



MARINHA DO BRASIL
INSTITUTO DE ESTUDOS DO MAR ALMIRANTE PAULO MOREIRA
PROGRAMA DE PÓS-GRADUAÇÃO EM ACÚSTICA SUBMARINA

DETECÇÃO E CLASSIFICAÇÃO AUTOMÁTICA DE SINAIS ACÚSTICOS DE
BAIXA FREQUÊNCIA DE *BALAENOPTERA BOREALIS* (BALEIA SEI)
UTILIZANDO TÉCNICAS DE VISÃO COMPUTACIONAL

Vinícius Reveilleau

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Acústica Submarina, PPGAS, do Instituto de Estudos do Mar Almirante Paulo Moreira, como parte dos requisitos necessários à obtenção do título de Mestre em Acústica Submarina.

Orientadores: Rogério de Moraes Calazan
Fábio Contrera Xavier

Arraial do Cabo
Outubro de 2023



**MARINHA DO BRASIL
INSTITUTO DE ESTUDOS DO MAR ALMIRANTE PAULO MOREIRA
SUPERINTENDÊNCIA DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ACÚSTICA SUBMARINA**

**ATA DA DEFESA DE DISSERTAÇÃO DE MESTRADO Nº 04
DISSERTAÇÃO PARA OBTENÇÃO DO TÍTULO DE
MESTRE EM ACÚSTICA SUBMARINA**

Aos trinta dias do mês de outubro do ano de dois mil e vinte e três, às nove horas, pelo Programa de Pós-Graduação em Acústica Submarina, em Arraial do Cabo, RJ, realizou-se, de forma presencial, a prova de Defesa de Dissertação, intitulada: "*Detecção e classificação automática de sinais acústicos de baixa frequência de Balaenoptera borealis utilizando técnicas de computação visual*" de autoria do candidato **Vinicius Reveilleau**, aluno do Programa em Acústica Submarina, em nível de Mestrado. A Banca Examinadora esteve constituída pelos professores: Dr. Rogério de Moraes Calazan (Instituto de Estudos do Mar Alte. Paulo Moreira – IEAPM) – Orientador – Presidente, Dr. Fábio Contrera Xavier (Instituto de Estudos do Mar Alte. Paulo Moreira – IEAPM) – Coorientador, Dr. Marcus Vinicius da Silva Simões (Instituto de Estudos do Mar Alte. Paulo Moreira – IEAPM) – Membro Interno, Dr. Alexandre Geddes Lemos Guarino (Instituto de Estudos do Mar Alte. Paulo Moreira – IEAPM) – Membro Interno e Dr. Fernando Moraes de Oliveira (Instituto Federal do Rio de Janeiro – IFRJ) – Membro Externo. Concluídos os trabalhos de apresentação e arguição, o candidato foi aprovado pela Banca Examinadora. Foi concedido um prazo de 30 dias para o candidato efetuar as correções sugeridas pela Banca Examinadora, apresentar o trabalho em sua redação definitiva, atender ao requisito de publicação estabelecido nas normas do PPGAS e entregar a documentação necessária para elaboração do Diploma sob pena de não expedição do Diploma. E, para constar, foi lavrada a presente ata, que vai assinada pelos membros da Banca.

Prof. Ph.D. Rogério de Moraes Calazan – IEAPM

Prof. D.Sc. Fábio Contrera Xavier – IEAPM

Prof. D.Sc. Marcus Vinicius da Silva Simões – IEAPM

Prof. Ph.D. Alexandre Geddes Lemos Guarino – IEAPM

Prof. D.Sc. Fernando Moraes de Oliveira – IFRJ

Prof. D.Sc. Eduardo Barros Fagundes Netto – IEAPM

DETECÇÃO E CLASSIFICAÇÃO AUTOMÁTICA DE SINAIS ACÚSTICOS DE
BAIXA FREQUÊNCIA DE *BALAENOPTERA BOREALIS* (BALEIA SEI)
UTILIZANDO TÉCNICAS DE VISÃO COMPUTACIONAL

Vinícius Reveilleau

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA
DE PÓS-GRADUAÇÃO EM ACÚSTICA SUBMARINA DO INSTITUTO DE
ESTUDOS DO MAR ALMIRANTE PAULO MOREIRA COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ACÚSTICA SUBMARINA.

Orientadores: Prof. Rogério de Moraes Calazan, D.Sc.
Prof. Fábio Contrera Xavier, D.Sc.

Aprovada por: Prof. Fernando Moraes de Oliveira, D.Sc.
Prof. Marcus Vinícius da Silva Simões, D.Sc.
Prof. Alexandre Geddes Lemos Guarino, Ph.D.

ARRAIAL DO CABO, RJ – BRASIL
OUTUBRO DE 2023

Reveilleau, Vinícius

Detecção e Classificação Automática de Sinais Acústicos de Baixa Frequência de *Balaenoptera borealis* (baleia sei) utilizando Técnicas de Visão Computacional/Vinícius Reveilleau. – Arraial do Cabo: IEAPM, 2023.

XVI, 54 p.: il.; 29, 7cm.

Orientadores: Rogério de Moraes Calazan

Fábio Contrera Xavier

Dissertação (mestrado) – IEAPM/Programa de Pós-Graduação em Acústica Submarina, 2023.

Referências Bibliográficas: p. 48 – 54.

1. Acústica de cetáceos. 2. Processamento de imagem.
3. Aprendizado de máquina. I. de Moraes Calazan, Rogério *et al.* II. Instituto de Estudos do Mar Almirante Paulo Moreira, Programa de Pós-Graduação em Acústica Submarina. III. Título.

Agradecimentos

O fim deste ciclo não teria sido possível sem o apoio, orientação e incentivo de muitas pessoas e instituições. Gostaria de expressar minha profunda gratidão a todos que contribuíram de alguma forma para este trabalho.

Primeiramente, quero agradecer aos meus orientadores, Dr.Fábio Contrera Xavier e Dr.Rogério de Moraes Calazan, pela orientação dedicada e pelos valiosos *insights* ao longo de todo o processo de pesquisa.

À minha família, que sempre acreditou em mim e me apoiou incondicionalmente. Seu apoio emocional e encorajamento foram a âncora que me manteve focado durante os momentos desafiadores. Em especial à minha namorada Ariela Vulcano, que acreditou em mim em momentos que eu duvidava das minhas escolhas e capacidade. Obrigado por estar ao meu lado por incontáveis horas de estudo. Todo seu carinho e cuidado em pequenas ações do dia-dia me davam forças para superar todos desafios.

Aos meus amigos e colegas de turma Lima, Pedro e Rodrigo. Muito mais que colegas nos tornamos uma unidade que voltava suas forças ao inimigo mais próximo, que foram muitos. Armados com grafite e tinta, treinados por horas em salas de estudo compartilhando conhecimentos, e com orgulho, afirmo que todas batalhas foram vencidas. Tenho certeza que amizades formadas nas adversidades são duradouras e sempre lembraremos com carinho dos anos de mestrado.

À instituição IEAPM e a todos os professores que contribuíram para minha formação, agradeço por todos os ensinamentos. Estendo os agradecimentos a Sócio Ambiental Consultores Associados Ltda, assim como a toda tripulação e pesquisadores a bordo do Sea Route.

Agradeço à PETROBRAS por ter cedido os dados do Projeto de Monitoramento de Cetáceos na Bacia de Santos (PMC-BS). Este projeto foi executado pela PETROBRAS para cumprimento de condicionante ambiental exigida no Licenciamento Ambiental conduzido pelo IBAMA das atividades de produção e escoamento de petróleo e gás natural no Pólo Pré-Sal da Bacia de Santos (processo no. 02001.114279/2017-80, ACCTMB no. 657/2015). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001

“O segredo para ir longe é começar”

Mark Twain

Resumo da Dissertação apresentada ao IEAPM como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

DETECÇÃO E CLASSIFICAÇÃO AUTOMÁTICA DE SINAIS ACÚSTICOS DE
BAIXA FREQUÊNCIA DE *BALAENOPTERA BOREALIS* (BALEIA SEI)
UTILIZANDO TÉCNICAS DE VISÃO COMPUTACIONAL

Vinícius Reveilleau

Outubro/2023

Orientadores: Rogério de Moraes Calazan
Fábio Contrera Xavier

Programa: Acústica Submarina

Estudos no campo dos mamíferos marinhos são tradicionalmente conduzidos por observadores a bordo de aeronaves ou embarcações, sendo altamente onerosas e limitadas por condições climáticas e oceanográficas. Nesse contexto, o monitoramento acústico passivo (MAP) se apresenta como uma alternativa eficaz para o estudo dos cetáceos mesmo em condições climáticas adversas. Com o crescente volume de dados acústicos armazenados nos últimos anos, tem surgido a necessidade de algoritmos capazes de detectar e classificar automaticamente sinais acústicos. Entre as baleias com barbatanas, a *Balaenoptera borealis* (baleia sei) é uma das espécies menos estudadas na comunidade científica. Assim, a escassez de dados acústicos de referência e a grande variabilidade das vocalizações produzidas por mamíferos marinhos representam desafios no campo da bioacústica. Desta forma, técnicas de aumento de dados, como espelhamento e distorção, para melhorar a qualidade e a quantidade dos dados acústicos foram empregadas. Neste estudo, uma metodologia foi proposta para detectar sinais acústicos de baixa frequência e classificar de forma binária chamados de baleia sei. Para detectar os sinais acústicos foram aplicadas ferramentas do campo da visão computacional. Por fim, algoritmos de aprendizado de máquina, como *Random Forest* (RF), *Multilayer Perceptron* (MLP), *Support Vector Machine* (SVM) e *Gaussian Naive Bayes* (GNB) foram empregados para classificação binária. Os resultados obtidos demonstram que o método alcançou uma probabilidade de detecção de 95.5%, enquanto a classificação obteve precisão e *f1-score* dos modelos RF, MLP e SVM superiores a 94%

Abstract of Dissertation presented to IEAPM as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

DISSERTATION TITLE

Vinícius Reveilleau

October/2023

Advisors: Rogério de Moraes Calazan

Fábio Contrera Xavier

Department: Underwater Acoustics

Studies on the field of marine mammals have traditionally been conducted by observers aboard of aircraft or vessels, which is highly costly and limited by weather and oceanographic conditions. Passive acoustic monitoring (PAM) is an effective alternative for studying cetaceans, even under rough weather conditions. With the increasing amount of recorded acoustic data in recent years, there is a growing need for algorithms capable of automatically detecting and classifying acoustic signals. However, the lack of large ground truth acoustic data from marine mammals calls is a challenge in the bioacoustic field. Among the baleen whales, the *Balaenoptera borealis* (sei whale) is one of the least studied in the scientific community. In this study, a low frequency signal detection system and a binary classifier for sei whales are proposed. Furthermore, data augmentation techniques such as flipping and distortion were employed to enhance the quality and the size of the acoustic data. Image processing techniques were used to detected acoustic signals, along with machine learning (ML) algorithms such as Random Forest (RF), Multilayer Perceptron (MLP), Support Vecctor Machine (SVM), and Gaussian Naive Bayes for binary classification of sei whale calls. Comparisons among the ML algorithms were carried out to determine the best binary classification method. The results show a probability of detection of 98.7% for the sei whale, and more than 94% of precision and *f1-score* for RF, MLP, and SVM.

Sumário

Lista de Figuras	x
Lista de Tabelas	xii
Lista de Símbolos	xiii
Lista de Abreviaturas	xv
1 Introdução	1
1.1 Os Cetáceos e acústica submarina	1
1.2 <i>Balaenoptera borealis</i> : a baleia esquecida	2
1.3 Monitoramento acústico Passivo	3
1.4 Desafios na bioacústica dos cetáceos	3
1.5 Objetivos	4
1.5.1 Objetivo Geral	5
1.5.2 Objetivo Específico	5
2 Revisão Bibliográfica	6
2.1 Sinais e Ruídos Acústicos	6
2.2 Repertório Acústico da baleia sei	7
2.3 Detecção e classificação sinais acústicos	9
2.4 Visão computacional para detecção de sinais acústicos	10
2.4.1 Análise Espectral	10
2.4.2 Limiarização	11
2.4.3 Detecção de contornos	12
2.5 Aumento de dados	13
2.6 Aprendizado de máquina para classificação	14
2.6.1 Rede Neural Artificial	14
2.6.2 <i>Random Forest</i>	16
2.6.3 <i>Support Vector Machine</i>	17
2.6.4 Naïve Bayes	19

3	Metodologia	20
3.1	Dados Acústicos	21
3.1.1	Aquisição de dados DTAG	22
3.1.2	Áudios de livre acesso	23
3.2	Sistema de detecção automática de sinais de baixa frequência	24
3.2.1	Representação gráfica dos dados acústicos	25
3.2.2	Pré-processamento	25
3.2.3	Segmentação de imagem	27
3.2.4	Extração de atributos	28
3.3	Aumento do conjunto de dados	30
3.4	Classificação Binária Supervisionada de Sinais de Interesse	32
3.4.1	Parâmetros dos algoritmos de classificação	33
3.4.2	Avaliação dos modelos preditivos	34
4	Resultados e Discussões	36
4.1	Resultados do sistema de detecção de sinais acústicos de baixa frequência	36
4.2	Efeitos do aumento de dados	38
4.3	Resultado comparativo entre os métodos de classificação binária	44
5	Conclusões	46
	Referências Bibliográficas	48

Lista de Figuras

2.1	Cinco classes de sinais acústicos FM de baixa frequência emitidos pela baleia sei e caracterizados no trabalho de CERCHIO e WEIR [1]. Fonte: adaptado de [1]	8
2.2	Arquitetura característica de uma MLP, composta por neurônios de entrada, uma ou mais camadas ocultas e neurônios de saída, também chamados de perceptrons. Fonte: http://medimedium.com/sina-ai-rede-neural-perceptron-multicamadas . Acessado em 05/09/2023	15
2.3	Arquitetura do algoritmo <i>Random Forest</i> composto por diversas árvores de decisão (<i>Decision Tree</i>). Fonte: adaptado de http://webdesignemfoco.com . Acessado em 05/09/2023	16
2.4	Representação do modelo SVM aplicado a classificação binária. Fonte: Retirado de [2]	17
2.5	Aplicação de um kernel dentro do algoritmo SVM causando um aumento do espaço dimensional do conjunto de dados. Fonte: adaptado de https://medium.com/@msremigio-maquinas-de-vetores-de-suporte-svm-77bb114d02fc . Acessado em 10/08/2023	18
3.1	Cadeia de processamento do algoritmo proposto, composto por três estágios: detecção de sinais acústicos de baixa frequência, aumento de dados e classificação binária.	21
3.2	a) Equipamento DTAG (Fonte: [3]) b) Método de fixação com copos de sucção e vara de carbono utilizados durante a campanha de telemetria do PMC-BS (Fonte: [3]) c) Equipamento DTAG fixado em uma baleia. (Fonte: [4])	23

3.3	Exemplos dos sinais acústicos das três espécies em estudo, apresentadas em imagens em escala de cinza com contornos extraídos pela metodologia apresentada na seção 3.2 do sistema de detecção de sinais acústicos de baixa frequência. a) Sinais DS de baleia sei dos áudios do DTAG b) sinal L-call de baleia sei de áudios do material suplementar de [1] c) Sinais ascendentes e descendentes-ascendentes de baleia fraca austral de áudios do repositório Mobysound d)Vocalização de baleia fin de <i>back beat pulse</i> e pulso-20Hz de áudios do Mobysound.	24
3.4	Resultado do pré-processamento e adaptação da imagem em escala de cinza para áudio do DTAG. A) Espectrograma com três sinais FM de baixa frequência emitidos pela baleia sei. O espectrograma foi gerado com taxa de amostragem de 2048 Hz, FFT com 1024 pontos, 80% de sobreposição e janela Hann. B) Imagem em escala de cinza após redução do ruído de fundo e suavização pelo método de média não local.	26
3.5	Segmentação de imagem e máscaras dos sinais de interesse. A)Sinal de frequência modulada decrescente com contorno detectado. B,C,D) Máscaras individuais de cada sinal de interesse utilizada para extração de características espectrais e temporais.	28
3.6	Exemplo das máscaras dos sinais acústicos de interesse (SAI) de vocalização DS, aumentadas a partir de transformações dos dados experimentais. A) Imagem original com contorno encontrado. B) Máscara binária original. C,D,E,F) Máscaras depois das técnicas de aumento de dados de distorção e/ou espelhamento.	31
4.1	Processo de aprendizado dos classificadores representado pela curva de acurácia no subconjunto de treinamento e de validação do conjunto de dados original e aumentado. As linhas representam os resultados da acurácia média em relação ao número de amostras. Enquanto a área sombreada representa o desvio padrão. (a) Conjunto original de dados b) Conjunto de dados aumentados.	39
4.2	Curva Precisão- <i>Recall</i> (Curva-PR) dos modelos de ML calculada a partir das previsões de classe do subconjunto teste (independente) (a) Curva-PR dos modelos treinados com o conjunto original de dados (b) Curva-PR dos modelos treinados com o conjunto de dados aumentado. 42	
4.3	Matriz de confusão dos classificadores de aprendizado de máquina sobre subconjunto independente de teste.(a) <i>Random Forest</i> b) <i>Multi-layer Perceptron</i> c) <i>Support Vector Machine</i> d) <i>Gaussian Naive Bayes</i> . 44	

Lista de Tabelas

3.1	Descrição da base dados acústicos utilizados para detecção automática de sinais acústicos e classificação binária.	22
3.2	Descrição das características espectrais e temporais extraídas de cada sinal de interesse que originam a matriz de atributos. Esta matriz serve como dados de entrada para a classificação por métodos de aprendizado do máquina.	29
3.3	Quantidade de sinais acústicos de interesse (SAI) detectados pelo sistema de detecção de sinais acústicos de baixa frequência, que formam o conjunto original de dados (segunda coluna) e o conjunto aumentado de dados (terceira coluna) após operações de espelhamento e distorção.	32
4.1	Número de sinais de interesse detectados manualmente e automaticamente, classificados por um analista como sinal acústico biológico das espécies e ruído.	37
4.2	Resultados dos quatro modelos de classificação no subconjunto de validação cruzada quando treinados com o conjunto de dados original e aumentado. São apresentadas as médias e os desvios padrões das métricas propostas.	41
4.3	Resultados da precisão média (AP) da curva-PR (PR-AP), que é uma aproximação da área sob a curva-PR (PR-AUC) para os modelos RF, MLP, SVM e GNB treinados com o conjunto de dados original (coluna da esquerda) e aumentado (coluna da direita). O valor mais próximo de um indica melhor desempenho.	43
4.4	Resultado das métricas estatísticas sobre subconjunto independente de teste.	45

Lista de Símbolos

$I_{(i,j)}$	Matriz da imagem inicial, p. 26
I_{Bin}	Matriz da imagem Binária, p. 27
$K(x, x_i)$	Função Kernel RBF, p. 18
$M_{(x,y)}$	Magnitude do vetor gradiente, p. 12
$P(C)$	Probabilidade a priori da classe C, p. 19
$P(\omega)$	Sinal no domínio da frequência, p. 10
P_i	Precisão no i -ésimo ponto, p. 35
R_i	<i>Recall</i> no i -ésimo ponto, p. 35
$Sm_{(i,j)}$	Matriz filtrada do espectrograma, p. 25
Z	Constante de normalização, p. 19
Δf	Passo de frequência, p. 25
Δt	Passo de tempo, p. 25
α_i	Coefficiente de treinamento, p. 18
$\alpha_{(x,y)}$	Direção do vetor gradiente, p. 12
γ	Hiperparâmetro do Kernel RBF, p. 18
μ	Média, p. 25
ω	Frequência angular, p. 10
ϕ	Fase, p. 6
ψ	Função de ativação, p. 15
σ	Desvio padrão, p. 25

a	Amplitude, p. 6
b	Termo de viés, p. 18
c	Constante, p. 27
$\max(Sm)$	Valor máximo da matriz filtrada, p. 26
$\min(Sm)$	Valor mínimo da matriz filtrada, p. 26
n_{sv}	Número de vetores de suporte, p. 18
$p(t)$	Pressão sonora no domínio do tempo, p. 6
w	Pesos da FNN, p. 14
$w(t - \tau)$	Função janela, p. 11
x_i	Dados de entrada, p. 14
y'_i	Resultado real , p. 15
y_i	Resultados preditos, p. 15
∇	Operador gradiente, p. 12
f	Frequência, p. 6

Lista de Abreviaturas

AM	Amplitude Modulada, p. 7
AP	Precisão média, p. 35
AUC	Área sob a curva (<i>Area Under de Curve</i>), p. 35
Adam	Estimativa do momento adaptativo (<i>Adaptive Moment Estimation</i>), p. 33
Curva-PR	Curva de Precisão- <i>recall</i> , p. 34
DFT	Transformada discreta de Fourier (<i>Discrete Fourier Transform</i>), p. 10
DTAG	Tag de Registro Acústico Digital, p. 21
DT	<i>Decision Tree</i> , p. 16
FFT	Transformada Rápida de Fourier (<i>Fast Fourier Transform</i>), p. 10
FM	Frequência Modulada, p. 7
FNN	Redes neurais artificiais de alimentação direta (<i>Feedforward Neural Networking</i>), p. 14
FN	Falso negativo, p. 34
FP	Falso positivo, p. 34
GNB	<i>Gaussian Naive Bayes</i> , p. 4
IUCN	União Internacional para a Conservação da Natureza e dos Recursos Naturais, p. 2
MAP	Monitoramento Acústico Passivo, p. 3
MLP	<i>Multilayer Perceptron</i> , p. 4

ML	Aprendizado de Máquina (<i>Machine Learning</i>), p. 6
PMC-BS	Programa de Monitoramento de Cetáceos - Bacia de Santos, p. 22
PSD	Densidade espectral de potência (<i>Power spectral density</i>), p. 25
RBF	Função de base radial(<i>Radial Basis Function</i>), p. 18
RF	<i>Random Forest</i> , p. 4
ReLu	Unidade Linear Retificada (<i>Rectified Linear Unit</i>), p. 15
SAI	Sinal acústico de interesse, p. 27
STFT	Transformada de Fourier de Tempo Curto (<i>Short Time Fourier Transform</i>), p. 11
SVM	<i>Support Vector Machine</i> , p. 4
TN	Verdadeiro negativo, p. 34
TP	Verdadeiro positivo, p. 34
WMMDB	Watkins Marine Mammal Sound Database, p. 23

Capítulo 1

Introdução

1.1 Os Cetáceos e acústica submarina

Os mamíferos marinhos surgiram através da adaptação dos mamíferos terrestres à vida aquática por volta de 50 milhões de anos atrás [5, 6] na era geológica Cenozoica no Eoceno. Eles são divididos em três ordens taxonômicas: Cetáceos (como baleias e golfinhos), Sirênios (como o peixe-boi e a vaca-marinha) e Carnívora (como leões-marinhos, lobos-marinhos e focas). Os cetáceos vivem todo o seu ciclo de vida em ambiente aquático e estão distribuídos em todos os oceanos do mundo [5].

Dentro da ordem dos cetáceos, há duas subordens, os mysticetos e os odontocetos. A principal diferença fisiológica entre eles é a presença de dentes nos odontocetos (como golfinhos, cachalotes e orcas), enquanto os mysticetos possuem estruturas filtradoras denominadas cerdas ou barbatanas (como jubarte, baleia-azul e baleia sei) [7]. Estima-se que existam cerca de 96 espécies de cetáceos no mundo [5].

Dentre as muitas adaptações dos mamíferos marinhos para a vida em ambiente aquático, desenvolveram-se habilidades acústicas. Acredita-se que todos os sinais acústicos emitidos por mamíferos marinhos estejam contidos na faixa de frequência de 7 Hz a 180 kHz [8, 9]. Os cetáceos utilizam sua capacidade de emissão de ondas mecânicas para fins de comunicação social e ecolocalização. No entanto, apenas os odontocetos possuem a capacidade de ecolocalização, utilizada para navegação e alimentação [5].

Os sinais usados para ecolocalização se caracterizam por serem impulsivos de curta duração com amplo espectro de frequência, denominados "cliques" [10, 11]. Utilizando o mesmo princípio do Sonar, esses animais são capazes de emitir ondas mecânicas que refletem no alvo, sendo captados por órgãos sensoriais que permitem inferir distância, direção, formas e até texturas [7].

Para comunicação os cetáceos utilizam sinais de frequência modulada (FM) com duração temporal na escala de segundos [11]. Assim, desenvolveram-se sinais este-

reotipados, com padrões temporais e espectrais específicos de cada espécie, que são de especial interesse em tarefas de classificação [10]. Esses sinais acústicos foram denominados de várias formas, como assobios, cantos, grunhidos, gemido [8, 10]. Os sinais acústicos emitidos pelos mysticetos se caracterizam por emitirem sinais em frequência mais baixa em comparação aos outros cetáceos, entre 7 Hz e 5 kHz[5].

Estudos mostram a capacidade de aprendizado de comunicação e transmissão cultural desses animais, acredita-se que animais da mesma espécie em regiões geográficas distintas do mundo podem apresentar variações nas vocalizações [7, 12]. Essa grande variabilidade no repertório acústico das espécies, tanto interespecífico como intraespecífico, é um desafio para classificação das vocalizações no campo da bioacústica.

1.2 *Balaenoptera borealis*: a baleia esquecida

Dentro da subordem dos mysticetos, o gênero *Balaenoptera* é composto por 7 espécies e 15 subespécie [13]. A baleia sei (*Balaenoptera borealis*) é a terceira maior da família Balaenopteridae [14]. Esta possui uma distribuição cosmopolita, habitando todas as bacias oceânicas, desde as regiões tropicais até as regiões polares [14, 15]. No hemisfério sul, ela é encontrada predominantemente em águas profundas, mas também pode ser avistada na região do talude continental [14].

As baleias sei realizam migrações sazonais, durante o verão migram para áreas de alimentação em altas latitudes nas zonas temperadas e subpolares, durante o inverno se deslocam para áreas de reprodução em baixas latitude nas zonas tropicais e subtropicais [13]. As áreas de reprodução das baleias sei são pouco conhecidas e seus locais de alimentação apresentam grande variabilidade interanual [16, 17]. Devido a pressão nos estoques exercida pela pesca comercial das baleias entre os séculos IXX e XX, desde 1976 a baleia sei encontra-se na lista vermelha da União Internacional para a Conservação da Natureza e dos Recursos Naturais (IUCN) das espécies ameaçadas.

No oceano atlântico sul, foram registrados avistamentos desde o noroeste do Suriname até Argentina, com maior ocorrência de avistamentos entre Rio Grande do sul, Tierra del Fuego (Argentina) e Falkland Islands [1, 13]. No Brasil foram sugeridas duas potenciais áreas de berçários das baleias sei; a Ilha de Trindade e o Arquipélago de Martin Vaz [18]. Atualmente é a espécie menos estudada dentre os Balenopteridea, com grande lacuna de dados desde o fim da pesca comercial ganhando status de "a baleia esquecida", devido ao menor número de pesquisas realizadas sobre a espécie [14]. Em relação a sua capacidade de comunicação acústica, considera-se que o conhecimento sobre seu repertório continua limitado[19].

1.3 Monitoramento acústico Passivo

O Monitoramento Acústico Passivo (MAP) se baseia no emprego de sensores capazes de monitorar a variação da pressão acústica ao longo do tempo de forma passiva. Para isso são usados hidrofones, que são transdutores acústicos compostos de material piezoelétrico, capazes de converter proporcionalmente deformação mecânica em sinal elétrico, causada pela variação do campo de pressão acústico [6, 10].

Existem muitas aplicações e instrumentações para o MAP, à metodologia aplicada é dependente de diversos fatores que envolvem o objeto de estudo. Para isso, diversos métodos foram desenvolvidos para gravar dados acústicos como: hidrofones móveis lançados de plataformas fixas, arranjos de hidrofones rebocados por embarcações e hidrofones acoplados em dispositivos móveis (e.g. glider, auv), entre outros [6]. Assim diversas informações sobre a paisagem acústica podem ser aplicadas para conservação, mitigação de impactos e ecologia das espécies [20].

A aplicação do MAP nos estudos dos mamíferos marinhos é amplamente discutida e desenvolvida no meio científico. Tradicionalmente estudos de mamíferos marinhos são realizados de forma visual por observadores de bordo. Porém este método é limitado pelas condições ambientais, como estado do mar, neblina, luminosidade, acessibilidade a locais remotos e condições comportamentais da espécie [8]. O MAP se apresenta como uma ferramenta capaz de eliminar algumas limitações do método de observação visual, principalmente relacionadas as condições climáticas. Este método é capaz de gravar amostras da paisagem acústica por períodos prolongados. Logo, um grande volume de dados precisa ser processado e analisado ao fim de cada campanha de monitoramento acústico [21]. Uma tarefa a ser realizada após campanhas de monitoramento acústico de cetáceos é a detecção e classificação de vocalizações. Identificar a fonte do sinal acústico seja ela biológica ou não, é essencial para que possam ser acessadas informações como distribuição das espécies, densidade populacional, assim como multianálises de padrões comportamentais [6, 7].

1.4 Desafios na bioacústica dos cetáceos

O desenvolvimento de novas tecnologias aumentou nossa capacidade de gravar e armazenar grandes volumes de dados acústicos [22], levando à necessidade de métodos automáticos de detecção e classificação confiáveis para discriminar sinais bioacústicos [21]. No entanto, a detecção automática e classificação de sons de mamíferos marinhos enfrentam vários desafios, incluindo variabilidade de sinais biológicos, incertezas na identificação da origem de cada sinal bioacústico, conjunto pequeno de dados com identificação da espécie emissora, ruídos ambientais e antropogênicos

[22, 23].

Além disso, discriminar vocalizações que se sobrepõem em frequência tem sido uma tarefa desafiadora para pesquisadores da bioacústica [24], como das baleias sei, baleias fin (*Balaenoptera physalus*) e baleias franca austral (*Eubalaena australis*).

O sinal de baleia sei mais descrito em todo o mundo é um sinal FM descendente de baixa frequência. Esses chamados (*calls*) tem uma faixa de frequência entre 15 Hz e 230 Hz e uma duração de 0.3 s a 1.9 s [e.g., 1, 25]. A baleia franca austral produz muitas variações de vocalizações FM de baixa frequência, incluindo sinais descendentes (*downsweep*), ascendentes (*upsweep*) e variações tonais (*tonal variable*), com frequência máxima entre 79 Hz e 648 Hz e frequência mínima variando de 22 Hz e 558 Hz [26]. O sinal acústico mais comumente relatado para baleias fin em todo o mundo é um sinal FM descendente com duração inferior a 1 segundo e faixa de frequência entre 15 Hz a 40 Hz, chamado de "pulso de 20 Hz" (*20Hz pulse*). A segunda classe de chamado foi nomeada de "pulso de batida de retorno" (*back beat pulse*) caracterizado por uma banda estreita de 5Hz e faixa de frequência entre 15 Hz e 20 Hz [27].

Diversos classificadores foram propostos porém muitos falham quanto a variabilidade de sinais avaliados, possuem poucos sinais acústicos com confirmação visual da espécie emissora, e até dificuldades quanto a implementação da detecção e classificação em uma mesma cadeia de processamento [7, 28]. Sendo assim, estudos sobre algoritmos capazes de generalizar detecções e predições de diversas fontes de dados e espécies ainda é um campo aberto de pesquisa [6, 7, 10].

1.5 Objetivos

Objetivo principal desta abordagem é desenvolver uma metodologia capaz de detectar sinais acústicos de baixa frequência e distinguir vocalizações da baleia sei do ruído de fundo, bem como dos sinais acústicos de baleia fin e baleia franca austral. Para tarefa de classificação, serão empregados algoritmos de aprendizado de máquina, como *Random Forest* (RF), *Multilayer Perceptron* (MLP), *Support Vector Machine* (SVM) e *Gaussian Naive Bayes* (GNB) para determinar o melhor modelo de classificação binária. De acordo com a revisão bibliográfica realizada durante a elaboração desta dissertação, não foi encontrado registro na literatura de um detector e classificador binário que abrange todos os tipos de sinais acústicos de baixa frequência de baleias sei. Além disso, para lidar com as limitações causadas por um conjunto de dados de referência reduzido, introduzimos uma abordagem de aumento de dados envolvendo a inversão e distorção das máscaras de contorno dos sinais acústicos.

1.5.1 Objetivo Geral

Desenvolver uma técnica de detecção automática de sinais acústico de baixa frequência e classificação binária de vocalizações de *Balaenoptera borealis*.

1.5.2 Objetivo Específico

- Implementar um método de detecção automática de sinais acústicos baseado em técnicas de visão computacional em espectrogramas.
- Avaliar o método de aprendizado de máquina mais adequado para classificação binária de sinais acústicos de baixa frequência de *Balaenoptera borealis*.
- Analisar o efeito do aumento de dados a partir transformações do conjunto dados reais de referência no desempenho dos modelos de classificação.

Capítulo 2

Revisão Bibliográfica

Neste capítulo, vamos apresentar as informações fundamentais para proporcionar um melhor entendimento dos métodos desenvolvidos ao longo deste trabalho. Na Secção 2.1 são introduzidas definições de propagação de onda sonora, sinais e ruídos acústicos. Na secção 2.2 aprofundamos o conhecimento sobre o repertório acústico da espécie alvo. A Secção 2.3 apresenta os métodos de classificação e detecção de vocalizações de cetáceos, baseado em estudos pretéritos que fundamentam o trabalho desenvolvido nesta dissertação. A Secção 2.4 aborda as técnicas de visão computacional e processamento de sinais para detecção de sinais acústicos. A Secção 2.5 apresenta algumas operações comuns para o aumento sintético a partir de um conjunto de dados reais. Por fim, a Secção 2.6 introduz os algoritmos de aprendizado de máquina (ML, *Machine Learning*) aplicados neste trabalho.

2.1 Sinais e Ruídos Acústicos

O som pode ser definida como uma onda de pressão que se propaga a partir de uma fonte emissora em um meio sólido, líquido ou gasoso [6, 10]. Ela é gerada a partir da vibração de um material produtor de som (por exemplo, a laringe dos mamíferos), capaz de promover compressão e rarefação do meio em que se encontra, criando alternadamente zonas de alta e baixa pressão [10].

A pressão sonora $p(t)$ de uma onda plana senoidal, pode ser expressa como [10]:

$$p(t) = a(t) \cos(2\pi f(t)t + \phi(t)) \quad (2.1)$$

Um sinal acústico pode ser definido como uma onda sonora que carrega informação [10]. No contexto do monitoramento acústico, um sinal é caracterizado como a mensagem de interesse que desejamos receber. Essas informações podem estar codificadas na amplitude $a(t)$, na frequência $f(t)$ ou na fase ϕ . Sinais nos quais as informações são contidas nas variações de amplitude $a(t)$, são conhecidos como

sinais de amplitude modulada (AM). Por outro lado, as informações podem estar contidas na frequência, sendo chamados de sinais de Frequência Modulada (FM) [10]. Quando um sinal FM aumenta a frequência ao longo do tempo, é chamado de ascendente (*upsweep*), e quando decresce em função do tempo é chamado de descendente (*downsweep*) [4].

Os ruídos acústicos podem ser descritos como uma grande soma de ondas sonoras de diversas fontes emissoras com flutuações aleatórias de fase, amplitude e frequência [10]. No contexto do monitoramento acústico, ruídos são todos sinais detectados que não são de interesse para o objeto de estudo. Então, para as tarefas de detecção e classificação, a premissa básica é discriminar os ruídos dos sinais de interesse.

2.2 Repertório Acústico da baleia sei

A baleia sei utiliza som para comunicação, porém a variabilidade dos sinais acústicos emitidos por esta espécie ainda é pouco conhecida [14, 29]. Nos primeiros registros acústicos na presença de baleias sei, em New Scotia (Canada), foram descritos sinais FM ascendentes entre 1.5 e 3.5 kHz, porém há possibilidade de sinais de baixa frequência terem sido mascarados [30, 31]. Na península Antártica foram identificados sinais FM descendentes e ascendentes, na faixa de frequência entre 100 Hz e 600 Hz e duração de 1 a 3s [32, 33].

Porém o sinal mais documentado em diferentes regiões geográficas do mundo são os sinais FM descendentes [e.g., 1, 19, 29, 32, 34]. No Hawaii (EUA) foram identificados dois tipos de vocalização FM descendentes: Primeiro, entre 21 Hz e 39.1 Hz, segundo, entre 44.6 Hz e 100.3 Hz com aproximadamente 1.2 segundos de duração[19]. Resultados semelhantes foram apresentados em um estudo na região de New England (EUA), em que foram caracterizados sinais de baleia sei com frequência mínima de 34 Hz e máxima de 82.3 Hz e duração temporal média de 1.38 s, onde as vocalizações poderiam ocorrer de forma individual, em duplas ou trincas[29]. Em outro trabalho realizado em New England (EUA) também foi atribuído sinais acústicos FM descendentes à baleia sei, classificando como A; os de faixa frequências entre 34 Hz e 82 Hz, enquanto classes B, C e D entre 30Hz e 50Hz [34].

No hemisfério sul nas ilhas Auckland, foram atribuídos sinais acústicos FM ascendentes, descendentes e ascendentes-descendentes à baleia sei, com faixa de frequência entre 34 Hz e 87 Hz e duração temporal média de 1.1 s [35]. Em outro trabalho realizado na Patagônia chilena, foram relatados sinais FM descendente que ocorreram de forma individual, em duplas ou em trincas, com faixa de frequência entre 35.6Hz e 105.3 Hz, e duração temporal média de 1.6 s [25]. Estas variações das características dos sinais acústicos atribuídos a baleia sei podem estar relacionadas a separação geográfica das populações, sugerindo que diferentes populações podem

produzir sinais estereotipados distintos, e/ou as baleias sei produzem distintos sinais acústicos em baixas latitudes nas áreas de reprodução e em altas latitudes nas áreas de alimentação [19].

Além disso, em um estudo recente nas ilhas Malvinas, CERCHIO e WEIR [1] relataram uma variedade de cinco classes de sinais FM de baixa frequência chamadas de 'DS', 'US', 'L', 'Arch' e 'LF-Variable call' emitidos pela baleia sei e apresentados na Figura 2.1, nomeados de acordo com suas características espectrais, temporais e de forma no espectrograma.

A vocalização da classe 'DS' foi definida como um sinal acústico FM descendente com faixa de frequência entre 30 Hz e 230 Hz. A segunda classe foi caracterizada como um sinal FM ascendente com duração temporal de aproximadamente 2 s e faixa de frequência entre 25.4 Hz e 70.6 Hz, denominada de 'U'. A terceira classe de sinal foi chamada de 'L', sendo nomeada pela sua forma no espectrograma devido a ter uma modulação de frequência inicialmente descendente entre 80Hz e 50Hz, seguido por um tom em 60 Hz e duração temporal de aproximadamente 1.5 s. A quarta classe de sinal foi denominada de '*LF-Variable-call*' com aproximadamente 2 s de duração temporal, na faixa de frequência entre 20 Hz e 80 Hz. A quinta classe foi nomeado de 'Arch' devido a sua característica côncava, com 0.91 s a 2.19 s de duração e faixa de frequência entre 19 Hz e 119 Hz. Todas as cinco classes de sinais acústicos de baleia sei descritas no trabalho de CERCHIO e WEIR [1] foram incluídas no conjunto de dados analisado neste estudo.

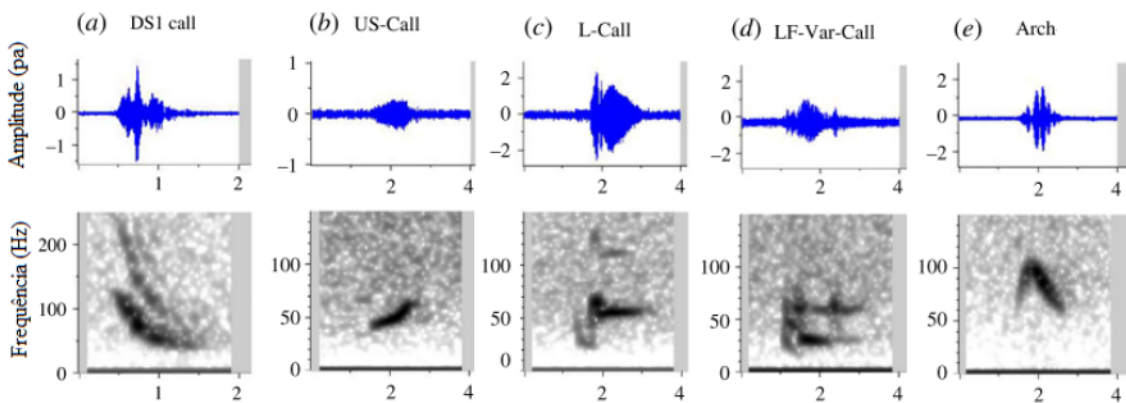


Figura 2.1: Cinco classes de sinais acústicos FM de baixa frequência emitidos pela baleia sei e caracterizados no trabalho de CERCHIO e WEIR [1]. Fonte: adaptado de [1]

2.3 Detecção e classificação sinais acústicos

Diversos detectores e classificadores automáticos para vocalizações de baixa frequência de baleias foram propostos nos últimos anos. A maioria deles opera no domínio do tempo x frequência, procurando por sinais biológicos em espectrogramas, enquanto apenas alguns detectam chamados no domínio do tempo [8]. BAUMGARTNER e MUSSOLINE [22] separaram os algoritmos no domínio de frequência em duas categorias. A primeira categoria envolve a convolução de um sinal desejável bem definido com o espectrograma para gerar uma função de detecção. Se o valor da função exceder um determinado limite, é considerado um sinal detectado. A segunda categoria seleciona todos os sinais dentro da faixa de frequência de interesse, extrai várias características e os classifica com base em suas similaridades.

Pertencendo a essa segunda categoria, alguns autores relataram o uso de técnicas de processamento de imagem para detectar vocalizações de cetáceos. No trabalho de GILLESPIE [36] técnicas de *edge detection* foram implementadas em espectrogramas para detectar sinais acústicos emitidos por baleias franca. A metodologia aplicada para detecção resultou em uma probabilidade de detecção de 90%. SÁNCHEZ-GARCÍA *et al.* [37] extraíram assovios de golfinhos por segmentação de regiões, baseado em operações morfológicas de imagens de espectrogramas, detectando 89% dos sinais acústicos. Em outro trabalho aplicado sobre sinais de baleia franca, MOHAMMAD e MCHUGH [38] apresentaram um método de extração de contornos que não é necessário o uso de um limiar global. O autor utilizou apenas três características temporais e espectrais (frequência mínima, frequência máxima e duração do sinal) como parâmetros de entrada para classificação realizada pelo modelo *Support Vector Machine*, resultando em 93% de probabilidade de detecção e classificação.

O método mais difundido para detectar e classificar mysticetos testados em sinais de baleia sei foi proposto por BAUMGARTNER e MUSSOLINE [22]. Eles usaram uma extração de características espectrais e temporais por meio do rastreamento da frequência do sinal de contorno (*Pitch tracking*), combinado com análise de função de discriminação quadrática.

Trabalhos mais recentes que consideraram as vocalizações da baleia sei por meio da distribuição *Pseudo Wigner-Ville* e classificação pelo modelo SVM alcançaram acurácia de 92% [24]. Em outro trabalho, um sistema baseado em Modelo de Markov Oculto associado à uma Rede Neural Profunda resultou em *f-1 score* de 80.95% [39].

No entanto, todos os algoritmos desenvolvidos para detectar vocalizações de baleia sei trabalharam apenas com sinais FM descendente e ascendentes, e com conjuntos de dados de referência reduzido. Para os algoritmos de classificação su-

pervisionada, os conjuntos de dados reduzidos e desbalanceados podem representar uma dificuldade para o processo de aprendizado e para a capacidade de generalização dos modelos [40].

2.4 Visão computacional para detecção de sinais acústicos

A visão computacional, que combina técnicas de processamento de imagens e análise de dados, tem sido aplicada com sucesso no contexto detecção de vocalização das espécies, permitindo uma análise mais precisa e automatizada dos sinais acústicos [7]. Para aplicar técnicas de visão computacional em dados acústicos, devemos representar dos dados acústicos em espectrogramas utilizando métodos de processamento de sinais. A detecção de sinais acústicos em espectrogramas envolve um conhecimento multidisciplinar e algumas técnicas serão apresentadas nas secções abaixo.

2.4.1 Análise Espectral

A Transformada de Fourier é uma ferramenta essencial na análise de sinais, permitindo-nos decompor um sinal complexo no domínio do tempo em suas componentes de frequência [10]. Esta técnica, nos proporciona uma visão detalhada do comportamento de um sinal no domínio da frequência. A expressão fundamental da Transformada de Fourier é dada por [10]:

$$P(\omega) = \int_{-\infty}^{\infty} p(t)e^{-j\omega t} dt \quad (2.2)$$

Onde $P(\omega)$ representa o sinal no domínio da frequência, $p(t)$ corresponde ao sinal no domínio do tempo e ω é a frequência angular ($\omega = 2\pi f$).

Porém esta solução para a transformada só é possível para funções contínuas no domínio do tempo [4]. Embora a propagação de uma onda sonora seja um fenômeno contínuo, quando registrado de forma digital, a pressão instantânea é gravada de forma discreta, o que resulta em um série temporal de dados discretos e finitos [4].

Quando temos uma série temporal de dados discretos podemos aplicar a Transformada Rápida de Fourier (FFT, *Fast Fourier Transform*) [41]. A FFT é um algoritmo eficiente que calcula a Transformada discreta de Fourier (DFT, *Discrete Fourier Transform*). A DFT é uma função matemática utilizada para analisar o conteúdo espectral de um sinal discreto, dado por:[10]

$$Pm = \frac{1}{f_s} \sum_{n=0}^{N-1} p_n e^{-j\omega_m n} \quad (2.3)$$

A DFT é aplicada em N amostras do sinal discreto $p[n]$ para $n = 0$ e $n = N - 1$. Onde a frequência discreta w_m é definida por $w_m = 2\pi \frac{m}{M}$ para $m = 0$ e $m = M - 1$ e fs representa a taxa de amostragem.

Uma ferramenta fundamental para o monitoramento acústico dos cetáceos é a representação de dados acústicos em espectrogramas. Esta ferramenta nos permite visualizar os sinais acústicos emitidos pelas espécies inferindo informações sobre frequência, tempo e magnitude e pode ser calculada pela DFT. Para analisar as componentes de frequência em relação ao tempo, utiliza-se a Transformada de Fourier de Tempo Curto (STFT, *Short Time Fourier Transform*). A STFT divide um sinal contínuo no domínio do tempo em segmentos menores por uma janela deslizante que calcula a DFT para cada segmento. Em cada segmento é aplicado uma função janela (eg. janela Hann, janela Hamming, etc.). A STFT é recomendável quando queremos examinar como a frequência de um sinal varia ao longo do tempo [4], sendo expressa por:[10]

$$X(\omega, t) = \int_{-\infty}^{\infty} p(t) \cdot w(t - \tau) \cdot e^{-j\omega t} dt \quad (2.4)$$

Em que, $X(\omega, t)$ é a resultante da STFT, $p(t)$ representa o sinal no domínio do tempo e $w(t - \tau)$ é uma janela definida por uma função que desliza no tempo t .

2.4.2 Limiarização

A segmentação de imagem permite extrair características específicas de uma imagem. Neste contexto, a limiarização desempenha um papel fundamental na segmentação de imagens e na detecção de objetos de interesse tendo em vista sua simplicidade de implementação e velocidade computacional [41]. A limiarização ou binarização, em sua essência, consiste na conversão de uma imagem em escala de cinza em uma imagem binária, onde os píxeis são divididos em dois grupos distintos com base em um valor de limiar (*threshold*) [42].

Fundamentalmente, as técnicas de limiarização podem ser resumidas em dois grupos: o primeiro é nomeado de limiarização global, onde apenas um valor pré definido de limiar (T) é aplicado para todos os valores dos píxeis de uma imagem. O segundo é nomeado de limiarização adaptativa (*adaptive threshold*), que busca múltiplos limiares ao longo da imagem analisando as propriedades dos píxeis vizinhos [41]. Esta técnica é mais recomendável quando temos mais de dois modos no histograma de frequência de uma imagem [41].

Dentre as principais aplicações de limiares adaptativos podemos citar o limiar adaptativo de média ou mediana, que busca valores médios ou medianas dos vizinhos próximos, determinado por uma janela espacial (x,y) em que buscam múltiplos limiares pela intensidade dos píxeis da imagem [41].

A abordagem adaptativa traz diversas vantagens para o processamento de imagem. Ela permite a segmentação precisa de objetos de interesse, mesmo em áreas com ruído não uniforme ou com gradientes de contraste abruptos. Isso é particularmente útil em áreas como reconhecimento de caracteres, onde os caracteres podem variar em tamanho, estilo e iluminação. Além disso, o limiar adaptativo é eficaz na detecção de contornos em imagens médicas e na análise de texturas em imagens industriais [42].

2.4.3 Detecção de contornos

Para o completo entendimento da detecção de contornos é necessário conhecer a definição de borda. Em uma imagem, uma borda pode ser definida como a região onde há uma mudança abrupta de intensidade ou cor da imagem. Assim, os algoritmos de detecção de bordas buscam as regiões com variações de intensidade significativas dentro de uma imagem [42].

Pela perspectiva da computação visual, o contorno pode ser definido com uma curva que interpola todos os pontos de mesma intensidade ou com propriedades do píxel semelhantes, sendo uma ferramenta muito útil para detecção e reconhecimento de objetos em uma imagem [42]. Dentre os diversos métodos de extração de contornos, este trabalho se dedica às técnicas que procuram por variações abruptas nas imagens baseadas no cálculo do gradiente da intensidade de cada píxel da imagem.

O operador gradiente é ideal para encontrar variações da intensidade e direção de uma borda no plano (x, y) de uma imagem [41], denotado por:

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (2.5)$$

A equação 2.5 apresenta o operador gradiente (∇), que utiliza as derivadas parciais de primeira ordem na função (f) em relação às coordenadas (x) e (y) sendo definido como um vetor coluna bidimensional [41]. O vetor do gradiente de f possui três propriedades essenciais para detecção de bordas, são elas: sentido, magnitude e direção. Este vetor aponta no sentido da maior variação de f , no ponto (x, y) .

A magnitude deste vetor representa o valor da taxa de variação da função f , definida por:

$$M_{(x,y)} = \sqrt{G_x^2 + G_y^2} \quad (2.6)$$

Por fim, à última propriedade do vetor gradiente para detecção de bordas é a direção. A direção da borda de uma imagem em um ponto arbitrário é ortogonal a direção do vetor gradiente no mesmo ponto $\alpha_{(x,y)}$, definida por:

$$\alpha_{(x,y)} = \tan^{-1} \left[\frac{G_y}{G_x} \right] \quad (2.7)$$

Porém o operador gradiente não fornece o contorno das bordas, apenas encontra os píxeis que pertencem a borda. Para isto, são aplicados algoritmos de ligação capazes de unir os pontos encontrados pelo operador gradiente criando então o contorno desejado. Uma maneira eficaz foi apresentada no trabalho de SUZUKI e BE [43] que utilizou imagens binárias para extração de contornos. O algoritmo começa localizando o primeiro píxel de um contorno e segue a cadeia de píxeis vizinhos com intensidade semelhante até chegar ao ponto inicial, formando um contorno fechado. Esse processo é repetido para todos os píxeis restantes, até que todos os contornos na imagem sejam detectados.

2.5 Aumento de dados

O aumento de dados (*data augmentation*) consiste em aplicar transformações controladas nos dados de treinamento, gerando novas instâncias de dados que são semanticamente equivalentes, mas apresentam variações que podem ocorrer naturalmente [41]. Isso não apenas enriquece o conjunto de dados, mas também ajuda a mitigar problemas relacionados à falta de diversidade e tamanho limitado dos dados experimentais.

Em muitos estudos de bioacústica, é comum encontrar um desequilíbrio entre as classes, ou seja, certas espécies podem estar sub-representadas nas amostras. Isso pode levar a resultados tendenciosos e modelos que são insuficientemente treinados para identificar espécies menos frequentes [44]. O aumento de dados pode ser empregado para criar instâncias sintéticas das classes menos representadas, equilibrando assim o conjunto de dados e melhorando a capacidade do modelo de reconhecer todas as espécies de maneira igualitária [41].

Na bioacústica e em outras áreas, existem diversas técnicas de aumento de dados que podem ser aplicadas para enriquecer os conjuntos de dados e melhorar a performance dos modelos. Alguns dos tipos mais comuns de aumento de dados incluem:

- a) Rotação e Espelhamento: Essas técnicas são frequentemente usadas em imagens e envolvem a rotação da imagem em diferentes ângulos ou a criação de versões espelhadas horizontalmente ou verticalmente [45].
- b) Deslocamento Temporal: Deslocar um sinal no tempo pode criar variações na duração e no início das vocalizações. Isso ajuda a melhorar a capacidade dos algoritmos em identificar os pontos iniciais das vocalizações em diferentes gravações [44].
- c) Adição de Ruído: A adição controlada de ruído ambiental pode simular condições reais de gravação em campo, onde o ruído ambiente não é estacionário.

Isso torna os modelos mais robustos em relação a interferências sonoras [44].

d) Transformações Geométricas: Para imagens de espectrogramas, transformações como zoom, corte e deformação elásticas podem ser aplicadas para simular diferentes condições de gravação[46].

A combinação dessas técnicas, de acordo com a natureza dos dados e do problema, pode fornecer um conjunto de dados mais abrangente e robusto, resultando em modelos de análise mais precisos e generalizáveis [41].

2.6 Aprendizado de máquina para classificação

O aprendizado de máquina pode ser dividido em dois grupos de métodos: supervisionado e não supervisionado. A principal diferença entre os grupos é em relação às amostras utilizadas para o treinamento. Os métodos supervisionados utilizam dados em que sabemos a priori suas classes de saída. Os métodos não supervisionados não precisam de nenhum conhecimento prévio dos dados. Estes métodos tem como objetivo caracterizar à estrutura natural dos dados sem uso de classes pré-determinadas [47].

Neste trabalho são aplicados métodos de aprendizado de máquina de classificação supervisionados, onde é necessário fornecer um conjunto de dados de entrada que estejam rotulados (classe que cada dado de entrada pertence). O aprendizado do modelo consiste em fornecer um número suficiente de dados de entrada com suas respectivas saídas desejadas. Assim o modelo de classificação visa encontrar regras discriminativas capazes de mapear as entradas e seus rótulos (classes) de saída [48]

2.6.1 Rede Neural Artificial

Resumidamente, as redes neurais artificiais de alimentação direta (FNN, *Feedforward Neural Networking*) são algoritmos inspirados nos sistemas neurais biológicos. As FNN são compostas por um sistema interconectado de neurônios e pesos. O sinal de entrada passa por cada neurônio que é multiplicado por um peso e sua resultante é somada e adicionada à uma função de ativação, produzindo um sinal de saída [47]. O resultado de cada saída de uma unidade de neurônio pode ser descrito como:

$$y = \psi \left(\sum_{i=0}^k x_i w_i \right) \quad (2.8)$$

Em que x_i são os dados de entrada da FNN, w representa os pesos que multiplicam o valores de entrada. O somatório dos pesos e dados de entrada são utilizados

na função de ativação ψ . Como principais funções de ativação podemos citar a unidade linear retificada (ReLU, *rectified linear unit*) e sigmoide [47].

A aprendizagem supervisionada de um modelo FNN é realizada por algoritmos de retro-propagação que promovem ajustes dos pesos. Este algoritmo calcula o erro por uma função de custo entre o resultado predito y_i o resultado real y'_i , ajustando os pesos da última camada à primeira [48]. Este processo iterativo acontece até que a função de custo seja minimizada, assim mapeando os dados de entrada com seus rótulos de saída.

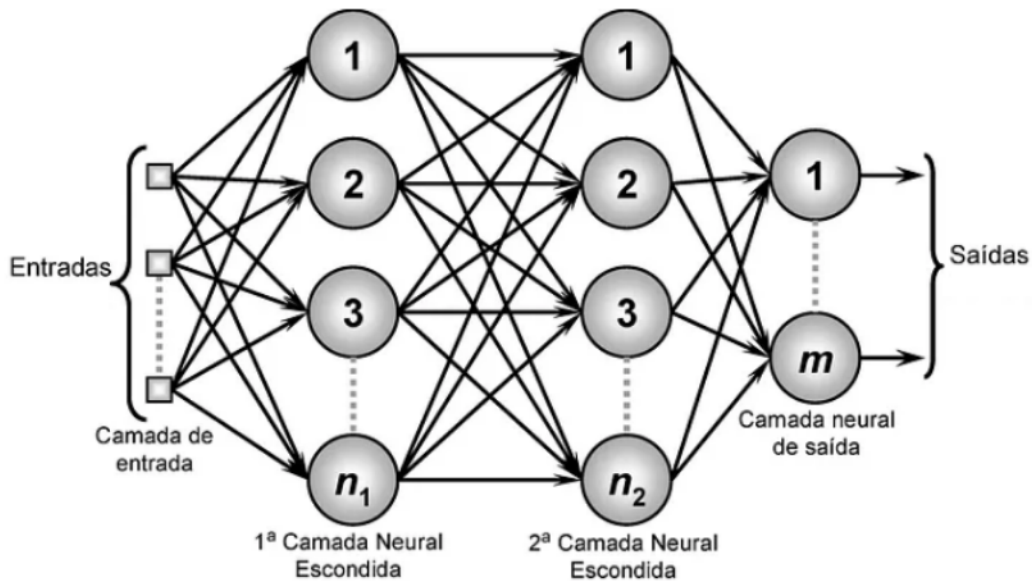


Figura 2.2: Arquitetura característica de uma MLP, composta por neurônios de entrada, uma ou mais camadas ocultas e neurônios de saída, também chamados de perceptrons. Fonte: <http://medimedium.comensina-airede-neural-perceptron-multicamadas>. Acessado em 05/09/2023

Dentre as diversas redes neurais, a *Multilayer Perceptron* (MLP) destaca-se pela simplicidade, versatilidade e capacidade de resolver problemas complexos e não lineares [48]. A MLP é alimentada por dados tabulares e rótulos de saída. Na figura 2.2, podemos ver a arquitetura de uma MLP, observa-se a presença de camadas ocultas entre as camadas de entrada e saída. Estas camadas são assim chamadas devido a não sabermos seus resultados preditos.

2.6.2 *Random Forest*

O algoritmo *Random Forest* (RF) é um método *ensemble*. Os métodos *ensemble* combinam diversos modelos de classificações que encontram um resultado único, por uma votação de majoritária. A Figura 2.3 apresenta o algoritmo RF, que é composto por repetidas arquiteturas de *Decosion Tree*, com variáveis selecionadas aleatoriamente, onde cada resultado de saída (classe) tem um peso em uma votação que decidirá o resultado de saída final (classe final) [47].

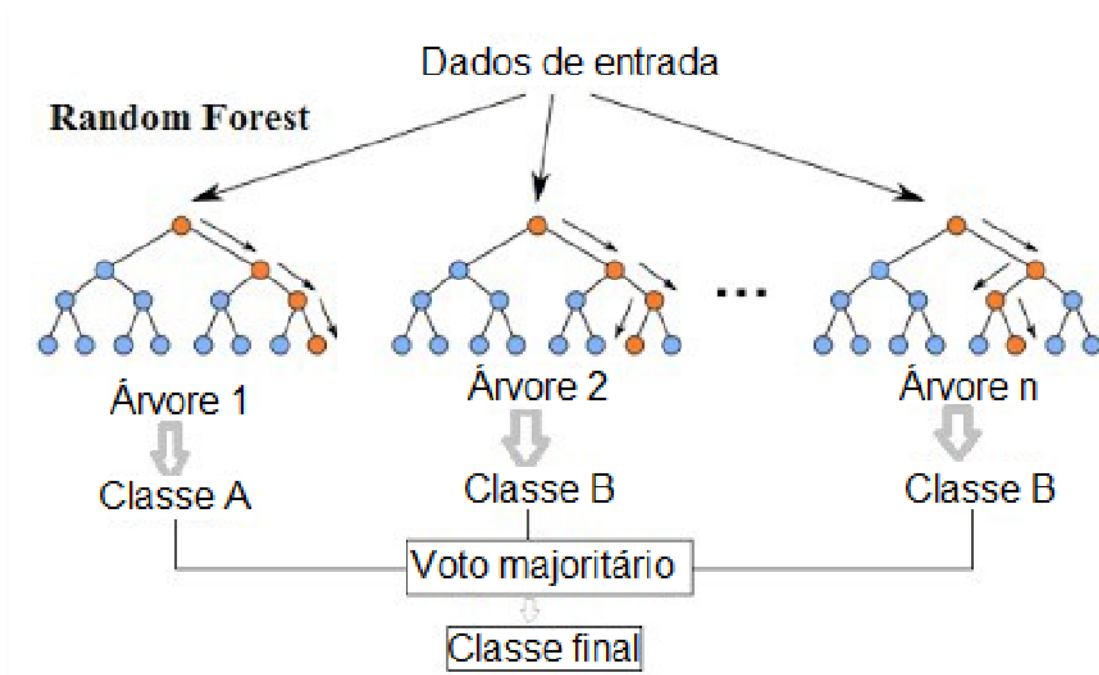


Figura 2.3: Arquitetura do algoritmo *Random Forest* composto por diversas árvores de decisão (*Decision Tree*). Fonte: adaptado de <http://webdesignemfoco.com>. Acessado em 05/09/2023

A *Decision Tree* (DT) assume uma sequência de decisões interconectadas. Em alusão a uma árvore, o algoritmo é composto por raízes, nós, ramos e folhas. Cada nó corresponde a um atributo de entrada que passa por um teste de condição, se atendida segue por um ramo descendente em que passará por outra condicional, até a folha que apresta o resultado de saída, indicando a qual classe pertence à amostra [49]. As decisões tomadas no percurso da raiz até à folha correspondem as

regras de classificação, que mapeiam um conjunto de dados com classes previamente determinadas [47].

2.6.3 *Support Vector Machine*

O *Support Vector Machine* (SVM) realiza um mapeamento não linear onde aumenta a dimensão dos dados de treino, buscando nesta nova dimensão, um hiperplano que separe os dados linearmente. Este hiperplano é encontrado usando vetores de suporte e margens, sendo capaz de separar as classes dos dados [47, 48].

Na figura 2.4 podemos ver os vetores de suporte que se caracterizam por serem os pontos mais próximos do hiperplano que separa as classes de um conjunto de dados. A distância entre o hiperplano e os vetores de suporte é chamada de margem. O SVM busca a máxima amplitude para as margens, pois assim será possível encontrar o melhor hiperplano para separação das classes [2].

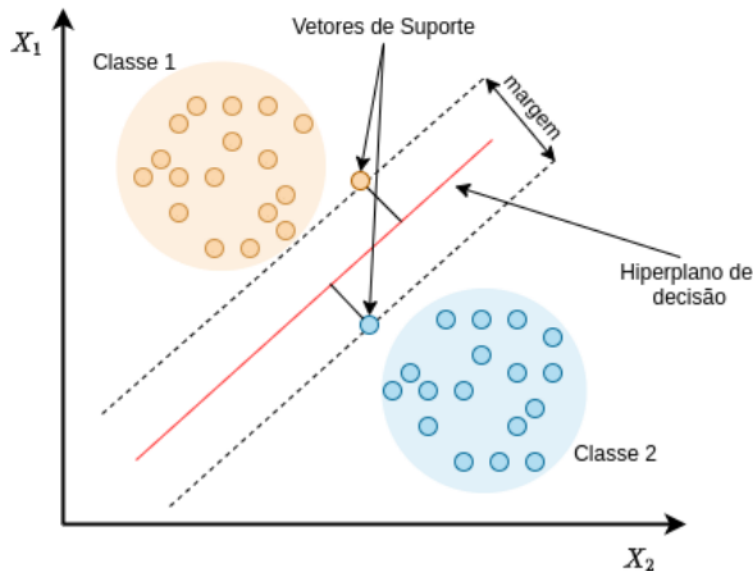


Figura 2.4: Representação do modelo SVM aplicado a classificação binária. Fonte: Retirado de [2]

Quando temos um conjunto de dados não lineares e complexos, o SVM utiliza funções denominadas kernel, que aplicam uma transformação do conjunto de dados para uma dimensão maior, sendo então possível buscar um hiperplano que separe as classes do conjunto de dados [2]. Na figura 2.5 podemos observar a aplicação de um kernel sobre um conjunto de dados não linear.

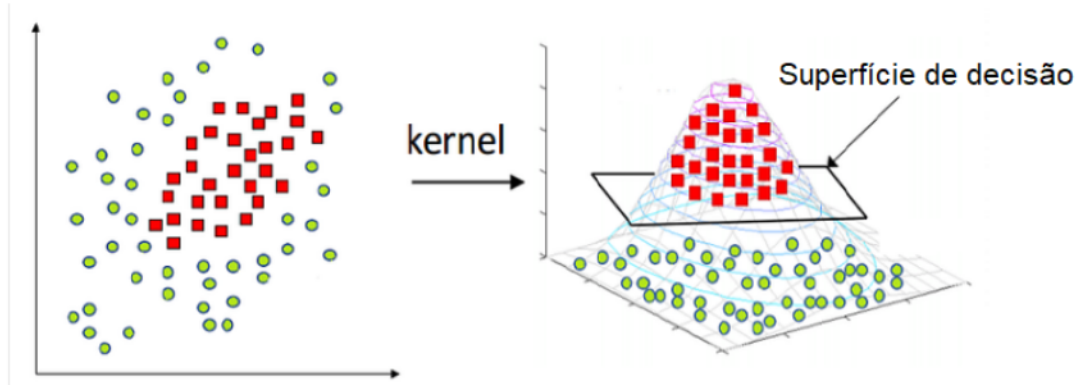


Figura 2.5: Aplicação de um kernel dentro do algoritmo SVM causando um aumento do espaço dimensional do conjunto de dados. Fonte: adaptado de <https://medium.com/@msremigio-maquinas-de-vetores-de-suporte-svm-77bb114d02fc>. Acessado em 10/08/2023

Exemplos de funções kernel que são aplicadas no modelos do SVM, as lineares, polinomiais e a função de base radial (RBF, *Radial Basis Function*). Podemos descrever matematicamente o SVM com Kernel RBF, também conhecido como kernel Gaussiano, por:

$$f(x) = \sum_{i=1}^{n_{sv}} \alpha_i y_i K(x, x_i) + b \quad (2.9)$$

Em que $f(x)$ é a função de decisão que determina a qual classe x pertence. O n_{sv} é o número de vetores de suporte, que são os pontos de treinamento mais próximos à margem de decisão. O α_i representa os coeficientes obtidos durante o treinamento do SVM. O y_i são os rótulos das classes dos vetores de suporte. Por fim, o $K(x, x_i)$ representa a função de kernel, que no caso do RBF é definida como:

$$K(x, x_i) = \exp(-\gamma \|x - x_i\|^2) \quad (2.10)$$

Onde γ é um hiperparâmetro que controla o grau de influência do kernel. O b é o termo de viés ou polarização. O kernel RBF é particularmente útil para capturar relações não lineares nos dados, permitindo que o SVM modele fronteiras de decisão mais complexas em espaços de alta dimensionalidade [48].

2.6.4 Naïve Bayes

Naïve Bayes é um método de classificação muito utilizado por necessitar de poucos dados de treinamento e tempo de processamento. Este modelo é embasado no teorema de Bayes, basicamente determina a probabilidade de um conjunto de dados pertencer a uma determinada classe, desconsiderando qualquer correlação entre os atributos do conjunto de dados. [50].

O *Gaussian Naive Bayes* (GNB) é uma variação do algoritmo *Naive Bayes*, o qual assume uma distribuição gaussiana para os atributos contínuos [47]. A probabilidade de uma amostra x pertencer a uma classe C usando o GNB é denotado por:

$$P(C|x) = \frac{1}{Z} P(C) \prod_{i=1}^n P(x_i|C) \quad (2.11)$$

Em que, $P(C|x)$ é a probabilidade condicional de C dado x , ou seja, a probabilidade de x pertencer à classe C dadas as informações dos atributos. $P(C)$ é a probabilidade a priori da classe C , que pode ser estimada a partir da frequência de ocorrência de C no conjunto de treinamento. $P(x_i|C)$ é a probabilidade condicional de x_i dado C , assumida como uma distribuição gaussiana. Por fim, (Z) é uma constante de normalização que garante que as probabilidades somem um.

Capítulo 3

Metodologia

A metodologia proposta tem por objetivo a detecção e classificação de sinais acústicos de baixa frequência de baleia sei. Na Secção 3.1 serão abordados os métodos de aquisição de dados. A Secção 3.2 apresenta os detalhes da cadeia de processamento para detecção automática de sinais acústicos por técnicas de visão computacional. A Secção 3.3 cita como foram utilizadas técnicas para o aumento de dados à partir de transformações de dados experimentais. A Secção 3.4 se remete aos métodos de classificação binária por meio de ML, aos parâmetros utilizados para otimização dos algoritmos, assim como as métricas utilizadas para avaliação dos modelos.

Exceto pelas representações dos áudios em espectrograma e redução de ruído tonal, todas as outras técnicas de processamento de imagem foram realizadas usando o software OpenCV em linguagem Python [51]. Os algoritmos de aprendizado de máquina foram implementados usando o software Scikit-Learn [52]. Para todos os processos, foi utilizado um processador Intel(R) Core(TM) i5-7200U, com velocidade de 2,5GHz e equipado com 20.0 GB de RAM. Cada estágio é descrito mais detalhadamente na subsecção a seguir e ilustrado pelo fluxograma na Figura 3.1.

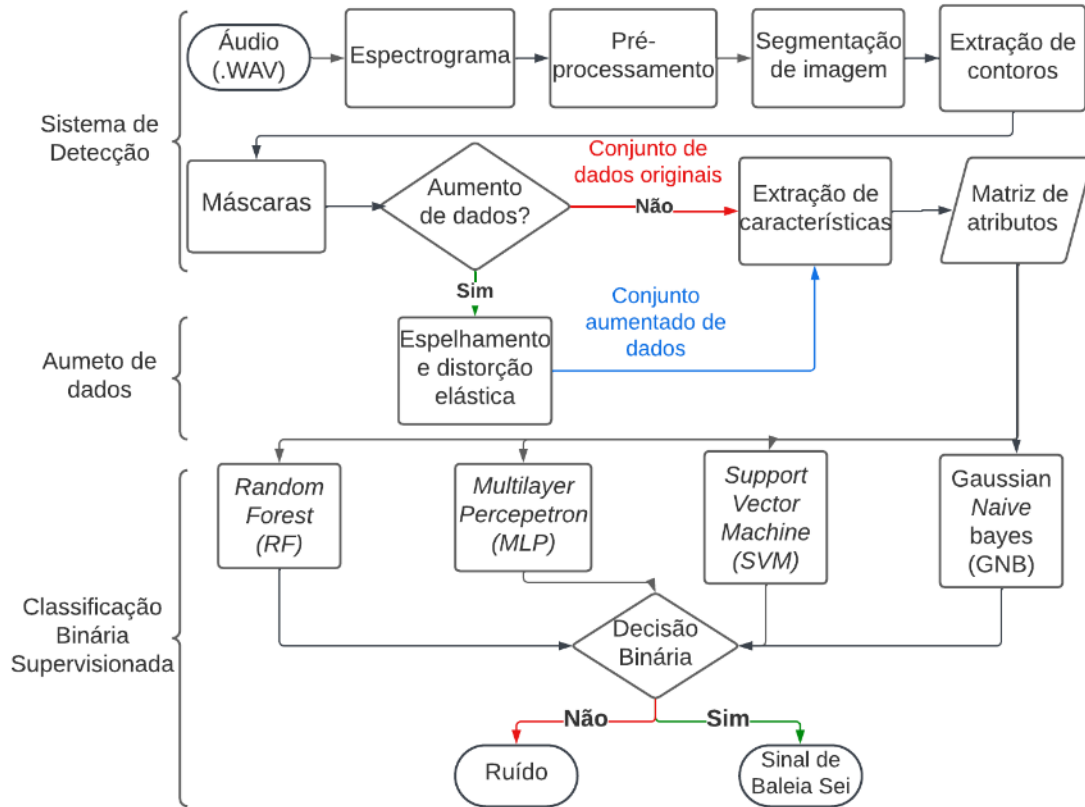


Figura 3.1: Cadeia de processamento do algoritmo proposto, composto por três estágios: detecção de sinais acústicos de baixa frequência, aumento de dados e classificação binária.

3.1 Dados Acústicos

Arquivos de áudio no formato WAV, contendo sinais acústicos de baleias fin, baleias franca austral e baleias sei, foram obtidos a partir de quatro diferentes fontes de dados. Vocalizações de baleias sei foram especificamente adquiridas utilizando Tag de Registro Acústico Digital (DTAG) e obtidos no material suplementar do trabalho publicado por CERCHIO e WEIR [1]. Enquanto que os áudios de baleia franca austral e baleia fin, foram obtidos em bibliotecas acústicas online de livre acesso. Detalhes de cada fonte de dados são descritos mais detalhadamente na Tabela 3.1 e alguns exemplos são ilustrados na Figura 3.3.

Tabela 3.1: Descrição da base dados acústicos utilizados para detecção automática de sinais acústicos e classificação binária.

Espécie (Fonte do dado)	Tempo de gravação (min)	Taxa de Amostragem (Hz)	Área geográfica	SNR (dB)	Tipo de Sinal
Baleia sei (DTAG)	364	196,000	Bacia de Santos Brasil	87.7 a 98.2	DS
Baleia Sei ([1])	6.5	11,000	Ilhas Malvinas	16.3 a 53.7	DS, L, US Arch, LF-variable
Baleia Fin (MobySound)	96	100	Atlântico Norte	1.1 a 8.5	Pulso de 20Hz <i>back beat pulse</i>
Baleia Franca Austral (MobySound e WMMDB)	7.2	8,000 5,120	Africa do Sul Golfo San Jose, Argentina	1.5 a 5.1	descendente-ascendente Variações tonais, ascendente

3.1.1 Aquisição de dados DTAG

. O programa de monitoramento de cetáceos - Bacia de Santos (PMC-BS) é uma condicionante ambiental para atender às exigências de licenciamento ambiental concedido à Petrobras para exploração de petróleo e gás na Bacia de Santos, Brasil.

Dentro do PMC-BS são realizadas Campanhas de Telemetria. Em 2016, durante a segunda campanha de telemetria, foi acoplado um DTAG em uma baleia sei. As Figuras 3.2a e 3.2b mostram um DTAG fixado por quatro copos de sucção (ventosas) e uma vara de fibra de carbono para fixação do equipamento nos animais. Na Figura 3.2c observa-se um DTAG fixado em uma baleia [4]. O equipamento possui um processador de sinal digital capaz de processar dados de áudio gravados com hidrofones e dados abióticos de hora, profundidade, direção magnética e inclinação. Possui sensores de pressão, magnômetros e acelerômetros [53]. Os DTAGs precisam ser recuperados após desprendimento das ventosas. Por meio de uma antena de rádio transmissão VHF (Very High Frequency) os marcadores são localizados, resgatados e seus dados recuperados [3]. O DTAG foi fixado na baleia sei às 9:29 h nas coordenadas 25,7192°S e 45,2261°W, a profundidade local no momento da marcação era de 237,4 metros. Às 14:33 h o instrumento se despreendeu e foi recuperado às 15:20 h nas coordenadas 25,7575°S e 45,2081°W, onde a profundidade local era de 458,1 metros. Totalizando 5,3 h de aderência e dados coletados. Observadores de bordo acompanharam a baleia sei marcada por 1,5 horas contínuas e tiveram observações oportunísticas com o bote de auxílio [3]. Além da baleia sei foram avistadas orcas nas imediações durante a coleta de dados do DTAG [3]. Esta foi a primeira vez que um indivíduo de baleia sei foi equipado com um DTAG no mundo.



(a)



(b)



(c)

Figura 3.2: a) Equipamento DTAG (Fonte: [3]) b) Método de fixação com copos de sucção e vara de carbono utilizados durante a campanha de telemetria do PMC-BS (Fonte: [3]) c) Equipamento DTAG fixado em uma baleia. (Fonte: [4])

3.1.2 Áudios de livre acesso

Os arquivos de áudio da baleia fin foram baixados da biblioteca acústica *Mobysound*, que é um repositório de dados de código aberto fornecido pelo Cooperative Institute for Marine Resources Studies (Oregon State University and NOAA/PMEL) [21]. Além disso, gravações de sons de baleias franca austral foram obtidas de duas fontes: *Mobysound* e do banco de dados acústico do *Watkins Marine Mammal Sound Database* (WMMDB). O WMMDB é fomentado pelo *Woods Hole Oceanographic Institution* e pelo *New Bedford Whaling Museum*.

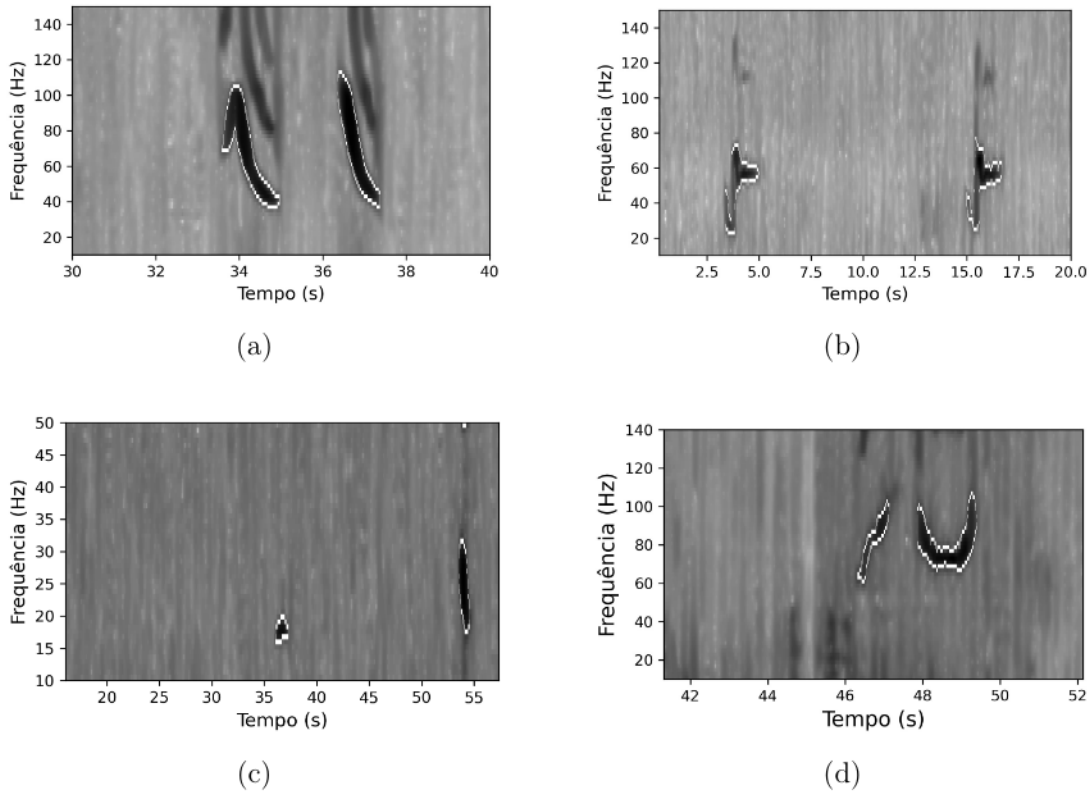


Figura 3.3: Exemplos dos sinais acústicos das três espécies em estudo, apresentadas em imagens em escala de cinza com contornos extraídos pela metodologia apresentada na seção 3.2 do sistema de detecção de sinais acústicos de baixa frequência. a) Sinais DS de baleia sei dos áudios do DTAG b) sinal L-call de baleia sei de áudios do material suplementar de [1] c) Sinais ascendentes e descendentes-ascendentes de baleia fraca austral de áudios do repositório Mobysound d) Vocalização de baleia fin de *back beat pulse* e pulso-20Hz de áudios do Mobysound.

3.2 Sistema de detecção automática de sinais de baixa frequência

Nesta seção serão descritos os métodos empregados para detecção de sinais acústicos de baixa frequência. Inicialmente, os arquivos de áudio foram transformados em espectrogramas para representar os dados. Uma etapa de pré-processamento dos dados, incluindo redução de ruído tonal, adaptação da imagem e suavização foi aplicada. Posteriormente, a segmentação da imagem foi realizada utilizando o limiar adaptativo médio para identificar o contorno de cada vocalização. Máscaras individuais foram criadas a partir dos contornos dos sinais acústicos detectados e 15 características tempo e frequência foram extraídas. Essas características deram origem a um conjunto de dados tabulares para servirem de dados de entrada no processo de classificação. Cada uma destas etapas estão descritas mais detalhadamente nas subseções abaixo.

3.2.1 Representação gráfica dos dados acústicos

Para estabelecer uma taxa de amostragem padrão, os arquivos de áudio foram decimados para 2048 Hz. A única exceção foram os arquivos de baleia fin, pois foram adquiridos e mantidos a uma taxa de amostragem de 100 Hz. Os arquivos de áudio que foram segmentados em arquivos de 60 segundos.

Para produzir espectrogramas a partir dos arquivos de áudio foi aplicada a STFT. A densidade espectral de potência (PSD, *Power spectral density*) foi calculada por uma FFT e extraída sua magnitude, resultando em uma matriz de amplitude em função do tempo e da frequência ($S(t, \omega)$). Os valores de amplitude foram convertidos para a escala decibel ($10 \log_{10}(A)$, onde A representa a amplitude do PSD) [22]. Os parâmetros escolhidos para produzir o espectrograma dos dados de baleia sei e de baleias franca austral foram: tamanho de quadro (*frame size* de 1024 amostras, com sobreposição de 80% e janela de Hann, resultando em um passo de frequência (Δf) de 2 Hz e um passo de tempo (Δt) de 2.44 ms. Além disso, para produzir o espectrograma das baleia fin, seguimos a recomendação do provedor dos dados de livre acesso com tamanho de quadro de 128 amostras, sobreposição de 87.5% e janela de Hann ($\Delta f = 0.781$ Hz e $\Delta t = 10$ ms).

3.2.2 Pré-processamento

O pré-processamento foi realizado em três etapas: redução de ruído tonal, normalização de imagem e suavização pela média não local. Para reduzir o ruído tonal da paisagem acústica local por ruídos antropogênicos (por exemplo, embarcações) e o ruído de fundo causado por ruídos ambientais (por exemplo, tempestades) [22], cada elemento $x_{i,j}$ do espectrograma $S(t, \omega)$ foi subtraído pela média de amplitude de cada bin de frequência (μ_i) e dividido pelo desvio padrão (σ).

$$Sm_{(i,j)} = \frac{x_{i,j} - \mu_i}{\sigma} \quad (3.1)$$

Onde $Sm_{(i,j)}$ corresponde à matriz filtrada resultante do espectrograma $S(t, \omega)$. Ao final desta etapa, os elementos da matriz filtrada que possuíam valor igual a média, assumirão valor zero e os elementos que correspondiam ao valor do desvio padrão, terão valor um. Esse processo também garante que todos os elementos que representam a amplitude sejam ajustados para uma mesma escala padronizada.

Para transformar os valores padronizados de amplitude do PSD em uma escala de cinza, cada elemento da matriz filtrada é normalizada para valores entre 0 e 255. Para isso, é realizada uma normalização min-max com a seguinte equação:

$$I_{(i,j)} = \left(\frac{Sm_{(i,j)} - \min(Sm)}{\max(Sm) - \min(Sm)} \right) \beta + \alpha \quad (3.2)$$

Onde $\min(Sm)$ e $\max(Sm)$ representam os valores mínimo e máximo da matriz filtrada Sm . Para que a matriz normalizada assuma o intervalo entre 0 e 255, a constante β corresponde a 255 e α é igual a 0. O processo de normalização resulta em uma matriz $I_{(i,j)}$, que é considerada como a imagem inicial representada em escala de cinza. Cada elemento na matriz é referido como um píxel com valor no intervalo entre 0 e 255.

A etapa final de pré-processamento consiste na suavização do ruído de fundo proposto por BUADES *et al.* [54]. Para reduzir o ruído de fundo e preservar os detalhes da imagem, foi aplicada a técnica de suavização por média não local. O algoritmo faz um busca por toda imagem na escala de cinza, encontrando valores dos píxeis que sejam semelhantes, calcula a média ponderada e substitui os píxeis similares pela média calculada [54]. Para a implementação e execução do algoritmo de média não local, foi utilizado a biblioteca OpenCV, empregando os parâmetros de configuração recomendados. [51].

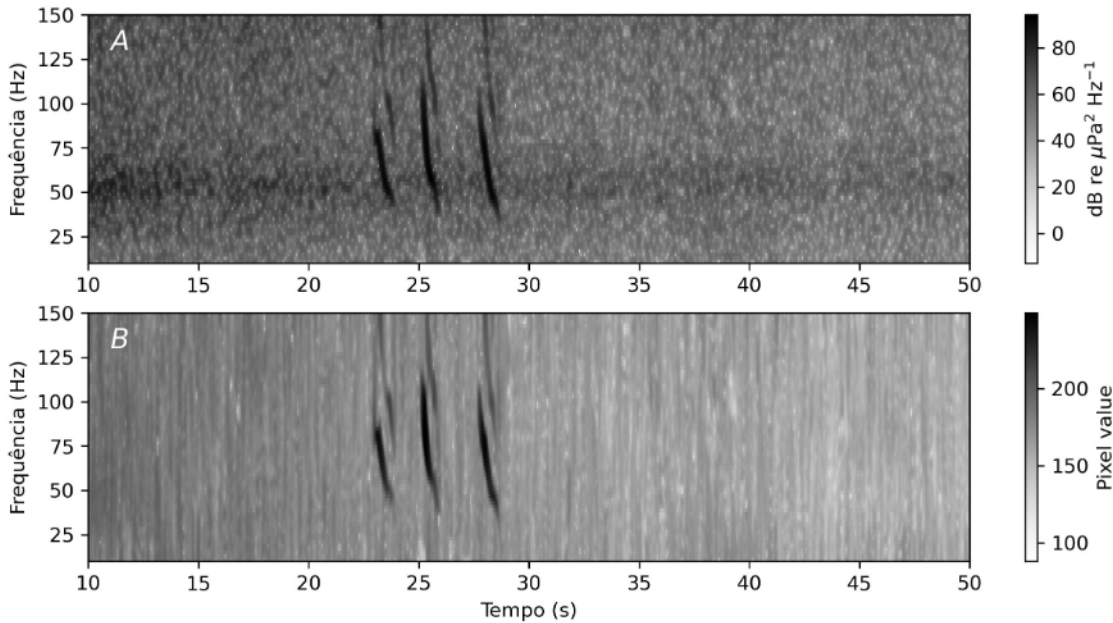


Figura 3.4: Resultado do pré-processamento e adaptação da imagem em escala de cinza para áudio do DTAG. A) Espectrograma com três sinais FM de baixa frequência emitidos pela baleia sei. O espectrograma foi gerado com taxa de amostragem de 2048 Hz, FFT com 1024 pontos, 80% de sobreposição e janela Hann. B) Imagem em escala de cinza após redução do ruído de fundo e suavização pelo método de média não local.

Assim, pode ser observado na Figura 3.4 a transformação dos espectrogramas

em imagens na escala de cinza e os efeitos do pré-processamento, resultando em uma redução do ruído de fundo e um suavização geral da imagem.

3.2.3 Segmentação de imagem

A segmentação de imagem envolve dividir uma imagem em várias regiões para facilitar a extração de informações da imagem original [37]. É uma técnica amplamente utilizada na detecção de objetos no campo da visão computacional [37]. Várias abordagens de segmentação de imagem tem sido relatadas para detecção de sinais acústicos de baleias, incluindo operações de limiarização, crescimento de região ativa e segmentação por aprendizagem profunda (*Deep Learning Segmentation*) [36, 38, 55]. A principal desvantagem da segmentação por limiar global é a sua suscetibilidade à variação do ruído de fundo [36].

Para superar esse problema, utilizamos um valor de limiar dinâmico utilizando uma técnica do campo de processamento de imagem, chamada de limiar adaptativo médio. Neste método cada píxel possui um valor de limiar calculado com base na média local dos valores de intensidade dos píxeis vizinhos ($\mu_{(x,y)}$). Essa abordagem converte imagens em escala de cinza para imagens binárias comparando o valor de cada píxel da imagem com a média dos píxel vizinhos menos uma constante, aplicando a seguinte regra de segmentação binária:

$$I_{Bin} = \begin{cases} 1, & \text{if } P_{(i,j)} > \mu_{(n,m)} - c \\ 0, & \text{if } P_{(i,j)} < \mu_{(n,m)} - c \end{cases} \quad (3.3)$$

Onde I_{Bin} é uma imagem binária com píxel (P) tendo valores de zero ou um. O tamanho da janela (n, m) para cálculo da média móvel, foi configurado como $n = 201$ e $m = 201$ píxeis. A constante (c) foi definida com valor de -24 ou -35, dependendo da fonte do áudio. Foram realizados teste empíricos com distintos valores para a constante e as imagens binárias foram avaliadas. Sendo definido para c , o valor que melhor segmentava a imagem em termos de destacar os sinais acústicos em relação ao ruído de fundo.

O contorno de cada sinal acústico de interesse (SAI) foi determinado a partir de imagens binárias utilizando o algoritmo proposto por SUZUKI e BE [43]. Este algoritmo foi implementados na biblioteca OpenCV e aplicado para encontrar contornos em imagens binárias.

A área de cada contorno foi calculada usando a função de Green. Esta função permite calcular a área delimitada por uma curva fechada em um plano [56]. Essa técnica utiliza integrais de linha para relacionar a geometria da curva com um campo vetorial associado, resultando em uma método que expressa a área em termos dessas integrais [56].

Para avaliar se as áreas dos contornos detectados das vocalizações de baleia sei possuem uma distribuição normal, foi realizado um teste de Shapiro-Wilk [57]. Este teste calcula a métrica estatística chamada de valor-p. Quando o valor-p for maior que 0.05 podemos assumir que às amostras em análise seguem uma distribuição normal. Após este teste, foi constatado que a área do contorno dos sinais de baleia sei possuem uma distribuição normal. Neste contexto, propomos um filtro de área que descarta contornos com áreas maiores ou menores do que a média da área \pm dois desvios padrões (σ), reduzindo o número de detecções falsas.

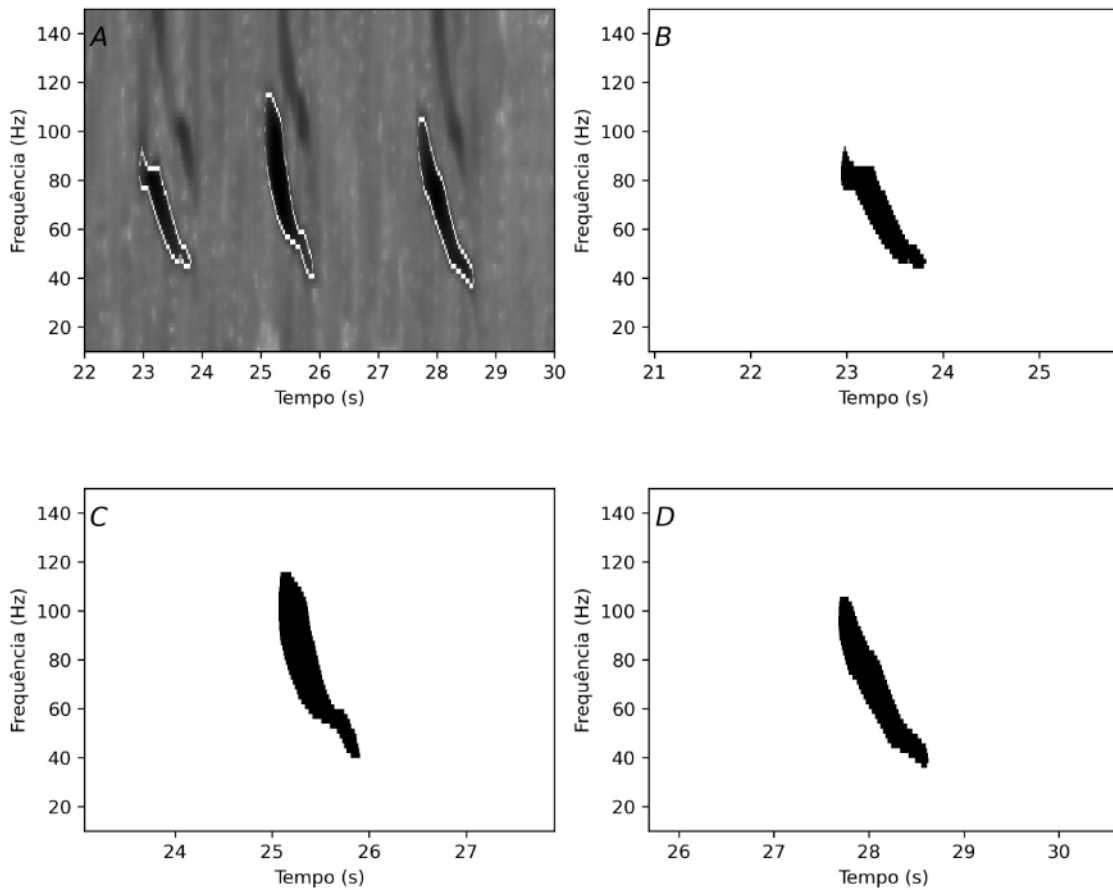


Figura 3.5: Segmentação de imagem e máscaras dos sinais de interesse. A) Sinal de frequência modulada decrescente com contorno detectado. B,C,D) Máscaras individuais de cada sinal de interesse utilizada para extração de características espectrais e temporais.

3.2.4 Extração de atributos

Para cada SAI identificado nas imagens, uma máscara de contorno individual foi gerada, a partir da qual as informações de tempo e frequência podem ser extraídas. As dimensões dos vetores de tempo e frequência resultantes da matriz $S(t, \omega)$, foram aumentadas para a mesma dimensão das máscaras binárias. Desta forma,

pode ser realizada uma multiplicação elemento a elemento, conhecida como produto *Hadamard*, entre as máscaras binárias e as matrizes de tempo e frequência, preservando o tamanho original das máscaras. Isso resultou em uma nova matriz preenchida com valores de tempo ou frequência apenas dentro da área da máscara detectada. Finalmente, 15 características (Tabela 3.2) foram extraídas, conforme sugerido na literatura como características chave para a classificação de sinais de mamíferos marinhos [58, 59].

Tabela 3.2: Descrição das características espectrais e temporais extraídas de cada sinal de interesse que originam a matriz de atributos. Esta matriz serve como dados de entrada para a classificação por métodos de aprendizado de máquina.

Características	Descrição
T	Tempo de duração do sinal acústico (s)
$F_{MAX}, F_{MIN}, F_{IN}, F_{FIN}$	Frequência máxima, mínima, inicial e final (Hz)
F_{MEAN}, F_{SD}	Frequência média e desvio padrão
F_B	Faixa de frequência ($F_{MAX} - F_{MIN}$)
F_c	Frequência central ($F_B/2 + F_{MIN}$)
F_r	Faixa de frequência relativa (F_B/F_c)
F_{MM}	Razão entre a frequência máxima e mínima (F_{MAX}/F_{MIN})
F_{IF}	Razão entre a frequência inicial e final (F_{IN}/F_{FIN})
P_F, P_S, P_T	Frequências do primeiro, segundo e terceiro percentil (Hz)

3.3 Aumento do conjunto de dados

Algoritmos de classificação supervisionada de aprendizado de máquina são suscetíveis a conjuntos de dados reduzidos e desequilibrados, o que pode afetar a capacidade do classificador e a avaliação do método proposto [40]. Considerando isso, é importante observar que o conjunto de dados original consiste em apenas algumas amostras de sinais acústicos de baleia sei ($n = 77$). O que poderia levar a problemas de sobre-ajuste (*overfitting*) ou avaliações imprecisas do desempenho dos modelos preditivos. Para lidar com esse problema, foi realizado o aumento de dados. Estudos anteriores demonstraram a eficácia de operações como recorte, redimensionamento, rotação, espelhamento e distorção elástica no aumento do desempenho de modelos preditivos [44, 46].

Para expandir e aumentar a variabilidade do conjunto de dados, foi utilizada a biblioteca Augmentor. Essa biblioteca emprega uma abordagem baseada em *pipeline* com natureza estocástica que pode realizar várias operações de aumento de dados simultaneamente [46]. Alguns parâmetros podem ser definidos pelo usuário, incluindo a probabilidade de executar uma determinada operação, a magnitude e a grade da operação.

As operações denominadas de espelhamento e distorção elástica, foram aplicadas sobre as máscaras dos SAI. A probabilidade de ocorrência de 30% foi definida para operações de espelhamento, enquanto a distorção elástica foi configurada com uma probabilidade de ocorrência de 90%. Uma preocupação com aumento de dados, é manter às amostras significativas e diversificadas. Neste contexto, a razão de aspecto média de 10:1 dos sinais de baleia sei, foi preservada. A operação de distorção foi aplicada com um tamanho de grade de 100 x 10 píxeis e uma magnitude de 5.

A Figura 3.6 apresenta o exemplo de um sinal acústico de baleia sei no processo de aumento de dados. Neste processo, a máscara do sinal original (B) é distorcida e/ou espelhada, resultando em um número maior de amostras requeridos para treinar modelos classificação robustos. Cada máscara dos sinais acústicos de interesse foi replicada com variações geométricas de forma aleatória, dentro dos parâmetros descritos acima, até que o conjunto de dados estivesse equilibrado.

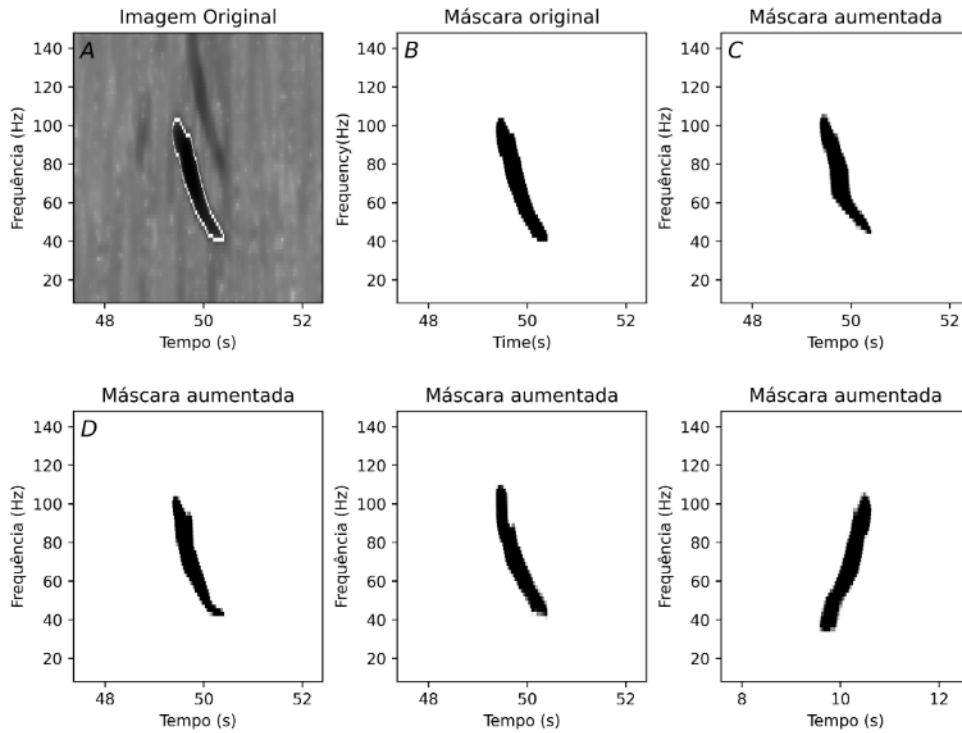


Figura 3.6: Exemplo das máscaras dos sinais acústicos de interesse (SAI) de vocalização DS, aumentadas a partir de transformações dos dados experimentais. A) Imagem original com contorno encontrado. B) Máscara binária original. C,D,E,F) Máscaras depois das técnicas de aumento de dados de distorção e/ou espelhamento.

Como resultado das operações de aumento sintético de dados (distorção elástica e espelhamento), o número de amostras do conjunto de dados original acresceu de $n = 1.231$ para $n = 10.746$. Para obter um conjunto de dados equilibrado, o número de chamados de baleia sei aumentou de $n = 77$ para $n = 4.956$, que foram rotuladas como classe um. Os sinais acústicos de baleia fin, baleias franca austral e ruído ($n = 5.800$) detectados automaticamente, foram rotulados como classe zero.

A partir deste ponto em diante, dois conjuntos de dados foram usados para treinar e testar os modelos de classificação propostos. Os conjuntos de dados consistem em um par ordenado de características extraídas dos SAI detectados e suas respectivas classes. Uma matriz de atributos foi criada com dimensões de $n \times m$, representando o número de SAI detectados, e as características da Tabela 3.2. Assim, o conjunto de dados original foi composto por uma matriz $OD_{x(1.231,15)}$, juntamente com um vetor contendo informações de classes binárias $OD_{y(1.231,1)}$. O conjunto de dados aumentado consistiu na matriz de atributos com dimensões $AD_{x(10.746,15)}$ e $AD_{y(10.746,1)}$. Os detalhes do processo de aumento de dados são resumidos na Tabela 3.3.

Tabela 3.3: Quantidade de sinais acústicos de interesse (SAI) detectados pelo sistema de detecção de sinais acústicos de baixa frequência, que formam o conjunto original de dados (segunda coluna) e o conjunto aumentado de dados (terceira coluna) após operações de espelhamento e distorção.

Sinais detectados Automaticamente	Conjunto Original de Dados	Conjunto Aumentado de Dados
Baleia Sei (DTAG)	24	2402
Baleia Sei (CERCHIO e WEIR 1)	53	2554
Baleia Fin	251	1506
Baleia Franca Austral	36	1512
Ruído	867	2772
Total	1231	10746

3.4 Classificação Binária Supervisionada de Sinais de Interesse

Quatro algoritmos de aprendizado de máquina foram considerados para discriminar chamados de baleia sei do ruído de fundo e dos sinais de baleia franca e baleia fin. Os algoritmos de classificação aplicados foram *Random Forest* (RF), *support vector machine* (SVM), *Gaussian Naive Bayes* e *Multilayer Perceptron* (MLP). Uma revisão de cada algoritmo de classificação pode ser encontrada em OSISANWO *et al.* [50] e WITTEN *et al.* [47].

O primeiro passo para aplicar algoritmos de aprendizado de máquina é dividir o conjunto de dados em subconjuntos de treinamento, validação e teste. Uma separação dos dados foi realizada para treinar, validar e testar todos os classificadores de aprendizado de máquina propostos. Inicialmente, 30% do conjunto de dados foi separado e denominado de subconjunto de teste. Este subconjunto não participará do processo de treinamento e validação do modelo. É um conjunto independente que terá um papel importante na avaliação de desempenho.

Em seguida, foi aplicado o método de validação cruzada *k-fold* com cinco *folds* para segmentar os 70% restantes dos dados [e.g., 44]. A técnica *k-fold* estratificada foi aplicada para garantir que cada um dos cinco segmentos tivesse praticamente a mesma quantidade de vocalizações de baleias sei. O método *K-fold* aplicado divide o conjunto de dados em cinco subconjuntos, sendo que quatro deles são usados para treinar os modelos e um é usado como subconjunto de validação. Esse método é crucial, pois desempenha um papel significativo na avaliação do desempenho dos modelos e na otimização dos hiperparâmetros.

3.4.1 Parâmetros dos algoritmos de classificação

Para otimizar o desempenho dos modelos, foi realizada uma busca em grade (*grid search*) para encontrar os melhores hiperparâmetros para a classificação binária. Os resultados da busca em grade foram analisados monitorando a acurácia média e o *F1 score* da validação cruzada de cada teste realizado com diferentes configurações. Os hiperparâmetros que resultaram nos valores mais altos para as métricas analisadas foram aplicados. Nesta seção serão apresentados apenas os hiperparâmetros que foram alterados em relação aos valores padrões definidos na biblioteca scikit-learn. Para garantir resultados consistentes, o mesmo valor de semente foi escolhido para todos os classificadores de aprendizado de máquina em cada teste.

No algoritmo RF, o número de árvores de decisão afeta o desempenho do modelo, pois possui um valor ideal que combina precisão do modelo, estabilidade e eficiência computacional. Em geral, um maior número de árvores melhora a precisão do modelo. No entanto, essa melhoria diminui à medida que o número de árvores cresce, atingindo um ponto de saturação onde passa a prejudicar o modelo em relação a eficiência computacional e ajuste. Neste estudo utilizamos 300 árvores de decisão no algoritmo RF. O hiperparâmetro de profundidade máxima determina a complexidade de cada árvore individual e foi definido como 7. Além disso, o parâmetro de divisão mínima de amostra estabelece um limite para o número mínimo de amostras necessárias para dividir um nó interno em nós filhos, sendo estabelecido como 7.

A arquitetura do MLP usada neste estudo consistiu em duas camadas ocultas com 128 e 256 neurônios, respectivamente. A função de ativação selecionada foi a Unidade Linear Retificada (ReLU). Durante o processo de treinamento, os pesos do modelo foram atualizados usando o algoritmo *Adaptive Moment Estimation* (Adam). Outros hiperparâmetros também foram definidos para otimizar o desempenho do classificador MLP. O tamanho do lote (*batch size*), que determina o número de amostras propagadas pela rede neural antes de atualizar os pesos, foi definido como 50. Além disso, um valor de parâmetro de regularização foi definido com valor de 1×10^{-4} . Esse parâmetro controla o equilíbrio entre ajustar os dados de treinamento e a evolução dos pesos do modelo, ajudando a prevenir sobre-ajuste.

O algoritmo SVM empregou a função de kernel de base radial. O parâmetro de regularização, denotado como C no scikit-learn, foi definido com valor 10. Esse parâmetro determina o equilíbrio entre maximizar a margem e minimizar a classificação incorreta das amostras.

Quanto ao algoritmo GNB, em que pressupõe que cada característica segue uma distribuição Gaussiana dentro de cada atributo. Um hiperparâmetro de suavização de variância com valor de 1×10^{-9} foi adicionado às variâncias das distribuições das características, para evitar divisão por zero nos cálculos de probabilidade.

3.4.2 Avaliação dos modelos preditivos

Para avaliar o desempenho dos modelos de classificação, foi utilizada uma matriz de confusão, que apresenta uma tabulação das classes reais e previstas geradas pelos modelos propostos [60]. A matriz de confusão consiste em quatro categorias: verdadeiro positivo (TP), verdadeiro negativo (TN), falso positivo (FP) e falso negativo (FN). Usando essas categorias, as métricas estatísticas de acurácia, precisão, *recall* e *F1 score*, podem ser calculadas.

A acurácia avalia a proporção geral das classificações corretas em relação a todas as observações. A precisão representa quão bem o modelo preditivo evita falsos positivos. A métrica *recall* avalia a capacidade de prever eventos verdadeiros positivos, e o *F1-score* é calculado pela média ponderada da precisão e do *recall* [61].

$$\text{Acuracia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.4)$$

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (3.5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.6)$$

$$F1 \text{ score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (3.7)$$

No entanto, quando o conjunto de dados possui classes desbalanceadas, a acurácia pode levar a avaliações de desempenho excessivamente otimistas [60, 62]. Isso ocorre porque o número de categorias negativas (por exemplo, ruído detectado) é muito maior do que as categorias positivas (chamados de baleias sei). Logo, a acurácia pode ser alta mesmo que o desempenho do modelo nas categorias positivas seja ruim [44]. Assim, nos casos onde o conjunto de dados é desbalanceado, as métricas de precisão, *recall*, *F1-score* e a curva de Precisão-*Recall* (Curva-PR) podem fornecer uma avaliação mais precisa do desempenho dos modelos [44, 60].

A curva-PR é uma representação gráfica das taxas de *recall* no eixo x e da precisão no eixo y , à medida que o limiar de probabilidade varia de 0.0 a 1.0 [62]. Um classificador com valores mais altos de precisão e *recall* resultará em uma curva-PR que ocupará o canto superior direito do gráfico [60]. A curva-PR é um método altamente recomendados para avaliar problemas de classificação binária e comparação das performances de modelos com conjuntos de dados desbalanceados [62].

Outro bom indicador do desempenho dos modelos de classificação é a área formada sob a Curva-PR, denominada de área sob a curva (AUC, *Area Under de Curve*). Porém o cálculo da AUC pelo método trapezoidal utiliza interpolação linear e quando aplicado sobre a Curva-PR pode superestimar os resultados [62, 63]. Na literatura, recomenda-se o cálculo da precisão média (AP, *Average precision*) para estimar a área sob a Curva-PR (PR-AP), denotado na equação 3.8.

$$AP = \sum_{i=1}^n (R_i - R_{i-1}) \cdot P_i \quad (3.8)$$

Na equação 3.8, n representa o número total de pontos da curva-PR, onde P_i é a precisão no i -ésimo ponto e R_i é o *recall* no i -ésimo ponto. A soma ocorre de $i = 1$ à $i = n$. Essa equação aproxima a AUC somando os produtos das diferenças do *recall* ($R_i - R_{i-1}$) e os valores de precisão correspondentes (P_i) em cada ponto da curva.

Capítulo 4

Resultados e Discussões

Os resultados reportados neste trabalho foram discutidos separadamente visando atender os objetivos citados na seção 1.5. Assim, na Seção 4.1 serão apresentados e discutidos os resultados alcançados pelo sistema de detecção de sinais acústicos de baixa frequência. A Seção 4.2 relata os efeitos causados nos modelos preditivos quando treinados com conjunto de dados aumentados. A Seção 4.3 demonstra as análises comparativas entre os quatro algoritmos de ML testados.

4.1 Resultados do sistema de detecção de sinais acústicos de baixa frequência

Para avaliar a eficácia do sistema de detecção utilizando técnicas de visão computacional, uma comparação entre os resultados da detecção automatizada e a revisão manual dos metadados anotados foi realizada. Para esta avaliação, foi considerada a análise manual como perfeita.

Os dados obtidos *online* em bibliotecas acústicas de livre acesso eram acompanhados de metadados, que continham anotações da quantidade de vocalizações detectadas e horários de ocorrência. Estas anotações foram realizadas por outros autores e revisadas manualmente para este estudo. Para validar as informações do número de detecções contida nas anotações, foram analisados cuidadosamente todos os espectrogramas dos arquivos de áudio, confirmando assim o número total de detecções. Ao aproveitar a experiência de outros pesquisadores quando utilizamos os metadados dos arquivos baixados como referências e conduzir revisões meticulosas, visamos garantir a validade das análises. Esse processo permitiu avaliar com precisão e confiabilidade o sistema de detecção automatizado.

A segunda etapa da avaliação do sistema de detecção envolveu uma análise abrangente dos espectrogramas e máscaras de contorno para todos os sinais identificados pelo sistema de detecção automatizado. Para facilitar essa análise, salvamos to-

das as máscaras dos sinais detectados como imagens. Um analista foi encarregado de rotular as imagens das máscaras em duas classes: "ruído" ou "Vocalizações" da espécie-alvo. Essa classificação foi realizada para os sinais de todas as fontes de dados. Os resultados desse processo estão apresentados na Tabela 4.1.

Tabela 4.1: Número de sinais de interesse detectados manualmente e automaticamente, classificados por um analista como sinal acústico biológico das espécies e ruído.

Espécie e fonte do dado	Deteção manual (Vocalizações)	Deteção automática (Vocalizações)	Deteção automática (Ruído)
Baleia Sei (DTAG)	24	24 (100%)	699
Baleia Sei (Cerchio et al,2022)	54	53 (98.2%)	12
Baleia Fin (Mobysounds)	265	251 (96.9%)	69
Baleia Franca Austral (Mobysounds e WMMB)	36	34 (94.4%)	87
Total de detecções	379	362 (95.5%)	867

Após a análise manual e revisão dos metadados anotados, um total de 379 vocalizações de três espécies foram identificadas. O sistema de detecção automatizado encontrou 362 sinais acústicos de baleias. Considerando ambas as fontes de dados de baleias sei, foram detectados manualmente um total de 78 sinais acústicos de baleias sei. O sistema de detecção automatizada encontrou 77 vocalizações, o que representa 98.7% do total. Além disso, o sistema de detecção automatizado detectou 867 sinais de interesse que foram considerados como ruído pelo analista, com 80.6% ($n = 699$) do total de ruídos provenientes do DTAG.

É importante notar que todos os sistemas foram originalmente desenvolvidos para detectar chamados de baleias sei. No entanto, com pequenos ajustes nos parâmetros de detecção, filtro de área e na constante C do limiar adaptativo, o método também chegou a um desempenho satisfatório na detecção de sinais acústicos de outras duas espécies, alcançando uma probabilidade de detecção de 96.9% para baleia fin e 94.4% para baleia franca austral.

Os resultados apresentados neste estudo demonstram a eficácia das técnicas de processamento de imagem em conjunto com processamento de sinais aplicadas na detecção de sinais acústicos FM de baixa frequência, alcançando uma probabilidade de detecção de 95.5% para o total de sinais acústicos das três espécies-alvo. Esse desempenho está em consonância com os resultados de autores anteriores que utilizaram métodos semelhantes, implementando a segmentação de imagem por abordagem baseada em limiar global que alcançou 90% de detecção de sinais biológicos, e o método de extração de contorno baseado em crescimento de região (*region growing threshlod*), que detectou 89% do total das vocalizações [36, 37]. A implementação

do limiar adaptativo médio na binarização da imagem ajudou o sistema a lidar com uma ampla gama de razões sinal-ruído (SNR, *signal to noise ratio*).

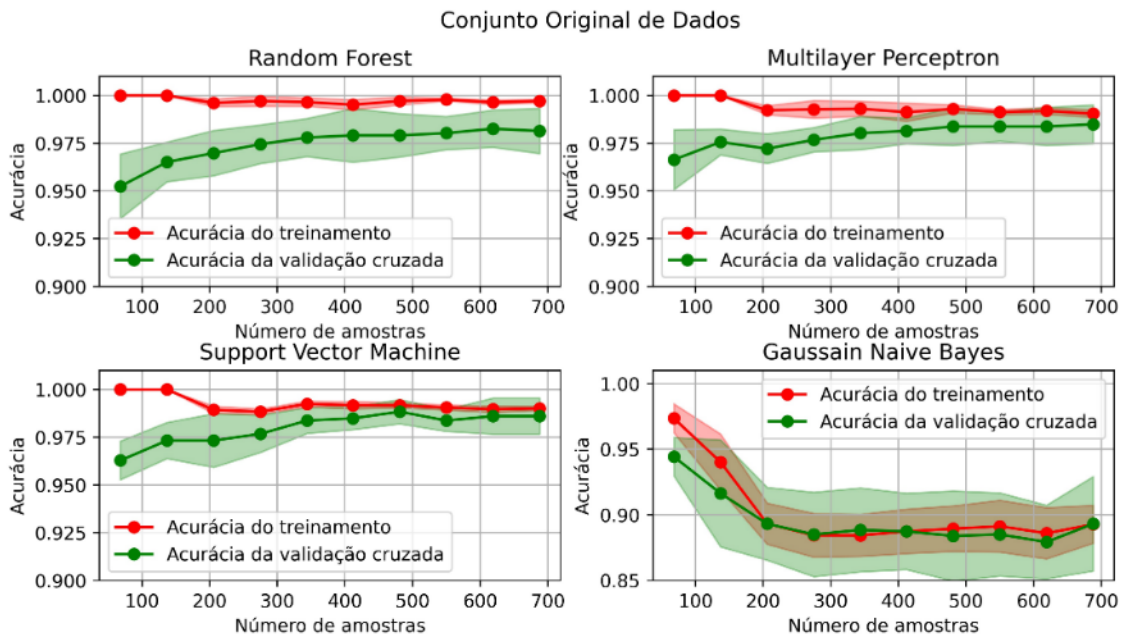
Além disso, a análise revelou que as detecções falsas foram predominantemente observadas nos dados DTAG das baleias sei. Isso pode ser atribuído ao ruído inerente gerado pelos dispositivos de gravação ligados aos mamíferos marinhos. Embora nenhuma medida específica tenha sido tomada para mitigar detecções falsas, exceto pela aplicação de um filtro de área do contorno. Portanto, a tarefa de discriminar os sinais de interesse detectados foi deixada para os classificadores de aprendizado de máquina utilizados no estudo.

4.2 Efeitos do aumento de dados

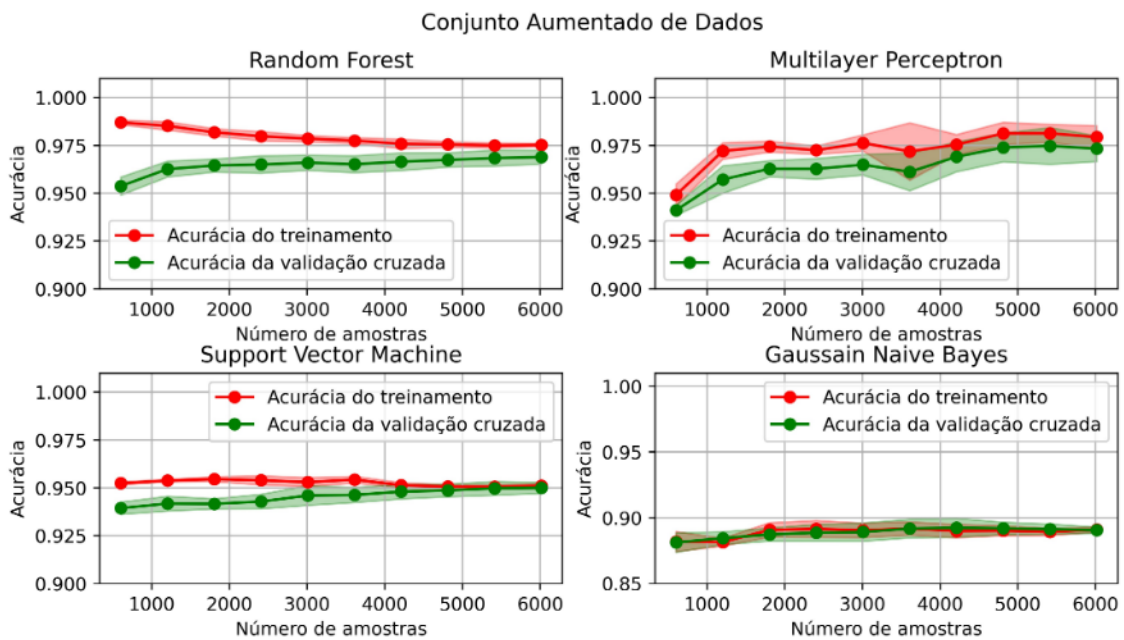
Para obter um modelo preditivo robusto, o método proposto deve evitar o sobreajuste e o subajuste. Neste contexto, as curvas de aprendizagem são indicadores eficazes para avaliar os ajustes dos modelos. A Figura 4.1 apresenta as curvas de aprendizagem para os quatro algoritmos de classificação. Estas curvas consistem na evolução da acurácia do subconjunto de dados de treinamento e subconjunto de validação a medida que um maior número de amostras é utilizado. Esta análise é realizada no conjunto dos dados originais e aumentados, a fim de avaliar os efeitos provocados pelo aumento de dados de forma computacional.

O sobreajuste ocorre quando um modelo de ML se ajusta muito bem aos dados de treinamento, a ponto de aprender o "ruído" dos dados em vez dos padrões reais [7]. Isso resulta em um modelo com bom desempenho nos dados de treinamento, mas possui um desempenho ruim em dados não vistos, como o subconjunto de validação ou teste. Na curva de aprendizagem, os modelos que sofrem sobreajuste apresentam um aumento contínuo na curva da acurácia no subconjunto de treinamento a medida que aumenta o número de amostras utilizadas, porém quando é usado o subconjunto de validação, a curva da acurácia diminui ou se mantém estagnada, resultando em menor convergência entre as duas curvas. Enquanto, o subajuste ocorre quando o modelo é muito simples para encontrar padrões no conjunto de dados, com isso a acurácia no treinamento e na validação terão resultados ruins. Nas curvas de aprendizagem de modelos com subajuste, a acurácia é baixa e pode diminuir a medida que mais amostras são utilizadas no treinamento [7]. O mesmo comportamento acontece nas curva de acurácia utilizando o subconjunto de validação.

O desempenho de um classificador pode ser influenciado por vários fatores, incluindo a qualidade e o tamanho do conjunto de dados usado para o treinamento [44]. A técnica de aumento de dados, que consistiu na inversões e distorções das máscaras dos sinais de interesse detectados, foi utilizada sobre o conjunto de dados original. As curvas de aprendizado da acurácia do conjunto de dados original, representado



(a)



(b)

Figura 4.1: Processo de aprendizado dos classificadores representado pela curva de acurácia no subconjunto de treinamento e de validação do conjunto de dados original e aumentado. As linhas representam os resultados da acurácia média em relação ao número de amostras. Enquanto a área sombreada representa o desvio padrão. (a) Conjunto original de dados b) Conjunto de dados aumentados.

na Figura 4.1a, apresentaram uma menor convergência das curvas de treinamento (em vermelho) e validação (em verde), sugestivo de ocorrência de sobre-ajuste nos modelos RF e MLP. Enquanto, o modelo GNB apresentou uma curva decrescente

da acurácia com aumento do número de amostras analisados do conjunto de dados originais, o que indica a ocorrência de subajuste.

No entanto, quando a avaliação foi realizada no conjunto de dados aumentado, representados na Figura 4.1b, observa-se uma melhor convergência das curvas da validação e treinamento para todos os modelos, sugerindo a mitigação dos problemas de sobre-ajuste e subajuste. As áreas sombreadas representam o desvio padrão da curva de aprendizagem. Comparando a Figura 4.1a e Figura 4.1b quando os modelos são treinados com conjunto de dados aumentados, há um queda no desvio padrão da curva de aprendizagem. Isto indica que os modelos estão tendo menor dispersão dos resultados da classificação, sugerindo que modelos treinados com dados aumentados possuem maior estabilidade.

Para obter uma estimativa quantitativa dos efeitos do aumento sintético de dados, e comparar os algoritmos de aprendizado de máquina aplicados a classificação binária de sinais de baleia sei. Os desempenhos dos modelos foram avaliados e apresentados na Tabela 4.2 utilizando métricas estatísticas após o método de validação cruzada por *k-fold*.

Devido ao desequilíbrio significativo no conjunto de dados original, com a classe de ruído (classe 0) muito mais numerosa do que a classe de vocalizações de baleia sei (classe 1), resultando em um grande número de verdadeiros negativos (matriz de confusão), a métrica de acurácia pode levar a uma interpretação incorreta dos resultados. Portanto, para a análise quantitativa com o objetivo de comparar os efeitos da ampliação de dados, a Tabela 4.2 apresenta os resultados médios e o desvio padrão apenas para precisão, *recall* e *F1 score*.

O aumento do tamanho do conjunto de dados para treinamento dos modelos, resultou em um melhor desempenho nas médias e desvio padrões das métricas de precisão, *recall* e *F1 score*, apresentados na Tabela 4.2. Isto sugere que os modelos de classificação treinados com o conjunto de dados aumentado apresentam uma melhor consistência e redução na dispersão em suas previsões de classe [44].

O *F1 score* representa o equilíbrio entre a capacidade do modelo de identificar corretamente todas as instâncias positivas (precisão) e evitar falsos positivos (*recall*). Esta métrica sumariza os resultados de *recall* e precisão. O *F1 score* apresentou um aumento significativo no modelo GNB com o conjunto de dados aumentado, passando de 55.3% para 89.9%, o que representa um aumento de aproximadamente 35.8%. Os modelos RF, MLP e SVM também apresentaram maior desempenho nesta métrica quando treinados com o conjunto de dados aumentado, resultando em aumento médio de 9% do *F1 score*.

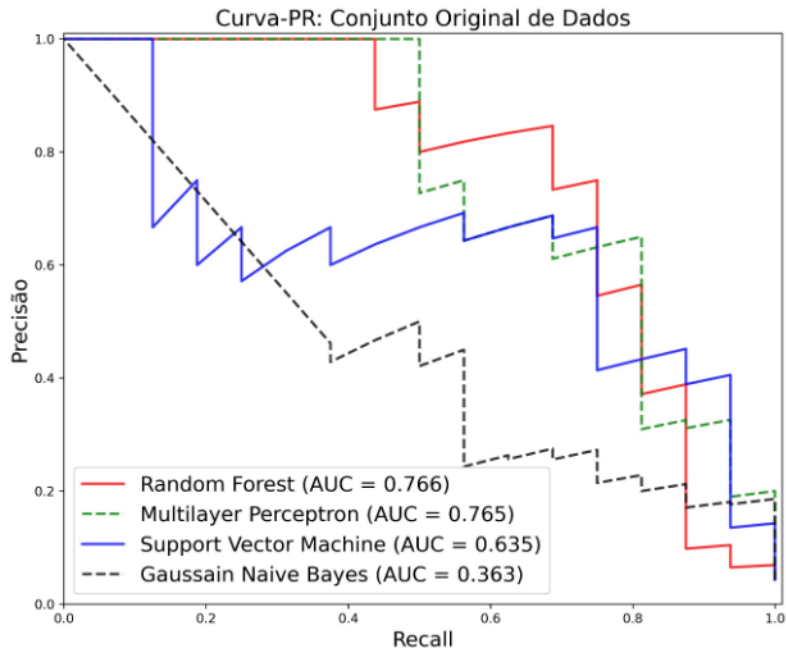
Tabela 4.2: Resultados dos quatro modelos de classificação no subconjunto de validação cruzada quando treinados com o conjunto de dados original e aumentado. São apresentadas as médias e os desvios padrões das métricas propostas.

Classificador	Conjunto de dados	Precisão (%)	Recall (%)	<i>F1 score</i> (%)
RF	Original	94.5 ± 5.9	77.1 ± 15.3	84.6 ± 11.3
	Aumentado	94 ± 0.7	99.6 ± 0.1	96.7 ± 0.4
MLP	Original	89.33 ± 10.3	86.9 ± 8.4	87.7 ± 6.9
	Aumentado	98.6 ± 0.4	97.8 ± 1	98.2 ± 0.5
SVM	Original	91.38 ± 10.6	90.3 ± 8	90.3 ± 6.4
	Aumentado	93.7 ± 0.5	95.7 ± 0.8	94.7 ± 0.3
GNB	Original	41.2 ± 8.7	88.6 ± 8.4	55.3 ± 7.3
	Aumentado	82.9 ± 0.7	97 ± 0.7	89.9 ± 0.2

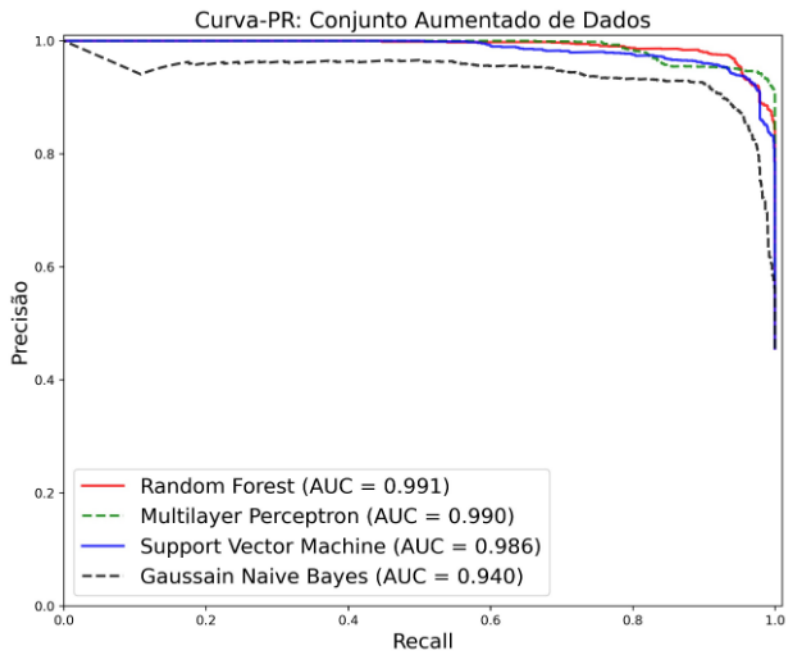
Os resultados obtidos neste estudo em relação aos efeitos do aumento de dados são consistentes com as descobertas relatadas por CAI *et al.* [44]. Os autores testaram dois modelos de classificação de aprendizado profundo e observaram uma melhoria de 6.5% e 1.5% na precisão ao aplicar a inversão de dados brutos como parte de seu método proposto.

A métrica que teve o maior crescimento médio em todos os modelos de classificação foi o *recall*, com destaque para o modelo RF que foi de 77.1 ± 15.3 à 99.6 ± 0.1 e o algoritmo MLP que passou de 86.9 ± 8.4 para 97.8 ± 1 . Este resultado sugere que modelos treinados com o conjunto de dados aumentado diminui consideravelmente a quantidade de falsos negativos.

Para a avaliação final dos efeitos do aumento de dados e a comparação dos modelos, foi utilizado um subconjunto de testes completamente independente (dados não vistos). Este subconjunto de teste foi obtido na primeira etapa do método de segmentação de dados e nunca participou do processo de treinamento ou validação, representando 30% do total dos dados. Outro método apropriado para medir o desempenho de um classificador em um conjunto de dados desequilibrado é a curva-PR [60]. As curvas-PR de ambos os conjuntos de dados foram calculadas, resultando nas Figuras 4.2a e 4.2b.



(a)



(b)

Figura 4.2: Curva Precisão-Recall (Curva-PR) dos modelos de ML calculada a partir das previsões de classe do subconjunto teste (independente) (a) Curva-PR dos modelos treinados com o conjunto original de dados (b) Curva-PR dos modelos treinados com o conjunto de dados aumentado.

Outro fator que vale ressaltar foi o aumento da área sob a curva-PR (PR-AUC) na Figura 4.2b. A PR-AUC pode ser aproximada pelo somatório da precisão média de n pontos sob a curva-PR (PR-AP), como visto na seção 3.4.2. Assim a Tabela 4.3

apresenta os resultados da métrica PR-AP.

A Figura 4.2 mostra uma evolução substancial nas curvas-PR quando o número da amostra é maior. No conjunto de dados original, o número de amostras positivas (vocalizações de baleia sei) é significativamente menor do que as amostras negativas (ruído e outras sinais biológicos). Como resultado, a precisão e o *recall* podem apresentar alta variação com diferentes limiares. Consequentemente, a classificação correta das amostras positivas é altamente sensível, levando a flutuações acentuadas na Curva-PR e menor área sob a curva-PR, como podemos ver na Figura 4.2a. A ampliação do conjunto de dados de forma computacional, expôs os modelos a padrões mais diversos da classe positiva, resultando em melhores previsões entre instâncias positivas e negativas. Isso levou a um aumento nos valores de precisão e *recall* e a um melhor desempenho na posição da Curva-PR aumentando a área sob a curva, como visto na Figura 4.2b e Tabela 4.3.

Tabela 4.3: Resultados da precisão média (AP) da curva-PR (PR-AP), que é uma aproximação da área sob a curva-PR (PR-AUC) para os modelos RF, MLP, SVM e GNB treinados com o conjunto de dados original (coluna da esquerda) e aumentado (coluna da direita). O valor mais próximo de um indica melhor desempenho.

Modelo de Classificação	Conjunto Original	Conjunto Aumentado
RF	0.77	0.99
MLP	0.77	0.99
SVM	0.64	0.98
GNB	0.36	0.94

Na Tabela 4.3, pode-se observar que os modelos RF, SVM e MLP, treinados com dados aumentados, tiveram crescimento no valores da PR-AP chegando a 0.99. Em comparação com os modelos treinados com conjunto de dados original, o aumento médio da PR-AP foi de 0.21.

É importante observar que qualquer pontuação de PR-AP abaixo de 0.5 implica que as previsões do modelo não são melhores do que um palpites aleatórios. O classificador GNB, usando o conjunto de dados original, obteve um PR-AP de 0.37. Enquanto que com o conjunto de dados aumentados, o PR-AP cresceu significativamente para 0.94. Isso indica que com um pequeno número de amostras e classes desbalanceadas do conjunto de dados original, o algoritmo GNB não foi capaz de criar um modelo preditivo útil. Isso pode ter ocorrido devido a simplicidade de modelo, que com pouco volume de dados, não foi capaz de encontrar padrões no conjunto de dados para correta classificação das amostras.

4.3 Resultado comparativo entre os métodos de classificação binária

Uma vez que o aumento de dados melhora o desempenho de todos os classificadores, uma análise quantitativa foi realizada no subconjunto de teste do conjunto aumentado de dados. Como visto na secção 3.4.2, a matriz de confusão consiste em uma avaliação tabular com quatro categorias da performance dos modelos de classificação, quanto maior o número de instância verdadeiras (verdadeiro positivo e negativo) e menor as instâncias falsas (Falso negativo e positivo) melhor o desempenho do modelo em questão. As métricas estatísticas da Tabela 4.4 utilizaram o subconjunto de teste e foram calculadas a partir das matrizes de confusão da Figura 4.3.

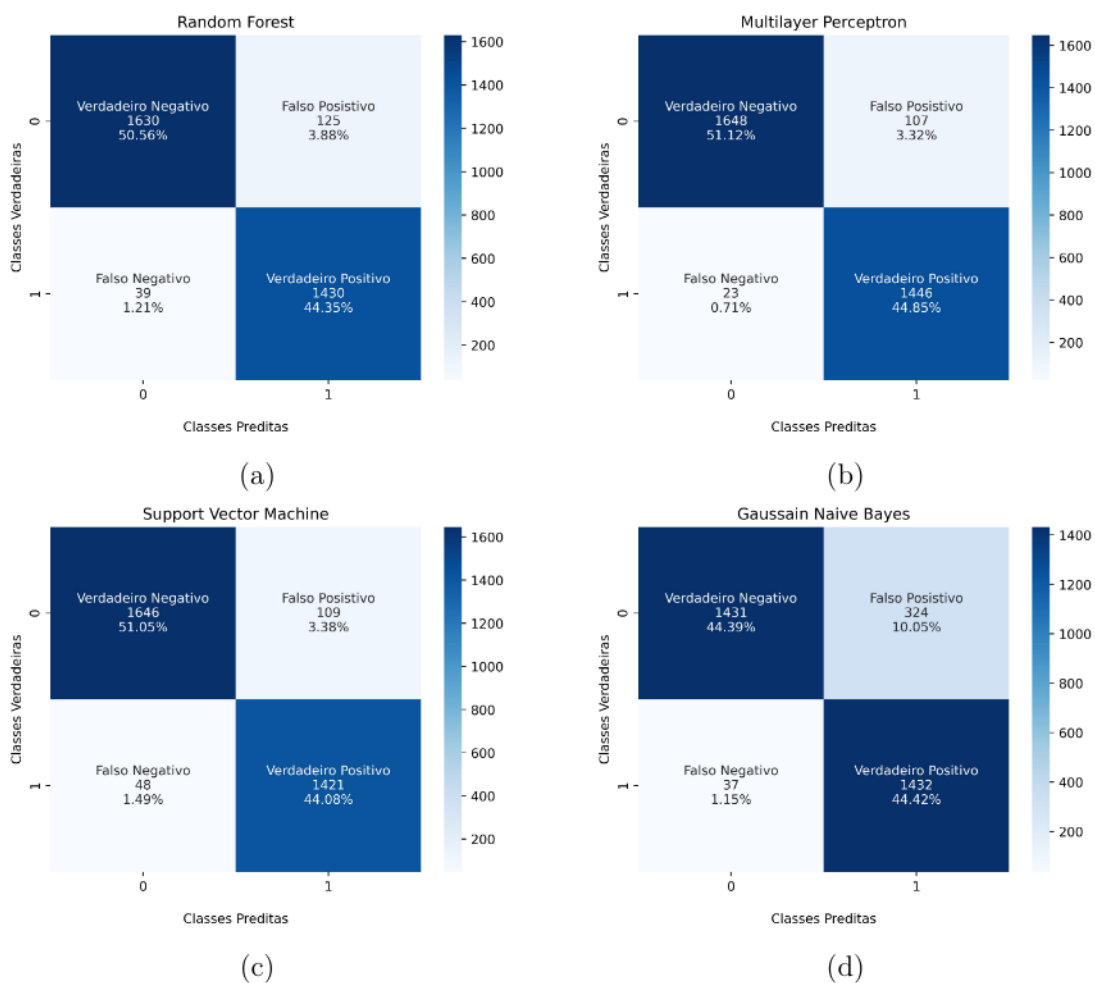


Figura 4.3: Matriz de confusão dos classificadores de aprendizado de máquina sobre subconjunto independente de teste. (a) *Random Forest* b) *Multilayer Perceptron* c) *Support Vector Machine* d) *Gaussian Naive Bayes*.

As métricas estatísticas apresentadas na Tabela 4.4 indicam que RF, MLP e SVM demonstraram um desempenho superior a 92% em todas as métricas propos-

tas. Entre os modelos testados, o algoritmo MLP se destacou, alcançando a maior acurácia (96%), precisão (93.1%), *recall* (98.4%) e *F1 score* (95.7%). Os resultados da matriz de confusão apresentados na Figura 4.3 corroboram a superioridade do MLP em todas as categorias, pois tem maior número nas classes verdadeiras e menor nas falsas.

Tabela 4.4: Resultado das métricas estatísticas sobre subconjunto independente de teste.

Nome do algoritmo	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)	Tempo de Treinamento (s)	Tempo de Teste (s)
RF	94.9	92	97.4	94.6	3.3	0.1
MLP	96	93.1	98.4	95.7	55.2	0.2
SVM	95.1	92.9	96.7	94.8	3.8	0.4
GNB	88.8	81.6	97.5	88.8	0.001	0.001

Vale ressaltar que o MLP também teve o tempo de treinamento mais longo, levando 55.15 segundos para treinar o algoritmo e 0.2 segundos para realizar as previsões das classes no subconjunto de teste. Em comparação, os algoritmos RF e SVM apresentaram tempo de treinamento de 3.3 s e 3.8 s e de teste de 0.1 s e 0.4 s, respectivamente. Sendo que, ambos apresentaram acurácia semelhante ao modelo MLP com 94.9% e 95.1%, respectivamente.

Em contraste, o modelo GNB apresentou a menor acurácia com 88.8%, assim como em todas as outras métricas da Tabela 4.4. Este algoritmo é conhecido por sua eficiência computacional e processamento rápido devido a sua simplicidade. O que resultou no menor tempo de treinamento e teste. No entanto, este modelo se mostrou sensível ao treinar com um conjunto de dados desbalanceados em relação ao número de amostras em cada classe.

Estes resultados estão em consonância com o trabalho anterior de OU *et al.* [24], que utilizou o SVM para a tarefa de classificação e alcançou uma probabilidade de classificação de 92% para sinais acústicos de baleias sei. Além disso, BUCHAN *et al.* [64] aplicaram Modelos de Markov Ocultos e redes neurais profundas para classificar sinais acústicos descendentes de baleias sei como uma categoria, apresentando um *F1 score* de 80.95%. No entanto, é essencial observar que seus estudos avaliaram múltiplas classificações envolvendo outras espécies. Em nosso estudo, focamos em avaliar a presença ou ausência de baleias sei considerando cinco tipos distintos de vocalizações.

Capítulo 5

Conclusões

Neste estudo, introduzimos uma metodologia de detecção de sinais acústicos de baixa frequência e classificação binária das vocalizações da baleia sei. Nosso sistema de detecção, que utiliza técnicas de processamento de imagem, demonstrou sua eficácia na detecção de sinais modulados de baixa frequência. Além disso, a adaptabilidade da cadeia de processamento permite a detecção potencial de várias espécies que produzem sinais acústicos de frequência modulada, com apenas pequenos ajustes.

A incorporação de técnicas de aumento de dados, como a inversão e distorção das máscaras de contorno do sinal, resultou em um melhor desempenho e facilitou o processo de aprendizado dos algoritmos de ML, tornando-a uma ferramenta valiosa para lidar com conjuntos de dados de referência reduzidos e desequilibrados.

Entre os modelos de aprendizado de máquina empregados, os modelos RF, MLP e SVM, apresentaram bom desempenho na tarefa de classificação binária, com todas as métricas de avaliação de performance acima de 92%. O classificador MLP, em particular, demonstrou resultados ligeiramente superiores.

No entanto, o algoritmo atual não aborda o desafio de detecção e classificação de sinais acústicos biológicos que se sobrepõem no tempo. Trabalhos futuros poderiam se concentrar em aprimorar o sistema de detecção para lidar com esses sinais sobrepostos. Além disso, estender a classificação para identificar simultaneamente múltiplos sinais acústicos de baleias de barbatanas seria uma direção valiosa para pesquisas futuras.

Em conclusão, o algoritmo proposto representa um avanço significativo no campo, fornecendo uma abordagem abrangente e eficaz. Considerando as referências pesquisadas para este estudo, este é o primeiro detector automatizado e classificador binário desenvolvido para as cinco classes distintas de vocalizações estereotipadas de baleias sei. O potencial de adaptabilidade a outras espécies e viabilidade do tratamento de conjuntos de dados desequilibrados, o tornam uma ferramenta promissora para pesquisas acústicas de mamíferos marinhos. Ao classificar os sinais de baleias, os pesquisadores podem acessar uma riqueza de informações, como densidade po-

pulacional, distribuição e uso do habitat. Essas informações podem levar a uma melhoria na gestão de áreas de conservação para espécies ameaçadas, cooperando para reduzir o impacto humano sobre os mamíferos marinhos.

Referências Bibliográficas

- [1] CERCHIO, S., WEIR, C. R. “Mid-frequency song and low-frequency calls of sei whales in the Falkland Islands”, *R. Soc. Open Sci.*, v. 9, n. 11, 2022. ISSN: 20545703. doi: 10.1098/rsos.220738.
- [2] BARBOSA, G., BEZERRA, G. M., DE MEDEIROS, D. S., et al. “Segurança em Redes 5G: Oportunidades e Desafios em Detecção de Anomalias e Predição de Tráfego Baseadas em Aprendizado de Máquina”, *Minicursos do XXI Simpósio Bras. Segurança da Informação e Sist. Comput.*, , n. October, pp. 145–189, 2021. doi: 10.5753/sbc.7165.8.4.
- [3] SOCIOAMBIENTAL. *PMC-BS - Relatório anual - Ciclos 1 e 2*. Relatório técnico, 2016.
- [4] MCFADDEN, S. L., SIMMONS, A. M., ERBE, C., et al. “Behavioral and physiological audiometric methods for animals”. In: *Explor. Anim. Behav. Through Sound Vol. 1 Methods*, pp. 355–387, 2022. ISBN: 9783030975401. doi: 10.1007/978-3-030-97540-1_10.
- [5] SIMPKINS, M. “Marine mammals”. In: *Chang. Arct. Backgr. Issues*, pp. 102–110, 2010. ISBN: 9781617610394.
- [6] BROWNING, E., GIBB, R., GLOVER-KAPFER, P., et al. *Passive acoustic monitoring in ecology and conservation*. Relatório Técnico 2, WWF Conservation Technology Series, 2017.
- [7] BEST, P. “Automated Detection and Classification of Cetacean Acoustic Signals”, *HAL Open Sci.*, 2022.
- [8] BITTLE, M., DUNCAN, A. “A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring”. In: *Annu. Conf. Aust. Acoust. Soc. 2013, Acoust. 2013 Sci. Technol. Amenity*, n. November, pp. 208–215, 2013. ISBN: 9781632662682.
- [9] AMBIENTAL, S. *PROJETO DE MONITORAMENTO DE CETÁCEOS NA BACIA DE SANTOS PMC-BS Relatório de Mobilização da 14^ª Campa-*

nha de Avistagem Embarcada e MAP Volume Único. Relatório técnico, 2021.

- [10] ZIMMER, W. M. *Passive acoustic monitoring of cetaceans*. 2011. ISBN: 9780511977107. doi: 10.1017/CBO9780511977107.
- [11] LAMMERS, M. O., AU, W. W. “Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal feature to cue direction of movement?” *Mar. Mammal Sci.*, v. 19, n. 2, pp. 249–264, 2003. ISSN: 08240469. doi: 10.1111/j.1748-7692.2003.tb01107.x.
- [12] JANIK, V. M. “Cetacean vocal learning and communication”, *Curr. Opin. Neurobiol.*, v. 28, pp. 60–65, 2014. ISSN: 18736882. doi: 10.1016/j.conb.2014.06.010. Disponível em: <<http://dx.doi.org/10.1016/j.conb.2014.06.010>>.
- [13] MILMANN, L., SICILIANO, S., MORAIS, I., et al. *A review of Balaenoptera strandings along the east coast of South America*, v. 37. Elsevier B.V., 2020. ISBN: 0000000264. doi: 10.1016/j.rsma.2020.101343. Disponível em: <<https://doi.org/10.1016/j.rsma.2020.101343>>.
- [14] PRIETO, R., JANIGER, D., SILVA, M. A., et al. “The forgotten whale: A bibliometric analysis and literature review of the North Atlantic sei whale *Balaenoptera borealis*”, *Mamm. Rev.*, v. 42, n. 3, pp. 235–272, 2012. ISSN: 13652907. doi: 10.1111/j.1365-2907.2011.00195.x.
- [15] NIEUKIRK, S. L., MELLINGER, D. K., DZIAK, R. P., et al. “Multi-year occurrence of sei whale calls in North Atlantic polar waters”, *J. Acoust. Soc. Am.*, v. 147, n. 3, pp. 1842–1850, 2020. ISSN: 0001-4966. doi: 10.1121/10.0000931.
- [16] CHRISTENSEN, I., HAUG, T., OIEN, N. “A review of feeding and reproduction in large baleen whales (Mysticeti) and sperm whales *Physeter macrocephalus* in Norwegian and adjacent waters”, *Fauna Nor. Ser. A*, v. 13, n. January, pp. 39–48, 1992.
- [17] PERRY, S. L., DEMASTER, D. P., SILBER, G. K. “The great whales: History and status of six species listed as endangered under the U.S. Endangered Species Act of 1973”, *Mar. Fish. Rev.*, v. 61, n. 1, pp. 1–74, 1999. ISSN: 00901830.
- [18] HEISSLER, V., FRAINER, G., SICILIANO, S. “Sei whale , *Balaenoptera borealis* , in the South-Western Atlantic Ocean : the discovery of a calving

ground in Brazilian waters”, *Rep. Int. Whal. Comm.*, v. SC/66b/SH2, n. January 2018, pp. 1–11, 2016.

- [19] RANKIN, S., BARLOW, J. “Vocalizations of the sei whale *Balaenoptera borealis* off the hawaiian islands”, *Bioacoustics*, v. 16, n. 2, pp. 137–145, 2007. ISSN: 21650586. doi: 10.1080/09524622.2007.9753572.
- [20] MELLINGER, D. K., MARTIN, S. W., MORRISSEY, R. P., et al. “A method for detecting whistles, moans, and other frequency contour sounds”, *J. Acoust. Soc. Am.*, v. 129, n. 6, pp. 4055–4061, 2011. ISSN: 0001-4966. doi: 10.1121/1.3531926.
- [21] MELLINGER, D. K., CLARK, C. W. “MobySound: A reference archive for studying automatic recognition of marine mammal sounds”, *Appl. Acoust.*, v. 67, n. 11-12, pp. 1226–1242, 2006. ISSN: 0003682X. doi: 10.1016/j.apacoust.2006.06.002.
- [22] BAUMGARTNER, M. F., MUSSOLINE, S. E. “A generalized baleen whale call detection and classification system”, *J. Acoust. Soc. Am.*, v. 129, n. 5, pp. 2889–2902, 2011. ISSN: 0001-4966. doi: 10.1121/1.3562166.
- [23] HALKIAS, X. C., PARIS, S., GLOTIN, H. “Classification of mysticete sounds using machine learning techniques”, *J. Acoust. Soc. Am.*, v. 134, n. 5, pp. 3496–3505, 2013. ISSN: 0001-4966. doi: 10.1121/1.4821203.
- [24] OU, H., AU, W. W. L., VAN PARIJS, S., et al. “Discrimination of frequency-modulated Baleen whale downsweep calls with overlapping frequencies”, *J. Acoust. Soc. Am.*, v. 137, n. 6, pp. 3024–3032, 2015. ISSN: 0001-4966. doi: 10.1121/1.4919304. Disponível em: <<http://dx.doi.org/10.1121/1.4919304>>.
- [25] ESPAÑOL-JIMÉNEZ, S., BAHAMONDE, P. A., CHIANG, G., et al. “Discovering sounds in Patagonia: Characterizing sei whale (*Balaenoptera borealis*) downsweeps in the south-eastern Pacific Ocean”, *Ocean Sci.*, v. 15, n. 1, pp. 75–82, 2019. ISSN: 18120792. doi: 10.5194/os-15-75-2019.
- [26] DOMBROSKI, J. R. G., PARKS, S. E., GROCH, K. R., et al. “Vocalizations produced by southern right whale (*Eubalaena australis*) mother-calf pairs in a calving ground off Brazil”, *J. Acoust. Soc. Am.*, v. 140, n. 3, pp. 1850–1857, 2016. ISSN: 0001-4966. doi: 10.1121/1.4962231. Disponível em: <<http://dx.doi.org/10.1121/1.4962231>>.

- [27] ŠIROVIĆ, A., OLESON, E. M., BUCCOWICH, J., et al. “Fin whale song variability in southern California and the Gulf of California”, *Sci. Rep.*, v. 7, n. 1, pp. 1–11, 2017. ISSN: 20452322. doi: 10.1038/s41598-017-09979-4.
- [28] ESFAHANIAN, M. *DETECTION AND CLASSIFICATION OF MARINE MAMMAL SOUNDS*. Tese de Doutorado, 2014.
- [29] BAUMGARTNER, M. F., VAN PARIJS, S. M., WENZEL, F. W., et al. “ Low frequency vocalizations attributed to sei whales (*Balaenoptera borealis*) ”, *J. Acoust. Soc. Am.*, v. 124, n. 2, pp. 1339–1349, 2008. ISSN: 0001-4966. doi: 10.1121/1.2945155.
- [30] THOMPSON, P. O., FINDLEY, L. T., VIDAL, O. “ 20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico ”, *J. Acoust. Soc. Am.*, v. 92, n. 6, pp. 3051–3057, 1992. ISSN: 0001-4966. doi: 10.1121/1.404201.
- [31] KNOWLTON, A., CLARK, C. W., KRAUS, S. “ Sounds recorded in the presence of sei whale, *Balaenoptera borealis* ”, *J. Acoust. Soc. Am.*, v. 89, n. 4B, pp. 1968–1968, 1991. ISSN: 0001-4966. doi: 10.1121/1.2029710.
- [32] MCDONALD, M. A., HILDEBRAND, J. A., WIGGINS, S. M., et al. “Sei whale sounds recorded in the Antarctic”, *J. Acoust. Soc. Am.*, v. 118, n. 6, pp. 3941–3945, 2005. ISSN: 0001-4966. doi: 10.1121/1.2130944.
- [33] GEDAMKE, J., ROBINSON, S. M. “Acoustic survey for marine mammal occurrence and distribution off East Antarctica (30-80°E) in January-February 2006”, *Deep. Res. Part II Top. Stud. Oceanogr.*, v. 57, n. 9-10, pp. 968–981, 2010. ISSN: 09670645. doi: 10.1016/j.dsr2.2008.10.042. Disponível em: <<http://dx.doi.org/10.1016/j.dsr2.2008.10.042>>.
- [34] TREMBLAY, C. J., VAN PARIJS, S. M., CHOLEWIAK, D. “ 50 to 30-Hz triplet and singlet down sweep vocalizations produced by sei whales (*Balaenoptera borealis*) in the western North Atlantic Ocean ”, *J. Acoust. Soc. Am.*, v. 145, n. 6, pp. 3351–3358, 2019. ISSN: 0001-4966. doi: 10.1121/1.5110713.
- [35] CALDERAN, S., MILLER, B., COLLINS, K., et al. “ Low-frequency vocalizations of sei whales (*Balaenoptera borealis*) in the Southern Ocean ”, *J. Acoust. Soc. Am.*, v. 136, n. 6, pp. EL418–EL423, 2014. ISSN: 0001-4966. doi: 10.1121/1.4902422.

- [36] GILLESPIE, D. “Detection and classification of right whale calls using an ‘edge’ detector operating on a smoothed spectrogram”, *Can. Acoust.*, v. 32, n. 2, pp. 39–47, 2004. ISSN: 07116659.
- [37] SÁNCHEZ-GARCÍA, A., MUÑOZ-ESPARZA, P., SANCHO-GOMEZ, J. L. “A novel image-processing based method for the automatic detection, extraction and characterization of marine mammal tonal calls”, *J. Mar. Biol. Assoc. United Kingdom*, v. 90, n. 8, pp. 1667–1684, 2010. ISSN: 14697769. doi: 10.1017/S0025315409000927.
- [38] MOHAMMAD, B., MCHUGH, R. “Automatic detection and characterization of dispersive north atlantic right whale upcalls recorded in a shallow-water environment using a region-based active contour model”, *IEEE J. Ocean. Eng.*, v. 36, n. 3, pp. 431–440, 2011. ISSN: 03649059. doi: 10.1109/JOE.2010.2060790.
- [39] BUCHAN, S. J., GUTIÉRREZ, L., BAUMGARTNER, M. F., et al. “Distribution of blue and sei whale vocalizations, and temperature - salinity characteristics from glider surveys in the Northern Chilean Patagonia mega-estuarine system”, *Front. Mar. Sci.*, v. 9, n. October, pp. 1–13, 2022. ISSN: 22967745. doi: 10.3389/fmars.2022.903964.
- [40] PRATI, R. C., BATISTA, G. E., SILVA, D. F. “Class imbalance revisited: a new experimental setup to assess the performance of treatment methods”, *Knowl. Inf. Syst.*, v. 45, n. 1, pp. 247–270, 2015. ISSN: 02193116. doi: 10.1007/s10115-014-0794-3.
- [41] GONZALES, R. C., WOODS, R. E. *Processamento Digital De-Imagens*. 3ed ed. São Paulo, Person, 2010. ISBN: 9788581435862.
- [42] SZELISKI, R. *Computer Vision Algorithms and Applications*. Seattle, USA, Springer US, 2022.
- [43] SUZUKI, S., BE, K. A. “Topological structural analysis of digitized binary images by border following”, *Comput. Vision, Graph. Image Process.*, v. 30, n. 1, pp. 32–46, 1985. ISSN: 0734189X. doi: 10.1016/0734-189X(85)90016-7.
- [44] CAI, W., ZHU, J., ZHANG, M., et al. “A Parallel Classification Model for Marine Mammal Sounds Based on Multi-Dimensional Feature Extraction and Data Augmentation”, *Sensors*, v. 22, n. 19, 2022. ISSN: 14248220. doi: 10.3390/s22197443.

- [45] TAYLOR, L., NITSCHKE, G. “Improving Deep Learning with Generic Data Augmentation”, *Proc. 2018 IEEE Symp. Ser. Comput. Intell. SSCI 2018*, pp. 1542–1547, 2019. doi: 10.1109/SSCI.2018.8628742.
- [46] BLOICE, M. D., ROTH, P. M., HOLZINGER, A. “Biomedical image augmentation using Augmentor”, *Bioinformatics*, v. 35, n. 21, pp. 4522–4524, 2019. ISSN: 14602059. doi: 10.1093/bioinformatics/btz259.
- [47] WITTEN, I. H., FRANK, E., HALL, M. A. *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*. 4rd ed. Cambridge, United States, The Morgan Kaufmann Publishers, 2011. ISBN: 9780123748560. doi: 10.1016/C2009-0-19715-5.
- [48] BIANCO, M. J., GERSTOFT, P., TRAER, J., et al. “Machine learning in acoustics: Theory and applications”, *J. Acoust. Soc. Am.*, v. 146, n. 5, pp. 3590–3628, 2019. ISSN: 0001-4966. doi: 10.1121/1.5133944.
- [49] ESCOVEDO, T., KOSHIYAMA, A. *Introdução a Data Science Algoritmos de Machine Learning e Métodos de Análise by Tatiana Escovedo Adriano Koshiyama (z-lib.org) (1)*. Casa do Código, 2020. ISBN: 9788572540544.
- [50] OSISANWO, F., AKINSOLA, J., AWODELE, O., et al. “Supervised Machine Learning Algorithms: Classification and Comparison”, *Int. J. Comput. Trends Technol.*, v. 48, n. 3, pp. 128–138, 2017. doi: 10.14445/22312803/ijctt-v48p126.
- [51] ZELINSKY, A. *Learning OpenCV—Computer Vision with the OpenCV Library*, v. 16. Sebastopol, US, O’Reilly Media, Inc., 2009. doi: 10.1109/MRA.2009.933612.
- [52] PEDREGOSA FABIAN, MICHEL, V., VAROQUAUX, G., et al. “Scikit-learn: Machine Learning in Python”, *J. Mach. Learn. Res.*, v. 12, pp. 2825–2830, sep 2011. Disponível em: <<http://scikit-learn.sourceforge.net>>.
- [53] JOHNSON, M. P., TYACK, P. L. “A digital acoustic recording tag for measuring the response of wild marine mammals to sound”, *IEEE J. Ocean. Eng.*, v. 28, n. 1, pp. 3–12, 2003. ISSN: 03649059. doi: 10.1109/JOE.2002.808212.
- [54] BUADES, A., COLL, B., MOREL, J.-M. “Non-Local Means Denoising”, *Image Process. Line*, v. 1, pp. 208–212, 2011. doi: 10.5201/ipol.2011.bcm_nlm.
- [55] RASMUSSEN, J. H., ŠIROVIĆ, A. “Automatic detection and classification of baleen whale social calls using convolutional neural networks”, *J. Acoust.*

Soc. Am., v. 149, n. 5, pp. 3635–3644, 2021. ISSN: 0001-4966. doi: 10.1121/10.0005047.

- [56] ARFKEN, G. B., WEBER, H. J., HARRIS, F. E. *Mathematical Methods for Physicists, Sixth Edition: A Comprehensive Guide*. 2005. ISBN: 0120598760.
- [57] SHAPIRO, S. S., WILK, M. B. “An Analysis of Variance Test for Normality (Complete Samples)”, *Biometrika*, v. 52, n. 3/4, pp. 591, 1965. ISSN: 00063444. doi: 10.2307/2333709.
- [58] LIMA, I. M. S., ANDRADE, L. G., BITTENCOURT, L., et al. “Whistle comparison of four delphinid species in Southeastern Brazil”, *J. Acoust. Soc. Am.*, v. 139, n. 5, pp. EL124–EL127, 2016. ISSN: 0001-4966. doi: 10.1121/1.4947310.
- [59] DELARUE, J., MARTIN, B., MOUY, X., et al. *Northeastern Chukchi Sea, Joint Acoustic Monitoring Program 2009–2010*. Relatório Técnico December, JASCO, Dartmouth, Nova Scotia, 2011.
- [60] HILDEBRAND, J. A., FRASIER, K. E., HELBLE, T. A., et al. “Performance metrics for marine mammal signal detection and classification”, *J. Acoust. Soc. Am.*, v. 151, n. 1, pp. 414–427, 2022. ISSN: 0001-4966. doi: 10.1121/10.0009270.
- [61] GARCIA, H. A., COUTURE, T., GALOR, A., et al. “Comparing performances of five distinct automatic classifiers for fin whale vocalizations in beam-formed spectrograms of coherent hydrophone array”, *Remote Sens.*, v. 12, n. 2, pp. 1–25, 2020. ISSN: 20724292. doi: 10.3390/rs12020326.
- [62] DAVIS, J., GOADRICH, M. “The relationship between precision-recall and ROC curves”, *ACM Int. Conf. Proceeding Ser.*, v. 148, pp. 233–240, 2006. doi: 10.1145/1143844.1143874.
- [63] FLACH, P. A., KULL, M. “Precision-Recall-Gain curves: PR analysis done right”, *Adv. Neural Inf. Process. Syst.*, v. 2015-Janua, pp. 838–846, 2015. ISSN: 10495258.
- [64] BUCHAN, S. J., DURAN, M., ROJAS, C., et al. “An HMM-DNN-Based System for the Detection and Classification of Low-Frequency Acoustic Signals from Baleen Whales, Earthquakes, and Air Guns off Chile”, *Remote Sens.*, v. 15, n. 10, 2023. ISSN: 20724292. doi: 10.3390/rs15102554.