

IDENTIFICAÇÃO DE ELASMOBRÂNQUIOS MARINHOS AMEAÇADOS DE EXTINÇÃO POR MEIO DE APRENDIZADO DE MÁQUINA PROFUNDO

Cristopher de Oliveira Silva¹, Rodrigo Machado², Merisandra Côrtes de Mattos³

Resumo: A detecção de espécies de peixes ameaçadas de extinção, especialmente em ambientes de pesca comercial, representa um desafio para a preservação da biodiversidade nos ecossistemas marinhos. A pesca acidental é responsável pela morte de animais marinhos não-alvo, e os métodos manuais de identificação falham em atender às demandas das operações pesqueiras modernas devido à sua baixa velocidade e precisão. Nesse contexto, a automação por meio de aprendizado de máquina profundo surge como uma possível solução. O objetivo desta pesquisa é analisar o desempenho de uma abordagem de aprendizado profundo, baseada em aprendizado por transferência e ajuste fino, para a identificação de elasmobrânquios marinhos ameaçados de extinção. Para isso, o estudo implementa uma estratégia de pré-processamento e aumento de dados e desenvolve o modelo utilizando a arquitetura *MobileNetV2* pré-treinada. O modelo proposto *MobileNetV2* alcançou uma acurácia média de 81,66%, um *F1-score* de 74,73% e uma precisão média de 69,12%, ao longo de 11 testes realizados que permitiram a obtenção de médias e desvio padrão para comprovação da consistência do modelo. A análise dos resultados indicou que a performance está diretamente associada à distinção visual entre as classes, concluindo-se que a alta similaridade morfológica presente no conjunto de dados foi um fator determinante para o desempenho da arquitetura na tarefa de classificação. Para futuras pesquisas, a expansão da base de dados com maior variabilidade de amostras é indicada como um passo fundamental para aprimorar os resultados e viabilizar a implementação de arquiteturas mais robustas.

¹Ciência da Computação, Grupo de Pesquisa em Inteligência Artificial Aplicada, Universidade do Extremo Sul Catarinense (UNESC). E-mail: cristopherspqr@unesc.net

²Coorientador. Curso de Ciências Biológicas, Universidade do Extremo Sul Catarinense (UNESC). E-mail: rodrigomachado@unesc.net

³Orientadora. Curso de Ciência da Computação, Grupo de Pesquisa em Inteligência Artificial Aplicada, Universidade do Extremo Sul Catarinense (UNESC). E-mail: mem@unesc.net

Palavras-chave: visão computacional; aprendizado por transferência; mobileNetV2; classificação de espécies; captura accidental; conservação marinha.

ABSTRACT: The detection of endangered fish species, especially in commercial fishing environments, represents a challenge for the preservation of biodiversity in marine ecosystems. Bycatch is responsible for the death of non-target marine animals, and manual identification methods fail to meet the demands of modern fishing operations due to their low speed and accuracy. In this context, automation through deep machine learning emerges as a possible solution. The objective of this research is to analyze the performance of a deep learning approach, based on transfer learning and fine-tuning, for the identification of endangered marine elasmobranchs. To this end, the study implements a data preprocessing and augmentation strategy and develops the model using the pre-trained MobileNetV2 architecture. The proposed model, MobileNetV2, achieved an average accuracy of 81.66%, an F1-score of 74.73%, and an average precision of 69.12% across 11 tests that allowed for the calculation of averages and standard deviations to verify the model's consistency. Analysis of the results indicated that performance is directly associated with the visual distinction between classes, concluding that the high morphological similarity present in the dataset was a determining factor in the architecture's performance in the classification task. For future research, expanding the database with greater sample variability is indicated as a fundamental step to improve results and enable the implementation of more robust architectures.

Keywords: computer vision; transfer learning; mobileNetV2; species classification; bycatch; marine conservation.

1 INTRODUÇÃO

A detecção de espécies de peixes ameaçadas de extinção, especialmente em ambientes de pesca comercial, representa um desafio para a preservação da biodiversidade nos ecossistemas marinhos e para a conformidade com as regulamentações ambientais.

De acordo com a World Wildlife Fund (2024), a pesca accidental é responsável pela morte de milhões de animais marinhos não-alvo, o que agrava as condições de espécies já ameaçadas. “A captura accidental ou incidental, denominadas *ByCatch*, é o emalhamento não intencional de

animais não alvos da pesca em artefatos pesqueiros. Mais de 1 milhão de animais vertebrados marinhos morrem anualmente em equipamentos de pesca.” (Viva, 2015). Além de comprometer a manutenção dos ecossistemas, esse tipo de captura também dificulta o trabalho de fiscalização ambiental, já que a identificação incorreta ou tardia das espécies reduz a capacidade de órgãos como a Polícia Ambiental e equipes de monitoramento de detectar e coibir práticas ilegais (Woodill et al., 2020). A pesca acidental ocorre muitas vezes devido à dificuldade em identificar as espécies no momento da captura, o que pode acarretar sanções para as empresas envolvidas e agravar a situação dessas espécies.

Nesse contexto, os impactos da pesca acidental não se restringem apenas às espécies capturadas, mas se estendem ao equilíbrio ecológico como um todo. Essa perspectiva está alinhada ao paradigma da Saúde Única (*One Health*), que reconhece a interdependência entre a saúde humana, animal e ambiental. Conforme discutido por Brown, Horton et al. (2024), a abordagem de Saúde Única destaca que ações que afetem a fauna marinha, incluindo espécies ameaçadas, repercutem diretamente na estabilidade dos ecossistemas e, conseqüentemente, na saúde das populações humanas que dependem desses ambientes.

Segundo Hasegawa, Kondo e Senou (2024), métodos manuais de identificação falham em atender às demandas das operações pesqueiras modernas devido à sua baixa velocidade e precisão. Dessa forma, a automação desses processos, por meio de visão computacional e aprendizado profundo, tem se mostrado uma solução promissora e eficaz nesse controle. Tornando essa abordagem especialmente importante para enfrentar o problema da pesca acidental de espécies de peixes ameaçadas de extinção.

Esses padrões podem ser identificados por algoritmos de Redes Neurais Convolucionais, do inglês Convolutional Neural Network (CNN), que treinam sistemas para reconhecer diferentes espécies de peixes com base em suas características visuais. Iniciativas neste sentido têm sido realizadas em diferentes pesquisas, como por exemplo, a de Peruchi, Machado e Mattos (2024) que analisou o desempenho dos modelos de redes neurais profundas ResNet-50 e AlexNet na identificação de espécies de peixes. Esta pesquisa dá continuidade ao trabalho realizado por Peruchi, Machado e Mattos (2024), propondo outra abordagem de aprendizado profundo a ser aplicada na mesma problemática, a fim de se identificar por meio de medidas de qualidade quais modelos apresentam melhores resultados.

A pesquisa de Ju e Xue (2020), aponta que técnicas de aprendizado por transferência⁴ são eficazes na adaptação de modelos pré-treinados para novos contextos, como na identificação de espécies de peixes em diferentes regiões geográficas, levando em consideração variações de aparência entre as espécies, conforme as condições ambientais locais. Essas técnicas, quando aplicadas a grandes bases de imagens, têm mostrado resultados promissores na detecção de espécies ameaçadas, com alta precisão em diferentes ecossistemas.

Neste contexto, a presente pesquisa compreende a análise de diferentes modelos de redes neurais convolucionais para identificar espécies de peixes ameaçadas. Para tanto, a base de imagens utilizada nesta pesquisa será a mesma empregada na pesquisa de Peruchi, Machado e Mattos (2024), que contempla sete espécies de peixes de três famílias de elasmobrânquios marinhos e foi construída com imagens disponibilizadas pelas seguintes plataformas de ciência: *FishBase*, *Shark References*, *BOLD Systems v3* e *IUCN Red List of Threatened Species*. Dessa forma, esta pesquisa busca contribuir para a comunidade acadêmica ao fornecer dados relevantes para futuras pesquisas em visão computacional, possibilitando o desenvolvimento de novas soluções e o aprofundamento em outros estudos na área.

A escolha do tema relacionado à identificação de espécies de elasmobrânquios marinhos⁵ ameaçados de extinção utilizando tecnologias de aprendizado profundo, é motivada pela necessidade de preservar a biodiversidade marinha, cada vez mais impactada pela pesca acidental e mudanças ambientais (World Wildlife Fund, 2024).

O uso de redes neurais convolucionais, amplamente reconhecida no campo da visão computacional, tem demonstrado potencial para processar e analisar imagens em grande escala, facilitando a automação do reconhecimento de espécies em ambientes subaquáticos, onde as condições podem dificultar a identificação manual (Zhang et al., 2022).

Os pontos positivos da abordagem proposta incluem a possibi-

⁴O aprendizado por transferência (*transfer learning*) consiste na reutilização de modelos previamente treinados em bases de dados para resolver problemas semelhantes em novos contextos, reduzindo a necessidade de grandes volumes de dados rotulados. (Pan; Yang, 2010)

⁵Segundo Nelson, Grande e Wilson (2016), os Elasmobrânquios (subclasse *Elasmobranchii*) são peixes que pertencem à classe *Chondrichthyes* (peixes cartilaginosos). O grupo é caracterizado por possuir um esqueleto composto de cartilagem calcificada, e não de osso verdadeiro. A subclasse agrupa as duas linhagens de peixes cartilaginosos modernos: *Selachii* (tubarões) e *Batoidea* (arraias e raias).

lidade de aumentar a eficiência na coleta de dados sobre populações de peixes, permitindo que os gestores pesqueiros realizem um monitoramento mais preciso e em tempo real. Ao utilizar aprendizado por transferência, torna-se viável adaptar modelos previamente treinados em grandes conjuntos de dados para novas tarefas específicas onde há escassez de dados, o que reduz significativamente o tempo de treinamento e a quantidade de dados rotulados necessários (Donges, 2024). De acordo com Christin, Hervet e Lecomte (2019), essa abordagem pode aumentar a acurácia da identificação das espécies, mas também facilitar a implementação de estratégias de conservação.

Os benefícios desta pesquisa voltam-se para o desenvolvimento de futuros estudos na área de conservação marinha e monitoramento automatizado de espécies. O uso de técnicas de aprendizado de máquina, como o aprendizado profundo, poderá contribuir para a redução da pesca accidental em contextos práticos no futuro, ajudando a formular estratégias de conservação mais eficazes. A presente investigação busca expandir esse escopo, incorporando abordagens mais recentes como a transferência de aprendizado, com o objetivo de ampliar a capacidade de adaptação dos modelos em contextos mais dinâmicos e realistas voltados à conservação marinha. Assim, o estudo se alinha com as exigências contemporâneas de conservação da biodiversidade, fornecendo uma base teórica e técnica que poderá ser explorada em aplicações posteriores.

O desenvolvimento de ferramentas capazes de auxiliar na identificação de espécies ameaçadas está diretamente alinhado ao Objetivo de Desenvolvimento Sustentável 14 (Vida na Água). Em especial, contribui para a meta 14.2, que propõe a gestão sustentável e a proteção dos ecossistemas marinhos, e para a meta 14.4, que reforça a necessidade de regular coleta e pesca, eliminando práticas ilegais e reduzindo a sobrepesca. Assim, ao oferecer suporte tecnológico para monitoramento da fauna marinha, este trabalho também contribui com estratégias globais de conservação e manejo sustentável dos recursos oceânicos. (United Nations Department of Economic and Social Affairs, 2025)

2 TRABALHOS CORRELATOS

Para embasar o desenvolvimento desta pesquisa, foram analisados trabalhos correlatos nacionais e internacionais, com o objetivo de compreender a área de estudo e identificar contribuições relevantes. Essa revisão permitiu adquirir conhecimento sobre métodos, resultados e limi-

tações de estudos anteriores, servindo de base para o direcionamento da proposta apresentada.

Tabela 1 - Síntese dos trabalhos correlatos analisados.

Autor(Ano)	Espécies	Metodologia	Principais Resultados
Peruchi et al. (2024)	7 elasmobrânquios	ResNet-50, AlexNet, AlexNet Modificada	AlexNet Modificada: 92,8%; ResNet-50: 85%; desafio em espécies semelhantes
Ju & Xue (2020)	20 espécies	AlexNet otimizada (dropout, L2, attention)	98,3% de acurácia; superou VGG-16 e GoogleNet
Hasegawa et al. (2024)	2.826 espécies possíveis	Modelo TFI (transfer learning e mascaramento)	Alta generalização; +10% acurácia com fine-tuning; sem re-treinamento completo
Rathi et al. (2018)	Espécies variadas	CNN para vídeos subaquáticos	96,29% de acurácia; robusto em baixa luminosidade; tempo real
Alshahrani et al. (2024)	Corais	VGG16, YOLOv5l, YOLOv8l	YOLOv8l: 97,8%; melhor velocidade e precisão
Zhao et al. (2019)	Espécies variadas	CNN embarcada em FPGA	25 FPS em HD; baixo consumo; robusto a ruído

Fonte: O autor (2025).

Peruchi, Machado e Mattos (2024) desenvolveram o estudo “Reconhecimento e Classificação de Espécies de Peixes Utilizando Redes Neurais Convolucionais”. O estudo teve como principal objetivo avaliar o desempenho de modelos de redes neurais convolucionais, especialmente a ResNet-50, a AlexNet e uma versão modificada da AlexNet, na tarefa de classificar sete espécies de peixes ameaçados de extinção, que são comuns no sul do Brasil. Para isso, os autores construíram uma base de dados com 241 imagens provenientes de plataformas científicas como *Fish-Base*, *BOLD Systems*, *IUCN Red List* e do CEPSUL. Essa base foi aumentada artificialmente para 2.399 imagens por meio de técnicas de *data augmentation*. Os modelos foram treinados e avaliados em diferentes cenários, variando o número de épocas (100 e 130) e o tamanho do lote (64, 96 e 128), utilizando métricas como acurácia, precisão, *recall* e *F1-score*. Os resultados demonstraram que, embora a ResNet-50 tenha se destacado frente à AlexNet original, o modelo modificado da AlexNet superou ambas as arquiteturas em alguns cenários. No melhor cenário de 130 épocas e lote de 64 imagens a versão modificada da AlexNet alcançou 92,8% de acurácia, 93% de precisão, 92,8% de *recall* e *F1-score*, sendo o modelo mais eficiente da pesquisa. O trabalho reforça o potencial das CNNs na identificação automatizada de espécies marinhas, especialmente ao empregar arquiteturas adaptadas a domínios específicos. Este trabalho se relaciona diretamente com a presente pesquisa, configurando-se como uma continuação direta da pesquisa realizada por Peruchi, Machado e Mattos (2024), que também teve como foco o reconhecimento por meio de redes neurais

profundas de espécies marinhas ameaçadas.

Ju e Xue (2020) desenvolveram uma pesquisa com o objetivo de aprimorar o desempenho da rede AlexNet na tarefa de reconhecimento automático de espécies de peixes, propondo uma nova arquitetura baseada na AlexNet com melhorias voltadas à redução do sobreajuste e ao aumento da precisão. Para atingir esse objetivo, os autores propuseram uma versão modificada da AlexNet composta por quatro camadas convolucionais (ao invés das cinco da arquitetura original), aplicando camadas otimizadas de *dropout*⁶, regularização L2⁷ e uma técnica chamada *soft attention*⁸ para dar mais foco a regiões relevantes nas imagens. O treinamento foi realizado com a base de dados Fish-Pak, que contém imagens de 20 espécies diferentes de peixes. A base passou por técnicas de *data augmentation*, como rotação, espelhamento e ajuste de brilho, para ampliar a diversidade visual e evitar o sobreajuste. Os resultados mostraram que o modelo proposto alcançou uma acurácia de 98,3%, superando tanto a AlexNet tradicional quanto outras redes neurais convolucionais utilizadas para comparação, como VGG-16 e GoogleNet. A principal conclusão do estudo é que a inclusão de mecanismos de atenção e técnicas avançadas de regularização são eficazes para melhorar significativamente a capacidade de generalização das CNNs no domínio da classificação de espécies marinhas. Este trabalho também se relaciona fortemente com a presente pesquisa, pois investiga formas de melhorar arquiteturas existentes para o reconhecimento de espécies, especialmente em conjuntos de dados com diversidade visual considerável.

Hasegawa, Kondo e Senou (2024) desenvolveram um modelo de aprendizado profundo que pudesse ser reutilizado em diferentes regiões costeiras, sem a necessidade de reconfiguração completa ou novo treinamento a cada localidade. Para isso, os autores propuseram o *Transferable Fish Identification* (TFI), um modelo baseado em redes neurais convolucionais treinado previamente com dados de diversas áreas de pesca. O

⁶O *Dropout* é uma técnica que consiste em descartar aleatoriamente neurônios, juntamente com suas conexões, durante a fase de treinamento. Essa abordagem impede que os neurônios aprendam a depender uns dos outros de forma muito específica nos dados de treino, mitigando assim o problema de sobreajuste. (Srivastava et al., 2014)

⁷A regularização L2 é uma técnica padrão para mitigar o *overfitting*, que funciona adicionando a soma dos quadrados de todos os pesos à função de perda original. Durante o treinamento, essa penalidade força os pesos a decair em direção a zero, desencorajando o modelo de aprender pesos excessivamente grandes. (Zhang et al., 2023)

⁸A *Soft Attention* é um mecanismo que, em vez de focar em uma única localização da imagem, calcula um vetor de importância relativa para a imagem, garantindo a atenção aos pontos mais importantes. (Xu et al., 2015)

modelo foi adaptado a novos ambientes por meio de duas abordagens: a primeira utilizando aprendizado por transferência, com readequação das camadas finais a partir de pequenas quantidades de dados locais; e a segunda aplicando mascaramento da camada de saída, que consiste em filtrar apenas as espécies relevantes para cada novo local, sem necessidade de novo treinamento. Os resultados demonstraram que o modelo TFI apresentou alta capacidade de generalização, alcançando acurácia elevada mesmo sem retreinamento. Na abordagem com mascaramento de saída, o modelo foi capaz de identificar até 2.826 espécies sem perda significativa de desempenho. Já com o uso do aprendizado por transferência, foi observada uma melhora de cerca de 10% na acurácia em relação a modelos tradicionais, como a ResNet-50 treinada com ImageNet. As métricas utilizadas incluíram acurácia, taxa de erro e tempo de adaptação. O estudo concluiu que modelos transferíveis como o TFI são eficazes em cenários com variações ambientais e limitações de dados locais, tornando-se úteis para aplicações em sistemas de monitoramento pesqueiro automatizado e de baixo custo. A pesquisa destaca a importância da reutilização de modelos treinados em larga escala para promover escalabilidade e eficiência em projetos de identificação de espécies marinhas.

Rathi et al. (2018) desenvolveram um sistema automatizado capaz de identificar espécies de peixes em ambientes subaquáticos, utilizando redes neurais convolucionais para realizar a classificação em tempo real com alta precisão. Para isso, os autores implementaram uma rede neural convolucional tradicional composta por múltiplas camadas convolucionais e camadas de *pooling*, seguida por camadas totalmente conectadas para realizar a etapa de classificação. O sistema foi treinado com uma base de dados de vídeos subaquáticos contendo imagens de diferentes espécies de peixes capturadas em ambientes reais, com ampla variação de luminosidade, ruído de fundo e movimento da água.

Durante os testes, o modelo alcançou uma acurácia de 96,29%, demonstrando robustez na identificação das espécies mesmo em condições desafiadoras, como iluminação irregular e presença de partículas em suspensão. A arquitetura foi capaz de realizar a classificação em tempo real, o que reforça seu potencial para ser aplicado em sistemas de monitoramento subaquático automatizado. A pesquisa conclui que as redes neurais convolucionais são altamente eficazes para a tarefa de reconhecimento visual de peixes em vídeos naturais, e destaca que o uso de estruturas profundas permite maior capacidade de generalização, mesmo sem técnicas

avanzadas de pré-processamento ou *data augmentation*.

Alshahrani et al. (2024) realizaram uma pesquisa em que o objetivo foi classificar espécies de corais presentes em recifes marinhos por meio de algoritmos de visão computacional baseados em redes neurais profundas. Para isso, os autores avaliaram o desempenho de diferentes arquiteturas de redes neurais convolucionais, entre elas a VGG16 e duas variantes da arquitetura YOLO (versões YOLOv5l e YOLOv8l), conhecidas por sua eficiência na detecção de objetos em imagens. As imagens utilizadas foram obtidas em ambientes reais de recifes de coral, considerando diferentes condições de iluminação, textura e cor. Os resultados mostraram que a versão mais recente, YOLOv8l, superou as demais arquiteturas, alcançando uma acurácia de 97,8% na classificação das espécies de coral.

A análise também evidenciou que as arquiteturas mais modernas apresentaram não apenas melhor desempenho em termos de precisão, mas também maior velocidade de processamento. O estudo conclui que redes neurais convolucionais são ferramentas eficazes para a identificação automatizada de organismos marinhos em imagens submarinas, e que modelos como o YOLOv8l têm potencial para aplicações práticas em monitoramento ambiental e preservação de ecossistemas marinhos.

Zhao et al. (2019) conceberam a pesquisa “Reconhecimento de imagens subaquáticas em tempo real com sistema embarcado em FPGA (*Field Programmable Gate Array*) para rede neural convolucional”. O objetivo do estudo foi implementar uma solução embarcada eficiente para reconhecimento de imagens sub- marinas, utilizando circuitos programáveis do tipo FPGA. A abordagem consistiu na construção de uma rede neural convolucional otimizada, adaptada para funcionar em tempo real em um dispositivo FPGA (circuito integrado programável), a fim de reduzir o consumo de energia e aumentar a velocidade de processamento em ambientes com restrição de recursos computacionais.

Nos testes realizados, o sistema foi capaz de processar imagens subaquáticas de alta definição a uma taxa de 25 quadros por segundo, mantendo desempenho estável e com baixo uso de energia. Essa capacidade permitiu a identificação eficiente de objetos e formas marinhas em tempo real, mesmo em cenários com interferência visual, como partículas suspensas e iluminação variável. No estudo concluíram que a combinação de redes neurais profundas com plataformas embarcadas programáveis é uma solução viável para aplicações de campo que exigem baixo consumo energético, alta velocidade e reconhecimento confiável de imagens

subaquáticas em tempo real.

3 MATERIAIS E MÉTODOS

Esta pesquisa compreende a análise de desempenho de arquiteturas de redes neurais convolucionais na classificação de sete espécies de elasmobrânquios marinhos. O estudo dá seguimento ao trabalho de Peruchi, Machado e Mattos (2024), que investigou as arquiteturas ResNet-50 e AlexNet, e propôs um modelo modificado (AlexNet Modificada) para esta mesma tarefa.

Atendendo a uma das sugestões para trabalhos futuros de Peruchi, Machado e Mattos (2024), que indicava a exploração de outras arquiteturas de CNNs e o uso de transferência de aprendizado, esta pesquisa foca na análise da abordagem baseada em aprendizagem por transferência (*transfer learning*) e ajuste fino (*fine-tuning*), na mesma problemática.

A abordagem da pesquisa é descrita como quantitativa, sendo quanto à natureza aplicada e de base tecnológica. Em relação aos objetivos, caracteriza-se como descritiva e, no que diz respeito aos procedimentos, compreende uma pesquisa experimental, visto que envolve a aplicação e avaliação prática de algoritmo de aprendizado profundo na classificação das espécies de peixes.

A Figura 1 representa o fluxo metodológico adotado na pesquisa, estruturado em seis etapas sequenciais voltadas à classificação das espécies de elasmobrânquios marinhos. O processo teve início com a análise da pesquisa de Peruchi, Machado e Mattos (2024), que serviu como base para contextualizar e fundamentar o presente estudo. Em seguida, procedeu-se à análise da base de imagens e à definição da estratégia metodológica a ser empregada. Na terceira fase, foram implementadas abordagens de pré-processamento e aumento de dados, com o objetivo de otimizar o conjunto de treinamento. Posteriormente, desenvolveu-se o modelo de classificação por meio da técnica de aprendizado por transferência. A quinta etapa concentrou-se no treinamento e na análise do modelo proposto e, por fim, realizou-se a interpretação dos resultados obtidos, pelo modelo na tarefa de classificação dos elasmobrânquios marinhos.

Figura 1 - Etapas de desenvolvimento da pesquisa.



Fonte: Do autor (2025).

3.1 BASE DE IMAGENS DE ELASMOBRÂNQUIOS MARINHOS

A base de imagens utilizada nesta pesquisa é a mesma consolidada por Peruchi, Machado e Mattos (2024), proveniente de plataformas científicas como *FishBase*⁹, *Shark References*¹⁰, *BOLD Systems v3*¹¹ e *IUCN Red List*¹², além de contar com a parceria do Centro Nacional de Pesquisa e Conservação da Biodiversidade Marinha do Sudeste e Sul (CEPSUL).

A base é composta por 241 imagens originais distribuídas em sete espécies, apresentando um volume de dados restrito para o treinamento de redes neurais profundas.

Tabela 2 - Espécies de elasmobrânquios marinhos na base de imagens.

Família	Espécie	IUCN Red List	Red List Brasil	Número de imagens
Arhynchobatidae	Atlantoraja castelnaui	CR	EN	39
Arhynchobatidae	Atlantoraja cyclophora	EN	VU	70
Arhynchobatidae	Rioraja agassizi	VU	VU	34
Rhinobatidae	Pseudobatos horkelli	CR	CR	54
Squatinae	Squatina argentina	CR	CR	8
Squatinae	Squatina guggenheim	EN	CR	23
Squatinae	Squatina occulta	CR	CR	13
Total de imagens:				241
Total de imagens pré-processadas:				2399

VU - Vulnerable; EN = Endangered; CR = Critically Endangered

Fonte: Peruchi, Machado e Mattos (2024).

Embora Peruchi, Machado e Mattos (2024) tenham aplicado um primeiro aumento de dados, expandindo a base para 2.399 imagens, uma análise preliminar indicou que este volume ainda era insuficiente e propenso a *overfitting*, especialmente ao se considerar arquiteturas mais com-

⁹FishBase:<https://www.fishbase.se/search.php>
¹⁰Shark References:<https://www.sharkreferences.com/>
¹¹BOLD Systems v3: <https://v3.boldsystems.org/>
¹²IUCN Red List of Threatened Species: <https://www.iucnredlist.org/>

plexas. O próprio trabalho de Peruchi identificou desafios de generalização, com algumas espécies possuindo um número muito baixo de amostras de validação.

Diante disso, para viabilizar o treinamento de um modelo mais profundo (com milhões de parâmetros) e mitigar os efeitos do *overfitting*, foi necessário implementar uma estratégia mais robusta de pré-processamento e aumento de dados.

3.2 PRÉ-PROCESSAMENTO E AUMENTO DE DADOS

O desenvolvimento desta etapa foi realizado no ambiente *Google Colaboratory*, utilizando a linguagem Python e bibliotecas como *Scikit-learn* e *TensorFlow* (Keras)¹³. Primeiramente, para corrigir o desbalanceamento entre as classes em que a maior classe possuía 70 imagens e a menor apenas 8, foi aplicada uma técnica de *oversampling* (sobreamostragem). Utilizando a função *resample* da biblioteca *Scikit-learn*, as classes minoritárias foram replicadas aleatoriamente, com reposição, até que todas as sete classes tivessem o mesmo número de amostras da classe majoritária. O resultado foi um *DataFrame* (Pandas) balanceado, contendo os caminhos dos arquivos e seus respectivos rótulos, usado exclusivamente para o treinamento.

Em seguida, foi utilizada a classe *ImageDataGenerator* do Keras para aplicar um aumento de dados, realizado em tempo real durante o treinamento (*on-the-fly*). As transformações aplicadas no gerador de treino incluíram a re-escala dos valores de *pixel* para o intervalo $[0, 1]$ (dividindo por 255); rotação aleatória das imagens em até 30 graus; translação aleatória de largura e altura de até 30%; distorção por cisalhamento de até 30%; zoom aleatório de até 30%; inversão horizontal aleatória; variação aleatória de brilho entre 70% e 130% (*brightness range*= $[0.7, 1.3]$); e, deslocamento aleatório nos canais de cor de até 30 (*channel shift range*=30). Para o conjunto de validação, composto por 20% dos dados originais, foi criado um gerador separado aplicando a re-escala dos *pixels* ($1/255$), sem qualquer aumento de dados, visando garantir uma avaliação fidedigna do desempenho do modelo em dados não vistos. Todas as imagens de entrada foram redimensionadas para 224×224 *pixels*, e o tamanho do lote foi

¹³Para desenvolvimento desta pesquisa foram utilizadas as bibliotecas gratuitas de Python: *Scikit-learn* e *TensorFlow* (Keras). Porém devido ao volume de processamento foi necessária a utilização do *Google Colaboratory Pro* versão paga. Disponível em:
Scikit-learn: <https://scikit-learn.org>;
TensorFlow (Keras): <https://www.tensorflow.org>
Google Colaboratory (Pro): <https://colab.research.google.com>

definido como 32. Inicialmente, a base de imagens utilizada na pesquisa de Peruchi, Machado e Mattos (2024) continha 2.399 imagens. Após o novo processo de aumento de dados e balanceamento das classes, passaram a compor o conjunto 4.402 imagens.

3.3 IMPLEMENTAÇÃO DE MODELO COM APRENDIZADO POR TRANSFERÊNCIA

Esta pesquisa emprega a técnica de aprendizado por transferência, utilizando a arquitetura *MobileNetV2* pré-treinada no conjunto de dados *ImageNet*¹⁴ como base extratora de características. Essa escolha configura uma abordagem intencional de *overengineering*¹⁵, aplicando um modelo com mais de 3 milhões de parâmetros a um problema com um conjunto de dados original reduzido.

A arquitetura *MobileNetV2* foi proposta por pesquisadores da Google em 2018 (Sandler et al., 2018). Esta rede neural convolucional foi projetada para otimizar o desempenho em aplicações de visão computacional que rodam em dispositivos com recursos computacionais limitados, como dispositivos móveis e sistemas embarcados.

A arquitetura utiliza blocos de resíduos invertidos e gargalos lineares. Os blocos de resíduos invertidos, primeiro expandem o número de canais, aplicam filtros leves nesse espaço expandido e, em seguida, comprimem os canais de volta. Já os gargalos lineares referem-se à remoção da função de ativação não-linear na última camada de compressão do bloco. Segundo Sandler et al. (2018), isso evita que informações importantes sejam perdidas em espaços de baixa dimensão. Essa estrutura permite que a rede seja eficiente em termos de memória e processamento, reduzindo o número de parâmetros e a complexidade computacional (medida em FLOPs), ao mesmo tempo em que mantém uma alta acurácia em tarefas de classificação, detecção de objetos e segmentação. Conforme Sandler et al. (2018), essa eficiência torna a *MobileNetV2* ideal para tarefas de inferência em tempo real diretamente no dispositivo, sem a necessidade de

¹⁴O *ImageNet* é um banco de dados visual de imagens rotuladas, é utilizado para o pré-treinamento de modelos de visão computacional. Este processo permite que o modelo aprenda a extrair um conjunto base de características visuais. O modelo pré-treinado é então adaptado para uma tarefa específica, geralmente com um conjunto de dados diferente. (Deng et al., 2009)

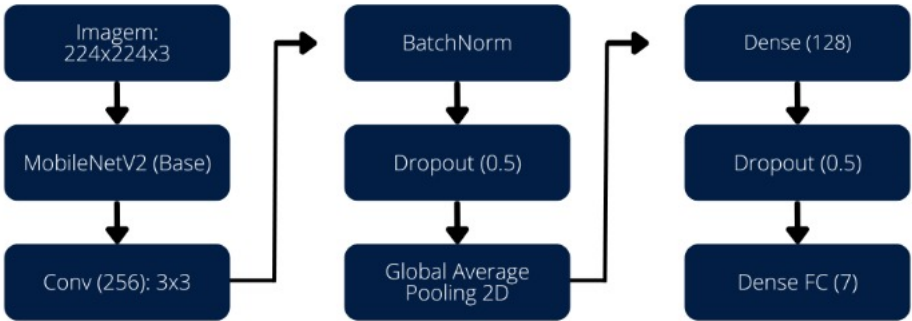
¹⁵O excesso de engenharia (*overengineering*) é a prática de projetar e implementar soluções desnecessariamente complexas para um problema. É descrita como uma complexidade desnecessária, ocorre quando os desenvolvedores adicionam funcionalidades ou camadas de abstração que não são exigidas pelos requisitos atuais, muitas vezes na tentativa de antecipar necessidades futuras hipotéticas (Martin, 2008).

um servidor robusto.

Nesta pesquisa foram investigadas as técnicas de regularização, como *Dropout*, combinadas com o intenso aumento de dados previamente descrito, permitiriam a convergência de um modelo tão complexo em comparação com outras abordagens.

A arquitetura foi definida carregando-se a base convolucional *MobileNetV2* com pesos da *ImageNet* e sem suas camadas superiores de classificação (*include top=False*), configurada para receber imagens de entrada de $224 \times 224 \times 3$. Inicialmente, todos os pesos desta base foram congelados (*base model trainable = False*) para a primeira fase de treinamento. Ao topo da base convolucional, adicionou-se um módulo de classificação customizado (*classification head*), composto sequencialmente por: uma camada convolucional (*Conv2D*) com 256 filtros e ativação ReLU, seguida por uma camada de normalização em lote para estabilização, e uma camada *Dropout* com taxa de 0.5 para regularização. Após estas, uma camada *GlobalAveragePooling2D* reduziu a dimensionalidade dos mapas de características antes de passar por uma camada densa com 128 neurônios e ativação ReLU, seguida por outra camada *Dropout* de 0.5. Finalmente, uma camada densa de saída com 7 neurônios, correspondendo às classes de elasmobrânquios marinhos, e a ativação *softmax* foi utilizada para produzir as probabilidades de classificação.

Figura 2 - Modelo *MobileNetV2* com Aprendizagem por Transferência.



Fonte: Do autor (2025).

3.4 TREINAMENTO E ANÁLISE DE DESEMPENHO

Os experimentos foram executados na plataforma *Google Colaboratory*, utilizando uma máquina virtual equipada com GPU (NVIDIA T4) e a biblioteca *TensorFlow* (versão 2.19.0).

O processo de treinamento do modelo seguiu uma abordagem de ajuste fino, dividida em duas fases. Inicialmente, com a base *Mobile-*

NetV2 congelada, apenas o módulo de classificação customizado foi treinado. Nesta primeira fase, o modelo foi compilado utilizando o otimizador *Adaptive Moment Estimation* (Adam) que consiste em um otimizador de gradiente estocástico, bastante empregado em *deep learning* (Kingma; Ba, 2015). O objetivo deste otimizador é ajustar os pesos da CNN durante o processo de treinamento, a fim de minimizar a função de perda (*loss*) (Goodfellow; Bengio; Courville, 2016). A sua aplicação se deu com uma taxa de aprendizado de 1×10^{-3} , objetivando permitir que as novas camadas se ajustassem aos dados sem perturbar os pesos pré-treinados da base. Este treinamento foi executado por até 50 épocas, sendo monitorado por mecanismos de *EarlyStopping*, com paciência de 10 épocas na perda de validação, e *ModelCheckpoint*, para salvar o modelo com a melhor acurácia de validação.

Posteriormente, após a convergência inicial, procedeu-se ao ajuste fino (*fine tuning*). Para esta segunda fase, a base *MobileNetV2* foi parcialmente "descongelada", permitindo que as últimas 50 camadas (*fine tune at* = -50) também fossem treinadas. O modelo foi então recompilado com o otimizador Adam, utilizando uma taxa de aprendizado significativamente reduzida para 5×10^{-6} , a fim de evitar que atualizações de gradiente muito grandes prejudicassem os pesos pré-treinados já ajustados. O treinamento continuou por mais 50 épocas, novamente empregando *EarlyStopping*, com paciência de 15 épocas, e *ModelCheckpoint* para controle e salvamento do melhor modelo.

3.5 ANÁLISE DOS RESULTADOS POR MEIO DE MEDIDAS DE QUALIDADE

Após a implementação do modelo de rede neural convolucional *MobileNetV2*, utilizando a abordagem de aprendizado por transferência e ajuste fino, procedeu-se à análise detalhada de seu desempenho na tarefa de classificação das sete espécies de elasmobrânquios marinhos selecionadas.

A análise foi conduzida com base nas métricas padrão para problemas de classificação, as quais são Acurácia, Precisão, *Recall* e *F1-Score*.

A Acurácia, segundo Zhang et al. (2023), corresponde à fração de previsões corretas sobre o número total de exemplos avaliados, fornecendo uma medida da eficácia de um classificador. Para análises específicas por classe, a Precisão indica a proporção de previsões corretas para

uma classe em relação ao número total de vezes que o modelo previu aquela classe. Por sua vez, o *Recall* reflete a proporção de predições corretas para uma classe em relação ao número total de exemplos que realmente pertencem àquela classe.

O *F1-score* consolida as duas métricas anteriores ao funcionar como a média harmônica da Precisão e do *Recall*. Diferente de uma média aritmética simples, a média harmônica penaliza valores extremos. Este cálculo assegura que uma *F1-score* elevada só é atingida quando ambas, Precisão e *Recall*, também apresentam valores elevados, representando assim um equilíbrio entre as influências dos Falsos Positivos e dos Falsos Negativos (Zhang et al., 2023).

Na análise dos resultados do modelo gerado utilizou-se um conjunto de dados de validação composto por 482 imagens, que não foram vistas pelo modelo durante as fases de treinamento ou ajuste fino. Para garantir a robustez e a confiabilidade estatística das conclusões, cada cenário de treinamento foi executado 11 vezes, seguindo metodologia similar à de Peruchi, Machado e Mattos (2024).

4 RESULTADOS E DISCUSSÃO

O modelo *MobileNetV2* final, após o processo de ajuste fino, alcançou uma acurácia média de validação de 81,66%. Este resultado serve como ponto de partida para a comparação com trabalhos anteriores e para a análise aprofundada do comportamento do modelo.

Para a validação estatística do desempenho, realizamos a análise utilizando acurácia, *recall*, *F1-score*, precisão e perda, obtendo a média ao longo de 11 execuções. Também calculamos o desvio padrão dos resultados, com o objetivo de avaliar a consistência do desempenho do modelo ao longo das execuções. A Tabela 2 apresenta os resultados obtidos deste processo.

Tabela 3 - Resultados estatísticos obtidos nas 11 execuções.

Métrica	Média	Desvio Padrão
Acurácia	81,66%	0,00702
Recall	81,66%	0,00702
F1-Score	74,73%	0,00646
Precisão	69,12%	0,00694
Perda (Loss)	1,0809	0,12084

Fonte: O autor (2025).

Conforme detalhado na Tabela 2, o p-valor resultante para todas

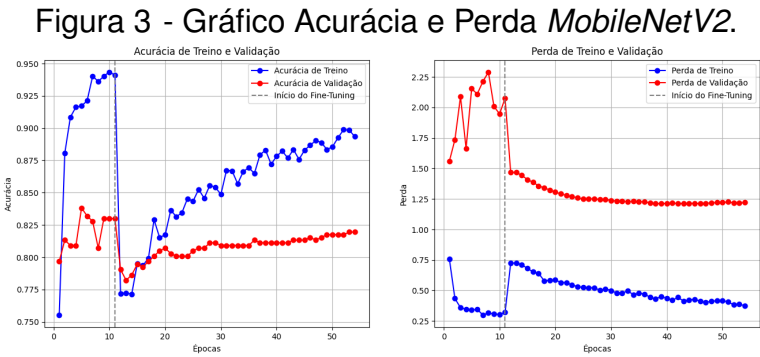
as métricas foi $p < 0,001$. Este resultado rejeita a hipótese nula, indicando que a probabilidade de o desempenho do modelo *MobileNetV2* ser fruto do acaso é inferior a 0,1%. Adicionalmente, podemos observar os valores de desvio padrão obtidos, como 0,00702 para a acurácia e 0,00646 para o *F1-Score*. Esta variabilidade demonstra que o modelo apresentou um desempenho estável e consistente entre as 11 execuções.

A presente pesquisa utilizou como principal referência o trabalho de Peruchi, Machado e Mattos (2024), que explorou as arquiteturas *ResNet-50* e *AlexNet*, além de uma versão modificada desta última, para a mesma tarefa de classificação. Peruchi, Machado e Mattos (2024) reportaram acurácias máximas de 85% para a *ResNet-50* e 68% para a *AlexNet* original, estabelecendo um *benchmark* notável de 92% de acurácia com sua arquitetura *AlexNet* modificada.

Na presente pesquisa buscou-se investigar se a *MobileNetV2*, combinada com aprendizado por transferência, poderia atingir ou superar os valores obtidos no estudo de Peruchi, Machado e Mattos (2024). O resultado de 81,66% de acurácia de validação obtido pela *MobileNetV2*, embora represente um desempenho significativamente superior ao da *AlexNet* original de Peruchi (68%), não alcançou o patamar da *AlexNet* modificada (92%) nem superou marginalmente a *ResNet-50* (85%).

Esta diferença percentual, embora coloque o modelo *MobileNetV2* em um patamar competitivo, sugere que fatores além da escolha da arquitetura base podem estar influenciando decisivamente o desempenho final, como as características intrínsecas do conjunto de dados ou as nuances no processo de ajuste fino.

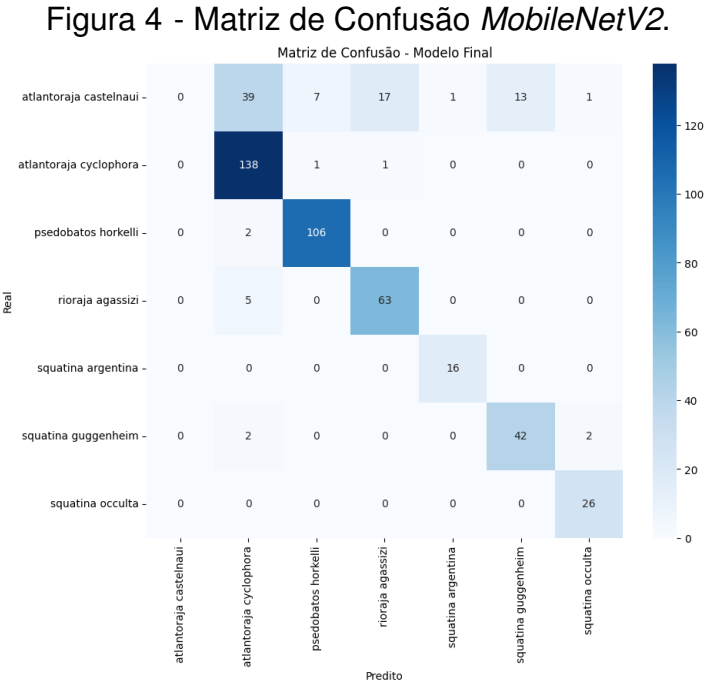
Os gráficos gerados ilustram a dinâmica do aprendizado ao longo das épocas combinadas da Fase 1 (treinamento do classificador) e Fase 2 (ajuste fino).



Fonte: Do autor (2025).

A curva de acurácia de treinamento ascende rapidamente, atingindo valores superiores a 94%, enquanto a acurácia de validação, após um bom desempenho inicial na Fase 1 (chegando a $\approx 83\%$ na época 11), apresenta ganhos mais modestos e certa instabilidade durante o ajuste fino, estabilizando em $\approx 82\%$. A disparidade acentuada entre a acurácia de treino e a de validação é um claro sinal de *overfitting*. O modelo aprendeu a representar os dados de treinamento, mas não conseguiu generalizar esse aprendizado com a mesma eficácia para o conjunto de validação. A curva de perda indica que a perda de treinamento continua a diminuir na Fase 2, mas a perda de validação, após atingir um mínimo por volta da época 27 (restaurada pelo *Early Stopping*¹⁶ da Fase 2), mostra uma tendência geral de aumento e oscilação, indicando que o modelo estava começando a se ajustar ao ruído dos dados de treino em detrimento da generalização.

A matriz de confusão, gerada a partir das predições no conjunto de validação (482 imagens), oferece uma visão granular da performance interclasse.



Fonte: Do autor (2025).

Observa-se um desempenho de 100% na classificação de *Squatina occulta* (26 acertos) e *Squatina argentina* (16 acertos). O modelo tam-

¹⁶O *Early Stopping*, configurado para restaurar os melhores pesos com base na perda de validação (paciência de 10 e 15 épocas), foi crucial para evitar que o desempenho de validação degradado das épocas finais fosse o resultado final, garantindo que o modelo reportado (com 82% de acurácia) representasse o melhor ponto de generalização alcançado durante o processo.

bém demonstrou alta performance com *Atlantoraja cyclophora* ($\approx 98,6\%$ de acerto, 138/140) e *Pseudobatos horkelli* ($\approx 98,1\%$ de acerto, 106/108). O desempenho foi inferior para *Rioraja agassizi* ($\approx 92,6\%$ de acerto, 63/68) e *Squatina guggenheim* ($\approx 91,3\%$ de acerto, 42/46). A principal dificuldade reside na classificação de *Atlantoraja castelnaui*, que não obteve nenhuma classificação correta (0 acertos em 78 amostras), sendo frequentemente confundida com *Atlantoraja cyclophora* (39 vezes), *Rioraja agassizi* (17 vezes), *Squatina guggenheim* (13 vezes) e *Pseudobatos horkelli* (7 vezes).

Essa complexidade na classificação, especialmente a confusão massiva de *A. castelnaui* com outras espécies da mesma família *Arhynchobatidae* (*A. cyclophora* e *R. agassizi*), aponta para a semelhança morfológica como um fator crítico. Além disso, a base original continha 39 imagens de *A. castelnaui*; embora o *oversampling* tenha aumentado a representatividade numérica no treino, ele não introduz nova variabilidade visual, o que pode ter limitado a capacidade do modelo de aprender características verdadeiramente distintivas para esta espécie. Este padrão de confusão entre espécies morfológicamente próximas também foi observado, embora em menor grau, por Peruchi, Machado e Mattos (2024), reforçando a ideia de que este é um desafio inerente ao *dataset*.

O desafio do *overfitting*, mesmo com a aplicação de *Dropout* e regularização L2, também foi abordado por Ju e Xue (2020). Em seus resultados, a regularização e mecanismos de atenção foram eficazes para melhorar uma arquitetura AlexNet, alcançando 98,3% de acurácia. Contudo, na presente pesquisa, a aplicação de técnicas similares de regularização na arquitetura *MobileNetV2*, que obteve 81,66%, não foi suficiente para superar o *benchmark* da AlexNet Modificada de Peruchi, Machado e Mattos (2024) de 92%. Isso sugere que as modificações estruturais de Peruchi, Machado e Mattos (2024) foram mais impactantes para este *dataset* específico do que a aplicação de regularização na arquitetura *MobileNetV2*.

O desempenho de 81,66% também se contrasta com os 96,29% de acurácia de Rathi et al. (2018). A principal diferença metodológica reside na fonte de dados, Rathi et al. (2018) utilizaram vídeos subaquáticos, que fornecem múltiplos quadros, ângulos e informações de movimento para a classificação. Isso representa uma vantagem significativa sobre o *dataset* de imagens estáticas utilizado nesta pesquisa, que oferece menos variabilidade visual por amostra.

A base de dados utilizada, com 2399 imagens antes do balanceamento e aumentada dinamicamente durante o treino com técnicas como

rotação, zoom, deslocamentos, inversão, brilho e ajustes de cor, representa um esforço para fornecer ao modelo informações suficientes e variadas. O balanceamento via *oversampling* (700 imagens/classe no treino) visou evitar que o modelo se tornasse enviesado para as classes majoritárias. Apesar disso, e da aplicação de técnicas de regularização como *Dropout* (taxa de 0.5 nas camadas adicionadas) e regularização L2 (nas camadas *Conv2D* e *Dense* adicionadas), além do congelamento parcial das camadas do modelo base durante o ajuste fino e o uso de uma taxa de aprendizado muito baixa (Adam com taxa de aprendizado =5e-6), o *overfitting* persistiu, como evidenciado pelas curvas de aprendizado (Figura 3). Isso sugere que a complexidade do modelo *MobileNetV2*, mesmo com ajuste fino, ainda pode ser excessiva para a quantidade de informação única e generalizável presente no conjunto de dados, especialmente considerando a dificuldade intrínseca da tarefa.

5 CONCLUSÃO

A abordagem utilizando *MobileNetV2* com aprendizado por transferência e ajuste fino resultou em um modelo capaz de classificar as espécies de elasmobrânquios marinhos com uma acurácia média de 81,66%. A análise realizada indica que, apesar da aplicação de múltiplas técnicas para otimização e combate ao *overfitting*, o desempenho foi limitado principalmente pelas características do conjunto de dados, em função da alta similaridade morfológica entre certas espécies e a possível falta de diversidade representacional em classes com menor número de amostras originais. A dificuldade parece residir mais na natureza dos dados do que na capacidade da arquitetura ou na metodologia de treinamento empregada.

Os resultados desta pesquisa indicam caminhos promissores para investigações futuras. A principal limitação identificada reside na quantidade e variabilidade da base de imagens, especialmente para espécies morfológicamente similares. Portanto, um passo fundamental seria a expansão significativa do *dataset*, buscando maior diversidade de amostras em diferentes condições.

Adicionalmente, sugere-se a exploração de outras arquiteturas de CNN, como *EfficientNet* ou *Vision Transformers* (condicionado a um *dataset* maior), comparando seu desempenho com a *MobileNetV2*. O refinamento das estratégias de treinamento, incluindo regularização avançada e validação cruzada, também pode aprimorar a generalização do modelo. Embora não implementado nesta pesquisa devido ao volume de dados, o

aprendizado incremental permanece uma abordagem relevante, bem como a realização de um estudo para determinar o número ideal de imagens por classe, visando otimizar a relação entre custo computacional e acurácia. A aplicação de técnicas de explicabilidade poderia esclarecer as causas das confusões entre espécies, guiando melhorias futuras. Por fim, recomenda-se a avaliação do modelo em cenários mais realistas, como vídeos subaquáticos, assim como a expansão da base de dados com maior variabilidade de amostras, passo fundamental para viabilizar a implementação de arquiteturas mais robustas.

REFERÊNCIAS

ALSHAHRANI, A. et al. Classification of coral reef species using computer vision and deep learning techniques. **Engineering, Technology & Applied Science Research**, v. 14, n. 5, p. 16478–16485.

BROWN, H. L.; HORTON, D. L. et al. One health: a structured review and commentary on trends and themes. **One Health Outlook**, v. 6, n. 1, p. 1–15. Acesso em: 6 dez. 2025. Disponível em: <<https://onehealthoutlook.biomedcentral.com/articles/10.1186/s42522-024-00111-x>>.

CHRISTIN, S.; HERVET, E.; LECOMTE, N. Applications for deep learning in ecology. **Methods in Ecology and Evolution**, British Ecological Society, v. 10, n. 10, p. 1632–1644. Acesso em: 5 nov. 2025. Disponível em: <<https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13256>>.

DENG, J. et al. ImageNet: A large-scale hierarchical image database. In: **2009 IEEE Conference on Computer Vision and Pattern Recognition**. [S.I.]: IEEE, 2009. p. 248–255.

DONGES, N. **Transfer Learning: A Practical Guide**. 2024. Acesso em: 9 nov. 2025. Disponível em: <<https://builtin.com/data-science/transfer-learning>>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press, 2016. Acesso em: 8 nov. 2025. ISBN 9780262035613. Disponível em: <<https://www.deeplearningbook.org>>.

HASEGAWA, T.; KONDO, K.; SENOU, H. Transferable deep learning model for the identification of fish species for various fishing grounds. **Journal of Marine Science and Engineering**, v. 12, n. 3, p. 415.

JU, Z.; XUE, Y. Fish species recognition using an improved alexnet model. **Optik**, v. 223, p. 165499. ISSN 0030-4026. Acesso em: 9 nov. 2025. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0030402620313358>>.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. In: **Proceedings of the 3rd International Conference on Learning Representations (ICLR)**. [s.n.], 2015. Acesso em: 9 nov. 2025. Disponível em: <https://openreview.net/forum?id=ryQu7f-RZ>.

MARTIN, R. C. **Clean Code: A Handbook of Agile Software Craftsmanship**. Upper Saddle River, NJ: Prentice Hall, 2008. 464 p. ISBN 978-0132350884.

NELSON, J. S.; GRANDE, T. C.; WILSON, M. V. H. **Fishes of the World**. 5th. ed. Hoboken, New Jersey: John Wiley & Sons, 2016.

PAN, S. J.; YANG, Q. A survey on transfer learning. **IEEE Transactions on knowledge and data engineering**, IEEE, v. 22, n. 10, p. 1345–1359.

PERUCHI, M. S.; MACHADO, R.; MATTOS, M. C. **Reconhecimento e classificação de espécies de peixes utilizando redes neurais convolucionais**. 2024. 23 p. Acesso em: 9 nov. 2025. Disponível em: <http://repositorio.unesc.net/handle/1/10984>.

RATHI, R. et al. Underwater fish species classification using convolutional neural network and deep learning. **arXiv preprint arXiv:1811.01494**.

SANDLER, M. et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2018. p. 4510–4520.

SRIVASTAVA, N. et al. Dropout: A simple way to prevent neural networks from overfitting. **Journal of Machine Learning Research**, v. 15, n. 1, p. 1929–1958. Acesso em: 8 nov. 2025. Disponível em: <http://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf>.

United Nations Department of Economic and Social Affairs. **Goal 14: Conserve and sustainably use the oceans, seas and marine resources for sustainable development**. 2025. <https://sdgs.un.org/goals/goal14>. Acesso em: 6 dez. 2025.

Viva. **Captura Acidental (Bycatch)**. 2015. Acesso em: 13 out. 2025. Disponível em: <https://www.viva.bio.br/captura-acidental-bycatch/>.

WOODILL, A. J. et al. Predicting illegal fishing on the patagonia shelf from oceanographic seascapes. **arXiv preprint arXiv:2007.05470**. Acesso em: 6 dez. 2025. Disponível em: <https://arxiv.org/abs/2007.05470>.

World Wildlife Fund. **Illegal Fishing: Which Fish Species Are at Highest Risk from Illegal and Unreported Fishing**. 2024. Acesso em: 7 nov. 2025. Disponível em: <https://www.worldwildlife.org/publications/illegal-fishing-which-fish-species-are-at-highest-risk-from-illegal-and-unreported-fishing>

XU, K. et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In: **Proceedings of the 32nd International Conference on Machine Learning**. [s.n.], 2015. (JMLR Workshop and Conference Proceedings, v. 37), p. 2048–2057. Acesso em: 1 nov. 2025. Disponível em: <<http://proceedings.mlr.press/v37/xuc15.html>>.

ZHANG, A. et al. **Dive into Deep Learning**. [S.l.]: Cambridge University Press, 2023. 574 p.

ZHANG, Z. et al. Large-scale underwater fish recognition via deep adversarial learning. **Marine Pollution Bulletin**, v. 64, p. 353–379. Acesso em: 1 nov. 2025. Disponível em: <<https://link.springer.com/article/10.1007/s10115-021-01643-8>>.

ZHAO, Z. et al. Real-time underwater image recognition with fpga embedded system for convolutional neural network. **Sensors**, v. 19, n. 2, p. 350.