intenção de obter o menor erro possível nessa base (XU; GOODACRE, 2018).

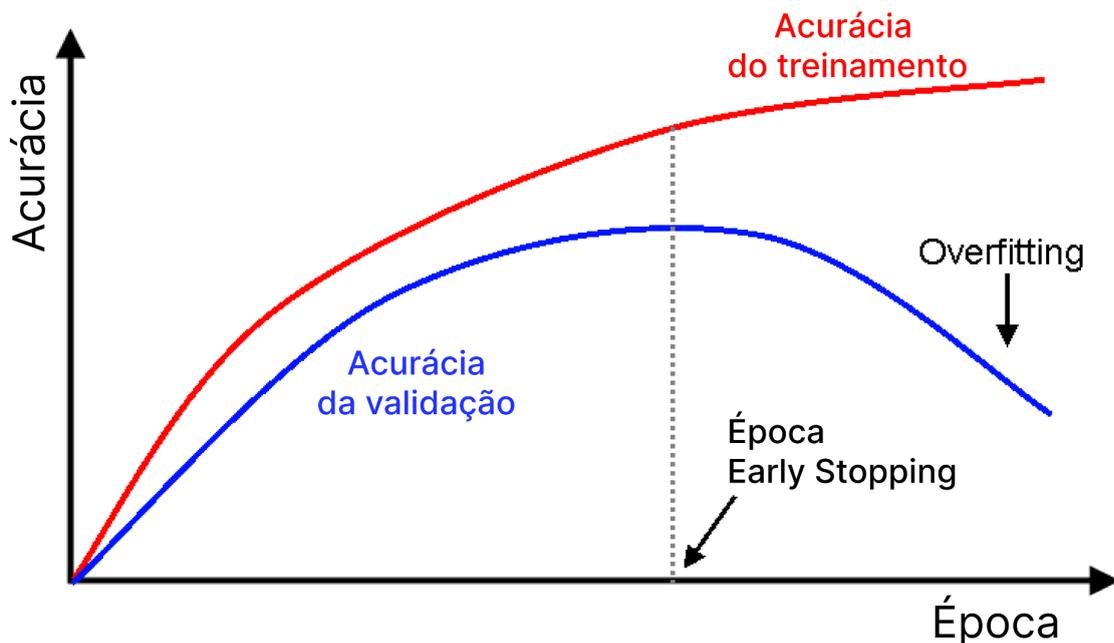
Segundo (XU; GOODACRE, 2018), até algumas décadas atrás, era comum utilizar o desempenho do modelo na base de validação como um avaliador imparcial do seu desempenho. Porém, estudos como (WESTERHUIS et al., 2008) e (HARRINGTON, 2017) demonstram que essa suposição nem sempre se sustenta, destacando a importância de se utilizar um conjunto de teste que não é exposto ao modelo durante o treinamento para avaliar o seu desempenho e sua capacidade de generalização.

Mesmo com a utilização de um conjunto de testes, é importante ressaltar que nem sempre o desempenho obtido será o mesmo em uma situação real. Isso porque os dados reais podem ser imprevisíveis e nem sempre seguem o mesmo padrão da base de dados utilizada (XU; GOODACRE, 2018). Por isso, a avaliação do modelo depende também que os dados utilizados nos três conjuntos simulem uma situação real com a maior confiabilidade possível.

Durante o treinamento, o conjunto de dados e validação é exposto ao modelo em uma quantidade de vezes definida por um hiperparâmetro chamado de época. Uma época se refere a uma passagem completa dos dados de treinamento e validação pelo modelo. Um número muito baixo de épocas pode não extrair toda a capacidade de aprendizagem do modelo sobre os dados de entrada, enquanto um número muito alto de épocas pode levar o modelo a apresentar overfitting, ou seja, perder a capacidade de generalização (DATA SCIENCE ACADEMY, 2021).

Uma técnica utilizada para se obter o número de épocas ideal é a *Early Stopping*, onde a cada época se calcula alguma métrica, como a acurácia, obtida na base de validação. A partir do momento em que essa métrica para de evoluir de forma positiva, o treinamento é interrompido, como mostrado na Figura 8 (DATA SCIENCE ACADEMY, 2021).

FIGURA 8 – DEMONSTRAÇÃO DA TÉCNICA EARLY STOPPING

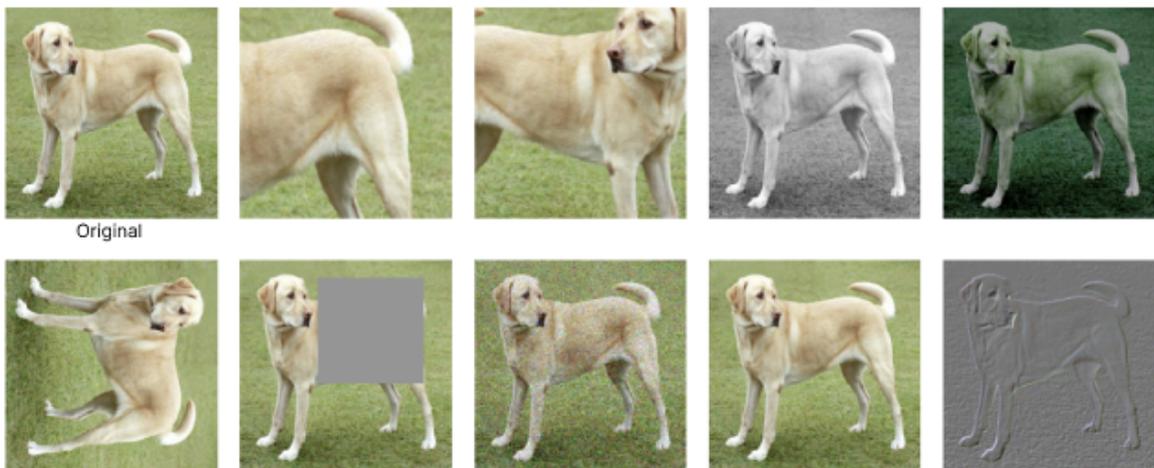


FONTE: Adaptado de (DATA SCIENCE ACADEMY, 2021)

2.2.9.1 Data augmentation

Uma técnica muito utilizada para melhorar o desempenho de modelos com dados de entrada limitados é a *data augmentation* (SHORTEN; KHOSHGOFTAAR, 2019). Essa técnica consiste em aplicar nas imagens de entrada existentes uma série de filtros, que podem realizar diversas operações, como rotação, transformações geométricas, alterações de cor, modificação de pixels aleatórios, entre outros. O objetivo é gerar novas imagens a partir das já existentes, sendo possível aumentar o tamanho da base de treinamento sem precisar de novas amostras (SHORTEN; KHOSHGOFTAAR, 2019).

FIGURA 9 – EXEMPLO DE DATA AUGMENTATION



FONTE: Adaptado de (CHEN et al., 2020).

2.2.10 Pré-treinamento

O uso de redes neurais pré-treinadas é baseado em uma área chamada de aprendizagem por transferência, na qual se avalia a utilização de uma rede neural treinada para um problema P_1 na resolução de um problema P_2 , sendo que ambos os problemas possuem entradas relacionadas. Se P_1 possui significativamente mais amostras de treinamento que P_2 , então a utilização da rede treinada em P_1 para resolver P_2 pode ajudar a melhorar o aprendizado de características úteis e a generalização do segundo problema (GOODFELLOW et al., 2016).

Em redes neurais convolucionais, essa técnica se tornou popular para resolução de problemas sem uma quantidade grande o suficiente de amostras, utilizando redes pré-treinadas em grandes bancos de imagens, como o ImageNet, para aprimorar os resultados (GOODFELLOW et al., 2016).

Uma das formas de aplicação de aprendizagem por transferência é a utilização de *fine-tuning*, uma técnica onde se aplica ajustes em uma rede pré-treinada para que ela se adapte a um novo problema. Esses ajustes geralmente incluem a substituição da camada de saída original por uma camada de saída que se adeque ao novo problema, além da inserção de novas camadas totalmente conectadas e técnicas de regularização (ROSEBROCK, 2017).

Além de realizar as alterações na estrutura da rede, é realizado um congelamento total ou parcial das camadas pré-existentes, tornando-as incapazes de sofrerem modificações nos seus pesos. Esse procedimento tem o objetivo de impedir que a rede tenha os seus pesos antigos modificados com os novos dados de treinamento, e conseqüentemente, perca o seu aprendizado robusto já existente (ROSEBROCK, 2017).

O treinamento com os novos dados é realizado e somente as camadas da rede que não estão congeladas são atualizadas e aprendem os recursos do novo problema. Opcionalmente, é possível descongelar algumas ou todas as camadas após uma sequência de treinamento e realizá-lo novamente com uma taxa de aprendizagem baixa, com a intenção de aprimorar os resultados (ROSEBROCK, 2017).

2.2.11 Arquiteturas referências

Atualmente, existem diversas arquiteturas de CNNs populares consideradas como estado-da-arte, pois conseguem atingir um bom desempenho em diferentes tarefas de classificação (ALOM et al., 2019).

A comparação de desempenho de arquiteturas aplicadas em grandes bancos de imagens têm sido fundamental para a orientação do desenvolvimento de modelos aplicados em tarefas variadas (TSIPRAS et al., 2020). Uma competição muito utilizada nesse sentido é a ILSVRC (RUSSAKOVSKY et al., 2014), que utiliza o banco de imagens ImageNet composto por mais de 14 milhões de imagens. Algumas arquiteturas que ganharam destaque nesta competição serão descritas a seguir.

2.2.11.1 VGG-16

A arquitetura VGG (Visual Geometry Group) definida em (SIMONYAN; ZISSERMAN, 2014) é um modelo treinado que foi destaque na competição em 2014. Ela possui seis tipos de camadas e um total de 13 camadas convolucionais e 3 camadas *fully-connected*. Cada camada convolucional é seguida de uma camada ReLu, aumentando a não-linearidade dos modelos. Entradas válidas para esse modelo são imagens de tamanho 224 x 224 x 3 (WU, 2017).

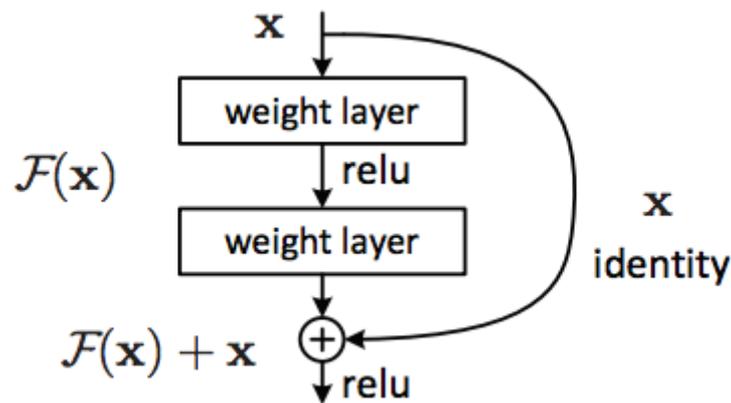
Suas variações consistem em dois modelos que se diferem pela quantidade de camadas convolucionais, e são denominados VGG-16 e VGG-19. Ambos os modelos são finalizados com uma camada de saída que utiliza a função softmax.

2.2.11.2 ResNet50

Em teoria, as redes neurais aumentam a acurácia das predições com o aumento do número de camadas na sua estrutura (DONAHUE et al., 2014). Porém, em alguns estudos como o de (HE et al., 2016), foi observado que o excesso de camadas pode afetar negativamente o desempenho da CNN. Isso acontece pois a função de perda acaba se otimizando para mínimos locais, causando um problema chamado degradação do gradiente (HE et al., 2016).

O modelo ResNet foi introduzido em (HE et al., 2016) com o objetivo de resolver esse problema. A técnica *Residual Learning*, aplicada pelo modelo, consiste em treinar as camadas com uma entrada residual ao invés de treiná-las com a saída da camada adjacente, como é feito na forma padrão. Com isso, a entrada passa a ser $F(x) + x$, como pode ser observado na Figura 10.

FIGURA 10 – ESTRUTURA DO RESIDUAL LEARNING



FONTE: (HE et al., 2016).

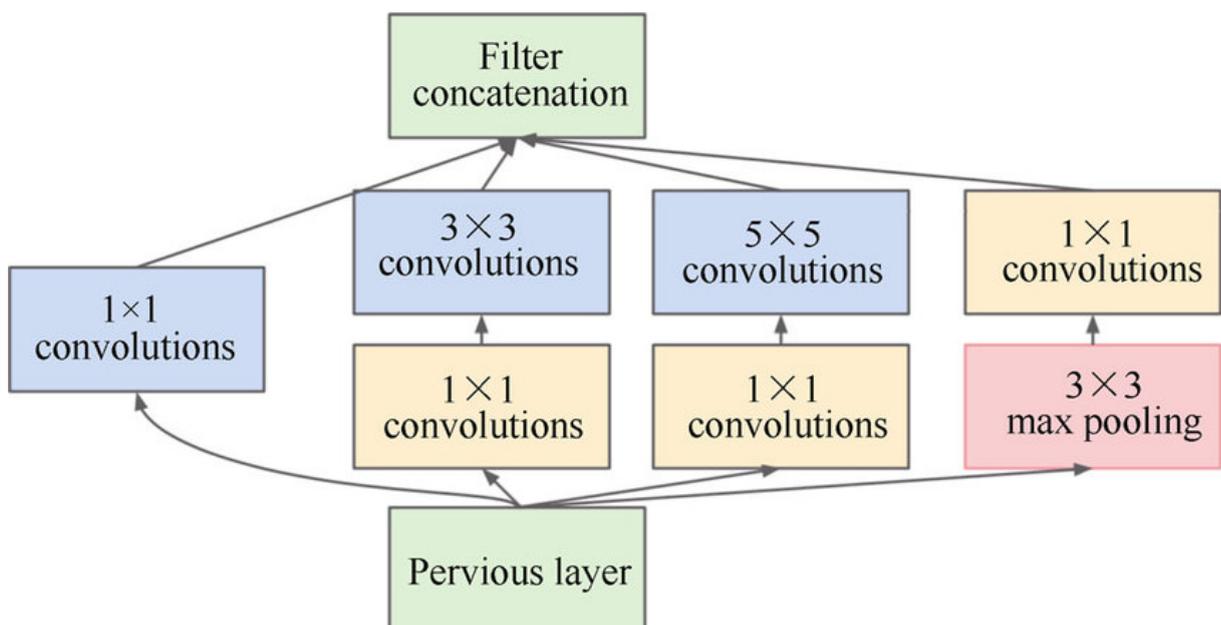
A arquitetura ResNet foi a vencedora do desafio ILSVRC em 2015 nas categorias de classificação, detecção e localização de imagens, além de outros desafios reconhecidos (TSANG, 2018). Ela pode ser desenvolvida com diversas quantidades de camadas, mas alguns de seus padrões mais populares são o ResNet50, com 50 camadas, e o ResNet101, com 101 camadas.

2.2.11.3 InceptionV3

Definida em (SZEGEDY et al., 2015), a arquitetura Inception foi criada com a intenção de otimizar a utilização dos recursos computacionais, de forma que a rede consiga se tornar mais complexa sem se tornar tão custosa computacionalmente. A arquitetura possui algumas variações, como a InceptionV1, InceptionV2 e InceptionV3.

O principal diferencial da arquitetura Inception é a utilização de um módulo Inception, mostrado na Figura 11. Seu objetivo é conseguir extrair mais de uma característica por camada, realizando convoluções com filtros de tamanho 1×1 , 3×3 e 5×5 simultaneamente. Essa operação otimiza a busca por características, já que em redes comuns é aplicado apenas um filtro por camada.

FIGURA 11 – MÓDULO INCEPTION



FONTE: (SZEGEDY et al., 2015).

Além disso, também é possível combinar a arquitetura Inception com outras arquiteturas, criando arquiteturas híbridas, como a ResNet-Inception, que mistura características da Inception com a ResNet.

2.2.11.4 Considerações

Além das arquiteturas apresentadas, existem diversos outros modelos que obtiveram resultados satisfatórios no desafio ILSVRC e são consideradas referências. Os modelos selecionados para o presente estudo foram escolhidos por, além de referências em desempenho, serem facilmente implementadas e possuírem estruturas bastante distintas entre si. A intenção dessa seleção é avaliar o comportamento de cada modelo aplicado na base de dados para entender qual deles melhor se encaixa no problema apresentado.

2.3 MÉTRICAS DE AVALIAÇÃO

Algumas das formas mais populares de se avaliar um modelo classificador é através de fórmulas que se baseiam na matriz de confusão gerada a partir dos resultados do modelo. Levando em consideração um classificador binário que classifica cada entrada como positiva ou negativa, existem quatro categorias de resultados possíveis: se a entrada é positiva e é classificada como positiva, então o resultado é considerado um verdadeiro positivo (VP); se é classificada como negativa, então o resultado é um falso negativo (FN); se a entrada é negativa e é classificada como negativa, então o resultado é um verdadeiro negativo (VN); se é classificada como positiva, então o resultado é um falso positivo (FP) (FAWCETT, 2006). A partir dessa categorização dos resultados é possível gerar a matriz de confusão exemplificada na Figura 12, em que a diagonal principal possui os valores classificados corretamente e a diagonal secundária possui os valores classificados incorretamente (FAWCETT, 2006).

FIGURA 12 – MATRIZ DE CONFUSÃO DE UM CLASSIFICADOR BINÁRIO

		Valores reais	
		Positivo	Negativo
Valores previstos	Positivo	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Negativo	Falso Negativo (FN)	Verdadeiro Negativo (VN)

FONTE: Adaptado de (FAWCETT, 2006).

Uma das métricas de desempenho mais básicas gerada a partir da matriz de confusão é a acurácia, que mede a porcentagem total de acertos cometidos pelo classificador, definida pela fórmula:

$$acurácia = \frac{VP + VN}{P + N}$$

Em que P se refere a quantidade total de amostras positivas e N a quantidade total de amostras negativas. Outra métrica comumente usada é chamada de recall, que tem o objetivo de mensurar a quantidade de classes positivas classificadas corretamente, definida na fórmula abaixo:

$$recall = \frac{VP}{P}$$

A fórmula da precisão, definida abaixo, mede a quantidade de resultados classificados como positivos que são realmente corretos.

$$precisão = \frac{VP}{VP + FP}$$

A última métrica apresentada nessa seção é F1 Score, que utiliza as métricas de precisão e recall em sua fórmula, com a intenção de medir a qualidade geral do modelo.

$$F1\ Score = 2 \times \frac{precisão \times recall}{precisão + recall}$$

2.4 CONSIDERAÇÕES

Neste capítulo foram apresentados a base teórica de Redes Neurais Profundas, na seção 2.1, e das Redes Neurais Convolucionais, na seção 2.2, que engloba também as arquiteturas que serão utilizadas na classificação das imagens de caravela-portuguesa. As métricas utilizadas na análise experimental para avaliar o desempenho das diferentes arquiteturas estão apresentadas na seção 2.3.

3 TRABALHOS RELACIONADOS

Neste capítulo serão apresentados os principais tópicos de pesquisas em trabalhos relacionados que fundamentam o presente trabalho. Primeiramente, será apresentada a pesquisa relacionada às caravelas-portuguesas, com foco em suas características visuais. A seguir, serão exibidos os trabalhos encontrados com foco em aplicações de aprendizagem profunda em imagens de águas-vivas.

3.1 CARACTERÍSTICAS DAS CARAVELAS-PORTUGUESAS

Para a tarefa de classificação de imagens, é importante realizar uma análise das principais características físicas do objeto de estudo e em quais ambientes ele é comumente fotografado, com a intenção de entender como a base de dados está estabelecida e os possíveis erros que podem ser cometidos pelos modelos de redes neurais.

As caravelas-portuguesas, *Physalia physalis*, são organismos que vivem em colônia e pertencem ao grupo dos cnidários, da ordem dos sifonóforos. Os indivíduos que compõem a colônia são dependentes entre si, e cada um possui uma função específica (MUNRO, 2019). O maior destaque visual de sua composição é o seu pneumatóforo, uma estrutura em formato de balão preenchida com gás utilizada na navegação, pois possui cores vívidas em tons rosados e azulados (BARDI e MARQUES, 2007).

Apesar de ser bastante característica, as cores e o formato da caravela-portuguesa podem ser confundidas com outro cnidário: a *Velella velella*, popularmente conhecida como veleiro. Ambas as espécies são oceânicas e flutuam livremente na superfície do mar, e eventualmente são carregadas para a costa e encalham na areia (BADALAMENTI et al., 2021; GHILARDI, 2019).

Analisando a Figura 13, é possível perceber que as imagens em que as caravelas-portuguesas possuem uma coloração mais azulada e estão em grupos podem ser facilmente confundidas com as imagens do veleiro, até mesmo por seres humanos.

FIGURA 13 – IMAGENS COMPARATIVAS DE PHYSALIA PHYSALIS E VELELLA VELELLA



FONTE: Compilação do autor (2022)¹

3.2 ÁGUAS-VIVAS E APRENDIZAGEM PROFUNDA

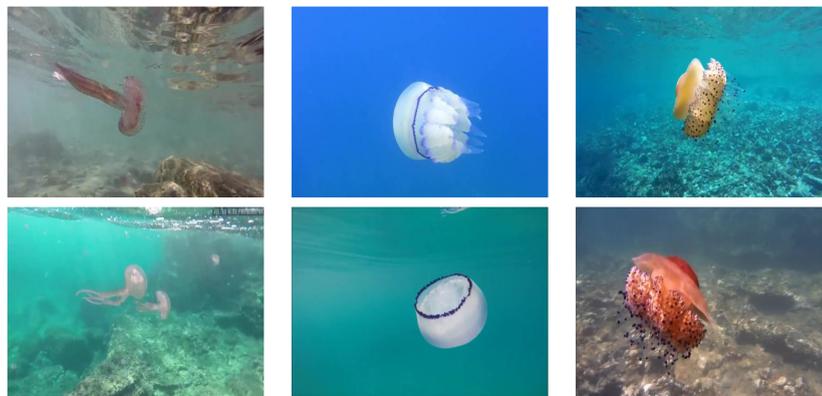
Essa seção apresenta os trabalhos encontrados que relacionam organismos gelatinosos com aprendizagem profunda. O objetivo deste estudo é buscar pesquisas que servissem de base para o estudo atual, apresentando tecnologias utilizadas e seus resultados. Como não foram encontradas muitas pesquisas dessa categoria que estude especificamente a caravela-portuguesa, foram buscados trabalhos que tenham como objeto de estudo qualquer espécie de água-viva ou organismo gelatinoso, animais fisicamente semelhantes à caravela-portuguesa.

Em (MARTIN-ABADAL et al., 2020) é apresentado o sistema Jellytoring, que tem como objetivo detectar e quantificar diferentes espécies de águas-vivas através

¹ Montagem a partir de imagens coletadas do site iNaturalist.

de uma CNN, permitindo a geração de um histórico de presença desses animais. As imagens de entrada desse modelo são provenientes de gravações de vídeos subaquáticos, e são referentes a três espécies: *Pelagia noctiluca*, *Cotylorhiza tuberculata* e *Rhizostoma pulmo*. A arquitetura utilizada foi a Inception-Resnet-v2 pré-treinada com o banco de dados COCO, que foi comparada com as arquiteturas ResNet101 e InceptionV2, e chegou a obter resultados da métrica F1 Score maior que 95%.

FIGURA 14 – AMOSTRA DE IMAGENS DO DATASET UTILIZADO NO SISTEMA JELLYTORING



FONTE: (MARTIN-ABADAL et al., 2020).

O estudo apresentado em (GAO et al., 2021) analisa o problema de detecção e classificação de sete espécies de águas-vivas e peixes em imagens subaquáticas, com a intenção de promover o monitoramento dessas espécies. A arquitetura empregada foi a YOLOv3, uma CNN utilizada para detectar objetos em uma imagem. Os autores realizam uma série de aprimoramentos na aplicação da CNN e conseguem atingir o resultado da métrica acurácia em 95,53%.

(MCLIWAINE; CASADO, 2021) apresenta o JellyNet, um modelo de CNN treinado para detectar afloramentos (*bloom*, em inglês) de águas-vivas a partir de imagens de alta resolução obtidas por veículos aéreos não tripulados (VANT), também conhecidos como drones. A arquitetura utilizada foi baseada na VGG-16, e chegou a atingir a acurácia de 90%.

Em (RECORD et al., 2018), é testado um sistema de previsão de ocorrência de três espécies de águas-vivas, *Cyanea capillata*, *Aurelia aurita* e *Staurostoma mertensii*, no Golfo do Maine. O sistema utilizou dados provenientes de satélites e também dados enviados por cidadãos através do e-mail e da rede social Twitter para treinar um modelo de predição. Para engajar os cidadãos a coletarem e enviarem dados para a pesquisa, foram realizados anúncios através de redes sociais e redes institucionais solicitando o envio. No total, foram coletados 259 registros via e-mail e 29 via Twitter. Em geral, o modelo apresentou resultados promissores. No entanto, os autores ressaltam que esses resultados não implicam necessariamente em um bom desempenho de previsão real, e que a generalidade do modelo ficará mais clara conforme previsões em períodos maiores de tempo forem incluídas.

O estudo apresentado em (CORREIA et al., 2020) tem o objetivo de desenvolver um sistema de baixo custo para detecção, reconhecimento e classificação de diversos tipos de incidentes em áreas costeiras através de imagens obtidas por drones. Os incidentes incluem desde derramamentos de petróleo até o aparecimento de criaturas marinhas perigosas, como a caravela-portuguesa. A primeira parte da solução é composta por uma CNN para processar as imagens obtidas pelos drones, e a segunda parte é um aplicativo de celular para gerenciar os drones. A solução ainda está sendo desenvolvida pelos autores.

3.2.1 Considerações

Apesar da semelhança do objeto de estudo do presente trabalho com os trabalhos citados, devido às características físicas das caravelas-portuguesas serem parecidas com a de outras águas-vivas, quase nenhum dos modelos poderia ser aplicado no problema de classificação aqui apresentado, pois possuem outra função. Em (MARTIN-ABADAL et al., 2020) e (MCLIWAIN; CASADO, 2021), o objetivo é detectar águas-vivas, ou seja, a intenção é encontrar onde e quantos animais se encontram na imagem. Em (GAO et al., 2021), o objetivo é também detectar águas-vivas e classificá-las, mas utilizando imagens subaquáticas e entre espécies

diferentes da aqui estudada. Em (RECORD et al., 2018), a função do modelo é prever o aparecimento de águas-vivas em uma região. Em (CORREIA et al., 2020), o modelo tem um objetivo mais abrangente do que o aqui apresentado, mas ainda assim utiliza imagens de caravelas-portuguesas. No entanto, não é possível validar se o modelo pode servir de base para o problema deste estudo pois ele ainda está em desenvolvimento.

Além disso, a utilização de imagens subaquáticas ou imagens aéreas para treinamento é um fator bastante relevante, uma vez que a maior parte das imagens de caravelas-portuguesas obtidas do Instagram são dos animais encalhados na areia, tornando-as bastante distintas.

A coleta de dados, na maior parte dos trabalhos citados, foi realizada de maneira ativa, onde se emprega um trabalho para realizar essa tarefa. No presente trabalho, se faz uso da grande quantidade de dados presentes nas redes sociais, mais especificamente no Instagram. Tais dados não são coletados especificamente para esse fim, mas são uma fonte útil de informações sobre a biodiversidade, além de serem baratos e fáceis de serem coletados.

É interessante observar que, os trabalhos de (MARTIN-ABADAL et al., 2020), (MCLIWAINE; CASADO, 2021), (GAO et al., 2021) e (CORREIA et al., 2020), recebem imagens como objeto de entrada e utilizam alguma CNN como arquitetura base. Isso confirma a popularidade dessa tecnologia no tratamento de imagens complexas.

Por fim, entre os trabalhos apresentados, quase todos são relacionados a espécies de organismos gelatinosos diferentes da aqui estudada, apenas (CORREIA et al., 2020) tem algum envolvimento com imagens de caravelas-portuguesas.

4 CONSTRUÇÃO DA BASE DE DADOS

Neste capítulo são apresentados os métodos utilizados para a coleta de imagens e construção da base de dados.

4.1 ANÁLISE DE DADOS OBTIDOS DO INSTAGRAM

Para realizar a coleta dos dados que servem de entrada para as redes neurais, primeiramente foi feita uma análise de uma amostra de 170 imagens obtidas através de busca da hashtag #caravelaportuguesa na rede social Instagram. O objetivo dessa busca é entender quais os tipos de dados a rede neural deve aprender para conseguir obter bons resultados de classificação das imagens.

As imagens foram contabilizadas por categorias de forma manual e os resultados obtidos podem ser vistos na Tabela 1.

TABELA 1 – CATEGORIZAÇÃO DE IMAGENS ENCONTRADAS NOS RESULTADOS DA BUSCA DA HASHTAG #CARAVELAPORTUGUESA NO INSTAGRAM

Tipo de imagem	Quantidade de imagens	Porcentagem
Caravela-portuguesa	102	60%
Pessoa	15	8,82%
Embarcação	14	8,23%
Ilustração	9	5,52%
Tatuagem	5	2,94%
Outros cnidários	10	5,88%
Outros	15	8,82%

FONTE: O autor (2022).

As imagens de caravelas-portuguesas foram as que apareceram com mais frequência. Dentre as 102 imagens observadas, em 97 a caravela-portuguesa se encontrava encaçada na areia, em 3 o animal estava sendo segurado por uma

pessoa, que também pode ser vista na imagem, e em 2 a caravela-portuguesa estava flutuando no mar.

Imagens de uma ou mais pessoas aparecem em cerca de 8,8% da amostra observada, sem contar as imagens de pessoas que seguram uma caravela-portuguesa, e geralmente são fotografadas em praias ou outras paisagens naturais.

Caravela é o nome de um tipo de embarcação criada pelos portugueses (DE SOUZA, 2007), e isso faz com que a hashtag #caravelaportuguesa seja utilizada também em imagens de embarcações desse tipo, em tamanho real ou em réplicas em miniaturas, que compõem cerca de 8,2% da amostra analisada.

Tatuagens e ilustrações somam quase 8,5% dos resultados, e são geralmente encontradas como representações tanto da *Physalia physalis* quanto da embarcação caravela.

Dentre os outros cnidários que aparecem nos resultados, cerca de 3 aparições são de veleiros e os outros 7 de espécies de águas-vivas diversas.

Imagens de uma mesma categoria que não somaram 5 ou mais resultados foram agrupadas em "Outros". Esse grupo é composto por imagens de artesanatos de diferentes formatos e técnicas, canecas, paisagens, aves, cachorros, entre outros.

4.2 COLETA DOS DADOS

A partir da análise feita na seção anterior, a base de dados foi estruturada em duas classes: a classe Caravela, composta por imagens que contém alguma aparição da caravela-portuguesa, e a classe Não-Caravela, composta por imagens que não contém nenhuma aparição da caravela-portuguesa. Com a intenção de simular um comportamento similar ao dos resultados obtidos do Instagram, a classe Caravela foi baseada principalmente em imagens de caravelas encalhadas na areia, e a classe Não-Caravela foi criada com base em sete tipos de imagens: pessoas (majoritariamente fotografadas em praias), embarcações (majoritariamente do tipo caravela), ilustrações, tatuagens, veleiros, águas-vivas de espécies diversas, e imagens aleatórias.

Para coleta dos dados da classe Caravela, foram utilizadas duas fontes de dados. A primeira fonte foi os dados obtidos do Instagram por (NASCIMENTO, 2020), através da busca pela hashtag #caravelaportuguesa, onde foram coletadas 426 imagens. A segunda fonte foi a plataforma iNaturalist, que permite que pessoas do mundo todo cadastrem e compartilhem o seu registro de diversas espécies de animais e plantas. A plataforma permite a exportação dos dados com a utilização de filtros, como espécie e tipo de dado, o que permitiu a obtenção de 5.731 imagens de caravelas-portuguesas de forma fácil e eficiente. No total, foram obtidas 6.157 imagens para a classe Caravela.

Para a coleta de dados da classe Não-Caravela, foram utilizadas três fontes de dados. A primeira foi a plataforma iNaturalist², onde foram obtidas 2.876 imagens de veleiros e 649 imagens de águas-vivas aleatórias (sem incluir a caravela-portuguesa) registradas na América do Sul. A segunda fonte de dados foi a plataforma Bing³, um mecanismo de pesquisa que permite o retorno de imagens através da busca de um termo. Com o objetivo de coletar o maior número de imagens em um menor tempo, foi utilizado um script escrito em linguagem Python que realiza a busca na plataforma Bing e salva as imagens resultantes. A quantidade de imagens e os termos buscados podem ser observados na Tabela 2.

TABELA 2 – IMAGENS OBTIDAS ATRAVÉS DA PLATAFORMA BING

Tipo de imagem	Termos buscados	Quantidade de imagens
Pessoa	"selfie", "pessoa na praia"	453
Embarcação	"caravela portuguesa embarcação", "embarcação"	471
Ilustração	"ilustração caravela-portuguesa", "ilustração água-viva"	585
Tatuagem	"tatuagem caravela-portuguesa", "tatuagem"	697
Aleatória	"animais", "artesanato", "praia"	602

FONTE: O autor (2022).

² Disponível em: <<https://www.inaturalist.org>>

³ Disponível em: <<https://www.bing.com>>

A última fonte de dados utilizada foi a própria rede social Instagram, a partir dos resultados obtidos pela busca da hashtag #caravela. Esse termo foi utilizado pois possui resultados bastante relacionados ao da hashtag #caravelaportuguesa, porém, majoritariamente composto por imagens que não são um registro da *Physalia physalis*: embarcações, ilustrações, tatuagens, pessoas e aleatórias. Esse método obteve um total de 2.329 imagens válidas para a classe Não-Caravela.

Com a intenção de manter a base o mais similar possível dos resultados negativos obtidos para a busca da #caravelaportuguesa, foram selecionadas apenas 500 imagens de veleiros e 500 imagens de águas-vivas aleatórias para compor a base de dados final, que somadas com as imagens obtidas através das plataformas Bing e Instagram totalizam 6.137 imagens.

4.3 CONSIDERAÇÕES

Todas as imagens passaram por um pré-processamento e foram redimensionadas para um tamanho de 224x224, com a intenção de reduzir o custo computacional. Esse tamanho foi escolhido por ser a entrada padrão das arquiteturas VGG-16 e ResNet50. O tamanho padrão da arquitetura InceptionV3 é 299x299, o que faz com que haja alteração de tamanho dos mapas de características gerados pela rede.

A coleta de imagens da classe Não-Caravela realizada em fontes como as plataformas Instagram e Bing não possui uma confiança segura para garantir que as imagens pertençam realmente às categorias desejadas. Para manter a integridade da classe, foi realizada uma limpeza manual, removendo as imagens que representassem uma caravela-portuguesa. No entanto, como qualquer tarefa manual, esse é um processo sujeito a falhas, sendo possível que algumas imagens inválidas possam ter sido mantidas como *outliers* na base.

5 ANÁLISE EXPERIMENTAL

Neste capítulo são apresentados os métodos utilizados para a implementação do código das redes neurais utilizadas, o desempenho de cada rede nos testes realizados, e a análise dos resultados obtidos.

5.1 IMPLEMENTAÇÃO DAS REDES NEURAIAS

As redes neurais selecionadas foram implementadas em linguagem Python, amplamente utilizada na resolução de problemas de aprendizado de máquina por sua grande quantidade de bibliotecas na área. O código foi escrito utilizando a ferramenta Jupyter Notebook, para facilitar a visualização dos resultados e geração das métricas. A principal ferramenta utilizada foi a API Keras, pertencente à biblioteca Tensorflow, que fornece a implementação de diversos modelos populares de CNNs com ou sem pré-treinamento.

O treinamento das redes foi executado em uma máquina com processador Apple M1 com CPU 8-core e GPU 8-core e 16GB de RAM. Cada arquitetura foi executada 2 vezes, a primeira sem nenhum tipo de pré-treinamento, e a segunda utilizando o pré-treinamento na base de dados ImageNet. A avaliação dos resultados obtidos foi feita utilizando as métricas acurácia, precisão, recall e F1 Score.

A base utilizada nos experimentos foi dividida em treinamento, contendo cerca de 60% da base total, validação, com 20%, e teste, também com 20%. Cada parcela foi dividida manualmente de forma que a classe Não-Caravela permanecesse relativamente balanceada entre os seus tipos de imagens. A quantidade exata de imagens de cada classe para cada conjunto de dados pode ser vista na Tabela 3. Uma amostra das imagens presentes na base de treinamento pode ser observada na Figura 14.

TABELA 3 – QUANTIDADE DE IMAGENS PRESENTES NAS BASES DE TREINAMENTO, VALIDAÇÃO E TESTE

Classe	Treinamento	Validação	Teste
--------	-------------	-----------	-------

Caravela	3.705	1.231	1.230
Não-caravela	3.681	1.228	1.228

FONTE: O autor (2022).

FIGURA 15 – AMOSTRA DE IMAGENS PRESENTES NA BASE DE TREINAMENTO



FONTE: O autor (2022).

Para aumentar a quantidade de imagens presente na base de treinamento, foi utilizada a técnica de *data augmentation*, aplicando rotação, zoom e espelhamento horizontal de forma aleatória nas imagens.

A mesma abordagem de *fine-tuning* foi utilizada em todas as arquiteturas testadas, e portanto, todas sofreram as mesmas modificações. Uma das principais alterações foi a substituição da camada de saída por uma nova camada de saída adequada ao problema proposto, composta por uma camada de *average pooling*, uma camada *flattening*, uma camada *fully-connected* que utiliza a função de ativação ReLu, e uma camada de regularização do tipo *dropout* com porcentagem de 30%. Além disso, todas as camadas pré-existentes nas arquiteturas foram congeladas durante o treinamento das redes que utilizam o pré-treinamento da ImageNet.

A função de perda utilizada foi a entropia cruzada, mais especificamente, a entropia cruzada binária, função fornecida pelo Keras própria para problemas de classificação binária.

O treinamento foi executado com um limite de 50 épocas, atrelado a um *Early Stopping* baseado na perda obtida na base da validação.

5.2 APRESENTAÇÃO DOS RESULTADOS

Os dados coletados para realizar a comparação das arquiteturas estão apresentados nas três tabelas a seguir. As tabelas seguem um mesmo formato: as três primeiras linhas representam as arquiteturas sem nenhum pré-treinamento, e as três últimas linhas representam as arquiteturas pré-treinadas com a base de dados ImageNet.

A Tabela 4 apresenta a acurácia obtida nos dados de treinamento e nos dados de validação durante o treinamento de cada arquitetura. A Tabela 5 apresenta as métricas gerais de acurácia, precisão, recall e F1 Score obtidas na base de testes de cada arquitetura. A Tabela 6 apresenta as mesmas métricas que a Tabela 5, mas com os valores separados por classe Caravela e Não-Caravela. As linhas destacadas representam os dados da arquitetura com melhor desempenho.

TABELA 4 – ACURÁCIA OBTIDA DURANTE O TREINAMENTO DE CADA ARQUITETURA

Arquitetura	Acurácia nos dados de Treinamento	Acurácia nos dados de Validação
VGG-16	0.77	0.80
ResNet50	0.87	0.87
InceptionV3	0.88	0.89
VGG-16 + ImageNet	0.96	0.95
ResNet50 + ImageNet	0.98	0.96
InceptionV3 + ImageNet	0.95	0.95

FONTE: O autor (2022).

TABELA 5 – MÉDIA DOS RESULTADOS OBTIDOS NOS TESTES DE CADA ARQUITETURA

Arquitetura	Acurácia	Precisão	Recall	F1 Score
VGG-16	0.7482	0.75	0.75	0.75
ResNet50	0.8360	0.84	0.84	0.84
InceptionV3	0.8238	0.82	0.82	0.82
VGG-16 + ImageNet	0.9308	0.93	0.93	0.93
ResNet50 + ImageNet	0.9475	0.95	0.95	0.95
InceptionV3 + ImageNet	0.9386	0.94	0.94	0.94

FONTE: O autor (2022).

TABELA 6 – RESULTADOS POR CLASSE OBTIDOS NOS TESTES DE CADA ARQUITETURA

Arquitetura	Classe	Acurácia	Precisão	Recall	F1 Score
VGG-16	Caravela	0.78	0.73	0.79	0.76
	Não-Caravela	0.70	0.77	0.71	0.74
ResNet50	Caravela	0.89	0.80	0.89	0.84

	Não-Caravela	0.77	0.88	0.78	0.83
InceptionV3	Caravela	0.82	0.83	0.82	0.82
	Não-Caravela	0.82	0.82	0.83	0.82
VGG-16 + ImageNet	Caravela	0.93	0.93	0.93	0.93
	Não-Caravela	0.93	0.93	0.93	0.93
ResNet50 + ImageNet	Caravela	0.95	0.94	0.96	0.95
	Não-Caravela	0.93	0.96	0.94	0.95
InceptionV3 + ImageNet	Caravela	0.95	0.93	0.95	0.94
	Não-Caravela	0.92	0.95	0.93	0.94

FONTE: O autor (2022).

5.3 ANÁLISE DOS RESULTADOS

Através da análise dos resultados apresentados na Tabela 4, Tabela 5 e Tabela 6, é possível perceber que as redes pré-treinadas com a base de dados ImageNet obtiveram resultados consideravelmente melhores que as redes sem nenhum pré-treinamento.

Dentre as redes sem pré-treinamento, o melhor desempenho foi obtido pela rede ResNet50, com cerca de 83% de acurácia, seguida pela InceptionV3, com 82%. A rede VGG-16 obteve o pior desempenho e teve resultados bastante inferiores às outras redes, com apenas 74% de acurácia.

Dentre as redes pré-treinadas, todas obtiveram métricas bastante consistentes em ambas as classes, o que demonstra uma qualidade generalizada dos modelos. Apesar de bastante próximos, o melhor desempenho foi obtido pela rede ResNet50, com quase 95% de acurácia e 95% obtidos nas métricas de precisão, recall e F1 Score. Em vista disso, seus resultados serão analisados com maior profundidade a seguir.

Os resultados da ResNet50 pré-treinada obtidos na base de testes geraram a matriz de confusão mostrada na Figura 16. A acurácia obtida na classe Caravela foi

ligeiramente melhor, com uma taxa de erro de cerca de 4,30%, enquanto a classe Não-Caravela obteve uma taxa de erro de cerca de 6,18%, ou seja, quase 2% de diferença da anterior.

FIGURA 16 – MATRIZ DE CONFUSÃO OBTIDA NO TESTE DA REDE RESNET50 PRÉ-TREINADA

		Valores reais	
		Caravela	Não-caravela
Valores previstos	Caravela	1177	76
	Não-caravela	53	1152

FONTE: O autor (2022).

Dentre os 53 falsos negativos obtidos na classe Caravela, cerca de 11 possuem a presença de uma pessoa na foto, e 14 possuem a presença de uma mão humana, segurando ou não a caravela. A presença humana pode ser um fator de impacto na classificação, uma vez que a classe Não-Caravela é composta também por imagens de pessoas. Além disso, 12 imagens dentre o total foram classificadas pelo autor como difíceis de serem identificadas, pois são fotos tiradas de longe do animal, com baixa qualidade ou apenas de seus tentáculos. Uma amostra dos falsos negativos podem ser observadas na Figura 17.

FIGURA 17 – AMOSTRA DE FALSOS NEGATIVOS DA CLASSE CARAVELA



FONTE: O autor (2022).

Dentre os 76 falsos positivos obtidos na classe Caravela, 46 são imagens de veleiros, 25 são imagens de águas-vivas, 5 imagens são de ilustrações e uma imagem é de um artesanato. Visto que a base de testes está aproximadamente balanceada entre os tipos de imagem, é possível perceber que a maior taxa de erros do modelo da classe Não-Caravela está nas imagens de veleiros e águas-vivas, que somam mais de 93% dos erros totais. Uma amostra dos falsos positivos pode ser observada na Figura 18.

FIGURA 18 – AMOSTRA DE FALSOS POSITIVOS DA CLASSE CARAVELA



FONTE: O autor (2022).

Uma possível tentativa de melhoria no desempenho pode ser realizada apresentando ao modelo uma maior quantidade de exemplos de imagens que façam parte dos grupos citados com maior ocorrência nos falsos positivos e falsos negativos da classe Caravela. No entanto, é necessário observar se essa alteração não causa impactos negativos na classificação das outras imagens, e consequentemente, piora os resultados.

6 CONCLUSÃO

O presente estudo teve como objetivo avaliar a utilização de redes neurais convolucionais para automatizar uma etapa da metodologia proposta por (NASCIMENTO, 2020). Com isso, foram testados três tipos de arquiteturas diferentes para medir o desempenho destas em uma tarefa de classificação de imagens binária para a identificação da caravela-portuguesa.

Os experimentos realizados no capítulo anterior mostram que esse tipo de tecnologia apresenta resultados promissores na realização dessa tarefa, chegando a obter uma acurácia de mais de 94% pela rede com melhor desempenho, a ResNet50 pré-treinada no banco de dados ImageNet.

Os resultados obtidos demonstram que a aplicação dessa tecnologia em um fluxo adequado da metodologia apresentada por (NASCIMENTO, 2020) pode ser bastante benéfica para a coleta de evidências.

Em possíveis trabalhos futuros podem ser aplicadas diferentes técnicas para aprimorar ainda mais os resultados, como outras abordagens de *fine-tuning*, *data augmentation* e regularização, além do refinamento e ampliação do banco de dados, com foco no tratamento dos tipos de imagens que apresentaram maior taxa de erro. Além disso, outras arquiteturas referências também podem ser validadas nesse problema.

Outra proposta para trabalhos futuros é analisar a automatização de outras partes da metodologia de (NASCIMENTO, 2020), como a coleta de dados e a análise textual das publicações coletadas, e até mesmo um sistema que integre estas e a tecnologia aqui apresentada como parte de um fluxo completo.

Um fluxo completo poderia consistir em um sistema que realize a coleta de dados de forma periódica e automática, analise os dados presentes em cada publicação, tanto imagens como texto, e classifique como um registro válido ou inválido para a caravela-portuguesa. Para casos de registros válidos, o sistema salva as informações relevantes em um banco de dados. É interessante também o desenvolvimento de um painel visual que possibilite o gerenciamento dos dados e do sistema por algum usuário.

REFERÊNCIAS

TOTTON, A.; MACKIE, G. Studies on *Physalia physalis*, **Discovery Reports**, 30, 301–407, 1960.

PRIETO, L. et al. Portuguese Man-of-War (*Physalia physalis*) in the Mediterranean: A permanent invasion or a casual appearance?. **Scientific Reports**, v. 5, 2015. Disponível em: <https://doi.org/10.1038/srep11545>. Acesso em: 21 Jun. 2019.

RASKOFF, K. A. Collection and Culture Techniques for Gelatinous Zooplankton. **The Biological bulletin**, Chicago, v. 204, n. 1, p. 64-80, 2002.

CANEPA, A. et al. Massive strandings of pleustonic Portuguese Man-of-War (*Physalia physalis*) related to ENSO events along the southeastern Pacific Ocean, *Lat. Am. J. Aquat. Res.*, Valparaíso, 2020, v. 48, n. 5, p. 806-817

FERRER, L.; PASTOR A. The Portuguese man-of-war: Gone with the wind. **Regional Studies in Marine Science**, v. 14, p. 53-62, 2017. Disponível em <https://doi.org/10.1016/j.rsma.2017.05.004>. Acesso em: 3 jan. 2022.

BADALAMENTI, R. et al. The Portuguese Man-of-war, One of The Most Dangerous Marine Species, Has Always Entered The Mediterranean Sea: Strandings, Sightings And Museum Collections. 2021. Disponível em <http://dx.doi.org/10.21203/rs.3.rs-828581/v1>. Acesso em: 3 jan. 2022.

NASCIMENTO, L. **Monitoring jellyfish population by social media**. Relatório (Pós-Graduação em Sistemas Costeiros e Oceânicos) - Setor de Ciências da Terra, Universidade Federal do Paraná, Pontal do Paraná (PR), 2020.

ALBAWI, S. et al. Understanding of a Convolutional Neural Network, **International Conference on Engineering and Technology (ICET)**, 2017. Disponível em <http://dx.doi.org/10.1109/ICEngTechnol.2017.8308186>. Acesso em: 4 jan. 2022.

MUNRO, C. et al. Morphology and development of the Portuguese man of war, *Physalia physalis*. **Scientific Reports**, v. 9, 2019. Disponível em: <https://doi.org/10.1038/s41598-019-51842-1>. Acesso em: 4 jan. 2022.

BARDI, J.; MARQUES A. Taxonomic redescription of the Portuguese man-of-war, *Physalia physalis* (Cnidaria, Hydrozoa, Siphonophorae, Cystonectae) from Brazil. **Ilheringia. Série Zoologia**, v. 97, n. 4, p. 425-433, 2007. Disponível em: <https://doi.org/10.1590/S0073-47212007000400011>. Acesso em: 4 jan. 2022.

GHILARDI, M. *Velella velella* (Cnidaria: Hydrozoa) in the Bay of Ranobe, south-west Madagascar. **Plankton and Benthos Research**, v. 14, p. 206-209, 2019. Disponível em: <http://dx.doi.org/10.3800/pbr.14.206>. Acesso em: 4 jan. 2022.

MOHRI M. et al. **Foundations of Machine Learning**. The MIT Press, 2018.

MITCHELL, T. **Machine Learning**. McGraw-Hill Education, 1997.

BURKOV, A. **The Hundred-Page Machine Learning Book**. Andriy Burkov, 2019.

RANJAN, C. **Understanding Deep Learning: Application in Rare Event Prediction**. Publicação independente, 2020.

NIELSEN, M. **Neural Networks and Deep Learning**. Determination Press, 2015. E-book. Disponível em: <http://neuralnetworksanddeeplearning.com>. Acesso em: 7 jan. 2022.

MOOLAYIL, J. **A Layman's Guide to Deep Neural Networks**, 2019. Disponível em: <https://towardsdatascience.com/a-laymans-guide-to-deep-neural-networks-ddcea24847fb>. Acesso em: 7 jan. 2022.

ALOM, M. et al. A State-of-the-Art Survey on Deep Learning Theory and Architectures. **Electronics**, v. 8, n. 3, 2019. Disponível em <http://dx.doi.org/10.3390/electronics8030292>. Acesso em: 7 jan. 2022.

RECORD, N. et al. The jelly report: Forecasting jellyfish using email and social media. **Anthropocene Coasts**. v. 1, n. 1, p. 34-43, 2018. Disponível em: <https://doi.org/10.1139/anc-2017-0003>. Acesso em: 10 jan. 2022.

MIGOTTO, A. et al. Skin lesions in envenoming by cnidarians (Portuguese man-of-war and jellyfish): etiology and severity of accidents on the Brazilian coast. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 52, p. 47-50, 2010. Disponível em: <https://doi.org/10.1590/S0036-46652010000100008>. Acesso em: 10 jan. 2022.

Data Science Academy. **Deep Learning Book**, 2021. E-book. Disponível em: <https://www.deeplearningbook.com.br/>. Acesso em: 10 jan. 2022.

WU, J. Introduction to Convolutional Neural Networks. **National Key Lab for Novel Software Technology**, Nanjing University, 2017. Disponível em: <https://cs.nju.edu.cn/wujx/paper/CNN.pdf>. Acesso em: 13 fev. 2022.

MARTIN-ABADAL, M. et al. Jellytoring: Real-Time Jellyfish Monitoring Based on Deep Learning Object Detection. **Sensors**, v. 20, n. 6, 2020. Disponível em: <https://doi.org/10.3390/s20061708>. Acesso em: 13 fev. 2022.

BOUVRIE, J. **Notes on Convolutional Neural Networks**. Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge (MA), 2006.

SKALSKI, P. **Gentle Dive into Math Behind Convolutional Neural Networks**, 2019. Disponível em: <https://towardsdatascience.com/gentle-dive-into-math-behind-convolutional-neural-networks-79a07dd44cf9>. Acesso em: 13 fev. 2022.

DRAELOS, R. **Connections: Log Likelihood, Cross Entropy, KL Divergence, Logistic Regression, and Neural Networks**, 2019. Disponível em: <https://glassboxmedicine.com/2019/12/07/connections-log-likelihood-cross-entropy-kl-divergence-logistic-regression-and-neural-networks/>. Acesso em: 12 fev 2022.

GOODFELLOW, I. et al. **Deep Learning**. MIT Press, 2016.

EMMERT-STREIB, F. et al. An Introductory Review of Deep Learning for Prediction Models With Big Data. **Frontiers in Artificial Intelligence**, v. 3. 2020. Disponível em: <https://doi.org/10.3389/frai.2020.00004>. Acesso em: 16 fev 2022.

DONAHUE, J. et al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. **UC Berkeley & ICSI**, 2014. Disponível em: <https://doi.org/10.48550/arXiv.1310.1531>. Acesso em: 16 fev 2022.

HE, K. et al. Deep Residual Learning for Image Recognition. **Conference on Computer Vision and Pattern Recognition**, 2016. Disponível em: <https://doi.org/10.1109/CVPR.2016.90>. Acesso em: 18 fev 2022.

BEZERRA, E. Introdução à aprendizagem profunda. **Tópicos em Gerenciamento de Dados e Informações**, Porto Alegre: Sociedade Brasileira de Computação, 2016.

RUSSAKOVSKY, O. et al. ImageNet Large Scale Visual Recognition Challenge. **International Journal of Computer Vision**, v. 115, p. 211-252, 2014.

RAUBER, T. **Redes Neurais Artificiais**. Departamento de Informática, Universidade Federal do Espírito Santo, 2005.

MCCULLOCH, S.; PITTS, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. **Bulletin of Mathematical Biophysics**, vol. 5. p. 115-133, 1943.

ROSENBLATT, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. **Psychological Review**, vol. 65, p. 386-408, 1958.

SHYAM, R. Convolutional Neural Network and its Architectures. **Journal of Computer Technology and Applications**, vol. 12, n. 2, 2021. Disponível em: https://www.researchgate.net/publication/357876345_Convolutional_Neural_Network_and_its_Architectures. Acesso em: 5 mar. 2022.

KAGAYA, H.; AIZAWA, K. Highly Accurate Food/Non-Food Image Classification Based on a Deep Convolutional Neural Network. **ICIAP Workshops**, p. 350-357, 2015.

JEONG, N.; CHO, S. Instagram Image Classification with Deep Learning. **Journal of Internet Computing and Services**, vol. 15, p. 61-67, 2017.

HU, C. et al. **Detection of Illicit Drug Trafficking Events on Instagram: A Deep Multimodal Multilabel Learning Approach**. West Virginia University, 2021. Disponível em: <https://arxiv.org/abs/2108.08920>. Acesso em: 16 mar. 2022.

SIMONYAN, K.; ZISSERMAN, A. **Very Deep Convolutional Networks For Large-Scale Image Recognition**. Visual Geometry Group, Department of Engineering Science, University of Oxford, 2014.

SZEGEDY, C. et al. Going Deeper with Convolutions. **Conference on Computer Vision and Pattern Recognition**, 2015. Disponível em: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298594>. Acesso em: 23 mar. 2022.

TSIPRAS, D. et al. From ImageNet to Image Classification: Contextualizing Progress on Benchmarks. **Proceedings of Machine Learning Research**, vol. 119. 2020.

GAO, M. et al. Real-Time Jellyfish Classification and Detection Based on Improved YOLO3 Algorithm. **Sensors**, v. 21, n. 23, 2021.

MCLIWAINE, B.; CASADO, M. JellyNet: The Convolutional Neural Network Jellyfish Bloom Detector. **International Journal of Applied Earth Observation and Geoinformation**, vol. 97, 2021.

DE SOUZA, W. **As Grandes Navegações e o Descobrimento do Brasil**, 2007. Disponível em: <http://www.fafich.ufmg.br/pae/apoio/asgrandesnavegacoeseodescobrimentodobrasil.pdf>. Acesso em: 2 abr. 2022.

FAWCETT, T. An Introduction to ROC Analysis. **Pattern Recognition Letters**, vol. 27, p. 861-874. 2006.

SHORTEN, C; . KHOSHGOFTAAR, T. A survey on Image Data Augmentation for Deep Learning. **Journal of Big Data**, v. 6, 2019.

CHEN, T. et al. **A Simple Framework for Contrastive Learning of Visual Representations**. 2020. Disponível em: <https://doi.org/10.48550/arxiv.2002.05709>. Acesso em: 3 abr. 2022.

XU, Y; GOODACRE, R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. **Journal of Analysis and Testing**, vol. 2, p. 249-262, 2018.

WESTERHUIS, J. et al. Assessment of PLSDA cross validation. **Metabolomics**, vol. 4, p. 81–89. 2008.

HARRINGTON P. Multiple versus single set validation of multivariate models to avoid mistakes. **Critical Reviews in Analytical Chemistry**, vol. 48, p. 33-46, 2017.

ROSEBROCK, A. **Deep Learning for Computer Vision with Python**. PyImageSearch, 2017.