UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO OCEANOGRAFIA

Flora Medeiros Sauerbronn

Classificação Automática de Cliques de Delfinídeos em Projetos de Pesquisa Sísmica

Florianópolis
2025

Flora Medeiros Sauerbronn

Classificação Automática de Cliques de Delfinídeos em Projetos de Pesquisa Sísmica

Dissertação de mestrado apresentada para o Programa de Pós-Graduação em Oceanografia da Universidade Federal de Santa Catarina, para a obtenção do grau de Mestre em Oceanografia.
Orientador: Antonio Fernando Härter Fetter Filho.
Coorientadora: Andrea Dalben Soares.

Florianópolis
2025

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Sauerbronn, Flora
Classificação Automática de Cliques de Delfinídeos em Projetos de Pesquisa Sísmica / Flora Sauerbronn ; orientador, Antonio Fernando Fetter Filho, coorientadora, Andrea Dalben Soares, 2025.
47 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências Físicas e Matemáticas, Programa de Pós-Graduação em Oceanografia, Florianópolis, 2025.

Inclui referências.

1. Oceanografia. 2. Aprendizado de Máquina. 3. Mamíferos Marinhos. 4. Monitoramento Acústico Passivo. 5. Pesquisa Sísmica. I. Fetter Filho, Antonio Fernando. II. Soares, Andrea Dalben. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Oceanografia. IV. Título.

Flora Medeiros Sauerbronn

Classificação Automática de Cliques de Delfinídeos em Projetos de Pesquisa Sísmica

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Antonio Fernando Härter Fetter Filho
Universidade Federal de Santa Catarina

Andre Silva Barreto
Universidade do Vale do Itajaí

Fabio Contrera Xavier
Instituto de Estudos do Mar Almirante Paulo Moreira

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Oceanografia.

ICP-Edu
Documento assinado digitalmente
Pedro de Souza Pereira
Data: 12/11/2025 15:30:55-0300
CPF: ***.759.524-**
Verifique as assinaturas em https://v.ufsc.br

Pedro de Souza Pereira
Pós-Graduação

ICP-Edu
Documento assinado digitalmente
Antonio Fernando Harter Fetter Filho
Data: 12/11/2025 14:42:23-0300
CPF: ***.266.980-**
Verifique as assinaturas em https://v.ufsc.br

Antonio Fernando Härter Fetter Filho
Orientador

Florianópolis, Dia 26 do mês Setembro de 2025

Dedico este trabalho a duas mulheres incríveis, cuja presença foi essencial para sua conclusão: Denise, minha mãe, e Andrea, minha amiga.

AGRADECIMENTO

Primeiramente, agradeço à existência da universidade pública brasileira, que possibilita o desenvolvimento contínuo de ciência de alta qualidade no país. Com isso, estendo minha gratidão a todos que estiveram ao meu lado durante essa jornada.

Agradeço ao meu orientador, Fetter, e à minha coorientadora, Andrea, pelo apoio e pela oportunidade de dar vida a este trabalho. Aos meus colegas de laboratório, Daniel, Gabriel Medon, Gabriel Coutinho, Pamela e Rafa, sou grata pela companhia que tornou nosso dia a dia mais leve. E aos meus colegas de mestrado Gabi, Mariane, Bárbara, Camila, Larissa, Laís, Luiza e muitos outros, por cafés de desabafo, risadas de conspirações acadêmicas e surtos coletivos.

Um agradecimento especial ao Tomas Carlotto, por ajustar sua rotina para que eu pudesse utilizar seu computador para rodar meus experimentos e a Ingridy Severino, por revisar todos os audios do trabalho se tornando uma peça chave para a qualidade desta pesquisa.

Aos meus amigos, que me fizeram rir e aproveitar Florianópolis durante esses dois anos acadêmicos, e à minha família, que sempre apoiou minhas decisões e escolhas de vida, mesmo quando pareciam questionáveis.

Por fim, agradeço ao meu namorado Pedro Igor de Araújo Oliveira, pelo suporte técnico, emocional e mental ao longo dessa caminhada.

Vamos para o próximo capítulo!

RESUMO

O Monitoramento Acústico Passivo (MAP) é uma técnica amplamente utilizada para o estudo de mamíferos marinhos; entretanto, sua aplicação em larga escala apresenta desafios significativos devido à necessidade de classificação manual por especialistas — um processo demorado e laborioso. Com os avanços em inteligência artificial, modelos de aprendizado de máquina tornaram-se ferramentas promissoras para automatizar essa tarefa. Este estudo propõe uma abordagem baseada em aprendizado de máquin para análise de dados acústicos coletados durante campanhas de prospecção sísmica, onde o monitoramento ocorre em condições de baixa relação sinal-ruído para detecção de cetáceos. Os dados foram obtidos dessas campanhas realizadas no Brasil, seguindo regulamentações do Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis que exigem o arquivamento de registros acústicos quando fontes sísmicas são desativadas devido à presença de espécies protegidas na zona de exclusão. Construiu-se um conjunto de dados para classificação binária, discriminando cliques de ecolocalização de golfinhos de ruído de fundo. Geraram-se espectrogramas a partir dos áudios, divididos em conjuntos de treinamento, validação e teste. Para abordar o desbalanceamento de classes, empregaram-se técnicas de subamostragem. Avaliou-se o desempenho de oito arquiteturas de redes neurais convolucionais (MobileNetV1, MobileNetV2, AlexNet, VGG16, VGG19, EfficientNetB0, DenseNet121 e ResNet18), pré-treinadas no ImageNet e adaptadas para esta tarefa. A MobileNetV2 obteve os melhores resultados, com acurácia de 0,9159, recall de 0,9230, AUC de 0,9188 e F1-score de 0,7878 no conjunto de teste. Demonstrou-se ainda como o ruído da embarcação afeta o MAP de cetáceos, destacando-se a importância do posicionamento dos hidrofones em relação à popa do navio. Este trabalho contribui com um conjunto de dados robusto e comprova o potencial de modelos de inteligência artificial para aprimorar processos de auditoria ambiental. A abordagem desenvolvida possibilita maior fiscalização do MAP durante atividades de pesquisa sísmica, apoiando esforços na conservação marinha.

Palavras-chave: Mamíferos Marinhos, Aprendizado de Máquina, Levantamentos Sísmicos, Monitoramento Acústico Passivo de Delfinídeos

ABSTRACT

Passive Acoustic Monitoring (PAM) is a widely used technique for studying marine mammals; however, its large-scale application presents significant challenges due to the need for manual classification by specialists, a time-consuming and labor-intensive process. With advances in artificial intelligence, machine learning models have emerged as promising tools to automate this task. This study proposes a machine learning-based approach for analyzing acoustic data collected during seismic surveys, where monitoring occurs under low signal-to-noise conditions for cetacean detection. Data were obtained from seismic surveys conducted in Brazil, following regulations from the Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis that require archiving acoustic records when sound sources are deactivated due to the presence of protected species in exclusion zones. A binary classification dataset was constructed to distinguish dolphin echolocation clicks from background noise. Spectrograms were generated from audio recordings and divided into training, validation, and test sets. Class imbalance was addressed through undersampling techniques. Eight convolutional neural network architectures (MobileNetV1, MobileNetV2, AlexNet, VGG16, VGG19, EfficientNetB0, DenseNet121, and ResNet18) were evaluated, using ImageNet pre-trained models fine-tuned for this specific task. MobileNetV2 achieved the best performance, with 0.9159 accuracy, 0.9230 recall, 0.9188 AUC, and 0.7878 F1-score on the test set. The analysis further demonstrated how vessel noise affects cetacean PAM, highlighting the importance of hydrophone positioning relative to the ship's stern. This work contributes a robust dataset and demonstrates the potential of deep learning models to enhance environmental auditing processes. The developed approach enables improved PAM oversight during seismic research activities, supporting marine conservation efforts.

Keywords: Marine Mammals, Machine Learning, Seismic Surveys, Passive Acoustic Monitoring of Delphinids

Lista de Figuras

Lista de Tabelas

Lista de Siglas

CNN Convolutional Neural Network

EZ Exclusion Zone

IBAMA Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis

MAP Monitoramento Acústico Passivo

PAM Passive Acoustic Monitoring

ROC AUC Receiver Operating Characteristic curve

ML Machine Learning

DL Deep Learning

MFCC Mel-frequency Cepstral Coefficients

Sumário

1 Introdução

A indústria petrolífera desempenha um papel estratégico na economia brasileira, respondendo por 13% do Produto Interno Bruto (PIB) nacional e 50% da matriz energética primária (PETRÓLEO, 2023), com operações distribuídas ao longo de toda a costa do país. Contudo, essas regiões costeiras abrigam ecossistemas marinhos biodiversos, incluindo mamíferos aquáticos que exercem funções ecológicas críticas, como a ciclagem de nutrientes via *whale pump* (ROMAN; MCCARTHY, 2010), mecanismo que incrementa a produtividade primária oceânica. Essa coexistência gera conflitos socioambientais, nos quais as atividades antrópicas podem comprometer a conservação da biodiversidade marinha.

A prospecção de hidrocarbonetos *offshore* utiliza métodos geofísicos baseados em levantamentos sísmicos, empregando arranjos de canhões de ar comprimido como fontes acústicas. Esses equipamentos emitem pulsos sonoros de alta intensidade em intervalos regulares (tipicamente a cada 4-10 segundos (ICMBIO, 2020)), os quais são refletidos pelas camadas geológicas e captados por hidrofones (rebocados ou fixos no leito marinho). Os dados adquiridos permitem a caracterização da estratigrafia submarina e a identificação de reservatórios de óleo e gás economicamente viáveis.

Essas operações, que se estendem por meses, ou mesmo anos, constituem uma das principais fontes de poluição sonora antropogênica em ambientes marinhos. Estudos demonstram que os pulsos sísmicos podem induzir desde alterações comportamentais até danos fisiológicos em organismos aquáticos, incluindo desorientação espacial, perda auditiva temporária ou permanente e em casos extremos, óbito (WRIGHT; COSENTINO, 2015; CARROLL et al., 2017; MCCAULEY et al., 2017; STONE et al., 2017; MERCHANT, 2019; SOUTHALL et al., 2019; SARNOCIŃSKA et al., 2020).

Diversos países possuem diretrizes próprias visando mitigar os efeitos da atividade de pesquisa sísmica, como Estados Unidos, Inglaterra, Canada, Rússia entre outros (COMPTON et al., 2008), sendo que cada um possui especificidades. Entretanto, uma diretriz comum entre certos países é o monitoramento acústico passivo prévio ao início das operações das fontes sísmicas para minimizar as chances de acionar as fontes com alguma espécie protegida de cetáceo dentro da área de impacto. Esse monitoramento é chamado de varredura.

O contexto regulatório brasileiro para pesquisas sísmicas marinhas foi estabelecido pela Resolução CONAMA nº 350/2004 e consolidado pelo Guia de Monitoramento da Biota Marinha, cuja primeira edição foi publicada em 2005 pelo Instituto Brasileiro do Meio

Ambiente e dos Recursos Naturais Renováveis Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA). Este marco normativo tem evoluído mediante sucessivas atualizações, sendo a versão mais recente datada de 2018. O guia determina a obrigatoriedade de equipes de monitoramento ambiental a bordo de navios sísmicos, compostas por: (1) observadores visuais capacitados para monitoramento diurno de aves, mamíferos e tartarugas marinhas por meio de métodos diretos (binóculos), e (2) especialistas em Monitoramento Acústico Passivo Monitoramento Acústico Passivo (MAP) responsáveis pela detecção contínua de cetáceos com equipes especializadas realizando vigilância 24 horas/dia por meio de hidrofones específicos para esta atividade.

O arranjo de hidrofones empregado no MAP obedece às exigências do Guia de Monitoramento da Biota Marinha, que determina a utilização de quatro hidrofones distribuídos em dois pares. Cada par mantém uma distância fixa de 100 metros do outro, sendo rebocados em configuração linear a pelo menos 200 metros da popa do navio e em profundidades superiores a 20 metros durante toda a operação sísmica.

Ao detectarem a aproximação de cetáceos da zona de exclusão, definida com um raio de um quilômetro ao redor das fontes sísmicas, as mesmas devem ser imediatamente desativadas, sendo permitido seu religamento apenas após 30 minutos de ausência de detecções. As varreduras realizadas são submetidas ao IBAMA como comprovante de conformidade operacional.

Entretanto, estudos indicam que apenas uma pequena fração desses registros são auditados oficialmente (SIMÕES, 2022; DALBEN; AVILA, 2021), criando lacunas na fiscalização que podem facilitar não conformidades (SCHMITT, 2016) e ampliar pressões sobre espécies ameaçadas (PARENTE; ARAÚJO, 2011b). Nesse cenário, a automação da análise acústica por meio de técnicas de inteligência artificial emerge como solução promissora para escalar a capacidade de fiscalização governamental.

A classificação automática de sinais bioacusticos apresenta desafios técnicos significativos, decorrentes da variabilidade intrínseca dos sinais (em frequência, duração e padrão espectral) e da interferência de ruídos ambientais e operacionais. Avanços recentes em *machine learning*, particularmente em redes neurais convolucionais aplicadas a espectrogramas, têm demonstrado eficácia superior em tarefas análogas de classificação de sinais bioacústicos.

Este estudo propõe uma metodologia baseada em visão computacional para detecção automatizada de cliques de delfinídeos em registros de MAP, com três contribuições princi-

pais: (1) criação de um *dataset* anotado de vocalizações, (2) desenvolvimento de um *pipeline* de treinamento de modelos, e (3) avaliação comparativa de arquiteturas de redes neurais. O *framework* desenvolvido visa subsidiar auditorias ambientais pelo IBAMA, aumentando a eficácia fiscalizatória com métodos computacionais reprodutíveis.

1.1 Perguntas de Pesquisa

Este trabalho busca responder às seguintes questões científicas:

- Modelos de classificação baseados em redes neurais convolucionais podem ser operacionalizados como ferramentas de triagem automática de cliques de delfinídeos para auditorias de conformidade em levantamentos sísmicos?
- Quais arquiteturas de redes neurais apresentam melhor desempenho na identificação de cliques de delfinídeos em ambientes ruidosos, e quais características estruturais explicam sua eficácia diferencial?

1.2 Objetivos

1.2.1 Objetivo Geral

Avaliar o desempenho comparativo de arquiteturas de redes neurais convolucionais (CNN) na classificação automática de cliques de delfinídeos em registros de monitoramento acústico passivo oriundos de projetos sísmicos marítimos.

1.2.2 Objetivos Específicos

- Curar e anotar um conjunto de dados acústicos, provenientes de pesquisa de prospecção sísmica, representativo das condições operacionais da costa brasileira referente a delfinídeos.
- Implementar um fluxo de trabalho (*pipeline*) para pré-processamento e treinamento de modelos de aprendizado de máquina.
- Quantificar métricas de desempenho (AUC-ROC, F1-score, recall) de diferentes estruturas de Convolutional Neural Network (CNN).

1.2.3 Formato de Dissertação

Esta dissertação organiza-se como um artigo expandido, submetido ao periódico *Ecological Informatics* (Qualis A1 na área de Computação), intitulado "Automated Delphinid Click Identification in Seismic Acoustic Surveys for Environmental Impact Assessment". A estrutura compreende: (i) contextualização teórica e revisão bibliográfica; (ii) delineamento metodológico; (iii) análise de resultados; e (iv) discussão de implicações para políticas ambientais e direções futuras de pesquisa.

2 Automated Delphinid Click Identification in Seismic Acoustic Surveys for Environmental Impact Assessment.

2.1 Abstract

Passive Acoustic Monitoring (PAM) is a important technique for studying marine mammal behavior; however, it presents significant challenges due to the need for trained specialists to manually classify audio recordings—a process that is both time-consuming and labor-intensive. With the advancement of artificial intelligence, machine learning models have emerged as powerful tools to automate this task. In this study, we propose a machine learning-based approach tailored to analyze audio data collected during seismic surveys, where environmental conditions result in a low signal-to-noise ratio when considering cetacean signals. Our goal was to develop an algorithm capable of detecting delphinid echolocation clicks within acoustic pre-watches—a mandatory procedure in seismic surveys for environmental protection. The data were obtained from seismic research campaigns conducted in Brazil, which follow strict regulations defined by IBAMA and include four-channel audio recordings. To address these challenges, we constructed a dataset for binary classification, distinguishing dolphin echolocation clicks from background noise. Spectrograms were generated from the audio recordings and subsequently split into training, validation, and test sets. To deal with class imbalance, we employed various techniques, including undersampling. We evaluated the performance of eight well-established convolutional neural network architectures—MobileNetV1, MobileNetV2, AlexNet, VGG16, VGG19, EfficientNetB0, DenseNet121, and ResNet18—pretrained on ImageNet and fine-tuned for our specific task. Among them, MobileNetV2 achieved the best results, with an accuracy of 0.9159, recall of 0.9230, AUC of 0.9188, and F1-score of 0.7878 on the final test set. Our analysis also highlights how seismic equipment can affect PAM operations carried out onboard by environmental monitoring teams, due to the proximity of the PAM hydrophones to the vessel and the noise caused by cavitation. This study contributes to the development of a robust dataset and demonstrates the potential of deep learning models to improve environmental auditing processes. By automating the analysis of passive acoustic data, our approach enhances oversight of cetacean PAM during seismic research activities and supports broader conservation efforts.

2.2 Introduction

Machine learning techniques for classifying marine mammal acoustic signals have advanced considerably. Numerous studies have explored multiclass classification approaches to identify species or acoustic signals, demonstrating the potential of deep learning in bioacoustics (WHITE et al., 2022; SEYDI et al., 2022; MURPHY et al., 2022). Acoustic signal classification is inherently complex due to variability in intensity, duration, timbre, and frequency, which is further compounded by background noise. Machine learning techniques offer robust solutions for handling such variability in large datasets (BIANCO et al., 2019). Notably, many of these studies utilize data collected via hydrophones in passive acoustic monitoring (PAM) systems, often deployed in controlled, low-noise environments such as marine reserves or small vessels (e.g., sailboats).

In contrast, this study focuses on acoustic data acquired in high-noise environments, specifically during seismic surveys. Marine seismic surveys are geophysical methods that use high-pressure compressed air to generate sound pulses and map potential oil and gas exploration sites. The surveys are carried out using equipment known as air guns, which are towed behind the stern of a vessel. As a way of mitigating the potential effects of seismic surveys on marine mammals, several countries adopt a procedure known as pre-watch before starting the air guns. The purpose of this procedure is to monitor a so-called exclusion area, centered on air guns, where the presence of some species is not allowed at the start of activities. How this monitoring is carried out depends on the country (COMPTON et al., 2008). In Brazil, the guidelines are described in the Guide for Monitoring Marine Biota in Marine Seismic Surveys (IBAMA, 2018) and determine that monitoring of the exclusion area should be done both through visual observation and through passive acoustic monitoring. Recorded audios of these activities must be delivered to the environmental agency for auditing purposes.

Given the volume of data collected, only a subset of submitted surveys undergoes detailed review (SIMÕES, 2022; DALBEN; AVILA, 2021). This limitation may hinder enforcement of environmental regulations (SCHMITT, 2016) and increase risks for endangered species (PARENTE; ARAÚJO, 2011a). Automating delphinid click identification in these recordings could enable large-scale analysis and improve monitoring efficiency.

2.2.1 Delphinid Clicks

Delphinids employ echolocation through the emission of a series of pulses, known as click trains, which serve critical functions in navigation and foraging. These acoustic signals are generated as discrete pulses, with each emitted click followed by an analysis period for echo reception before subsequent pulse emission, thereby forming click sequences (ELLIOTT, 2011). The number of clicks per sequence exhibits considerable variability, with structural characteristics dependent on behavioral context (AU, 1993).

In most delphinid species, the inter-click interval (ICI) is typically less than 0.1 seconds (AU; HASTINGS, 2008). This temporal parameter demonstrates adaptive flexibility, varying according to target distance and being modulated to accommodate echo return times prior to subsequent pulse emission (AU, 1993).

Delphinid clicks exhibit spectral energy concentrated within the 15-130 kHz frequency band (MADSEN; WAHLBERG, 2007). These signals demonstrate pronounced directionality, being emitted in highly focused beams that optimize both energy efficiency and spatial resolution (Figure 2.2.1). This directional characteristic is quantified by the directivity index (DI), representing an acoustic adaptation functionally analogous to the visual focal mechanisms of terrestrial predators (AU; HASTINGS, 2008).

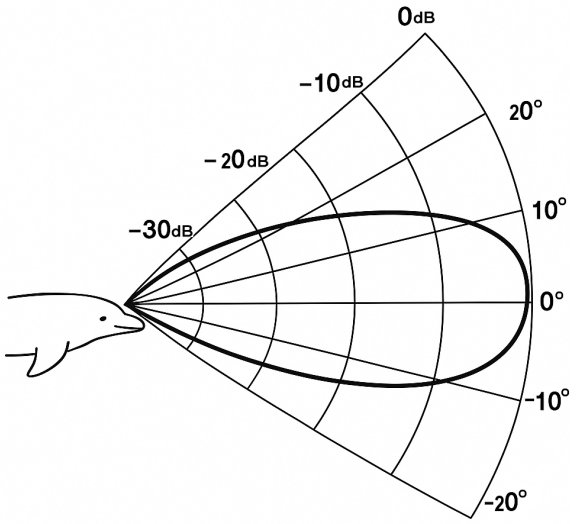


Figura 2.2.1: Far-field beam pattern of bottlenose dolphin echolocation clicks, adapted from (AU, 1993)

The propagation dynamics of these clicks through the marine environment are modulated by multiple physicochemical parameters, including water temperature, salinity, depth, and signal frequency (NUUTTILA et al., 2013). High-frequency signals such as delphinid clicks experience significantly greater attenuation compared to lower-frequency sounds, resulting in more limited effective ranges (ELLIOTT, 2011).

Empirical studies estimate maximum detection ranges for dolphin clicks under natural conditions spanning 600-1800 meters, with variation attributable to detection system characteristics and environmental conditions (REYES-ZAMUDIO, 2005). These empirical measurements provide critical biological benchmarks for understanding the functional ecology of delphinid echolocation.

The detection range of delphinid echolocation clicks can be modeled through the passive sonar equation:

$$SL - TL \geq NL + DT \tag{2.2.1}$$

where SL is source level, TL is transmission loss, NL is ambient noise, and DT is detection threshold. Studies by Parvin, Nedwell e Harland (2007) estimate maximum detection distances of 600–800 meters for small delphinids and 1,200–1,800 meters for larger species such as *Stenella* dolphins under typical seismic survey conditions. These ranges are primarily influenced by differences in source level and directional beam characteristics. For example, a large odontocete emitting a 100 kHz echolocation pulse with a source level of 220 dB re 1 μPa @ 1 m (peak-to-peak), an omnidirectional receiver, 2 kHz bandwidth, and ambient noise around 35 dB re 1 $\mu\text{Pa}/\sqrt{\text{Hz}}$ would achieve detection ranges up to approximately 1.5 km when facing the receiver, but below 500 meters when off-axis. In contrast, small odontocetes generally produce signals approximately 10 dB lower, resulting in maximum detection ranges around 1 km on-axis and approximately 300–500 meters off-axis. These biological detection limits directly inform the 1,000-meter exclusion zone established in IBAMA's marine fauna monitoring protocols.

2.2.2 Data Acquisition of PAM in Seismic Surveys

Brazil maintains a comprehensive legal framework for protecting marine mammals and sea turtles during seismic survey operations (COMPTON et al., 2008). The Marine Biota Monitoring Guide for Marine Seismic Surveys (IBAMA, 2018) specifically requires the implementation of PAM systems aboard seismic vessels for cetacean detection. During seismic operations, PAM operators are tasked with monitoring a one-kilometer exclusion zone

(EZ) surrounding the air guns. Upon detection of marine mammals within this zone, operators must immediately cease air guns activity. Operations may recommence only following a 30-minute monitoring period (designated as acoustic pre-watch), provided no further detections occur within the EZ. These pre-watch recordings constitute mandatory documentation for IBAMA compliance audits (IBAMA, 2018).

The PAM system configuration involves deploying an independent hydrophone array into the water, consisting of two pairs, with 100 meters of separation between each one. The first hydrophone is positioned 200 meters from the vessel's stern at a minimum depth of 20 meters to minimize vessel-generated noise. This setup complies with IBAMA regulations (IBAMA, 2018). Figure 2.2.2 illustrates the standard deployment configuration.

The most commonly used software for monitoring is Pamguard (GILLESPIE et al., 2009), which provides real-time acoustic data visualization through spectrograms and automated detection algorithms.

However, the directional characteristics and high-frequency nature of delphinid echolocation clicks frequently result in only a single hydrophone pair detection, reducing localization precision. Consequently, the dolphins' presence within the EZ is inferred through distance estimation, with IBAMA protocols attributing detected dolphin clicks to animals within the exclusion zone due to the signal's limited propagation beyond 1 kilometer (IBAMA, 2018).

2.2.3 Machine Learning and Marine Mammals

Machine learning algorithms have become increasingly prevalent in marine mammal acoustic detection and classification. In contrast to visual monitoring methods, which are constrained by marine environmental conditions and water opacity, PAM provides an effective observational approach through underwater sound capture and analysis, even under adverse weather conditions (VERFUSS et al., 2018). This methodology enables classification of both vocalizations and echolocation clicks, serving as a non-invasive and efficient alternative for studying marine mammal presence and behavior. Conventional acoustic data analysis has relied on signal processing techniques including power spectra, spectrograms, and Long Term Spectral Averages (LTSA), typically implemented through software platforms such as Triton (GRACIC; GUBNISKY; DIAMANT, 2024). Additional widely employed methods include time-frequency representations - encompassing Fourier Transform, Mel Frequency Cepstral Coefficients (MFCC) (LICCIARDI; CARBONE, 2024), Hilbert-Huang Transform, and Weyl Transform - along with

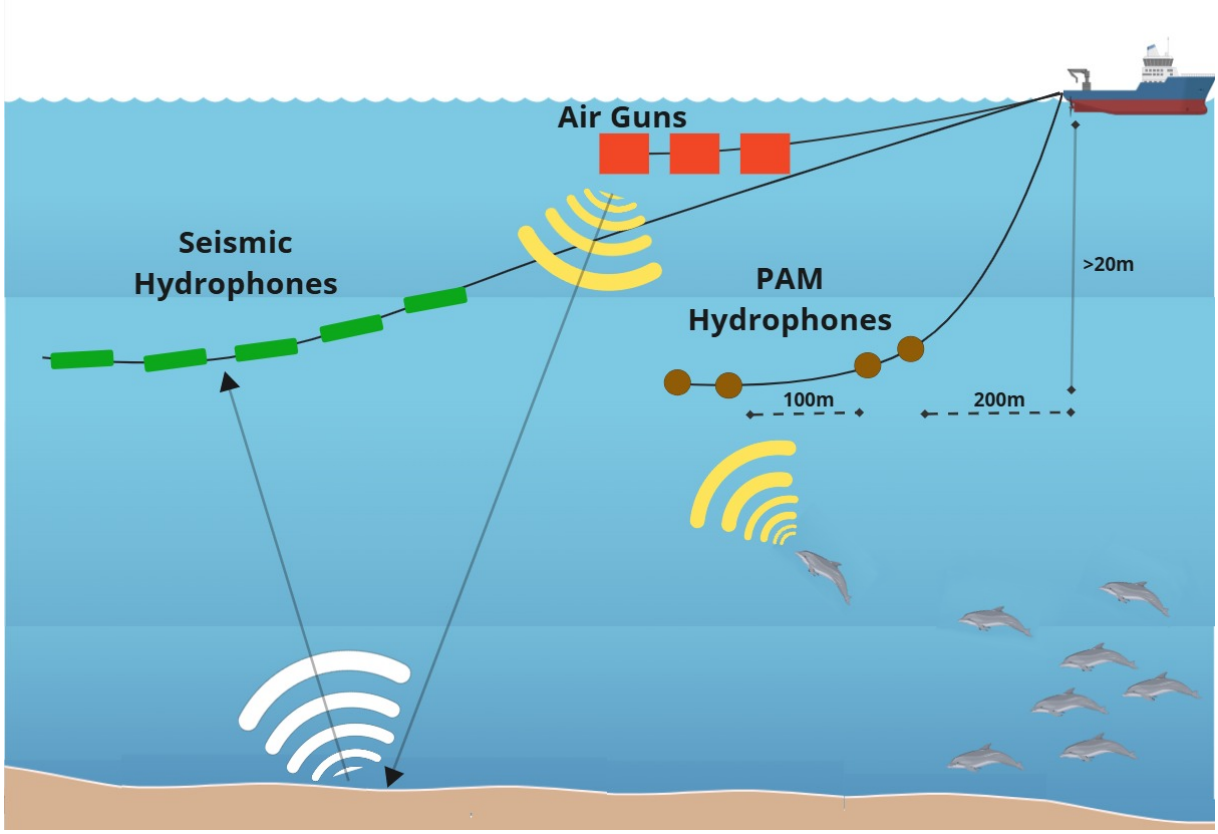


Figura 2.2.2: Schematic representation of the PAM hydrophone array configuration in seismic surveys according to IBAMA guidelines. Green rectangles denote seismic hydrophones, which capture the acoustic signals reflected from the seabed, illustrated as a red rectangle representing the air guns, while brown circles indicate PAM hydrophones dedicated to marine mammal detection.

Empirical Mode Decomposition techniques.

Recent years have witnessed a transition toward Machine Learning (ML) and Deep Learning (DL) approaches for marine mammal vocalization analysis (LICCIARDI; CARBONE, 2024). These techniques have exhibited superior performance in detection and classification tasks across multiple domains, including bioacoustics. Convolutional Neural Networks (CNNs) have demonstrated efficacy in detecting dolphin echolocation clicks, differentiating species or ecotypes, and identifying specific vocalization patterns (GRACIC; GUBNISKY; DIAMANT, 2024). The integration of DL architectures with advanced preprocessing methods is exemplified by tools like ORCA-SPOT and models trained on the Watkins Marine Mammal Sound Database (WMMD), which employ techniques such as Mel spectrograms and Wavelet Scattering Transform (WST) to improve classification accuracy (LICCIARDI; CARBONE; RONDONI, 2024). Transfer learning approaches have additionally been investigated to address the characteristic data scarcity challenges in marine acoustic datasets.

A critical consideration is that most datasets employed for training machine learning models in marine mammal acoustic detection are typically acquired under optimal conditions, such as in marine protected areas or during specialized research expeditions utilizing sailing vessels, where the primary focus is vocalization monitoring and recording. Conversely, the present research employs acoustic data obtained during seismic surveys, which are inherently characterized by elevated noise levels due to operational environment and anthropogenic activities - a context examined by only limited studies, including (SEYDI et al., 2022). This scenario introduces distinct challenges for signal detection and classification.

2.3 Methodology

2.3.1 Data Cataloging

Acoustic analysis was performed by a certified PAM specialist using Raven Pro bioacoustic software (Cornell Lab of Ornithology). The software generates spectrograms for each hydrophone channel with customizable parameters to optimize visualization of target signals, hereafter referred to as *calls*. For delphinid click and whistle analysis, spectrograms were configured with a 1024-point frame size, 512-sample hop size, and 96 kHz sample rate, using a light-to-dark blue rainbow color palette to enhance signal visibility.

During analysis, the specialist manually selected spectrogram framing regions containing biological signals, recording each call's temporal boundaries (start and end times) and spectral characteristics (minimum and maximum frequencies) in an output text file. As the recordings contained four discrete hydrophone channels, the software displayed quad-view spectrograms for simultaneous multi-channel inspection. Signals were annotated only on channels where visibly present, as calls typically appeared on channel subsets rather than all four channels.

2.3.2 Audio Segmentation

Following the data cataloging, a second processing step involves standardizing these recordings into uniform samples suitable for machine learning algorithms. Following methodologies established in Ziegenhorn et al. (2022) and Bermant et al. (2019), we segmented the audio into intervals (hereafter referred to as *chunks*) with a 0.25-second sliding window between consecutive samples (Figure 2.3.1). This approach maximized sample quantity while maintaining temporal resolution for dataset construction.

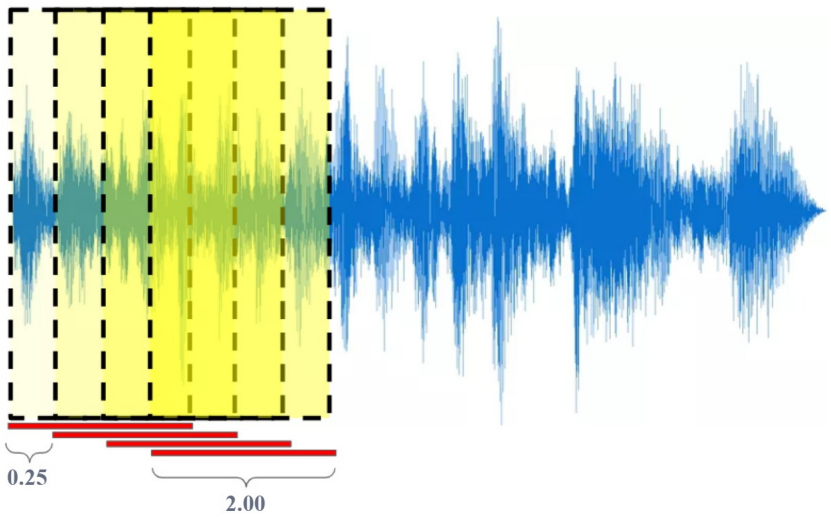


Figura 2.3.1: Audio segmentation schematic. Yellow rectangles depict sliding windows progressing through the acoustic signal (blue waveform), with red lines indicating temporal increments and segment durations.

Our experimental framework evaluated four dataset configurations based on parameters established in bioacoustic literature (ZIEGENHORN et al., 2022; BERMANT et al., 2019). We systematically compared two spectrogram durations (two-second versus five-second windows) and two distinct approaches for defining positive sample classification thresholds. This dual-parameter design enabled a comprehensive assessment of segmentation strategies while maintaining biological relevance.

The first classification methodology employed a spectrogram occupancy criterion, where segments were designated as positive when containing call signatures across $\geq 60\%$ of their temporal duration. This translates to minimum call lengths of 1.2 seconds for two-second spectrograms and three seconds for five-second spectrograms. The visual occupancy threshold ensured robust signal representation while accommodating natural variability in call duration and intensity across recording conditions.

The second approach implemented a biologically-grounded temporal structure criterion derived from fundamental properties of delphinid echolocation (AU, 1993). Valid click trains were required to exhibit at least four complete inter-click intervals (ICIs), with the species-typical ICI usually being less than 0.1 seconds (MADSEN; KERR; PAYNE, 2004), dictating a minimum 0.4-second call presence per segment. This criterion targeted the stereotypic patterning of odontocete click trains, providing species-specific validation of detected signals.

Segments were classified as positive if they either fully contained a call or met the specified threshold requirements. Those overlapping with calls but failing to meet the minimum threshold were excluded, while negative samples contained no call components whatsoever.

We assessed classification efficacy using the Average Precision Score (APS), which quantifies the area under the precision-recall curve. This metric is particularly suitable for imbalanced classification scenarios, as it evaluates precision maintenance across recall levels rather than relying on absolute accuracy. To enable fair model comparisons across datasets with varying positive/negative sample ratios, we standardized positive sample counts by randomly subsampling larger datasets to match the smallest dataset's positive count.

The optimal configuration employed two-second segments with a 0.4-second threshold (20% of segment duration), achieving superior performance with an APS of 0.6434 (Table 2.3.1). This configuration effectively distinguished dolphin vocalizations from background noise while maintaining biological relevance.

Tabela 2.3.1: Average Precision Score comparison across segmentation parameters

Duration (s)	**Call Threshold (s)**	**APS**
2	1.2	0.5859
2	0.4	0.6434
5	3	0.4892
5	0.4	0.4232

The final dataset comprised 313,445 .wav format chunks, categorized using timestamps from the cataloging process (Section 2.3.1).

Given the 4:1 ratio of clicks to whistles and the study's auditing objectives, we focused exclusively on click detection. This selective approach provided sufficient information for our purposes, with further justification provided in Section 2.5.

2.3.3 Splitting datasets and undersampling

Following the classification of chunks into positive and negative categories, we partitioned the data into training (70%), validation (20%), and test (10%) sets according to established machine learning practices (MURAINA, 2022). To prevent data leakage arising from acoustic similarities within individual audio files, we maintained strict segregation, ensuring all chunks from a given audio file were assigned to the same set. This approach mitigates overfitting risks and enhances model generalization (YAP et al., 2014).

Audio file allocation respected the specified percentage split while accounting for variable chunk counts per file. This resulted in 34 files for training, 10 for validation, and 3 for testing. The dataset exhibited inherent class imbalance, with negative samples (absence of animals) substantially outnumbering positives (presence), reflecting real-world environmental conditions.

Initial experiments revealed persistent overfitting despite hyperparameter optimization. To address this, training set undersampling was implemented, which balances class representation while preserving the original distribution in the validation and test sets. This technique retains all positive samples while randomly selecting an equal number of negatives, promoting robust learning while maintaining ecological validity during evaluation (YAP et al., 2014). Table 2.3.2 details the sample distributions before and after undersampling.

Tabela 2.3.2: Class distribution across datasets before and after undersampling

Set	Positive	Negative
Train	21,421	202,496
Train (undersampled)	21,421	21,421
Validation	6,637	59,837
Test	3,897	19,157
Total	31,995	281,490

2.3.4 Image processing

Following audio chunk cataloging, we transformed the acoustic data into spectrogram representations using the Short-Time Fourier Transform (STFT) implemented via Python's *librosa* library. The STFT parameters matched those employed in RavenPro and PAMGuard software: a 1024-point frame size, 512-sample hop length, and 96 kHz sample rate. Given

the demonstrated efficacy of click detection for delphinid monitoring within vessel exclusion zones, we focused exclusively on click signals, cropping spectrograms to display the biologically relevant 15-48 kHz frequency range rather than the full 0-48 kHz spectrum. This selective approach enhanced resolution while eliminating non-informative frequency bands that could impede classification performance.

To accommodate the input requirements of pretrained convolutional neural networks (CNNs), we formatted all spectrograms as 3×224×224 pixel arrays in CHW format (Channel × Height × Width). Since grayscale spectrograms contain only one channel, we triplicated the single-channel matrix to generate three identical RGB channels, as illustrated in Figure 2.3.2.

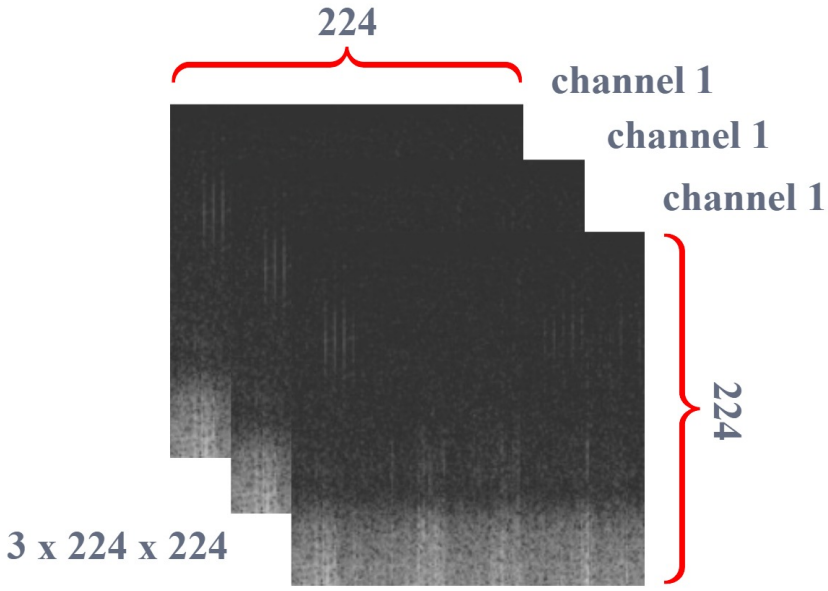


Figura 2.3.2: Data pipeline for CNN input preparation. Each 224×224 pixel grayscale spectrogram (representing acoustic data from one hydrophone channel) is converted to a three-dimensional RGB format through channel replication.

The four-channel acoustic recordings (from two hydrophone pairs) produced distinct spectrograms per channel. For positive samples, we exclusively used channels explicitly annotated during cataloging, while negative samples incorporated all available channels. This strategy optimized negative sample utilization, thereby improving model generalizability through exposure to diverse background noise conditions.

2.3.5 Experiments Configuration

Eight established convolutional neural network (CNN) architectures were evaluated for their classification performance: MobileNetV1, MobileNetV2, AlexNet, VGG19, VGG16, EfficientNetB0, DenseNet121, and ResNet18. These models were selected based on their demonstrated efficacy in image classification tasks (HUANG; LIAO, 2022; NEUPANE; ARYAL; RAJABIFARD, 2024; TRONG et al., 2022) and were initialized with ImageNet pretrained weights to leverage transfer learning benefits. All implementations utilized the PyTorch framework in Python.

A standardized preprocessing pipeline was applied to all input data. Spectrogram images underwent normalization using dataset-derived mean and standard deviation values, followed by conversion to PyTorch tensors. Binary labels were assigned (1 for positive samples, 0 for negatives) and organized into DataLoader objects for batch processing. Each model's final classification layer was modified from the original ImageNet configuration (1,000 classes) to a single-output neuron suitable for binary classification.

Preliminary experiments indicated model performance stabilized after approximately 15 training epochs, with negligible improvement thereafter. This observation informed our decision to fix the training duration at 15 epochs for all subsequent experiments.

The optimization framework employed Focal Loss (ROSS; DOLLÁR, 2017) to mitigate class imbalance effects by adaptively weighting difficult samples, complemented by the Adam optimizer for parameter updates. Given the inherent dataset imbalance, the Area Under the Receiver Operating Characteristic curve (ROC AUC) served as the primary evaluation metric, prioritizing discrimination capability over raw accuracy. Model selection was based on maximal validation set ROC AUC performance across all epochs, ensuring optimal classifier preservation.

2.4 Results

2.4.1 Exploratory Data Analysis

A comprehensive exploratory analysis was conducted on the audio dataset to characterize its structure and evaluate its suitability for training machine learning models. The dataset comprised 47 audio files of varying durations, with an average length of approximately seven minutes and 40 seconds.

Across these recordings, a total of 5,096 vocalization events were annotated, corresponding to acoustic signals such as clicks and whistles. Among these, 3,602 were labeled as click calls and 1,494 as whistles, indicating that click events occurred at more than twice the frequency of whistle events. In terms of duration, click signals cumulatively spanned approximately 4,157 seconds, whereas whistle events totaled 1,268 seconds.

Given the relatively limited number of whistle samples and the operational significance of each signal type, the analysis focused exclusively on click calls. According to acoustic monitoring protocols established by IBAMA, the detection of dolphin click trains by a PAM operator mandates the immediate suspension of air gun operations. This is because the directional nature of clicks enables the localization of delphinids within the EZ. In contrast, the presence of whistles does not permit accurate spatial localization and is therefore considered a less reliable indicator of animal presence within the exclusion zone.

The dataset also allowed for analysis of spatial signal distribution across hydrophones. Each audio channel corresponded to a hydrophone positioned along a towed array, with Channel 1 closest to the vessel and Channel 4 farthest. Figure 2.4.1 illustrates the number of detected calls per channel, categorized by signal type.

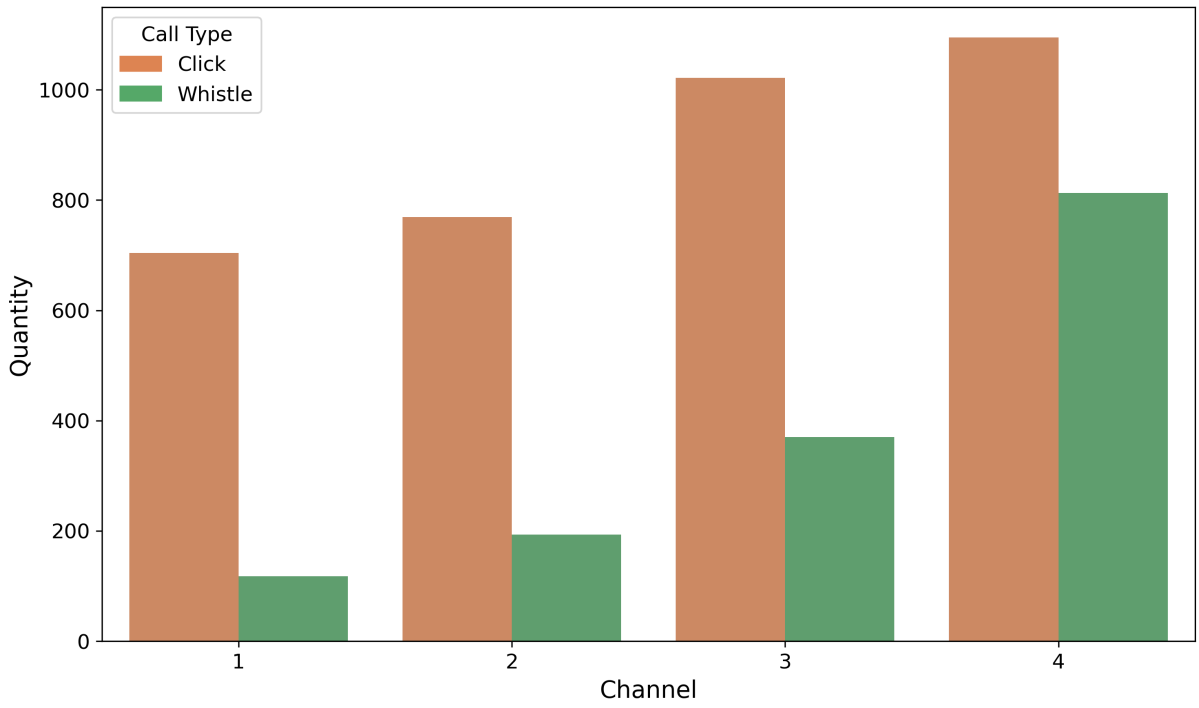


Figura 2.4.1: Number of call sessions per channel, categorized by type. Clicks are shown in orange, and whistles in green.

The results revealed a consistent increase in the number of detected click events across the array, from Channel 1 to Channel 4. This trend was even more pronounced for whistles, with the number of detections nearly doubling at each subsequent hydrophone. A detailed numerical summary is presented in Table 2.4.1.

Tabela 2.4.1: Distribution of call detections by type and channel.

Channel	Clicks	Whistles
1	704	118
2	773	193
3	1,025	370
4	1,100	813

In addition, the frequency distribution of call events was analyzed. Figure 2.4.2 presents box plots of the spectral range associated with each call type. Whistle events (right) predominantly occurred at frequencies below 20 kHz, whereas click events (left) were concentrated above 40 kHz. These patterns align with known acoustic characteristics of delphinid vocalizations and support the preprocessing strategies detailed in Section 2.3.4, particularly those involving frequency cropping.

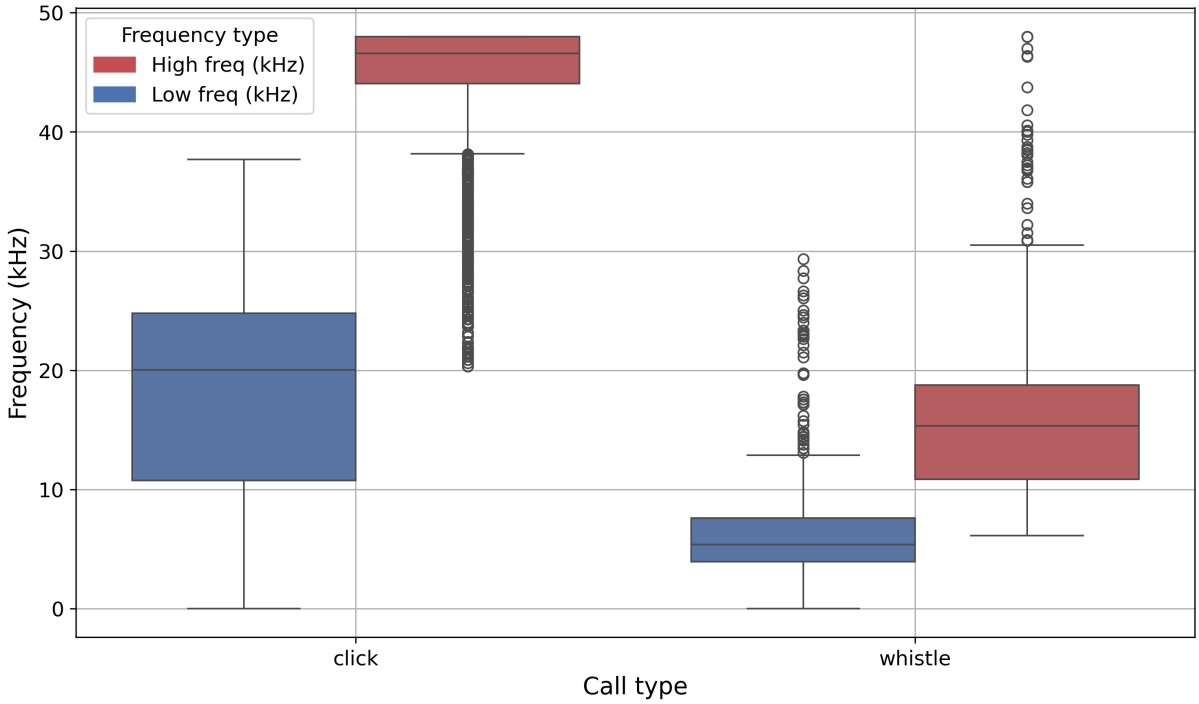


Figura 2.4.2: Box plots showing the minimum and maximum frequencies of clicks and whistles in the dataset.

2.4.2 Transfer Learning Experiments

Threshold Selection The output of the CNN models consisted of continuous values ranging from 0 to 1, representing the estimated probability that a given input image corresponded to a dolphin click train. Values close to 0 indicated low likelihoods of representing dolphin clicks, whereas values near 1 denoted high confidence. Due to the continuous nature of the output, it was necessary to define a classification threshold to distinguish between positive (click) and negative (non-click) predictions.

Given the primary objective of this study—to support environmental auditing—it was essential to emphasize the detection of true positive cases (i.e., actual dolphin clicks). This focus reflects the regulatory context in which missing a true event may imply failing to identify potential violations of environmental legislation. Therefore, the models were optimized to prioritize recall over precision, ensuring that as many true clicks as possible were identified, even at the cost of increasing false positives.

To formalize this trade-off, the F-Beta score (Equation 2.4.1) was employed, allowing the adjustment of the relative importance of recall and precision through the parameter β. When $\beta > 1$, the metric emphasizes recall; conversely, when $\beta < 1$, precision is prioritized. The F-Beta score ranges from 0 to 1, with higher values indicating better balance between recall and precision.

$$F_\beta = (1+\beta^2) \times \frac{\text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision}) + \text{Recall}} \tag{2.4.1}$$

In this work, the threshold was determined using $\beta = 2$, assigning twice as much importance to recall, since it represents the most critical metric for our objectives. From a monitoring and enforcement perspective, prioritizing the correct identification of positive samples is more relevant than achieving a balanced performance between positive and negative classifications. The F-beta scores were computed for each model across a range of threshold values on the validation set, and the results are presented in Figure 2.4.3.

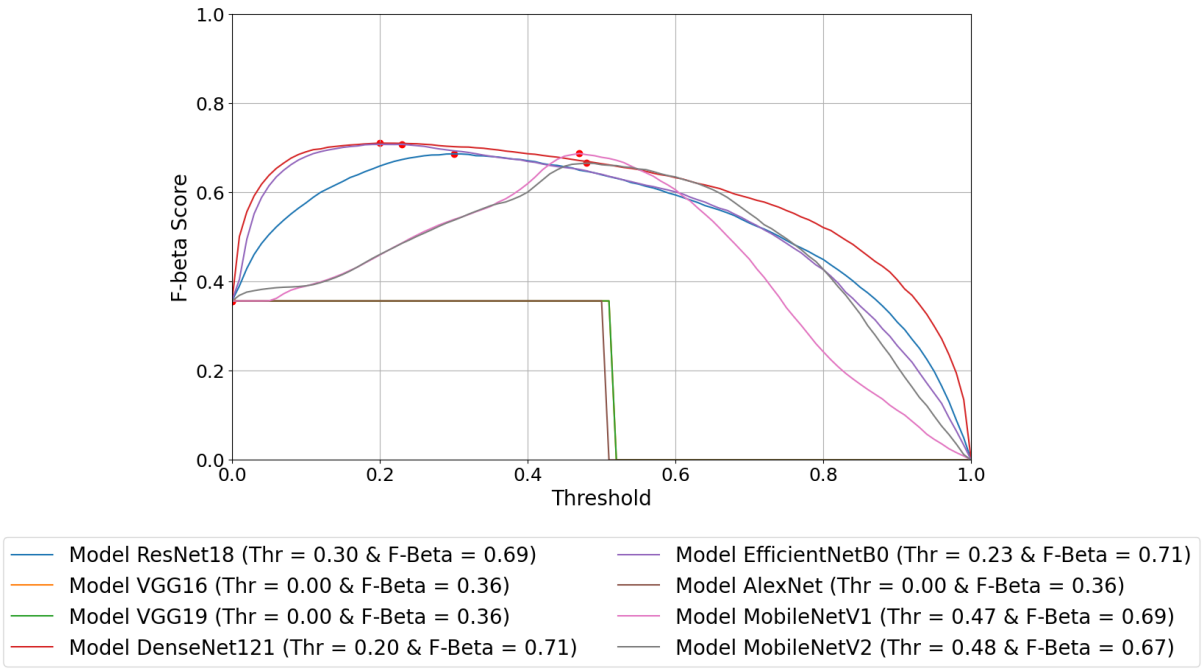


Figura 2.4.3: F-Beta score vs. classification threshold ($\beta = 2$) on the validation set. Each curve corresponds to a different model. The red circle marks the threshold at which each model achieved its highest F-Beta score.

DenseNet121 and EfficientNetB0 obtained the highest F-Beta scores (0.71) at relatively low optimal thresholds (0.23 and 0.20, respectively). MobileNetV1 and MobileNetV2 demonstrated slightly lower F-Beta scores (0.67 and 0.69), operating at more conservative thresholds (0.48 and 0.47). ResNet18 showed intermediate behavior, with performance and threshold values between those two groups. In contrast, AlexNet, VGG16, and VGG19 exhibited abnormal response profiles, with nearly constant F-Beta values across most thresholds, followed by abrupt performance drops near 0.5.

Performance Metrics Table 2.4.2 presents the evaluation metrics obtained on the test set using the thresholds selected via F-Beta optimization. As expected for an imbalanced dataset, accuracy alone proved insufficient for assessing model performance. For example, AlexNet, VGG16, and VGG19 all reported an accuracy and precision of 0.1690, alongside a recall of 1.0000. This resulted in a low F1-score of 0.2892 and AUC values of 0.5000—indicating random classification behavior. These models were saved after just one training epoch, suggesting premature termination and ineffective learning.

EfficientNetB0 achieved the highest overall performance, with an F1-score of 0.8071, recall of 0.8694, precision of 0.7532, and an AUC of 0.9057, having been saved at epoch 6. MobileNetV1 and ResNet18 also yielded strong results, with F1-scores of 0.7994 and 0.7913,

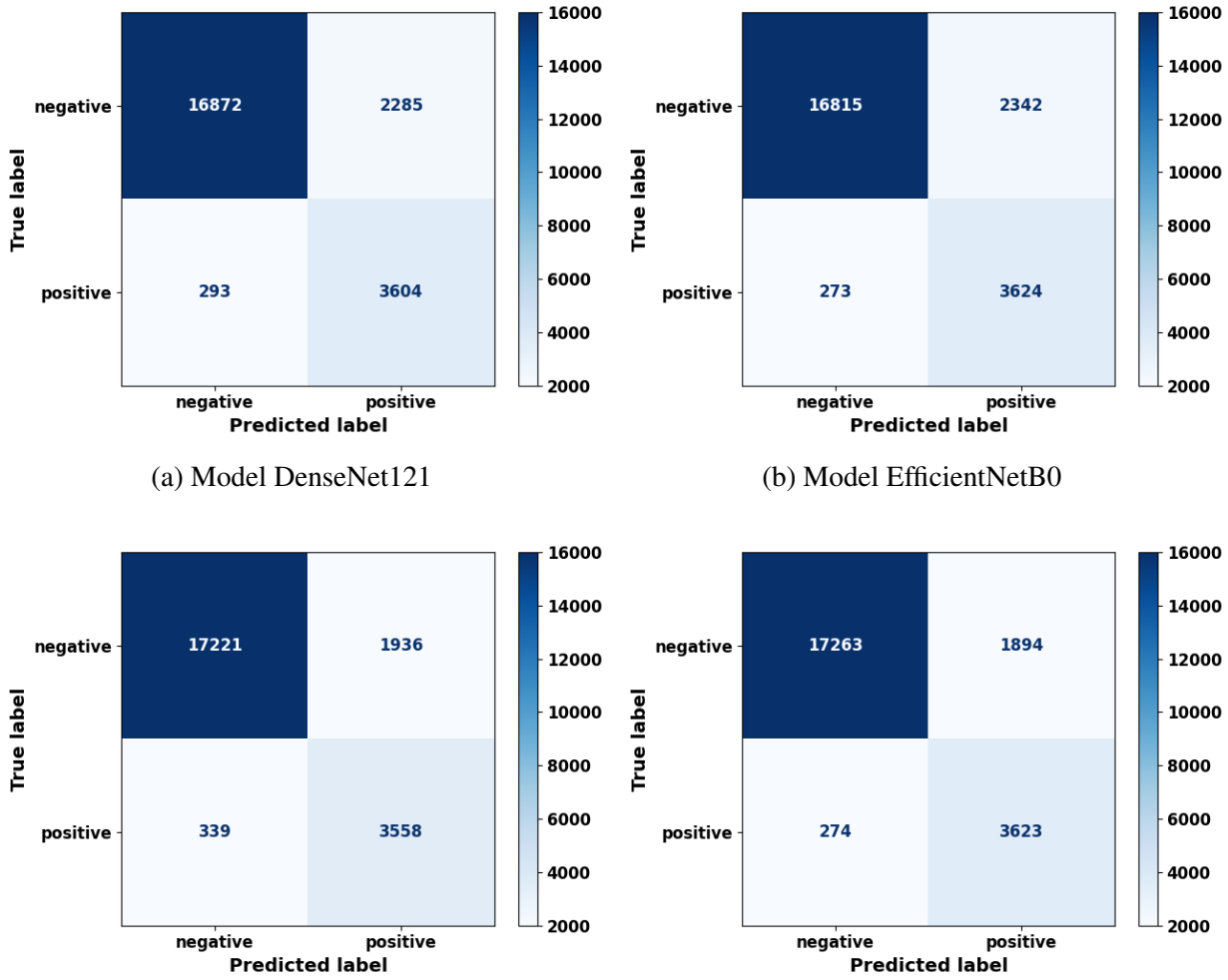
respectively. MobileNetV1 had the highest recall among the top-performing models (0.8927), albeit with lower precision (0.7237). ResNet18 exhibited a more balanced trade-off, with precision of 0.7561 and recall of 0.8299. MobileNetV2 demonstrated the highest recall overall (0.9230) and the highest AUC (0.9188), although its lower precision (0.6871) resulted in an F1-score of 0.7878. DenseNet121 reported the highest precision (0.7697), along with a recall of 0.8188 and an F1-score of 0.7935. This model was saved after 9 epochs—the latest among all tested models—potentially indicating a slower but more stable learning process.

Tabela 2.4.2: Performance metrics on the test set. Thresholds were determined by F-Beta optimization ($\beta = 2$). The highest values in each metric are highlighted in bold.

Model	**Accuracy**	**Precision**	**Recall**	**AUC**	**F1-Score**	**Epoch Saved Model**
AlexNet	0.1690	0.1690	1.0000	0.5000	0.2892	1
DenseNet	0.9280	**0.7697**	0.8188	0.8845	0.7935	**9**
EfficientNetB0	**0.9298**	0.7532	0.8694	0.9057	**0.8071**	6
MobileNetV1	0.9243	0.7237	0.8927	0.9117	0.7994	2
MobileNetV2	0.9159	0.6871	**0.9230**	**0.9188**	0.7878	2
ResNet18	0.9260	0.7561	0.8299	0.8877	0.7913	7
VGG16	0.1690	0.1690	1.0000	0.5000	0.2892	1
VGG19	0.1690	0.1690	1.0000	0.5000	0.2892	1

Confusion Matrix Analysis To further examine the classification performance, confusion matrix were generated for the four top-performing models: DenseNet121, EfficientNetB0, MobileNetV1, and MobileNetV2 (Figure 2.4.4). These matrices were constructed using the optimal thresholds derived from the F-Beta analysis with $\beta = 2$.

All models correctly classified over 90% of the positive samples and approximately 80% of the negative samples. This outcome suggests that the models exhibited strong discriminative capacity across both classes, with slight variations in error distribution.



(a) Model DenseNet121

(b) Model EfficientNetB0

(c) Model MobileNetV1

(d) Model MobileNetV2

Figura 2.4.4: Confusion matrices for four different models evaluated with thresholds optimized for $\beta = 2$.

2.5 Discussion

Exploratory Data Analysis The distribution patterns observed during the exploratory analysis can be attributed, in part, to the influence of background noise generated by the vessel. Hydrophones 1 and 2, being positioned closer to the stern, were more affected by low-frequency noise, which can mask acoustic signals—especially whistles, whose energy is predominantly concentrated in lower frequencies. Although dolphin presence was presumably uniform along the array, signal detectability was compromised in channels nearest to the vessel. In contrast, clicks—characterized by higher frequency content—were less susceptible to this masking effect and remained more reliably detectable even under elevated noise conditions.

These findings reinforce the rationale behind IBAMA's recommendation to position the first hydrophone at least 200 meters from the vessel's stern and the final hydrophone at approximately 300 meters. Increasing the distance between the acoustic sensors and the noise source enhances the detectability of biologically relevant signals and reduces the impact of anthropogenic interference. This configuration strengthens the role of echolocation click detection as a robust strategy for automated marine mammal monitoring within the EZ.

Threshold Selection Behavior The analysis of threshold behavior across models provided important insights for applications in environmental compliance monitoring. Although DenseNet121 and EfficientNetB0 achieved the highest F_β scores, their optimal thresholds were notably low (below 0.25), suggesting a predisposition to overclassification and, consequently, an increased rate of false positives. Such behavior may be acceptable—or even desirable—in regulatory contexts that prioritize sensitivity over specificity.

Conversely, MobileNetV1 and MobileNetV2 operated at more conservative thresholds (near 0.5), suggesting a more selective classification pattern with improved specificity. These models maintained adequate recall levels, making them promising candidates for applications requiring balanced detection strategies. ResNet18 exhibited intermediate behavior, offering a compromise between sensitivity and specificity. On the other hand, the flat response curves observed in VGG16, VGG19, and AlexNet suggest poor discriminative capacity and highlight their unsuitability for this task. These observations emphasize the importance of not only considering overall performance metrics but also evaluating the decision threshold dynamics when selecting models for deployment.

Model Performance Metrics The comparative evaluation of performance metrics underscored the superior suitability of EfficientNetB0, MobileNetV1, MobileNetV2, and ResNet18 for the classification of click events. These models demonstrated a capacity to learn discriminative features effectively, even under noisy and imbalanced conditions. EfficientNetB0, in particular, achieved the highest F1-score and a strong balance between recall and precision, indicating robust generalization and reliability in identifying true positive cases while minimizing false alarms.

MobileNetV1 attained the highest recall among all models, making it particularly advantageous in scenarios where missing a positive case (i.e., a dolphin click) is operationally critical. MobileNetV2, while achieving the highest recall and AUC, showed a modest drop in precision, suggesting a slight trade-off in specificity. In contrast, DenseNet121 exhibited the highest precision, coupled with stable performance and the longest training duration among the tested models—potentially reflecting a more conservative and gradual learning dynamic.

The poor performance of AlexNet, VGG16, and VGG19—manifested in uniformly low precision and accuracy, and a recall of 1.0—was likely a consequence of the models defaulting to classifying all inputs as positive. This behavior reflects both the challenge of learning from imbalanced datasets and the risk of relying solely on accuracy as a performance indicator. Moreover, these models were saved after only a single training epoch, suggesting insufficient learning and poor convergence. This highlights the need for robust training strategies, including early stopping criteria and careful monitoring of validation curves, especially in the context of noisy acoustic data.

Confusion Matrix Analysis The confusion matrix further clarified the error patterns across the top-performing models. DenseNet121 and EfficientNetB0 exhibited a higher number of false positives, likely due to their low decision thresholds and heightened sensitivity to positive instances. While this may be acceptable in precautionary monitoring contexts, it could result in a higher incidence of operational interruptions due to false alarms.

In contrast, MobileNetV1 and MobileNetV2 demonstrated slightly increased false negative rates, which may pose risks in scenarios where undetected positive cases could lead to environmental non-compliance. Notably, the difference between MobileNetV2 and EfficientNetB0 was minimal—only one additional false negative—suggesting practical equivalence in some use cases. When jointly considering sensitivity and specificity, MobileNetV2 emerged as the most balanced model, offering a reliable trade-off between over-detection and under-

detection.

2.6 Conclusion

The results presented in this study underscore both the challenges and the potential of applying machine learning techniques to PAM data collected under non-ideal conditions, specifically during seismic survey operations. Unlike most previous studies—which typically rely on data acquired during field expeditions explicitly designed for wildlife monitoring, such as the widely used Watkins Marine Mammal Sound Database—our dataset was obtained as a byproduct of industrial activity in the Southwestern Atlantic Ocean, a region underrepresented in the literature. Although considerably noisier due to the operational environment, this dataset contributes novel and regionally specific data to the field. Importantly, since the objective of our study was not behavioral analysis but rather the detection of delphinid echolocation clicks, the presence of background noise becomes a realistic and integral aspect of the classification challenge.

Prior research in the area frequently employs deep CNNs for the identification of marine mammals using spectrograms, particularly in grayscale and with two-second durations, as reported by Buchanan et al. (2021) and Jiang et al. (2019). Our findings align with this trend, as the best results were also achieved using two-second spectrograms. Preliminary tests with longer, five-second segments yielded inferior performance, which supports our methodological decision to prioritize shorter time windows.

Our results further highlight the importance of model selection when dealing with noisy acoustic environments. Among the architectures evaluated, MobileNetV2 achieved the highest F-Beta score while maintaining a lightweight structure of only 3.5 million parameters. In contrast, older and heavier architectures such as VGG16, VGG19, and AlexNet exhibited significantly lower performance, likely due to their high parameter counts (up to 144 million in the case of VGG19) and limited generalization capacity under noise-heavy conditions. These models struggled to adapt to the task even with extended training, indicating that they may require substantial fine-tuning and computational resources to be effective with our dataset. In comparison, more recent and efficient models—such as ResNet18, DenseNet121, EfficientNetB0, and other MobileNet variants—demonstrated superior robustness and classification accuracy.

It is important to emphasize that the goal of the proposed model is not species-level identification but rather to support environmental monitoring efforts by reliably flagging the

presence of delphinid click activity. This approach is consistent with real-world applications in seismic operations, where the presence of PAM operators is mandatory to mitigate environmental impacts. By addressing a gap in the literature related to machine learning applications for PAM in the South Atlantic, this study provides a foundation for further research using publicly accessible data from this region.

Furthermore, the ability to detect delphinid activity in acoustically challenging environments has implications beyond operational monitoring. Enhanced detection tools can contribute to the enforcement of environmental regulations and the management of marine protected areas, particularly in regions exposed to high levels of anthropogenic noise, such as near industrial ports. Ensuring that the presence of vulnerable species is accurately recorded—even under suboptimal conditions—and that the rate of false negatives remains low is essential, as missed detections could coincide with periods of increased risk to the animals. This reinforces the scientific basis for conservation policies and supports informed decision-making in marine management.

As future work, we propose the development of custom lightweight architectures specifically optimized for noisy acoustic datasets, with improved computational efficiency. Additionally, further studies could explore data augmentation techniques, self-supervised learning approaches, and the application of the proposed methodology to other geographic regions or signal types.

3 Discussão Geral

Este trabalho explorou o uso de modelos pré-treinados na tarefa de classificação binária de sinais acústicos de odontocetos. Embora essa abordagem tenha apresentado resultados promissores, o ideal seria o desenvolvimento de um modelo específico para essa aplicação, treinado a partir de dados brutos obtidos no Brasil e catalogados conforme a metodologia proposta. Tal modelo permitiria um ajuste mais preciso às características dos sinais acústicos locais, aumentando a acurácia da classificação e a aplicabilidade em estudos ambientais e de monitoramento.

Um desdobramento natural desta pesquisa seria a ampliação do escopo para a classificação de diferentes tipos de sinais de golfinhos — como os assobios — e também para outras espécies de cetáceos, expandindo o uso além do contexto de fiscalização ambiental. O desenvolvimento de um modelo capaz de identificar e diferenciar uma variedade de sinais acústicos fracos, emitidos por distintas espécies de mamíferos marinhos, especialmente em ambientes com baixa relação sinal-ruído, abriria novas possibilidades para a ecologia acústica. Essa abordagem poderia ser empregada em pesquisas sobre comportamento animal, impactos da poluição sonora nos oceanos e, principalmente, como ferramenta de fiscalização ambiental, com potencial de uso por diferentes países que realizam atividades de prospecção sísmica.

Além disso, recomenda-se que futuros trabalhos priorizem a criação de bases de dados públicas e abertas, contendo sinais acústicos de diferentes espécies coletados em contextos reais, como pesquisas sísmicas. A adoção de técnicas mais avançadas de aprendizado profundo — como modelos auto-supervisionados ou redes neurais convolucionais especializadas em áudio — também poderia contribuir para melhorar a acurácia e a capacidade de generalização dos modelos desenvolvidos.

4 Conclusões Gerais

O uso de modelos pré-treinados demonstrou ser uma estratégia viável e promissora para a classificação de sinais acústicos de odontocetos, servindo como um ponto de partida relevante para aplicações ambientais. No entanto, os resultados indicam que abordagens mais específicas, baseadas em dados locais e metodologias especializadas, têm potencial para gerar modelos mais eficazes e robustos.

A evolução dessa linha de pesquisa pode contribuir significativamente para a conservação marinha, especialmente se aliada a sistemas de monitoramento em tempo real. Ao permitir a identificação automatizada de espécies e a detecção precoce de impactos ambientais, essas soluções podem auxiliar na formulação de políticas públicas e no desenvolvimento de estratégias globais para proteção da biodiversidade oceânica.

5 Contribuições Científicas

As principais contribuições científicas deste trabalho são:

- Desenvolvimento de uma metodologia para a geração e organização de dados destinados ao treinamento de modelos de aprendizado de máquina, com foco em informações adquiridas em pesquisas sísmicas.
- Construção e disponibilização de uma base de dados estruturada, acessível publicamente, para fomentar investigações futuras sobre a detecção de sinais acústicos em ambientes de prospecção sísmica. Bases de dados disponibilizadas em:
  - Sauerbronn, Flora (2025), *"PAM Recordings of Seismic Surveys - IBAMA Regulations in Brazil - part1"*, Mendeley Data, V1, doi: 10.17632/w9b5z7znst.1
  - Sauerbronn, Flora (2025), *"PAM Recordings of Seismic Surveys - IBAMA Regulations in Brazil - part2"*, Mendeley Data, V1, doi: 10.17632/8srrb2rcw6.1
- Avaliação de diferentes modelos de aprendizado de máquina, destacando o modelo MobileNetV2 como a solução mais eficaz para a detecção automatizada de cliques em cenários de prospecção sísmica, utilizando dados obtidos conforme a legislação brasileira.

Referências

AU, W. W. Characteristics of dolphin sonar signals. In: *The sonar of dolphins*. [S.l.]: Springer, 1993. p. 115–139.

AU, W. W.; HASTINGS, M. C. *Principles of marine bioacoustics*. [S.l.]: Springer, 2008. v. 510.

BERMANT, P. C. et al. Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. *Scientific reports*, Nature Publishing Group UK London, v. 9, n. 1, p. 12588, 2019.

BIANCO, M. J. et al. Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America*, AIP Publishing, v. 146, n. 5, p. 3590–3628, 2019.

BUCHANAN, C. et al. Deep convolutional neural networks for detecting dolphin echolocation clicks. In: IEEE. *2021 36th International Conference on image and vision computing New Zealand (IVCNZ)*. [S.l.], 2021. p. 1–6.

CARROLL, A. et al. A critical review of the potential impacts of marine seismic surveys on fish & invertebrates. *Marine Pollution Bulletin*, Elsevier, v. 114, n. 1, p. 9–24, 2017.

COMPTON, R. et al. A critical examination of worldwide guidelines for minimising the disturbance to marine mammals during seismic surveys. *Marine Policy*, Elsevier, v. 32, n. 3, p. 255–262, 2008.

DALBEN, A.; AVILA, T. Monitoramento acústico passivo (map) como ferramenta de auditoria ambiental no ambiente marinho. In: *Congresso Brasileiro de Bioacústica*. [S.l.: s.n.], 2021.

ELLIOTT, R. G. *Passive acoustic monitoring of habitat use by bottlenose dolphins in Doubtful Sound*. Tese (Doutorado) — University of Otago, 2011.

GILLESPIE, D. et al. Pamguard: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans. *The Journal of the Acoustical Society of America*, Acoustical Society of America, v. 125, n. 4_Supplement, p. 2547–2547, 2009.

GRACIC, M.; GUBNISKY, G.; DIAMANT, R. Review of cetacean's click detection algorithms. *arXiv preprint arXiv:2402.04735*, 2024.

HUANG, M.-L.; LIAO, Y.-C. A lightweight cnn-based network on covid-19 detection using x-ray and ct images. *Computers in Biology and Medicine*, Elsevier, v. 146, p. 105604, 2022.

IBAMA. *Guia de monitoramento de biota marinha em atividade de aquisição de dados sísmicos*. 2018. Acesso em: 09 jan. 2023. Disponível em: ⟨http://www.ibama.gov.br/phocadownload/licenciamento/petroleo-e-gas/diretrizes/2018-11-01-ibama-guia_de_monitoramento_da_biota_marinha_outubro.pdf⟩.

ICMBIO. *Protocolo Sobre Diagnóstico e Avaliação dos Efeitos da Pesquisa Sísmica em Mamíferos Aquáticos*. 2020. Acesso em: 16 fev 2023. Disponível em: ⟨https://www.icmbio.gov.br/cma/images/stories/Publica%C3%A7%C3%B5es/Protocolo-Sismica-Mamiferos-Aquaticos.pdf⟩.

JIANG, J.-j. et al. Whistle detection and classification for whales based on convolutional neural networks. *Applied Acoustics*, Elsevier, v. 150, p. 169–178, 2019.

LICCIARDI, A.; CARBONE, D. Whalenet: a novel deep learning architecture for marine mammals vocalizations on watkins marine mammal sound database. *IEEE Access*, IEEE, 2024.

LICCIARDI, A.; CARBONE, D.; RONDONI, L. Wavelet scattering operators for multiscale processes: The case study. In: SPRINGER NATURE. *Proceedings of the 2nd International Conference on Nonlinear Dynamics and Applications (ICNDA 2024), Volume 3: Dynamical Models, Communications and Networks*. [S.l.], 2024. p. 173.

MADSEN, P.; KERR, I.; PAYNE, R. Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: false killer whales pseudorca crassidens and risso's dolphins grampus griseus. *Journal of Experimental Biology*, Company of Biologists, v. 207, n. 11, p. 1811–1823, 2004.

MADSEN, P. T.; WAHLBERG, M. Recording and quantification of ultrasonic echolocation clicks from free-ranging toothed whales. *Deep sea research part I: oceanographic research papers*, Elsevier, v. 54, n. 8, p. 1421–1444, 2007.

MCCAULEY, R. D. et al. Widely used marine seismic survey air gun operations negatively impact zooplankton. *Nature ecology & evolution*, Nature Publishing Group UK London, v. 1, n. 7, p. 0195, 2017.

MERCHANT, N. D. Underwater noise abatement: Economic factors and policy options. *Environmental science & policy*, Elsevier, v. 92, p. 116–123, 2019.

MURAINA, I. Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts. In: *7th international Mardin Artuklu scientific research conference*. [S.l.: s.n.], 2022. p. 496–504.

MURPHY, D. T. et al. Residual learning for marine mammal classification. *IEEE Access*, IEEE, v. 10, p. 118409–118418, 2022.

NEUPANE, B.; ARYAL, J.; RAJABIFARD, A. Cnns for remote extraction of urban features: A survey-driven benchmarking. *Expert Systems with Applications*, Elsevier, v. 255, p. 124751, 2024.

NUUTTILA, H. K. et al. Acoustic detection probability of bottlenose dolphins, tursiops truncatus, with static acoustic dataloggers in cardigan bay, wales. *The Journal of the Acoustical Society of America*, AIP Publishing, v. 134, n. 3, p. 2596–2609, 2013.

PARENTE, C. L.; ARAÚJO, M. E. de. A aquisição sísmica marítima no brasil e seus potenciais efeitos na ordem cetacea. *Natural Resources (1984-5901)*, v. 2, n. 1, 2011.

PARENTE, C. L.; ARAÚJO, M. E. de. Effectiveness of monitoring marine mammals during marine seismic surveys off northeast brazil. *Revista de Gestão Costeira Integrada-Journal of Integrated Coastal Zone Management*, Associação Portuguesa dos Recursos Hídricos, v. 11, n. 4, p. 409–419, 2011.

PARVIN, S.; NEDWELL, J.; HARLAND, E. Lethal and physical injury of marine mammals, and requirements for passive acoustic monitoring. *Subacoustech Report Reference: 565R0212, February*, Citeseer, 2007.

PETRÓLEO, G. N. e. B. BRASIL. Agência Nacional do. *Especial ANP 20 anos*. 2023. ⟨https://www.gov.br/anp/pt-br/acesso-a-informacao/institucional/especial-anp-20-anos⟩. Acesso em: 24 maio 2024.

REYES-ZAMUDIO, M. M. *T-POD detection and acoustic behaviour of bottlenose dolphins (Tursiops truncatus) in Cardigan Bay SAC: a comparison between T-POD recordings and visual observations*. Tese (Doutorado) — University of Wales Bangor, 2005.

ROMAN, J.; MCCARTHY, J. J. The whale pump: marine mammals enhance primary productivity in a coastal basin. *PloS one*, Public Library of Science, v. 5, n. 10, p. e13255, 2010.

ROSS, T.-Y.; DOLLÁR, G. Focal loss for dense object detection. In: *proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 2980–2988.

SARNOCIŃSKA, J. et al. Harbor porpoise (phocoena phocoena) reaction to a 3d seismic airgun survey in the north sea. *Frontiers in Marine Science*, Frontiers Media SA, v. 6, p. 824, 2020.

SCHMITT, J. Crime sem castigo: a efetividade da fiscalização ambiental para o controle do desmatamento ilegal na amazônia. 2016.

SEYDI, V. et al. The application of neural networks to classify dolphin echolocation clicks. *bioRxiv*, Cold Spring Harbor Laboratory, p. 2022–06, 2022.

SIMÕES, M. *Revisão de detecções acústicas de delfinídeos em projetos de prospecção sísmica: subsídios para a fiscalização ambiental*. Trabalho de Conclusão de Curso — Universidade Federal de Santa Catarina (UFSC), 2022.

SOUTHALL, B. L. et al. Marine mammal noise exposure criteria: Updated scientific recommendations for residual hearing effects. *Aquatic Mammals*, Aquatic Mammals, v. 45, n. 2, p. 125–232, 2019.

STONE, C. et al. The effects of seismic operations in uk waters: analysis of marine mammal observer data. *J. Cetacean Res. Manage.*, v. 16, n. 1, p. 71–85, 2017.

TRONG, T. N. et al. An empirical evaluation of feature extraction for vietnamese fruit classification. *Vietnam Journal of Science and Technology*, v. 60, n. 5, p. 837–852, 2022.

VERFUSS, U. K. et al. Comparing methods suitable for monitoring marine mammals in low visibility conditions during seismic surveys. *Marine Pollution Bulletin*, Elsevier, v. 126, p. 1–18, 2018.

WHITE, E. L. et al. More than a whistle: Automated detection of marine sound sources with a convolutional neural network. *Frontiers in Marine Science*, Frontiers Media SA, v. 9, p. 879145, 2022.

WRIGHT, A. J.; COSENTINO, A. M. Jncc guidelines for minimising the risk of injury and disturbance to marine mammals from seismic surveys: We can do better. *Marine Pollution Bulletin*, Elsevier, v. 100, n. 1, p. 231–239, 2015.

YAP, B. W. et al. An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In: SPRINGER. *Proceedings of the first international conference on advanced data and information engineering (DaEng-2013)*. [S.l.], 2014. p. 13–22.

ZIEGENHORN, M. A. et al. Discriminating and classifying odontocete echolocation clicks in the hawaiian islands using machine learning methods. *Plos one*, Public Library of Science San Francisco, CA USA, v. 17, n. 4, p. e0266424, 2022.

A APÊNDICE - Métricas

Nesta seção são apresentadas as principais métricas utilizadas para avaliação do desempenho dos modelos de classificação binária. As métricas são derivadas da matriz de confusão, que é composta por quatro elementos fundamentais:

- **VP (Verdadeiro Positivo)**: Amostras positivas corretamente classificadas como positivas.
- **VN (Verdadeiro Negativo)**: Amostras negativas corretamente classificadas como negativas.
- **FP (Falso Positivo)**: Amostras negativas classificadas incorretamente como positivas.
- **FN (Falso Negativo)**: Amostras positivas classificadas incorretamente como negativas.

Com base nesses elementos, as métricas são definidas conforme a Tabela A.0.1:

Tabela A.0.1: Métricas utilizadas para avaliação de desempenho de modelos de classificação binária

Métrica	Definição	Equação
acuracy	Proporção de previsões corretas entre todas as amostras	$\frac{VP+VN}{VP+VN+FP+FN}$
precision	Proporção de positivos corretamente identificados entre os previstos como positivos	$\frac{VP}{VP+FP}$
recall	Proporção de positivos corretamente identificados entre todos os reais positivos	$\frac{VP}{VP+FN}$
F1-Score	Média harmônica entre *precision* e *recall*	$\frac{2\cdot(precision\cdot recall)}{precision+recall}$
ROC-AUC	Área sob a curva ROC (TPR vs FPR)	$\int_0^1 TPR(FPR)\,d(FPR)$
PR-AUC	Área sob a curva *precision* vs *recall*	$\int_0^1 precision(recall)\,d(recall)$

A **curva ROC** (Receiver Operating Characteristic) é construída a partir da taxa de verdadeiros positivos (TPR = *recall*) contra a taxa de falsos positivos (FPR = $\frac{FP}{FP+VN}$). Já a **curva *precision-recall*** mostra a relação entre *precision* e *recall* para diferentes limiares de decisão. Quanto maior a área sob essas curvas, melhor o desempenho do modelo.

B APÊNDICE - Especificações Computacionais

Os experimentos de aprendizado de máquina descritos neste trabalho foram conduzidos em uma estação de trabalho de alto desempenho, configurada da seguinte forma:

- **Processador:** Intel Xeon W7-3455, 24 núcleos / 48 threads, frequência base de 2,5 GHz (turbo até 4,8 GHz), cache de 67,5 MB.
- **Memória:** 256 GB DDR5 ECC REG 4800 MHz (4 × 64 GB).
- **Placa de vídeo:** NVIDIA GeForce RTX 3060 com 12 GB de memória de vídeo e 3.584 núcleos CUDA.
- **Armazenamento:** SSD M.2 PCIe x4 NVMe de 1 TB (classe workstation).
- **Placa-mãe:** Chipset W790.
- **Fonte de alimentação:** 850 W 80 Plus com PFC ativo (tensão de entrada entre 90–240 V).
- **Sistema operacional:** Linux Ubuntu Desktop.
- **Gabinete e refrigeração:** Deepcool Macube 310, reforçado para configuração de um único processador.

Essa configuração garantiu desempenho computacional estável durante o treinamento dos modelos e o processamento dos dados, especialmente nos experimentos de aprendizado profundo que envolveram grandes volumes de dados.