

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

**BIOSSENSORES BASEADOS EM MARISCOS
AMARILLADESMA MACTROIDES E VISÃO
COMPUTACIONAL**

por

Je Nam Jun Junior

Dissertação para obtenção do Título de
Mestre em Modelagem Computacional

Rio Grande, julho, 2021

Je Nam Jun Junior

“Biossensores baseados em mariscos *Amarilladesma mactroides* e Visão Computacional”

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como requisito parcial para obtenção do Grau de Mestre. Área de concentração: Modelagem Computacional.

Aprovada em: 01/09/2022

BANCA EXAMINADORA

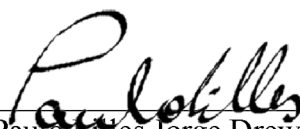


Profª. Dra. Silvia Silva da Costa Botelho (Orientadora – FURG)

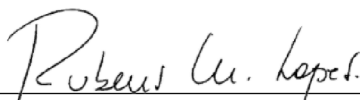


Documento assinado digitalmente
LEONARDO RAMOS EMMENDORFER
Data: 11/10/2022 08:52:16-0300
Verifique em <https://verificador.iti.br>

Prof. Dr. Leonardo Ramos Emmendorfer (UFSM/PPGMC-FURG)



Prof. Dr. Paulo Lúis Jorge Drews Junior (FURG)



Prof. Dr. Rubens Mendes Lopes (USP)

Rio Grande - RS
2022

Ficha Catalográfica

J95b Jun Junior, Je Nam.
 Biossensores baseados em mariscos *Amarilladesma mactroides* e
 visão computacional / Je Nam Jun Junior. – 2022.
 58 f.

 Dissertação (mestrado) – Universidade Federal do Rio Grande –
 FURG, Programa de Pós-Graduação em Modelagem Computacional,
 Rio Grande/RS, 2022.
 Orientadora: Dra. Silvia Silva da Costa Botelho.

 1. Bivalves 2. Visão Computacional 3. Análise Comportamental
 4. Classificadores I. Botelho, Silvia Silva da Costa II. Título.

CDU 564.185:004

Catálogo na Fonte: Bibliotecário José Paulo dos Santos CRB 10/2344

AGRADECIMENTOS

Agradeço à FURG por ter aberto as portas da ciência e da ética para mim. Minha formação superior só se deu graças ao ensino público de qualidade que as instituições federais são capazes de proporcionar.

Agradeço à professora Silvia pela orientação e toda ajuda necessária durante o mestrado. Da mesma forma, agradeço ao Adriano, ao Rafael e à professora Juliana pela ajuda na realização de experimentos e nos aspectos biológicos desse trabalho.

Agradeço ao Projeto ASTRAL pelo financiamento durante o período e por todo o suporte da equipe no desenvolvimento de tecnologias para aquacultura.

Agradeço à Bruna, por ser minha companheira em todos os momentos e todas ocasiões. Estar comigo nessa jornada permitiu que tudo isso fosse possível.

Por fim, agradeço aos meus amigos, que sempre se mantiveram tão perto mesmo estando tão longe. Agradeço aos que estão comigo e aos que partiram cedo demais. Obrigado.

RESUMO

O monitoramento de ambientes aquáticos tem importância primordial na garantia da qualidade da água para consumo e suporte à atividades econômicas bem como preservação de organismos e ecossistemas complexos. Biossensores baseados na análise comportamental de moluscos bivalves tem se destacado como potenciais ferramentas neste contexto. Entretanto, tecnologias atualmente disponíveis para a construção de biossensores não consideram padrões comportamentais inerentes aos mariscos, bivalves de grande importância socio-econômica, tendo em vista sua natureza móvel. Diante desse panorama, o presente trabalho propõe averiguar a possibilidade de construção de biossensores com base em mariscos *Amarilladesma mactroides* e parâmetros comportamentais inéditos. Um sistema biossensor é proposto com base na aquisição de dados, segmentação e calibração do sistema e inferência. Dois modelos de visão computacional são avaliados para extração de dados comportamentais através de segmentação de imagens experimentais, atingindo F_1 -score acima de 0,93. Através de métodos estatísticos, a atividade de exploração de mariscos foi identificada como um potencial *proxy* para monitoramento de temperatura e poluição por partículas nocivas. O uso de classificadores permitiu que modelos fossem treinados para inferência de condições ambientais com satisfatória performance, indicando viabilidade dos parâmetros propostos enquanto *proxies*. Através dos parâmetros comportamentais de mariscos propostos, foi possível calibrar um sistema biossensor para monitoramento de condições ambientais adversas.

Palavras-chaves: Bivalves, Visão Computacional, Análise Comportamental, Classificadores

ABSTRACT

Monitoring of aquatic environments has primordial importance in ensuring water quality for human use, to support economical activities and preserving organisms in complex ecosystems. Biosensors based on the behavioural analysis of bivalve mollusks has been highlighted as potential tools in this context. However, available technologies for biosensor construction do not take the particular behaviour of clams' mobile tissues into account. In this sense, the present work proposes to verify the possibility of constructing biosensors based on *Amarilladesma mactroides* clams and behavioural parameters. A biosensor system is proposed based in data acquisition, segmentation and calibration. Two computer vision models are evaluated for extracting behavioural data through the image segmentation of experimental data sets, achieving F₁-score over 0.93. By employing statistical methods, the exploration activity of clams was identified as a proxy of temperature variations and chemical pollution. Trained classifiers were assigned for inferring environmental conditions with satisfactory performance, indicating that the proposed parameters may be used as proxies. The use of the proposed behavioural parameters of clams has enabled the calibration of a biosensor system for the monitoring of adverse environmental conditions.

Keywords: Bivalves, Computer Vision, Behavioural Analysis, Classifiers

ÍNDICE

ÍNDICE	5
Lista de Figuras	7
Lista de Tabelas	10
1 INTRODUÇÃO	12
1.1 Justificativa	12
2 REVISÃO BIBLIOGRÁFICA	15
2.1 Moluscos Bivalves	15
2.1.1 <i>Amarilladesma mactroides</i>	15
2.1.2 Análise Comportamental	16
2.2 Medição Baseada em Visão Computacional	17
2.2.1 Calibração de Câmeras	17
2.2.2 Segmentação de Imagens	18
2.3 Classificadores	18
2.3.1 <i>K-Nearest Neighbors</i>	19
2.3.2 <i>Support Vector Machines</i>	20
2.4 Trabalhos Relacionados	21
2.5 Objetivo Geral	24
2.6 Objetivos Específicos	24
3 METODOLOGIA	25
3.1 Modelagem do Biossensor	25
3.2 Aquisição e Pré-Processamento de Imagens	26
3.2.1 Experimento 1	27
3.2.2 Experimento 2	27
3.2.3 Experimento 3	28
3.2.4 Pré-Processamento	28
3.3 Segmentação	29
3.3.1 Modelo de Segmentação com Algoritmo <i>Watershed</i>	30
3.3.2 Modelo de Segmentação baseado em Aprendizado de Máquina	31
3.4 Métricas de Avaliação	33
3.4.1 Curvas ROC	35
3.5 Calibração e Inferência	36
4 RESULTADOS	38
4.1 Aquisição e Pré-Processamento de Imagens	38
4.2 Segmentação	39
4.2.1 Modelo de Segmentação com Algoritmo <i>Watershed</i>	39
4.2.2 Modelo de Segmentação Baseado em Aprendizado de Máquina	40
4.2.3 Comparação dos Modelos	41
4.3 Calibração e Inferência	42

4.3.1	Avaliação da Temperatura	43
4.3.2	Avaliação de Contaminante	45
4.4	Discussões	47
5	CONCLUSÃO	51
6	REFERÊNCIAS	53

LISTA DE FIGURAS

Figura 2.1: Mariscos <i>Amarilladesma mactroides</i> em habitat natural com órgãos expostos, no qual (A) é o pé, (B) é o sifão exalante e (C) é o sifão inalante. (MARTIN, 2019, Adaptado).	16
Figura 2.2: Exemplo de segmentação semântica e de instância em imagens de microalgas, respectivamente. Na primeira imagem, há a detecção das partículas como pertencentes a um mesmo grupo, enquanto na segunda imagem há a categorização de cada ocorrência com uma cor distinta (SANTAQUITERIA et al., 2020).	18
Figura 2.3: Exemplo de discriminação linear bidimensional, no qual um hiperplano linear (representado pela linha tracejada) é transformado em um hiperplano bidimensional (representado pela parábola), permitindo a separação de suas observações em duas categorias (KUNCHEVA, 2014).	20
Figura 2.4: Exemplo gráfico da aplicação de SVM com seus elementos característicos, no qual o hiperplano ótimo está representado pela linha tracejada, as margens ótimas por linhas contínuas e os vetores de suporte como a cota entre as margens (CORTES; VAPNIK, 1995).	20
Figura 3.1: Configuração básica para aquisição de imagens de mariscos, constituída por uma câmera para imageamento, um reservatório fechado para acomodação dos animais, fundo escurecido para realce dos indivíduos nas observações e mariscos dispostos pela superfície.	26
Figura 3.2: <i>Pipeline do modelo proposto composto pelas etapas de aquisição e pré-processamento de imagens, segmentação dos mariscos, calibração do modelo biossensor e inferência de condições ambientais.</i>	26
Figura 3.3: Configuração do experimento 1 realizado em duplicata, contendo as seguintes concentrações de DCOIT: 0 mg/L (Controle), 0,01 mg/L (Tratamento 1), 0,13 mg/L (Tratamento 2).	28
Figura 3.4: Configuração do experimento 2 realizado em duplicata, contendo dois grupos em diferentes temperaturas: 20 °C e 17 °C.	28
Figura 3.5: Efeito de alagamento em bacia hidrográfica. A constante e uniforme queda das chuvas sobre a paisagem provoca o escoamento da água pelas diferentes regiões de captação e, conseqüentemente, o enchimento dos vales em diferentes tempos. O fenômeno serve como base para o algoritmo <i>watershed</i> , que identifica regiões de mesma intensidade dentro da imagem. (PREIM; BOTHA, 2014, Adaptado).	30
Figura 3.6: Esquematização de uma rede neural artificial. O processamento e transmissão individual de cada neurônio simples gera complexas redes multica- madas capazes de aprender padrões (ELGENDY, 2019, Adaptado).	32

Figura 3.7: Funcionamento de uma camada convolucional de uma rede neural. A imagem de entrada é lida como uma matriz e todas posições possíveis são varridas através de um filtro. O resultado da operação específica desse filtro permite que a região varrida possa ser compactada em um mapa de características de dimensão menor. (BROWNLEE, 2019, Adaptado). . . .	33
Figura 3.8: Funcionamento das camada de operação <i>pooling</i> , na qual cada sub-região pode ser resumida ao máximo valor apresentado ou pela média de todos valores contidos na sub-matriz (SAHA, 2018, Adaptado).	34
Figura 3.9: Estrutura da rede U-Net, composta por seções simétricas de convolução e deconvolução (RONNEBERGER; FISCHER; BROX, 2015).	35
Figura 3.10: A matriz de confusão relata o número de ocorrências de cada caso possível em uma classificação binária: ocorrências negativas classificadas corretamente (TN) ou erroneamente (FN) e ocorrências positivas classificadas corretamente (TP) e erroneamente (FP). Através desses quatro parâmetros, é possível calcular métricas de avaliação do modelo como a acurácia.	35
Figura 3.11: Exemplo de curva com ROC com área sob a curva (AUC) igual à 0,65. Linha em vermelho ilustra a curva de um classificador com saídas totalmente aleatórias.	36
Figura 4.1: Comparação entre <i>frame</i> extraído do vídeo com suas ROIs separadas, respectivamente. O procedimento transforma uma imagem de alta resolução (1920x1080p) em duas imagens de menor resolução (480x320p) além de corrigir a perspectiva e a distorção intrínseca da câmera empregada.	38
Figura 4.2: Comparação entre ROI extraída, imagem anotada e <i>ground truth</i> binarizado, respectivamente	39
Figura 4.3: Comparação entre histograma de imagem antes e depois de pré-processamento. A aplicação dessas operações causa o escurecimento da imagem, representada pelo deslocamento para a esquerda do gráfico, uniformizando o fundo preto.	39
Figura 4.4: ROI extraída antes e depois de pré-processamento	40
Figura 4.5: Exemplo de saída do algoritmo <i>watershed</i> utilizando imagens sem e com pré processamento, respectivamente. No primeiro caso, observa-se dificuldade de modelo em identificar todos os mariscos, enquanto na imagem tratada há a identificação das bordas de cada animal presente. . .	40
Figura 4.6: Performance do modelo ao longo das épocas de treinamento e validação, atingindo valor máximo de F ₁ -score (0,956) na décima época	41
Figura 4.7: Performance do modelo com <i>data set</i> sem pré-processamento, que atinge F ₁ -score máximo (0,968) na décima época.	42

Figura 4.8: F ₁ -score obtido por cada modelo de segmentação para cada intervalo de leitura. De forma geral, as performances entre modelos são próximas com exceção dos valores de F ₁ -score médio para Tratamento 1, no qual a rede U-Net sem pré-processamento nos dados atinge resultados melhores em comparação com os demais modelos.	43
Figura 4.9: Área ocupada por mariscos durante o experimento 2-A	43
Figura 4.10: Área ocupada por mariscos durante o experimento 2-B	44
Figura 4.11: Área explorada por mariscos durante o experimento 2-A	44
Figura 4.12: Área explorada por mariscos durante o experimento 2-B	44
Figura 4.13: Matriz de confusão e curvas ROC para o <i>data set</i> de temperatura. Os valores de AUC-ROC obtidos variam entre 0,91 e 0,93 indicando ótima performance dos três classificadores em inferir o grupo temperatura em novas observações.	46
Figura 4.14: Área ocupada por mariscos durante o experimento 3 frente a diferentes concentrações de DCOIT	47
Figura 4.15: Área explorada por mariscos durante o experimento 3 frente a diferentes concentrações de DCOIT	47
Figura 4.16: Séries de contaminante no domínio da frequência nos períodos 24 e 48 horas, respectivamente	48
Figura 4.17: Matriz de confusão e curvas ROC para o <i>data set</i> de contaminante. Os três classificadores propostos indicam performance adequada para inferir novas observações, apesar do maior número de classificações erradas em comparação com os resultados obtidos para temperatura. A causa dessa maior taxa de erros pode estar relacionada principalmente com o comportamento similar dos animais nos grupos Controle e Tratamento 1 durante as primeiras 24h.	49

LISTA DE TABELAS

Tabela 1:	Performance do modelo <i>Watershed</i>	41
Tabela 2:	F ₁ -score médio para cada aquário	42
Tabela 3:	Teste ANOVA ($\alpha = 5\%$) para área explorada nos <i>data sets</i> de temperatura	45
Tabela 4:	Acurácias atingidas para o <i>data set</i> de temperatura	45
Tabela 5:	Testes ANOVA ($\alpha = 5\%$) para <i>data set</i> DCOIT	46
Tabela 6:	Testes post-hoc para <i>data set</i> DCOIT	48
Tabela 7:	Acurácias atingidas para o <i>data set</i> de contaminante	49

LISTA DE ABREVIATURAS

FN	Falso negativo
FP	Falso positivo
pH	Potencial hidrogeniônico
KNN	<i>K-nearest neighbors</i>
RNA	Rede neural artificial
ROI	Região de interesse
SVM	<i>Support vector machines</i>
TN	Verdadeiro negativo
TP	Verdadeiro positivo
VBM	Medição baseada em visão

1 INTRODUÇÃO

Diversos Objetivos de Desenvolvimento Sustentável tem incentivado o desenvolvimento de sistemas de monitoramento aquático capazes de considerar a ampla variedade de perturbações possivelmente presentes nestes ambientes. Garantir a disponibilidade e a qualidade da água tem sido um desafio ao redor do mundo. O monitoramento aquático é uma necessidade crescente desde a água potável até o esgoto proveniente de cidades e indústrias (DONG et al., 2015). Para suprir essa demanda, técnicas de monitoramento inteligentes tem sido desenvolvidas para obter respostas rápidas frente a condições aquáticas indesejadas em rios, lagos, mares e oceanos (NAJAH et al., 2011).

A poluição destes corpos d'água pode se dar em diferentes formas. Por exemplo, o descarte de material orgânico pode levar ao desbalanceamento de compostos químicos nitrogenados (NAYAR, 2020). Por outro lado, o aumento da temperatura por descarga de correntes quentes pode reduzir a concentração de oxigênio dissolvido vital para a sobrevivência de organismos (VALLERO, 2019).

Diante da complexidade desses ecossistemas, a alteração de um parâmetro físico-químico pode ocasionar um efeito cascata de reações, atingindo organizações biológicas e cadeias tróficas (VALLERO, 2019). Metodologias de monitoramento para tais ambientes podem ser complexas e custosas devido à grande variedade de condições inadequadas para a vida. Nesse contexto, se evidencia a necessidade de soluções adequadas para realizar o monitoramento aquático tendo em vista esse conjunto de perturbações complexas. Com essa finalidade, os biossensores aparecem como alternativas sensíveis e subletais capaz de emitir sinais em condições ambientais adversas como *proxy* de poluições diversas.

1.1 Justificativa

O uso de biossensores para detecção de poluentes é uma alternativa que vem recebendo cada vez mais atenção. Seu funcionamento baseia-se na detecção de alterações de sinais biológicos em vários níveis (bioquímico, fisiológico ou comportamental) relacionados à exposição e efeitos tóxicos promovidos por compostos químicos e outras situações físico-químicas diversas. Eles permitem, dessa forma, a identificação de variações na organização biológica frente a um ambiente controlado. Dentre suas vantagens destaca-se a capacidade de biossensores em prover resultados confiáveis, fácil operação e baixo custo em comparação com outros métodos de detecção de poluentes (NEWTON; COPE, 2006).

Para a obtenção de sinais, os biossensores requerem um elemento biológico em sua construção, o qual será submetido ao ambiente controlado e ao ambiente contaminado. Moluscos bivalves são animais considerados adequados para essa aplicação, visto que seu comportamento está fortemente atrelado ao estresse toxicológico no ambiente em que está inserido (NEWTON; COPE, 2006). Dessa forma, utilizando-os em conjunto com métodos apropriados pode-se verificar como seus sinais biológicos refletem a qualidade da água presente (MORGAN et al.,

2005).

Por serem animais filtradores, os moluscos bivalves quando sujeitos à ambientes contaminados podem reter poluentes em seu interior como consequência de sua alimentação e respiração. A bioacumulação desses compostos pode ser observada através da análise dos tecidos dos animais, porém esse procedimento acarreta na morte dos indivíduos observados e exige longos períodos de exposição ao contaminante (KRAMER; JENNER; ZWART, 1989). Métodos baseados na bioacumulação de contaminantes são comumente empregados no monitoramento ambiental de locais de interesse (SCHÄFER et al., 2015). Entretanto, o monitoramento efetivo capaz de prover sinais de aviso prévio quanto à contaminação local exige o monitoramento em tempo real destes ambientes (GEETHA; GOUTHAMI, 2017). Biossensores destacam-se, portanto, como ferramentas sensíveis e subletais para o monitoramento de ambientes aquáticos.

Formas não-invasivas para emprego de moluscos bivalves como biossensores têm sido utilizadas por apresentarem respostas rápidas e permitirem a automação da aquisição de dados (LYUBIMTSEV; KHOLODKOVICH; DRUZHININ, 2019). Essas metodologias baseiam-se em sua análise comportamental através do monitoramento da amplitude de abertura de suas valvas (abertura e fechamento) ao longo do tempo (NEWTON; COPE, 2006; ANDRADE et al., 2016). Técnicas de valvometria permitem a aquisição de dados comportamentais e comumente baseiam-se na anexação de dispositivos nas conchas dos bivalves, de forma que seu sinal de saída varie de acordo com seu grau de abertura a cada instante de tempo. Sensores de efeito Hall (HARTMANN et al., 2015) em conjunto com ímãs são uma opção que vem demonstrando bons resultados quando aplicados em mexilhões (GUTERRES, 2021) e ostras (RANA et al., 2020). Em bivalves que se movimentam, como os mariscos, também é possível relacionar o estresse toxicológico com o grau de exposição de seus membros locomotores (BAMBER; WESTERLUND, 2016; NUÑEZ et al., 2014). Entretanto, técnicas para aquisição de dados comportamentais de mariscos através do movimento de seus sifões e pés ainda não estão disponíveis na literatura relacionada.

Sistemas de medição baseados em visão (*Vision Based Measurements - VBM*) são conjuntos de processos que visam obter determinadas características de uma cena física a partir do seu imageamento (HEIJDEN, 1994). Esses sistemas são compostos de etapas que compreendem a adequação das imagens geradas, extração das regiões de interesse, uso de inteligência computacional para identificação dos parâmetros a serem medidos e sua posterior medição com algoritmos apropriados (SHIRMOHAMMADI; FERRERO, 2014). O desenvolvimento desses métodos vem apresentando rápido crescimento e a disponibilidade destas tecnologias motiva o seu uso em diferentes aplicações (SONKA; HLAVAC; BOYLE, 2013), inclusive na análise comportamental de mexilhões (GUTERRES, 2021). Portanto, torna-se atraente a utilização desses sistemas em conjunto com biossensores baseados em moluscos bivalves.

O presente trabalho busca propôr uma alternativa para monitoramento de ecossistemas aquáticos. Até onde se tem conhecimento, esse é o primeiro trabalho a avaliar o grau de exposição de pés e sifões de mariscos como potencial ferramenta não-invasiva no monitoramento aquático através de técnicas de visão computacional. Para isso, um modelo de três etapas é

apresentado para aquisição de dados comportamentais, processamento através de métodos de visão computacional e análise de respostas do biossensor frente a diferentes condições ambientais.

2 REVISÃO BIBLIOGRÁFICA

2.1 Moluscos Bivalves

A classe *Bivalvia* compreende diversos grupos como os mexilhões, as ostras, os mariscos, os berbigões e as vieiras que variam de características como sedentarismo, condições necessárias para sobrevivência, anatomia e comportamento. Também é possível observar diferenças de cor, tamanho e formato entre as espécies até mesmo dentro de um mesmo grupo (GRAF; CUMMINGS, 2007).

Os moluscos bivalves se caracterizam pela presença de concha calcária composta por duas partes, ou valvas, normalmente simétricas. Elas atuam como exoesqueleto, dando proteção contra partículas externas e predadores. Essa concha pode transitar de totalmente fechada para parcialmente aberta devido à presença de músculos adutores no seu interior que realizam o movimento das valvas (GOSLING, 2003).

Diferente de outros bivalves sésseis (como mexilhões e ostras), mariscos não vivem aderidos a substratos. Chamados de "moluscos de areia", esses costumam se enterrar na terra para se proteger de ameaças em seu ambiente. Para isso, as espécies de mariscos apresentam um órgão desenvolvido chamado pé, que sai do interior da sua concha quando necessita escavar ou se movimentar. Além disso, para conseguirem se alimentar enquanto escondidos na areia, seus sifões alongados são expostos para fora do esconderijo para absorver e eliminar água do mar. Essa água passa pelos seus órgãos para assimilação de micro-organismos em seu processo de alimentação, caracterizando o marisco como um animal filtrador (CARVALHO, 2015).

2.1.1 *Amarilladesma mactroides*

O bivalve *Amarilladesma mactroides*, conhecido como "marisco branco" no Brasil e como "almeja amarilla" nos países hispânicos, é um molusco bivalve endêmico do litoral atlântico. Abundante desde a costa do Rio de Janeiro até o Rio Negro (Argentina), esse animal é considerado de alta importância para o ecossistema marinho dessas regiões (COSCARÓN, 1959).

O marisco branco possui concha oval em tom amarelado que mede de 6 a 8 centímetros em indivíduos adultos. Além disso, possuem um pé bem desenvolvido e um par de sifões alongados que permitem inalar e expelir água enquanto estão enterrados (HERRMANN, 2008). Na Figura 2.1, pode-se observar dois indivíduos e suas principais características físicas em um habitat natural.

O *Amarilladesma mactroides* habita em praias arenosas pois essas permitem fácil escavação e a possibilidade de se alimentarem filtrando a água do mar, condições que torna viável sua vida enterrado em zonas de entremarés. A profundidade que se enterram varia de acordo com seu tamanho, indo de 5 cm para pequenos indivíduos até 30cm para animais adultos (GIANUCA, 1975).

Outros mariscos da família *Mesodesmatidae* tem se mostrado promissores para aquicultura

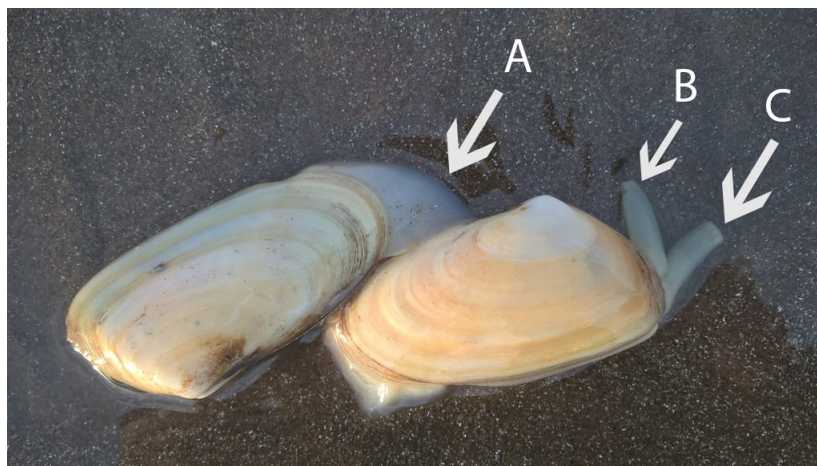


Figura 2.1: Mariscos *Amarilladesma mactroides* em habitat natural com órgãos expostos, no qual (A) é o pé, (B) é o sifão exalante e (C) é o sifão inalante. (MARTIN, 2019, Adaptado).

como as espécies *Mesodesma donacium* do Peru e Chile e *Paphies australis* da Nova Zelândia. Dessa forma, a densidade populacional do *Amarilladesma mactroides* e seu valor histórico e cultural o tornam bom candidato a ser estudado para aplicações como biossensor (SANTOS et al., 2016).

2.1.2 Análise Comportamental

Os moluscos bivalves são animais sensíveis às perturbações ambientais, tendo seu comportamento como um biomarcador dessas variações (LYUBIMTSEV; KHOLODKOVICH; DRUZHININ, 2019). Através da sua observação, é possível identificar períodos nas quais as condições do seu habitat não eram favoráveis para realizar suas funções biológicas apropriadamente (NEWTON; COPE, 2006).

A análise comportamental, portanto, possibilita que os moluscos bivalves possam ser aplicados como biossensores para monitoramento aquático (NEWTON; COPE, 2006). Essa análise compreende à leitura e registro de parâmetros comportamentais (como o grau de abertura das valvas) e o posterior estudo utilizando métodos estatísticos, de inferência e correlação com outros sensores paralelos (SCHÄFER et al., 2015).

Entretanto, por serem animais sensíveis, essa análise pode ser facilmente comprometida dependendo da metodologia utilizada. Idealmente, almeja-se utilizar técnicas não-invasivas, pois essas influenciam o comportamento animal da forma menos intensa possível através de menor manipulação do bivalve e menor variação do peso do animal por anexação de dispositivos (RANA et al., 2020). Nesse contexto, métodos baseados em visão computacional vem ganhando destaque por exigirem nenhuma conexão de equipamentos aos animais e, ainda assim, garantirem leituras com alta performance e confiabilidade (GUTERRES, 2021).

2.2 Medição Baseada em Visão Computacional

Sistemas de medição baseados em visão computacional (VBM) permitem a extração de características de uma cena através da aquisição e processamento de imagens (HEIJDEN, 1994). Através de tecnologia avançada, esses sistemas são capazes de obter informações ainda mais ricas que outros métodos para aplicações equivalentes, permitindo que se atinja alta performance de medição enquanto garante baixo custo operacional (MI et al., 2021).

Medições por sistemas VBM são aplicados em uma grande variedade de aplicações (SONKA; HLAVAC; BOYLE, 2013), inclusive na análise comportamental de mexilhões. Nessa, uma câmera é posicionada frontalmente aos animais e calcula o grau de abertura de suas valvas através do distanciamento de marcadores coloridos (GUTERRES, 2021). Em comparação com sensores de efeito Hall, tecnologia de alta performance ainda muito utilizada, os métodos baseados em visão computacional atingem resultados com alto grau de correlação (CALDERÓN; SIKORSKA; DOUDA, 2022; GUTERRES, 2021).

2.2.1 Calibração de Câmeras

O uso de lentes para a obtenção de imagens acarreta em alterações nas imagens resultantes. Essas alterações, denominadas distorções, fazem as representações de cenários serem diferentes do que seria visto pelo olho humano. A exemplo disso, pode-se citar os efeitos de curvaturas às linhas retas e variações na profundidade (JENKINS; WHITE, 1976). Nesse contexto, a calibração de câmeras é um procedimento essencial a partir do qual os parâmetros necessários à correção de imagens são obtidos e empregados (BRADSKI; KAEHLER, 2008), garantindo medições corretas em sistemas VBM.

Os parâmetros para calibração podem ser de origem extrínseca ou intrínseca à câmera. Os parâmetros extrínsecos são compostos por dados de rotação e translação da câmera em um sistema de coordenadas "do mundo real" (BURGER, 2016). Por outro lado, os parâmetros intrínsecos permitem o reconhecimento de distorções causados pela geometria da câmera e pelas características físicas de sua construção (BRADSKI; KAEHLER, 2008). Esses parâmetros podem ser apresentados pela matriz K , conforme Eq. 2.1, onde f é a distância focal da câmera e c é a posição do centro óptico no plano da imagem (BRADSKI; KAEHLER, 2008).

$$K = \begin{bmatrix} fx & 0 & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix} \quad (2.1)$$

Para a realização da calibração da câmera, também é necessário identificar os coeficientes de distorção radial (k_1, k_2, k_3) e tangencial (p_1, p_2) que são normalmente expressados através do vetor apresentado na Eq. 2.2 (BRADSKI; KAEHLER, 2008).

$$Coef = (k_1 \ k_2 \ p_1 \ p_2 \ k_3) \quad (2.2)$$

O procedimento realizado é descrito integralmente por Bradski e Kaehler (2008), o qual também apresenta algoritmo para aplicação do método em linguagem Python.

2.2.2 Segmentação de Imagens

Em visão computacional, segmentação é o processo de separar elementos de uma imagem em classes de forma que seu conteúdo possa ser compreendido. Esse reconhecimento pode ser feito utilizando diferentes técnicas, as quais se baseiam basicamente na identificação de descontinuidades e seções homogêneas. Dessa forma, é possível rotular cada pixel de uma imagem de acordo com a informação que carrega (BRADSKI; KAEHLER, 2008).

A classificação de pixels pode ser abordada de duas formas: (1) na segmentação semântica, os pixels que representam os mesmos elementos são rotulados em uma mesma classe; (2) na segmentação de instância há uma camada de processamento que diferencia cada ocorrência do mesmo tipo de objeto (SANTAQUITERIA et al., 2020). A Figura 2.2 ilustra a diferença entre os dois métodos aplicados em uma imagem de microalgas.

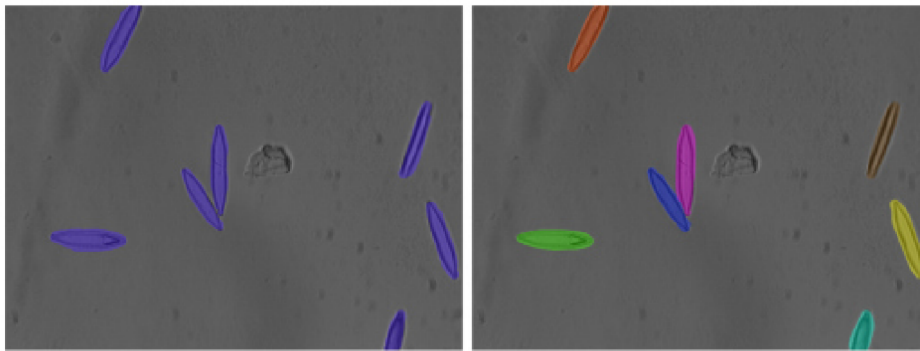


Figura 2.2: Exemplo de segmentação semântica e de instância em imagens de microalgas, respectivamente. Na primeira imagem, há a detecção das partículas como pertencentes a um mesmo grupo, enquanto na segunda imagem há a categorização de cada ocorrência com uma cor distinta (SANTAQUITERIA et al., 2020).

Segmentação de imagens é uma das áreas mais antigas e estudadas dentro da visão computacional havendo, portanto, diversas técnicas e modelos na literatura relacionada. A forma mais comum, então, de identificar métodos adequados para determinada aplicação é avaliando suas performances sobre *data sets* anotados por humanos (*ground truth*) (SZELISKI, 2010). Os diferentes métodos aplicados podem ser divididos em segmentação não-supervisionada, a qual leva em conta estruturas presentes nas imagens, ou supervisionada que incorpora conhecimento prévio através de amostras de treinamento (SEO et al., 2020).

2.3 Classificadores

Classificadores são ferramentas capazes atribuir classes a um conjunto de dados de entrada tendo em vista suas características e padrões determinados durante um processo de treinamento (KUNCHEVA, 2014). Estes objetivam realizar a inferência de classes frente a novos dados de

entrada. Dessa forma, múltiplos parâmetros podem ser avaliados simultaneamente e atinge-se uma resposta baseada em interações numéricas complexas (URBANOWICZ; BROWNE, 2017). Com essa finalidade, diversos algoritmos podem ser utilizados para o reconhecimento de padrões em dados como, por exemplo, o *K-Nearest Neighbors* (KNN) e o *Support Vector Machines* (SVM) (HASAN et al., 2020).

Embora existam metodologias não supervisionadas para a construção de classificadores, esses podem ser construídos com base em um *data set* anotados com classes conhecidas para seu treinamento bem como verificação de sua performance. É comum que ambos conjuntos de dados sejam partes de um mesmo *data set* original, fracionados em uma parte para treinamento e outra parte para teste (URBANOWICZ; BROWNE, 2017). Essa divisão permite que o modelo seja testado com dados não vistos durante o processo de treinamento, permitindo a avaliação de sua performance ao realizar novas inferências (KUNCHEVA, 2014).

Durante o treinamento supervisionado de classificadores, pode-se utilizar o processo de validação cruzada. Nessa etapa, o *data set* de treinamento é subdividido em K partições menores. Cada subconjunto K é utilizado para testar o modelo treinado com as demais $K-1$ partes. Dessa forma, performance de validação obtida ao final deste processo iterativo se dá pela média de todas iterações (KUNCHEVA, 2014).

Dentre outros fatores, a performance de classificadores inteligentes depende dos algoritmos utilizados em seu treinamento bem como da origem dos dados a serem analisados. Nesse contexto, visto que não é possível estabelecer um classificador absolutamente melhor que os demais, torna-se interessante a testagem de múltiplos modelos para a aplicação desejada (KUNCHEVA, 2014). Nesses casos, a escolha dos classificadores de melhor performance é feita analisando-se a acurácia e a área sob a curva de Característica de Operação do Receptor (AUC - ROC) dos diferentes modelos (HASAN et al., 2020; URBANOWICZ; BROWNE, 2017). As subseções a seguir versam sobre os classificadores *K-Nearest Neighbors* e Máquinas de vetores de suporte (*Support vector machine* - SVM).

2.3.1 *K-Nearest Neighbors*

Classificadores baseados em KNN classifica novos dados de acordo com a similaridade com dados de treinamento. Esse parâmetro é calculado através da distância no espaço entre a observação e os "k" dados mais próximos (vizinhos). Cada distância é considerada um voto e a classificação é feita de acordo com o voto majoritário da vizinhança (ZHANG, 2016).

Mudanças nos parâmetros operacionais definem diferentes tipos de classificadores KNN. Por exemplo, o número de vizinhos utilizados para votação definem os tipos *Fine*, *Medium* e *Coarse*, os quais utilizam poucos, médio e muitos vizinhos respectivamente. Outros tipos são definidos pela função de cálculo da distância entre as observações: O KNN cúbico utiliza função cúbica no cálculo da distância, enquanto o KNN *weighted* atribui pesos específicos para cada voto (GUO et al., 2003).

2.3.2 Support Vector Machines

Os SVMs baseiam-se na separação de dados através de um plano em um espaço n -dimensional. Nos casos que o conjunto de dados não permite essa separação, funções *kernel* são utilizadas para transformar as dimensões do *data set* e permitir que um hiperplano possa ser definido (CORTES; VAPNIK, 1995). A Figura 2.3 exemplifica este processo por meio da transformação de um conjunto de dados unidimensionais (reta cinza) em uma nova função no espaço bidimensional (parábola preta). Com isso, um hiperplano linear (reta tracejada) pode ser definido para separar as observações de cada classe.

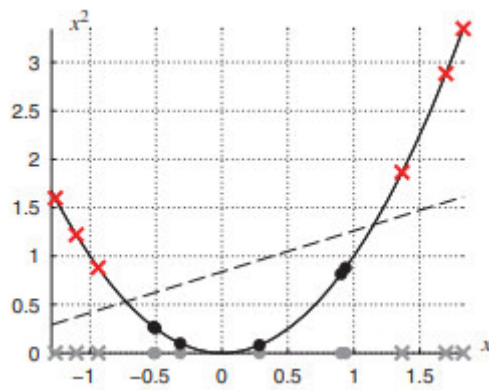


Figura 2.3: Exemplo de discriminação linear bidimensional, no qual um hiperplano linear (representado pela linha tracejada) é transformado em um hiperplano bidimensional (representado pela parábola), permitindo a separação de suas observações em duas categorias (KUNCHEVA, 2014).

O hiperplano é definido de forma a maximizar a distância entre as classes. Dessa forma, os vetores de suporte são determinados como os pontos que definem as margens com maior separação entre as diferentes classes em um processo de treinamento (HAYKIN, 2007) (Figura 2.4)

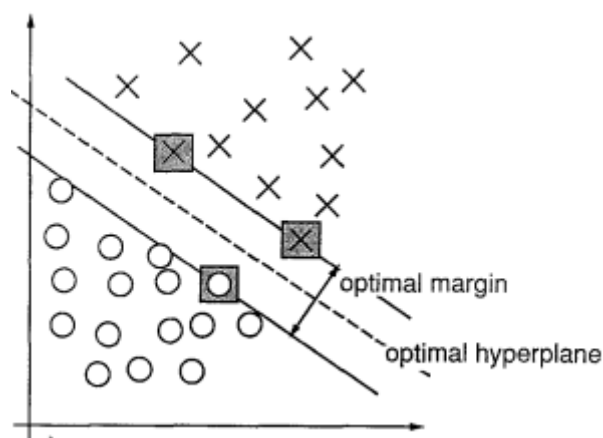


Figura 2.4: Exemplo gráfico da aplicação de SVM com seus elementos característicos, no qual o hiperplano ótimo está representado pela linha tracejada, as margens ótimas por linhas contínuas e os vetores de suporte como a cota entre as margens (CORTES; VAPNIK, 1995).

Dessa forma, as funções *kernel* empregadas definem os diferentes tipos de classificadores SVM. Funções polinomiais permitem o desenvolvimento dos SVMs linear, quadrático e cúbico, enquanto funções gaussianas permitem distinções detalhadas, médias ou grosseiras entre as classes (*fine*, *medium* e *coarse gaussian*, respectivamente) (BHATI; RAI, 2019).

2.4 Trabalhos Relacionados

A aplicação de moluscos bivalves como biossensores é promissora e vem sendo extensivamente estudada nos últimos anos (CAMPAS; PRIETO-SIMON; MARTY, 2007; GUDIMOV; KOMAROVA, 2021). Trabalhos publicados exploram diferentes técnicas de aquisição de dados, métodos de processamento e análise de *data sets*, emprego de diversas espécies de bivalves e efeitos toxicológicos de contaminantes aquáticos e outras condições inóspitas.

O uso de moluscos bivalves como elementos biológicos em biossensores tem como vantagem a abundância e a disponibilidade destes animais ao longo de todo ano. Além disso, maior parte de suas espécies são sedentárias ou apresentam baixa mobilidade. Por serem animais filtradores, a bioacumulação de contaminantes nos seus tecidos internos é um indicativo de poluição no ecossistema em que vivem. Por outro lado, essa análise requer longos períodos de exposição do molusco ao contaminante além de ser um método invasivo, pois o animal é sacrificado no processo (KRAMER; JENNER; ZWART, 1989). Em contraste, começam a ser propostos métodos não-invasivos que podem apresentar resultados de forma rápida e contínua, possibilitando a implantação de sistemas automáticos de leitura de sinais bem como monitoramento destes ambientes. A análise comportamental está inserida nesse grupo e se baseia na variação de sinais biológicos de bivalves quando sob estresse toxicológico (KRAMER; JENNER; ZWART, 1989).

As atividades biológicas de bivalves como respiração, alimentação, excreção, reprodução e movimentação (no caso de mariscos) está diretamente ligada à posição de sua concha. Em um ambiente saudável, os bivalves tendem a passar maior parte do tempo com as valvas abertas e fechá-las é a sua reação mais rápida a distúrbios. Dessa forma, técnicas de valvometria utilizam dessa correlação para identificar estresse toxicológico através da observação de padrões de movimento das valvas (BIALONCZYK; JERMACZ; ZIELSKA, 2019).

As primeiras técnicas de valvometria desenvolvidas incluíam a anexação dos bivalves em equipamentos analógicos com superfície fixa. Dessa forma, a movimentação da valva superior era registrada através de transdutores de deslocamento (BYRNE et al., 1990) ou extensômetros (HIGGINS, 1980). Essa metodologia, por fixar o animal aos dispositivos, acabava se tornando invasiva e provocando estresse aos animais e, portanto, prejudicando o seu comportamento. Outro exemplo de técnica invasiva é proposto por Kramer, Jenner e Zwart (1989), que anexa duas bobinas energizadas às valvas de mexilhões *Dreissena polymorpha* e *Mytilus edulis* junto a um oscilador de alta frequência.

A tecnologia avançou até a criação de sensores leves que podem ser aderidos às valvas dos moluscos sem causar o mesmo nível de estresse que as alternativas anteriores. Sensores de

efeito Hall, por exemplo, permitem a leitura de sinais comportamentais através do emprego de dois elementos: um sensor e um ímã, fixados em valvas opostas. O movimento de abertura e fechamento das conchas de mexilhões e ostras provoca diferença de campo magnético incidente no sensor, visto que a distância entre as partes aumenta e diminui. Dessa forma, passou a ser possível o uso de técnicas de valvometria em ambientes menos controlados (NAGAI et al., 2006).

Neste contexto, técnicas de valvometria tem sido utilizadas para detecção de mudanças comportamentais frente a várias condições ambientais. Trabalhos exploram com sucesso o comportamento moluscos bivalves frente a parâmetros físico-químicos como temperatura (MORGAN et al., 2005), pH (BAMBER; WESTERLUND, 2016), salinidade (HARTMANN et al., 2015) e pressão (RANA et al., 2020) e poluentes químicos como metais, hipoclorito (KRAMER; JENNER; ZWART, 1989), dióxido de carbono (BAMBER; WESTERLUND, 2016) e arsênico (LIAO et al., 2009). Esses trabalhos evidenciam comportamentos característicos desses animais como a redução da amplitude de abertura das valvas, redução do tempo com valvas abertas e mudança na frequência de micro-movimentos quando submetidos a condições inóspitas. Outros parâmetros comportamentais já foram propostos e se mostraram representativos ao comportamento de ostras e mexilhões como, por exemplo, frequência de transição de movimento (HARTMANN et al., 2015) e distância linear movimentada pelas valvas (BAMBER; WESTERLUND, 2016).

No entanto, essas abordagens se restringem a alguns grupos de moluscos bivalves. As ostras e os mexilhões são os mais adequados para tais técnicas de valvometria, visto que são animais sésseis. Nesses animais, a falta de mecanismos de locomoção tornam a operação mais fácil e restringem os padrões comportamentais à abertura e fechamento de suas valvas. Os mariscos, por outro lado, possuem tecidos desenvolvidos para se moverem. Esses animais possuem um pé que é utilizado para enterramento e dois sifões que permitem que o animal inale e exale enquanto escondido no sedimento de praias (COSCARÓN, 1959). Portanto, o comportamento desses animais pode ser observado, também, através da exposição e retração desses tecidos macios (RIISGARD, 2001).

Poucos trabalhos exploram o comportamento distinto dos mariscos. Jou et al. (2013) propõe a análise qualitativa manual dos sifões de mariscos *Corbicula fluminea* (expostos/recolhidos) para verificar estresse promovido por sensores Hall fixados às suas conchas. Bamber e Westerlund (2016) realiza uma tentativa de relacionar a extensão do pé dos animais com diferentes níveis de pH através de medição com paquímetro. Já (NUÑEZ et al., 2014) observa o comportamento de enterramento após cortar parte dos sifões de mariscos brancos *Amarilladesma mactroides*. Além dessas abordagens, novas técnicas específicas para a anatomia do marisco devem ser exploradas para o emprego de seu comportamento como *proxy* de qualidade ambiental.

Nos últimos anos, surgem as técnicas de valvometria baseadas em visão computacional. Guterres (2021) utiliza com sucesso um sistema de visão computacional para medição do grau de abertura das valvas do mexilhão *Perna perna*. Calderón, Sikorska e Douda (2022) comparam

resultados de valvometria de efeito Hall feita em mariscos com valvometria realizada por vídeos de alta frequência e obtém respostas similares nos dois métodos. Dessa forma, evidencia-se o uso da visão computacional como ferramenta para aquisição de dados comportamentais de bivalves.

Os dados comportamentais provenientes dessas técnicas são utilizados para calibração de métodos analíticos, os quais identificam *proxies* de qualidade da água. Testes estatísticos como Student (BAMBER; WESTERLUND, 2016) e ANOVA (KRAMER; JENNER; ZWART, 1989; LIAO et al., 2009; HARTMANN et al., 2015) se mostram adequados para identificação de mudança nos padrões comportamentais de bivalves. Alguns métodos estatísticos também são aplicados em forma de alarmes, como em soluções comerciais (AQUADECT, 2020). Além dessas metodologias, modelos baseados em inteligência artificial são apresentados como ferramentas para inferência de condições ambientais anormais baseados no comportamento de moluscos bivalves. Entre os modelos, destacam-se as redes neurais artificiais (RNA) (MORGAN et al., 2005) e os classificadores (GUTERRES, 2021; RANA et al., 2020). A utilização de modelos baseados em inteligência artificial possibilita o uso de parâmetros comportamentais como *proxy* para condições aquáticas.

É notável a existência de uma gama de metodologias aplicáveis para construção de biosensores baseados em moluscos bivalves. Entre essas, destaca-se o uso da visão computacional como método não-invasivo que permite a aquisição de dados sem que haja o contato físico do animal com qualquer dispositivo. O uso da visão, nesses casos, também permite a observação do animal de outras perspectivas além do movimento linear de valves que as atuais tecnologias se limitam. A aplicação de métodos não-invasivos baseados em visão computacional vai de encontro com as particularidades anatômicas e comportamentais dos mariscos, os quais realizam deslocamentos verticais e também horizontais. Dessa forma, métodos de visão podem ser capazes de contemplar a ocupação dos tecidos de mariscos em um plano de forma que as tecnologias disponíveis não são capazes. Essa medição, conforme apresentado, só é realizada na literatura disponível através de métodos que não permitem a automação do processo de aquisição de dados ou que manipulam os animais de forma invasiva, comprometendo a validade dos padrões comportamentais como biomarcadores do condições ambientais.

O processamento dos dados é um desafio estatístico visto a presença de forte auto-correlação (GUTERRES, 2021) entre as observações devido à influência do tempo nas séries. Esse fenômeno ocorre para leituras de abertura de valvas de moluscos sésseis e, portanto, é esperado para observações de ocupação espacial de mariscos. Formas de avaliar os resultados obtidos incluem transformar as leituras diretas em outros padrões comportamentais como documentado. Visto que a utilização do grau de exposição de tecidos dos mariscos é uma forma inédita de avaliar comportamento de moluscos, novos parâmetros devem ser propostos e avaliados para que se permita o uso dos métodos de avaliação apresentados na literatura, como análise de variância.

Nesse contexto, o presente trabalho busca aplicar metodologias não-invasivas baseadas em visão computacional para aquisição de dados comportamentais de mariscos *Amarilladesma mac-*

troides, considerando as particularidades desses animais em relação aos demais moluscos bivalves. Além disso, visa-se realizar a análise comportamental de mariscos através de parâmetros relacionados ao grau de exposição de seus pés e sifões ao ambiente, visto o comportamento desses animais como potencial *proxy* de qualidade da água. Também, investiga-se o desenvolvimento de biossensores através do uso de ferramentas baseadas em inteligência artificial e da medição baseada em visão de padrões de movimentação de pés e sifões de mariscos.

2.5 Objetivo Geral

A presente dissertação de mestrado tem como objetivo geral averiguar a possibilidade de se desenvolver biossensores de qualidade da água baseados no grau de exposição de mariscos brancos (*Amarilladesma mactroides*) a partir do desenvolvimento e utilização de modelos e metodologias não-invasivas baseados em visão computacional.

2.6 Objetivos Específicos

Como objetivos específicos, têm-se:

- Propôr modelos baseados em visão computacional para o monitoramento comportamental de mariscos brancos;
- Investigar a viabilidade da análise comportamental de mariscos através de imagens como base para a construção de biossensores;
- Calibrar o sistema biossensor utilizando sua análise comportamental como *proxy* de qualidade da água;
- Validar a empregabilidade dos animais estudados como biossensores de qualidade da água utilizando as técnicas propostas.

3 METODOLOGIA

Nessa seção, são apresentadas as metodologias que compreendem o desenvolvimento de um biossensor de qualidade da água baseado em mariscos. Cada etapa do desenvolvimento do modelo biossensor é abordado, incluindo o embasamento para proposta do modelo e as etapas que compreendem seu desenvolvimento.

3.1 Modelagem do Biossensor

O conceito básico de biossensores descreve sistemas capazes de emitir sinais baseados em um elemento biológico (HARTMANN et al., 2015). Nesse trabalho, busca-se expandir as aplicações de biossensores baseados em moluscos bivalves explorando-se formas inéditas para obtenção de sinais de mariscos, os quais são pouco explorados em trabalhos relacionados.

Em comparação com os demais moluscos bivalves, os mariscos diferem-se por sua anatomia única que permite seu estilo de vida distinto em regiões próprias para seu crescimento. Dessa forma, sua atividade biológica inclui movimentação de músculos para deslocamento, alimentação e excreção (COSCARÓN, 1959). As atividades específicas dos mariscos podem, portanto, ser exploradas como uma forma de identificar estresse ambiental sobre o animal (SANTOS et al., 2016).

Técnicas tradicionais de valvometria são apropriadas para aquisição do grau de abertura da concha de mariscos, de forma similar ao processo realizado em ostras e mexilhões. Entretanto, o grau de exposição dos músculos ao ambiente requer uma metodologia própria para ser medido. A atividade de pés e sifões pode ser observada como a ocupação espacial do animal em um ambiente controlado e sua variação ao longo do tempo. Com isso, torna-se possível a obtenção de sinais comportamentais a partir da observação dos mariscos e das regiões que os mesmos ocupam com suas conchas e seus músculos dentro de um aquário (CALDERÓN; SIKORSKA; DOUDA, 2022).

Visto que os pés e sifões dos mariscos são partes móveis que podem ser totalmente recolhidas, a anexação de dispositivos para obtenção de sinais torna-se inviável. Nesse contexto, a medição baseada em imagens se destaca por ser totalmente não-invasiva e ser capaz de registrar a ocupação espacial dos animais. Uma vez que os animais estão dispostos dentro de um reservatório, uma câmera fixa é capaz de obter dados referentes ao grau de ocupação da área disponível pelos pés e sifões dos mariscos (Figura 3.1). Nesse sistema, a área ocupada observada vai ser maior quanto mais expelidos estiverem os músculos do marisco. O grau de exposição desses tecidos e sua variação no tempo é potencialmente influenciado por diversas condições da natureza, como falta de nutrientes, temperatura, presença de contaminantes e concentração de oxigênio dissolvido no ambiente (CARVALHO, 2015).

Na configuração apresentada, a densidade de animais no aquário deve ser controlada, de forma que se evite a sobreposição de um animal sobre o outro. Em uma situação de superpopulação, a área ocupada registrada pela câmera pode ser subestimada. Além disso, essa

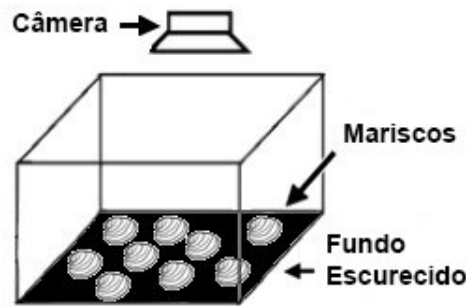


Figura 3.1: Configuração básica para aquisição de imagens de mariscos, constituída por uma câmera para imageamento, um reservatório fechado para acomodação dos animais, fundo escurecido para realce dos indivíduos nas observações e mariscos dispostos pela superfície.

interação pode provocar reação ao toque entre os animais envolvidos, apresentando comportamentos que não estão relacionados ao ambiente estudado em experimento. Além disso, a presença de muitos animais próximos pode levar ao rápido esgotamento de nutrientes essenciais no aquário (SANTOS et al., 2016).

Para a utilização do marisco branco *Amarilladesma mactroides* como biossensor, esse trabalho propõe um modelo de três etapas, conforme apresentado na Figura 3.2. Essa sequência de passos deve permitir a aquisição de imagens, o processamento dos dados para obtenção de dados comportamentais, a posterior análise comportamental ao longo do intervalo de aquisição para identificação de condições anormais no ambiente através de mudanças nos padrões de ocupação dos bivalves.

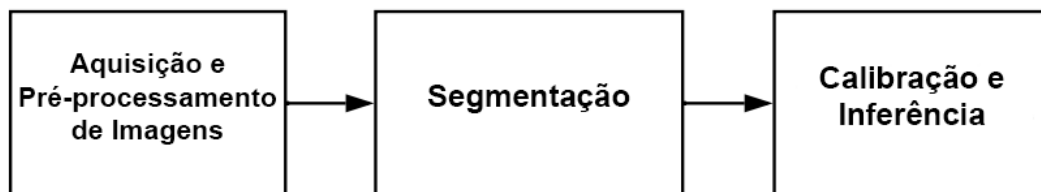


Figura 3.2: Pipeline do modelo proposto composto pelas etapas de aquisição e pré-processamento de imagens, segmentação dos mariscos, calibração do modelo biossensor e inferência de condições ambientais.

3.2 Aquisição e Pré-Processamento de Imagens

Uma série de experimentos laboratoriais foram realizados para desenvolver e validar o modelo computacional para aplicação de mariscos como biossensores. Cada experimento teve como objetivo a obtenção de *data sets* adequados para avaliação de diferentes parâmetros e condições.

Todos os experimentos empregaram mariscos brancos adultos com um comprimento médio de $53,2 \pm 1,9\text{mm}$. Os bivalves *Amarilladesma mactroides* foram extraídos da Praia do Mar Grosso, São José do Norte - RS, Brasil (-32,3107, -51,5926). Após coleta, os mariscos brancos foram mantidos em ambiente semelhante ao que foram recolhidos até início dos experimentos (temperatura de 20°C e salinidade de 30 PSU). Esses animais foram, então, alocados em

aquários de vidro translúcido de dimensões 12x18x12 centímetros. Todos reservatórios utilizados tiveram seu fundo coberto com papel *contact* preto, a fim de maximizar o contraste entre a cor do fundo e os animais.

Uma vez inseridos no interior dos aquários com água salgada e fornecimento contínuo de oxigênio, foi aguardado um período de 2 dias para aclimação dos mariscos ao novo ambiente. Periodicamente, os bivalves foram alimentados com diatomácea *Conticribra weissflogii* e tiveram sua água renovada de forma semi-estática com taxa de 50%. Durante aclimação e após o início do experimento, os animais foram submetidos a um fotoperíodo constante de 12 horas claro e 12 horas escuro. Conforme subseções a seguir, um total de três experimentos foram realizados com o intuito permitir o desenvolvimento e validação do sistema biossensor proposto bem como investigar o comportamento de mariscos brancos frente a diferentes temperaturas e níveis de contaminação ambiente.

3.2.1 Experimento 1

O primeiro experimento realizado teve como principal objetivo a obtenção de imagens para desenvolvimento e calibração dos modelos de segmentação necessários à implementação do sistema biossensor proposto. Embora o uso de câmeras distintas ocasione variações de características das imagens provocadas por diferenças intrínsecas aos equipamentos, o experimento em questão teve como intuito verificar a empregabilidade de câmeras *off-the-shelf* na construção de biossensores versáteis, customizáveis e de fácil construção.

O procedimento consistiu na exposição controlada de mariscos brancos ao contaminante DCOIT (4,5-dicloro-2-octil-isotiazolinona) em três concentrações distintas (0, 0,01 e 0,13 mg/L). Doze animais foram submetidos a cada um dos tratamentos, sendo eles Controle (0 mg/L), Tratamento 1 (0,01 mg/L) e Tratamento 2 (0,13 mg/L). O experimento foi realizado em duplicata. Três câmeras foram utilizadas para a aquisição de dados comportamentais ao longo do experimento, cada uma dedicada ao monitoramento comportamental de tratamentos individuais. Dessa forma, os equipamentos foram posicionados de forma a prover informações visuais acerca de ambos aquários em um tratamento. Os modelos utilizados do controle à maior concentração foram Nikon Coolpix S9700, GoPro Hero 6+ e GoPro Hero 3, respectivamente. A esquematização do experimento é apresentado na Figura 3.3

O experimento foi conduzido por 96 horas com as gravações sendo feitas de 12 em 12 horas. A cada turno, as câmeras registraram aproximadamente 20 minutos de atividade biológica dentro dos aquários totalizando 21 vídeos ao final do experimento (*Data set 1*).

3.2.2 Experimento 2

O segundo experimento teve como objetivo observar o comportamento de mariscos brancos sujeitos a diferentes temperaturas. Dois grupos foram investigados sendo eles submetidos à temperaturas constantes e controladas de 20 °C (temperatura ambiente) e 17 °C. Os reservatórios foram inseridos em uma caixa termicamente isolada a fim de manter as temperaturas

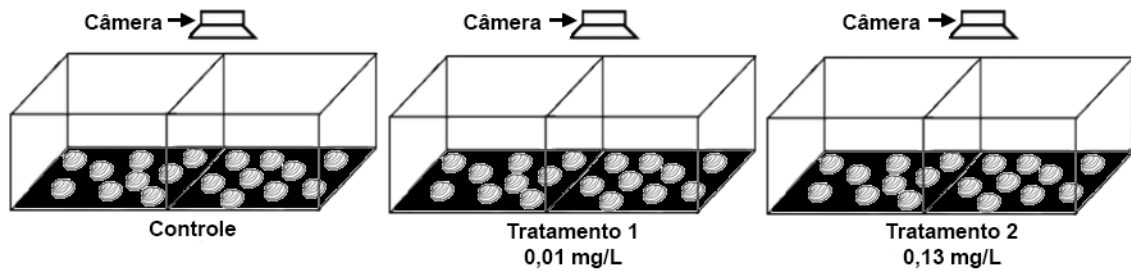


Figura 3.3: Configuração do experimento 1 realizado em duplicata, contendo as seguintes concentrações de DCOIT: 0 mg/L (Controle), 0,01 mg/L (Tratamento 1), 0,13 mg/L (Tratamento 2).

propostas ao longo do período experimental.

O procedimento foi realizado em duplicata (cada aquário contendo 6 mariscos adultos). Câmeras GoPro Hero 6+ e GoPro Hero 5 foram posicionadas em configuração semelhante ao experimento anterior com o intuito de possibilitar a aquisição de dados comportamentais conforme ilustrado através da Figura 3.4. O experimento foi conduzido por 96 horas, com filmagens de 60 minutos a cada 24 horas. Após a finalização, o experimento foi totalmente reproduzido considerando-se a mesma metodologia de exposição e coleta de dados resultando, portanto, em dois *data sets* (2A e 2B).

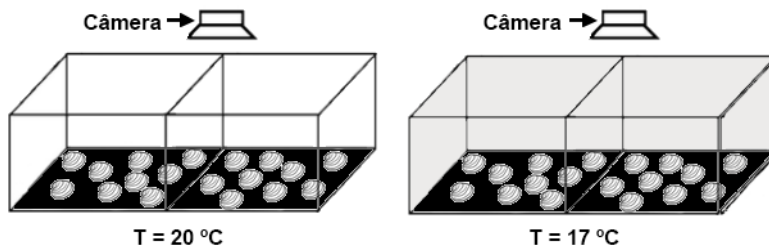


Figura 3.4: Configuração do experimento 2 realizado em duplicata, contendo dois grupos em diferentes temperaturas: 20 °C e 17 °C.

3.2.3 Experimento 3

O terceiro experimento teve como objetivo observar mudanças comportamentais de mariscos *Amarilladesma mactroides* frente a contaminantes químicos. Para isso, três concentrações de DCOIT (0, 0,01 e 0,13 mg/L) foram aplicados em aquários em configuração semelhante ao Experimento 1, porém com população reduzida de animais no ambiente (seis mariscos por aquário). Os animais foram filmados através de câmeras GoPro Hero 6+ e GoPro Hero 5. O experimento foi realizado durante 48 horas e coletas de dados (vídeos de 60 minutos) foram realizadas a cada 24 horas (*Data set 3*).

3.2.4 Pré-Processamento

Cada vídeo produzido durante os experimentos compreende informações de dois aquários pertencentes ao mesmo grupo de investigação e sujeitos às mesmas condições de interesse. O

processamento de dados através de algoritmos de segmentação é relativamente menos complexo se a informação é inserida em forma de imagem. Dessa forma, se torna atraente a conversão dos vídeos presentes nos *data sets* em um conjunto de imagens individuais. Ademais, torna-se necessária a determinação de regiões de interesse com o intuito de avaliar cada duplicata separadamente. Dessa forma, propõe-se a extração de quadros a cada 10 segundos para posterior processamento. Duas regiões de interesse (ROIs) são extraídas das imagens resultantes tendo em vista o posicionamento dos aquários em duplicata. Estas regiões são salvas como arquivos independentes após correção da perspectiva e da distorção proveniente das lentes das câmeras. Essa calibração foi realizada utilizando parâmetros intrínsecos e extrínsecos das câmeras calculados através do método de calibração com tabuleiro utilizando a biblioteca OpenCV. Obtém-se, dessa forma, um conjunto imagens na resolução 320x480 pixels.

Uma vez extraídas as regiões de interesse, as imagens são pré-processadas através aplicação de filtro de mediana (CHAKI; DEY, 2019), abertura e transformação Top Hat (BRADSKI; KAEHLER, 2008). O pré-processamento tem como objetivo a eliminação de ruídos provenientes de reflexos na superfície da água, eventuais bolhas causadas pelo fluxo de água, texturas no fundo do aquário causadas pela condensação da umidade do ar e pela cola utilizada na construção dos reservatórios. Além disso, essas operações buscam melhorar o contraste entre o marisco branco e o fundo escuro. Em resumo, o pré-processamento proposto visa otimizar a performance dos modelos de segmentação responsáveis pelo monitoramento comportamental dos animais ao longo do tempo.

3.3 Segmentação

Para o desenvolvimento de modelos de segmentação, é necessário o uso de imagens anotadas (*ground truth*), nas quais cada pixel é categorizado entre as possíveis classes de forma manual. O uso desses dados anotados possibilita comparar uma predição com sua segmentação fundamentalmente correta. Modelos supervisionados baseados em aprendizado de máquina também requerem essas anotações durante a etapa de treinamento para que consigam se adaptar em direção a melhores resultados. A obtenção dessa informação *ground truth* permite o uso de métricas apropriadas para avaliação da performance dos modelos de segmentação. Este *data set* é importante para a alteração dos parâmetros dos modelos ao longo do processo de treinamento sendo, portanto, chave para a aquisição de resultados satisfatórios sob o âmbito de um treinamento supervisionado.

A criação do *data set ground truth* inicia-se com a escolha de um subconjunto dos dados originais para posterior anotação bem como treinamento e validação dos modelos de segmentação. A anotação é realizada em 10% das imagens disponíveis para cada grupo, aquário e período de aquisição de dados tendo em vista a seleção das primeiras e últimas 6 imagens. Totaliza-se, dessa forma, um total de 432 anotações. A anotação é manualmente realizada através do *software* Adobe Photoshop que permite a inserção de uma camada vazia ao topo da imagem original, facilitando a identificação das posições dos mariscos. Cada animal é pintado com uma

cor específica utilizando a função pincel, enquanto o fundo do aquário é pintado da cor cinza e a imagem é exportada no formato .png.

Dessa forma, investiga-se o emprego de dois modelos de visão computacional distintos para a segmentação de mariscos em imagens. O primeiro, baseado em segmentação *Watershed*, é um método clássico de simples implementação. O segundo modelo utiliza uma rede neural profunda de aprendizagem de máquina para segmentação dos animais. O desenvolvimento desses modelos é avaliado através de métricas adequadas para imagens com desbalanceamento de classes, visto que há mais pixels de fundo em comparação com área ocupada por mariscos.

3.3.1 Modelo de Segmentação com Algoritmo *Watershed*

Um dos métodos clássicos de segmentação semântica é o algoritmo *Watershed*, o qual opera com inspiração no efeito topográfico recorrente em bacias hidrográficas. Ao atingir regiões de relevo mais alto, a chuva tende a escorrer para as regiões de captação, em direção aos vales mais baixos, e alagá-las conforme representado na Figura 3.5. A diferença de altura e a presença de estruturas separatórias fazem que os pontos de mínima sejam preenchidos em tempos diferentes ao longo do período de precipitação (PREIM; BOTHA, 2014).

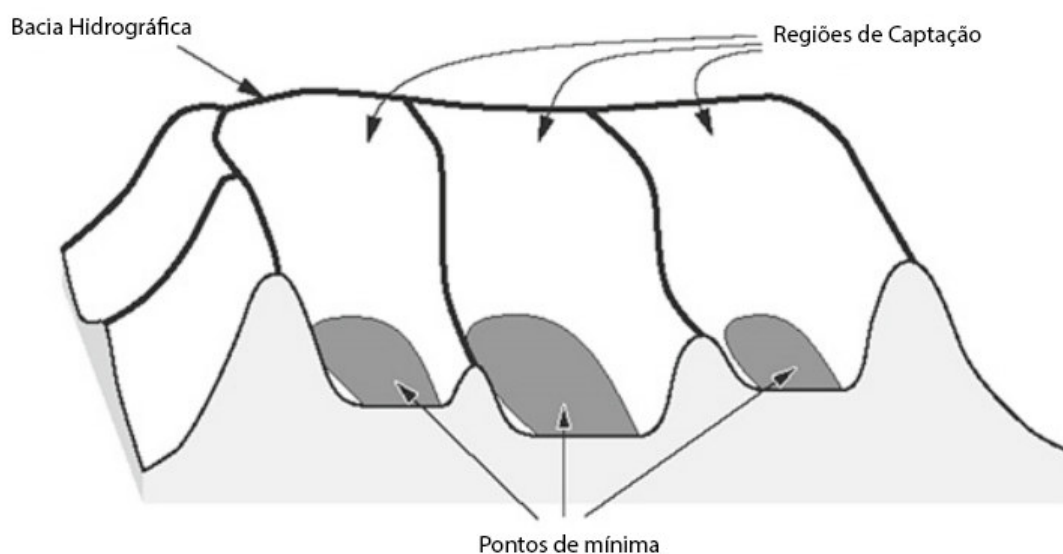


Figura 3.5: Efeito de alagamento em bacia hidrográfica. A constante e uniforme queda das chuvas sobre a paisagem provoca o escoamento da água pelas diferentes regiões de captação e, conseqüentemente, o enchimento dos vales em diferentes tempos. O fenômeno serve como base para o algoritmo *watershed*, que identifica regiões de mesma intensidade dentro da imagem. (PREIM; BOTHA, 2014, Adaptado).

De forma análoga, o algoritmo *Watershed* decompõe a imagem original a partir de um gradiente de intensidade. Regiões uniformes são identificadas como vales, enquanto as linhas dominantes são indicadas pelo usuário ou por marcadores gerados através de um algoritmo incorporado. Com o preenchimento das zonas baixas, o algoritmo classifica as áreas que se

conectam na região da bacia hidrográfica como pertencentes à mesma classe (BRADSKI; KAEHLER, 2008).

Entretanto, por se tratar de um método baseado em gradiente de intensidade, a segmentação *Watershed* é fortemente sujeita a ruído. A presença de texturas na imagem levam o algoritmo a supersegmentar, classificando os seus elementos erroneamente como classes existentes ou novas classes. Para diminuir a ocorrência desse fenômeno, técnicas de filtragem são necessárias para suavização dos objetos na imagem sem prejudicar no reconhecimento de suas bordas (BEUCHER; MEYER, 1993).

A supersegmentação também pode ser minimizada com o emprego de segmentação baseada em marcadores. Esse método permite que a segmentação seja realizada ao redor de contornos gerados automaticamente (região de incerteza) evitando-se, assim, que pontos de mínimo da imagem causem ruídos na saída do algoritmo. Além disso, a geração automática dos marcadores proporciona menor tempo para se processar conjuntos extensos (ZHANG et al., 2014). Por outro lado, o método de marcação manual não é obsoleto, sendo utilizado dentro da visão computacional como ferramenta assistiva em *softwares* de anotação de imagens (BRÉHÉRET, 2017), por exemplo.

Baseado na biblioteca OpenCV, o algoritmo de segmentação consiste em determinar a região da imagem em que há certeza de ser plano de fundo através de dilatação, seguido do uso de transformação de distância e limiarização para se obter uma região de mariscos. Adiciona-se, então, marcadores nos contornos e a região de incerteza é classificada com a função *Watershed*. A saída do algoritmo é uma imagem binária (máscara) contendo mariscos e fundo apresentados com valores diferentes.

3.3.2 Modelo de Segmentação baseado em Aprendizado de Máquina

A segmentação de imagens também pode ser conduzida através do uso de redes neurais. Tratam-se de estruturas matemáticas que aprendem a extrair características de um cenário ao ser treinada com exemplos. Esses modelos passaram a receber atenção pelos anos 90 com grande crescimento nos estudos a partir de 2015. Elas têm alcançado performance estado-da-arte para segmentação semântica de imagens naturais (SEO et al., 2020).

Uma rede neural artificial é composta por vários neurônios, estruturas menores que recebem informação e calculam um sinal de saída. A informação emitida é recebida por outros neurônios que passam pelo mesmo processo, formando assim uma estrutura complexa. Cada sinal nessa estrutura é recebido com um nível de significância (peso) diferente. De forma semelhante ao cérebro humano, uma rede multicamadas de operações promove um modelo capaz de aprender padrões (ELGENDY, 2019) (Figura 3.6).

A U-Net é uma arquitetura de rede desenvolvida por Ronneberger, Fischer e Brox (2015) com o objetivo de alcançar alta performance em segmentação de imagens biomédicas a partir de pequenos *data sets*. Sua aplicação, por outro lado, também vem apresentando bons resultados em outras áreas como na segmentação de imagens de satélite (RAKHLIN; DAVYDOW;

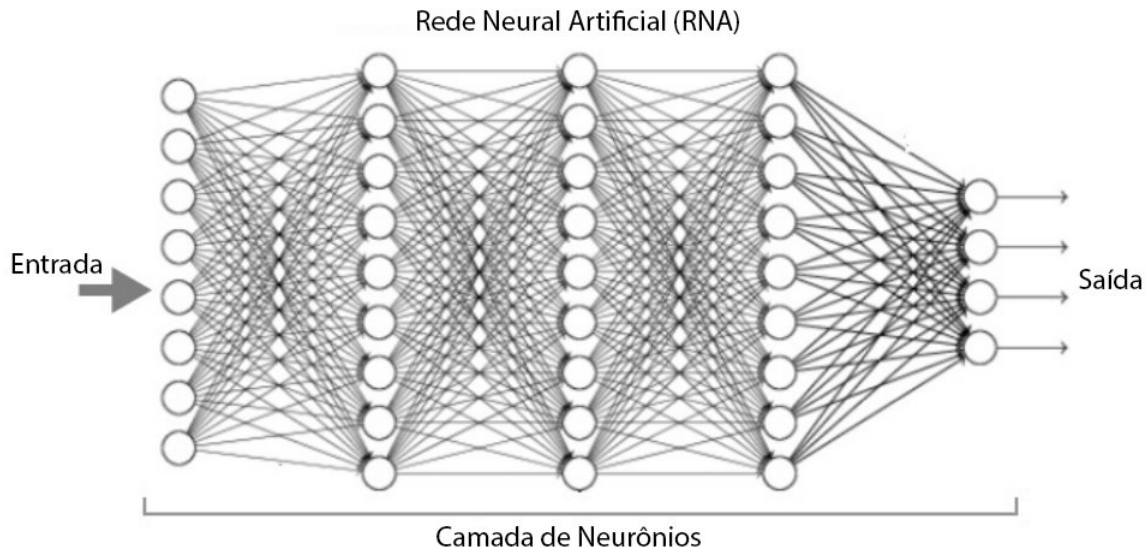


Figura 3.6: Esquematização de uma rede neural artificial. O processamento e transmissão individual de cada neurônio simples gera complexas redes multicamadas capazes de aprender padrões (ELGENDY, 2019, Adaptado).

NIKOLENKO, 2018), na detecção de rachaduras em estruturas de concreto (LIU et al., 2019) e na separação da faixa vocal em músicas no domínio da frequência (JANSSON et al., 2019).

Cada camada em uma rede neural convolucional pode ser classificada de acordo com os cálculos que essa exerce. Uma camada de convolução utiliza um filtro (matriz de pesos) de tamanho menor que o mapa de entrada, o qual é sobreposto em todas posições possíveis. Quando projetado, são realizadas operações matemáticas ponderadas na região e esse valor é inserido em um novo mapa de dimensão menor (*feature map*), conforme ilustrado na Figura 3.7. A realização dessas operações acarreta na preservação de características nas sub-regiões observadas (VENKATESAN; LI, 2018).

Da mesma forma, uma camada de operação *pooling* utiliza um filtro para varrer o mapa receptivo. Como exemplificado na Figura 3.8, esse método pode ser dos tipos *max pooling*, no qual é armazenado o maior valor das sub-regiões, ou *average pooling* que calcula e insere a média aritmética. Com isso, há redução nas dimensões das camadas que aceleram o tempo de processamento da rede neural.

A arquitetura da rede U-Net, apresentada na Figura 3.9, é composta por camadas de convolução e *max pooling* dispostas em uma etapa de contração e de expansão. Na primeira, são aplicadas convoluções 3x3 e operações *max pooling* 2x2 repetidamente, duplicando-se o número de canais a cada passo. As saídas de cada camada convolucional que antecede um *max pooling* são também concatenadas com as entradas equivalentes durante a etapa de expansão, que compreende convoluções 3x3 e 2x2. Essas transferências de informação entre as etapas da rede são chamadas (*skip connections*) e seu uso garante que detalhes finos e contexto da imagem inicial não sejam perdidos.

Para realizar a segmentação dos mariscos no *data set* obtido, é proposto o uso de modelos

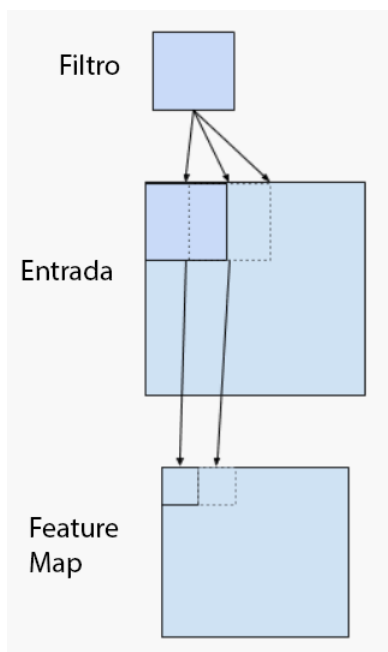


Figura 3.7: Funcionamento de uma camada convolucional de uma rede neural. A imagem de entrada é lida como uma matriz e todas posições possíveis são varridas através de um filtro. O resultado da operação específica desse filtro permite que a região varrida possa ser compactada em um mapa de características de dimensão menor. (BROWNLEE, 2019, Adaptado).

baseados em aprendizagem de máquina. Entre as arquiteturas de RNA disponíveis, a U-Net se mostra adequada para aplicação visto sua boa performance em casos com poucas imagens disponíveis no conjunto de dados (RONNEBERGER; FISCHER; BROX, 2015; MO et al., 2022). Assim, o processo de geração do modelo é realizado utilizando os *frameworks* Keras e Tensorflow para linguagem Python.

O procedimento consiste na fixação e manipulação de hiper-parâmetros configuráveis da rede para obtenção de melhores resultados. O conjunto de dados e suas anotações são divididos em 60% para treinamento, 20% para validação e o restante para teste, de forma que todos grupos contenham imagens referente aos três grupos toxicológicos na mesma proporção. Além disso, o treinamento da rede é conduzido utilizando a função de perda Dice (*Dice Loss*) visto a possibilidade de atribuir pesos a cada classe na imagem e melhorar a performance do modelo ao se processar dados com algum nível de desbalanceamento (JADON, 2020).

3.4 Métricas de Avaliação

Modelos de segmentação semântica envolvendo duas classes podem ser avaliados de forma semelhante a modelos de classificação binária, considerando que cada pixel na imagem está sendo rotulado. Nesses casos, cada pixel pode ser classificado como positivo (quando pertence à classe) ou negativo (quando não pertence à classe) além da predição estar correta ou errada, totalizando quatro possíveis resultados (ELGENDY, 2019)

Quando um pixel é predito corretamente como positivo têm-se um verdadeiro positivo (TP -

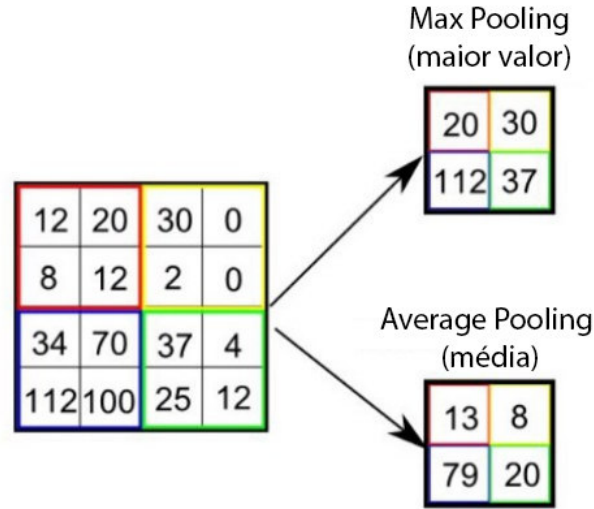


Figura 3.8: Funcionamento das camada de operação *pooling*, na qual cada sub-região pode ser resumida ao máximo valor apresentado ou pela média de todos valores contidos na sub-matriz (SAHA, 2018, Adaptado).

true positive); se incorretamente predito positivo têm-se um falso positivo (FP - *false positive*); quando há a classificação correta de um negativo têm-se um verdadeiro negativo (TN - *true negative*) e quando se aponta erroneamente um negativo têm-se um falso negativo (FN - *false negative*) (FAWCETT, 2005). A organização desses valores em uma matriz de confusão é exemplificada na Figura 3.10.

As métricas que relacionam os elementos da matriz de confusão descrevem a performance do modelo e possibilitam a sua quantificação. O *Recall* (Eq. 3.1) é calculado como a taxa de predições positivas pelo total de casos positivos no sistema. Da mesma forma, a *Precisão* (Eq. 3.2) faz referência à taxa de predições corretamente positivas dentro de todas predições positivas. Por fim, a *acurácia* (Eq. 3.3) relaciona o total de predições corretas pelo total de predições feitas (POWERS, 2007).

$$Recall = \frac{TP}{TP + FN} \quad (3.1)$$

$$Precisao = \frac{TP}{TP + FP} \quad (3.2)$$

$$Acuracia = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.3)$$

Em *data sets* em que o número de ocorrências de cada classe são muito diferentes (desbalanceados) a acurácia se torna uma métrica menos significativa. Nesses casos, a métrica pode apresentar valores altos mesmo quando o modelo tem baixa performance na identificação de classe menos frequentes. Dessa forma, o uso da métrica F_1 -score (ou coeficiente de Dice) - que relaciona a precisão e o *recall* - se torna mais apropriada (Eq. 3.4). Assim como a acurácia, essa métrica varia entre 0 e 1, indicando melhor performance do modelo quanto maior for o seu

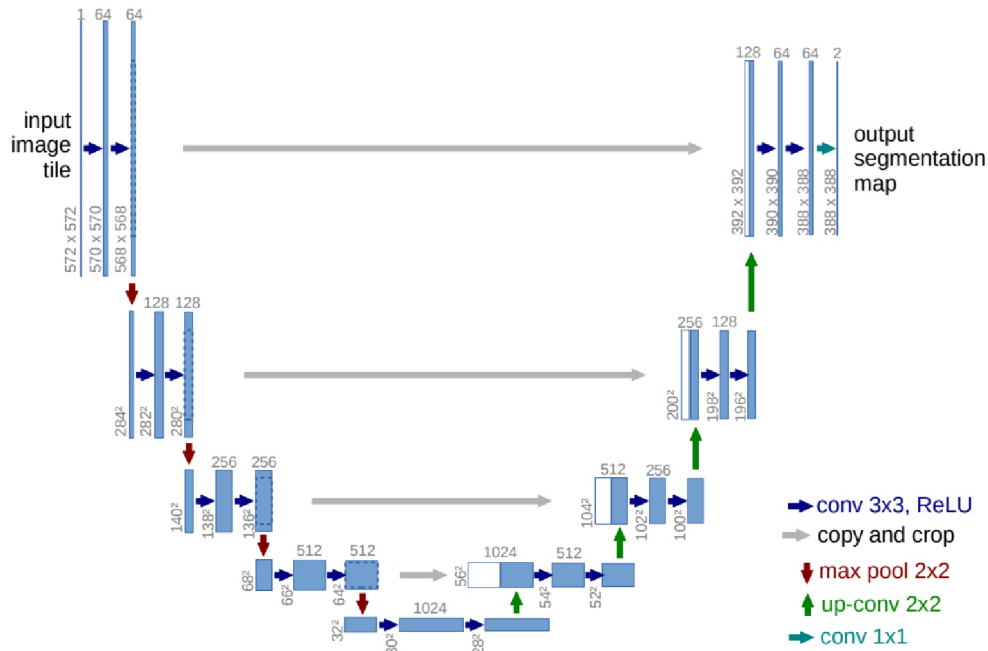


Figura 3.9: Estrutura da rede U-Net, composta por seções simétricas de convolução e deconvolução (RONNEBERGER; FISCHER; BROX, 2015).

		Classe Real	
		Negativo	Positivo
Classe Predita	Negativo	TN	FN
	Positivo	FP	TP

Figura 3.10: A matriz de confusão relata o número de ocorrências de cada caso possível em uma classificação binária: ocorrências negativas classificadas corretamente (TN) ou erroneamente (FN) e ocorrências positivas classificadas corretamente (TP) e erroneamente (FP). Através desses quatro parâmetros, é possível calcular métricas de avaliação do modelo como a acurácia.

valor (ELGENDY, 2019).

$$F_1 - Score = \frac{2 \cdot Precisão \cdot Recall}{Precisão + Recall} \quad (3.4)$$

3.4.1 Curvas ROC

A curva Característica de Operação do Receptor (ROC) de um classificador binário plota a taxa de verdadeiros positivos (sensibilidade) em função da taxa de falsos positivos conforme ilustrado na Figura 3.11.

Curvas ROC correspondem à uma demonstração bidimensional da *performance* de um modelo de classificação. Logo, a área sob a curva ROC (*Area Under the Curve* - AUC) resume estatisticamente a qualidade de um modelo de classificação e varia de 0 à 1 (CORTES; MOHRI,

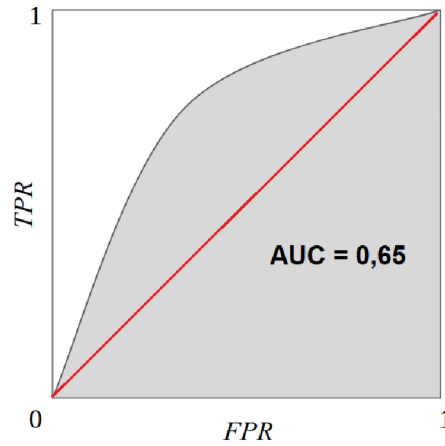


Figura 3.11: Exemplo de curva com ROC com área sob a curva (AUC) igual à 0,65. Linha em vermelho ilustra a curva de um classificador com saídas totalmente aleatórias.

2004). Um classificador provendo saídas aleatórias apresentaria AUC igual à 0,5 e curva ROC da origem ao ponto (1, 1) conforme linha vermelha ilustrada na Figura 3.11.

3.5 Calibração e Inferência

Os *data sets* obtidos nos experimentos 2 e 3 foram empregados para calibração do biosensor sob diferentes condições ambientais. Com esse intuito, observou-se o comportamento dos mariscos frente a diferentes temperaturas bem como presença de contaminante DCOIT em duas concentrações distintas.

A análise comportamental dos mariscos foi realizada com base em sua atividade de exploração. Esse parâmetro contempla a movimentação dos pés e sifões dos bivalves durante tanto sua exposição quanto recolhimento. A atividade de exploração de um grupo de mariscos em um aquário *Ex* é, portanto, definida segundo a Eq. 3.5 na qual A é a área ocupada no instante de tempo i .

$$Ex = \sum_{i=0}^i |A_{(i)} - A_{(i-1)}| \quad (3.5)$$

Os valores de área ocupada foram obtidos para cada 10 segundos conforme taxa de amostragem dos vídeos na etapa de pré-processamento. Esses dados foram agrupados como atividade de exploração a cada 5 minutos. Além disso, foram utilizados os valores médios de atividade entre cada replicata.

A viabilidade da atividade de exploração como *proxy* de qualidade da água é avaliada através da Análise de Variância (ANOVA) e teste *post-hoc* de Tukey-Kramer. A normalidade dos resíduos foi verificada através do teste de Shapiro-Wilk e a autocorrelação foi analisada com o teste de Durbin-Watson. Para os casos que as suposições da ANOVA não foram alcançadas, o teste de Kruskal-Wallis bem como o teste *post-hoc* de Dunn foram empregados. Os procedimentos estatísticos foram realizados nos *software* Microsoft Excel e RStudio e com nível de

significância de 5%.

Posteriormente, a área total ocupada e a atividade de exploração foram empregados na construção de classificadores de qualidade da água. Essa etapa teve como objetivo prover classificadores inteligentes treinados com ambos parâmetros comportamentais e a condição ambiental referente a cada observação (*target*). Ou seja, para cada par de valores de área total e atividade de exploração foi atribuído um valor numérico referente ao controle (classe "0") ou um número referente a condições com parâmetros ambientais alterados (classe "1"). Nos experimentos realizados, a classe "1" abrange a temperatura reduzida (experimento 2) ou a presença de contaminante DCOIT (experimento 3).

O *data set* utilizado foi dividido aleatoriamente em 80% e 20% para aplicação no treinamento e teste dos classificadores, respectivamente. O desenvolvimento desses modelos foi realizado utilizando o *software* MATLAB, que permite o treinamento paralelo de múltiplos classificadores através da aplicação *Classification Learner*. Foram testados 11 modelos de classificadores diferentes, sendo seis variações do tipo SVM (linear, quadrático, cúbico, *fine gaussian*, *medium gaussian*, *coarse gaussian*) e cinco do tipo KNN (*fine*, *medium*, *coarse*, *weighted* e cúbico). As validações foram realizadas através do método de validação cruzada com 10 *folds* (RANA et al., 2020; GUTERRES, 2021).

Os três classificadores de maior acurácia durante treinamento foram selecionados para teste. Nessa etapa, os 20% do *data set* não empregados no treinamento foram utilizados para inferência de condição ambiental. As acurácias e AUC-ROC desses classificadores também foram avaliados nessa etapa. Por serem treinados com base em parâmetros comportamentais de mariscos sujeitos a diferentes condições de qualidade de água, esperava-se que os classificadores seriam capazes de inferir corretamente se novas observações desses animais referem-se a situações de controle ou a condições alteradas, nas quais devem ser emitidos alarmes.

4 RESULTADOS

A seguir, serão discutidos os resultados obtidos após a execução da metodologia proposta. Também serão contempladas as discussões pertinentes sobre cada etapa do modelo de biossensor que esse trabalho propõe.

4.1 Aquisição e Pré-Processamento de Imagens

Os vídeos obtidos durante experimentação registraram satisfatoriamente os seus respectivos aquários. Através das gravações, foi possível extrair *frames* a cada 10 segundos de cada par de aquários. Além disso, as ROIs para cada reservatório foram obtidas conforme a metodologia. A Figura 4.1 ilustra um exemplo de *frame* original e suas respectivas ROIs.



Figura 4.1: Comparação entre *frame* extraído do vídeo com suas ROIs separadas, respectivamente. O procedimento transforma uma imagem de alta resolução (1920x1080p) em duas imagens de menor resolução (480x320p) além de corrigir a perspectiva e a distorção intrínseca da câmera empregada.

Utilizando as ROIs extraídas, o *data set ground truth* foi manualmente gerado para 10% do conjunto de dados. As imagens anotadas foram criadas em cores distintas para cada marisco e foi binarizada para uso na avaliação dos modelos de segmentação, conforme exemplificado na Figura 4.2.

Devido ao desbalanceamento entre pixels de marisco e fundo, os histogramas das imagens do *data set* apresentam dois sinos de alturas distintas. O pré-processamento das imagens permitiu uniformizar o fundo do aquário. O histograma da imagem resultante demonstrou uma redução da frequência de pixels intermediários e aumento da frequência de pixels escuros. Esse fato é evidenciado na Figura 4.3 que ilustra histogramas de uma imagem do *data set 1* antes e depois de seu pré-processamento.

O resultado promovido pelo pré-processamento também pode ser observado diretamente nas imagens, como exemplificado na Figura 4.4. Nessa, é evidente o escurecimento e uniformização do fundo, contrastando os mariscos presentes.



Figura 4.2: Comparação entre ROI extraída, imagem anotada e *ground truth* binarizado, respectivamente

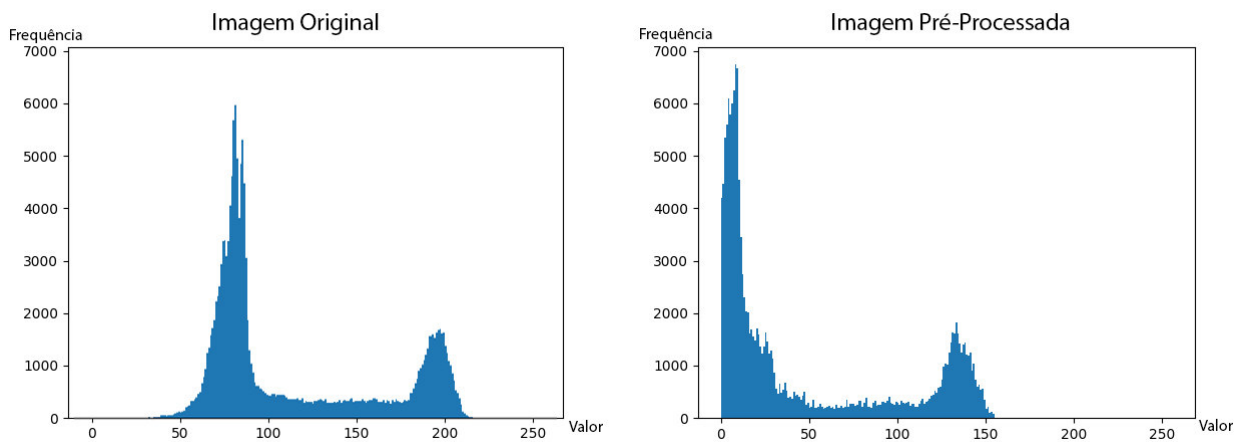


Figura 4.3: Comparação entre histograma de imagem antes e depois de pré-processamento. A aplicação dessas operações causa o escurecimento da imagem, representada pelo deslocamento para a esquerda do gráfico, uniformizando o fundo preto.

4.2 Segmentação

Conforme proposto na metodologia, o desenvolvimento dos modelos de segmentação foi realizado com base no *Data set 1*. As subseções a seguir demonstram os modelos propostos tendo em vista o algoritmo *watershed* e modelo baseado em aprendizado de máquina.

4.2.1 Modelo de Segmentação com Algoritmo *Watershed*

Devido às imperfeições provenientes de ruídos nas imagens sem pré-processamento, espera-se que o efeito de super-segmentação característico do método *watershed* ocorra. Para observar essa ocorrência, na Figura 4.5 é apresentado um exemplo de segmentação do algoritmo utilizando imagens sem e com pré-processamento, respectivamente. Nota-se na imagem à esquerda, que o algoritmo falha ao segmentar toda região superior da imagem, no qual classifica tudo como área de mariscos. Na região inferior, o algoritmo tem sucesso em segmentar três animais. Entretanto, ele rotula o fundo do aquário como duas classes distintas. A saída com pré-processamento, na imagem à direita, apresenta resultado mais próximo àqueles esperados,

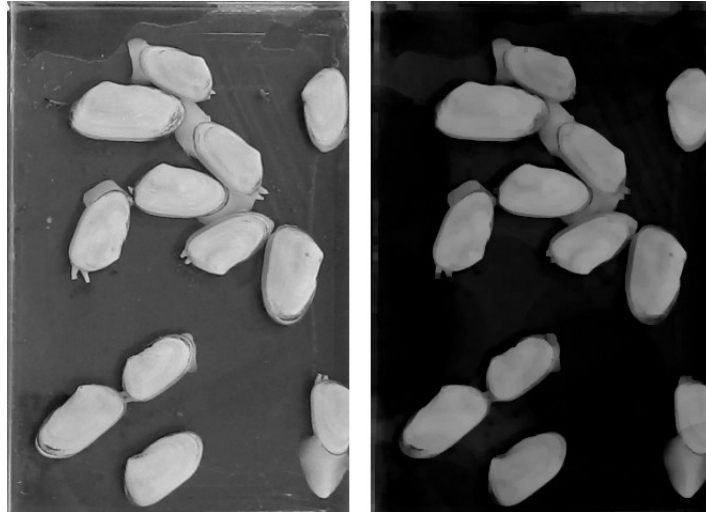


Figura 4.4: ROI extraída antes e depois de pré-processamento

com boa definição das bordas dos bivalves e alcançando F_1 -score igual à 0,945.

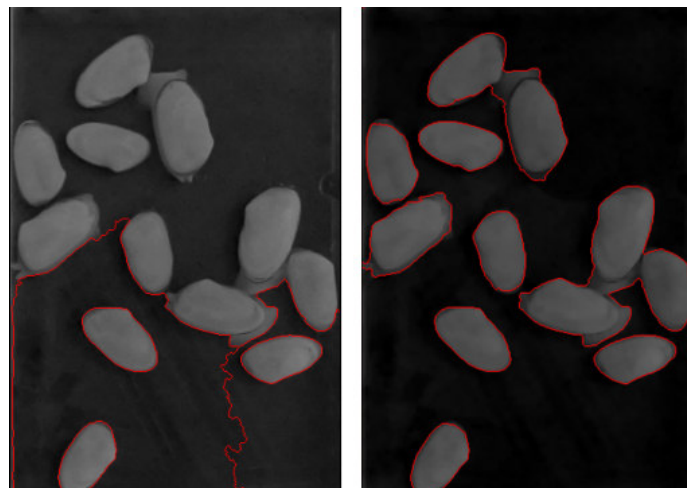


Figura 4.5: Exemplo de saída do algoritmo *watershed* utilizando imagens sem e com pré processamento, respectivamente. No primeiro caso, observa-se dificuldade de modelo em identificar todos os mariscos, enquanto na imagem tratada há a identificação das bordas de cada animal presente.

Dessa forma, as 432 imagens do *data set* que possuem anotações foram segmentadas por meio do algoritmo *watershed*. A saída do modelo para cada imagem foi utilizada junto ao seu respectivo *ground truth* para cálculo das métricas de precisão, *recall* e F_1 -score. A Tabela 1 apresenta os resultados médios obtidos para cada um dos seis aquários.

4.2.2 Modelo de Segmentação Baseado em Aprendizado de Máquina

Seguindo a metodologia proposta, o conjunto de dados pré-processados foram empregados para a geração do modelo. Eles foram divididos em 60% para treinamento, 20% para validação e 20% para teste. Os pesos utilizados para a função de perdas Dice foram de 0,7 para mariscos e 0,3 para fundo, visto a proporção média de cada classe no *data set*. A rede *backbone* utilizada

Tabela 1: Performance do modelo <i>Watershed</i>			
Aquário	Precisão	Recall	F ₁ -score
Controle - A	91,42 %	95,78 %	0,9355
Controle - B	93,11 %	95,69 %	0,9438
Tratamento 1 - A	91,69 %	92,75 %	0,9222
Tratamento 1 - B	93,45 %	92,48 %	0,9296
Tratamento 2 - A	92,58 %	96,36 %	0,9443
Tratamento 2 - B	92,35 %	95,71 %	0,9400
Média	92,43 %	94,80 %	0,9359

foi a Inception V3 com pesos pré-treinados no *data set* Imagenet. O treinamento foi conduzido utilizando taxa de aprendizagem 0,001 com otimizador Adam, *batch size* de 16 imagens e 20 passos por época por 15 épocas.

A performance da rede durante treinamento e validação pode ser observada na Figura 4.6. Observa-se o decaimento quase linear das perdas de validação com picos nas épocas 2, 5 e 7 que podem estar atrelados ao uso do otimizador Adam com bateladas pequenas (RUDER, 2017). A performance de validação tem rápido aumento a partir da época 5, alcançando seu valor máximo de F₁-score igual 0,956 na décima época.

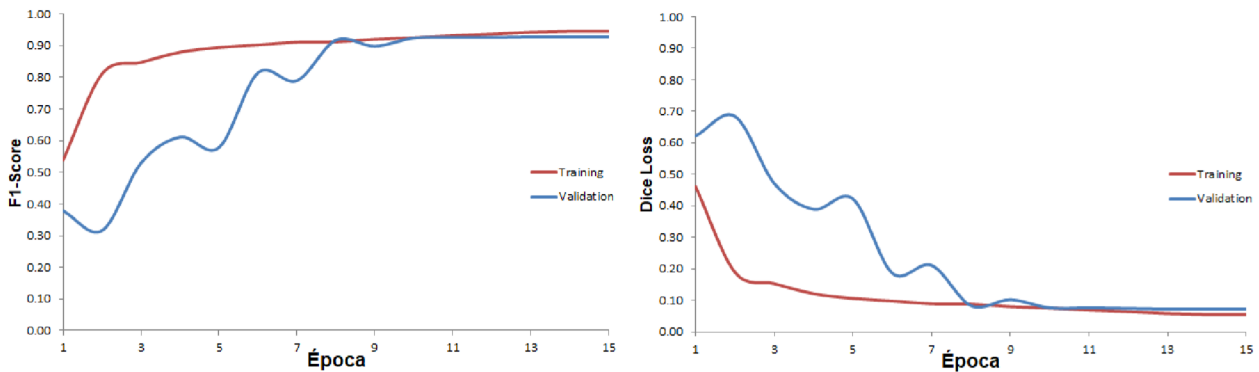


Figura 4.6: Performance do modelo ao longo das épocas de treinamento e validação, atingindo valor máximo de F₁-score (0,956) na décima época

De forma análoga, foi gerado outro modelo porém utilizando o *data set* sem pré-processamento. A performance de treinamento e validação dessa segunda rede é apresentada na Figura 4.7. Nota-se que essa também atinge seu máximo F₁-score (0,968) na décima época.

4.2.3 Comparação dos Modelos

o F₁-score médio obtido por cada método de segmentação é calculado para cada aquário conforme apresentado na Tabela 2. Os dados referentes ao modelo baseado em aprendizado de máquina incluem resultados para a rede treinada com dados originais e pré-processados. Além disso, a performance é avaliada para cada período individualmente como apresentado na Figura 4.8.

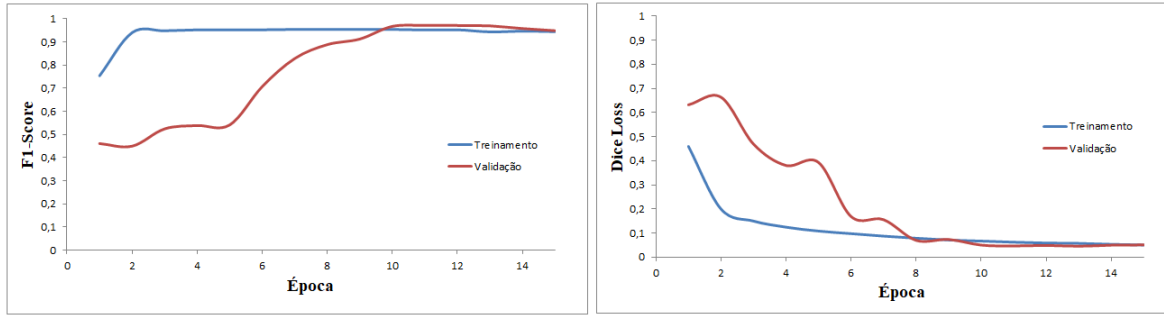


Figura 4.7: Performance do modelo com *data set* sem pré-processamento, que atinge F_1 -score máximo (0,968) na décima época.

Tabela 2: F_1 -score médio para cada aquário

Aquário	F-score		
	Watershed	U-Net C/ Pré-proc.	U-Net S/ Pré-proc.
Controle A	0,922	0,944	0,946
Controle B	0,940	0,953	0,956
Tratamento 1-A	0,933	0,917	0,940
Tratamento 1-B	0,926	0,926	0,959
Tratamento 2-A	0,952	0,956	0,953
Tratamento 2-B	0,943	0,957	0,958
Média	0,936	0,942	0,952

Observa-se que os modelos baseados em aprendizado de máquina atingiram performance levemente superior em relação ao modelo de segmentação *Watershed*. Esse, no entanto, já apresenta boa performance, com F_1 -score médio de 0,936. Esse fator apresenta o algoritmo clássico como uma alternativa viável para a etapa de segmentação do modelo de biossensor.

A performance obtida pelas redes U-Net apresentam diferenças, principalmente quando aplicadas aos dados do Tratamento 1-A e 1-B. O modelo treinado sem pré-processamento apresentou maior F_1 -score em todos períodos de filmagem quando comparado ao modelo com dados pré-processados. Essa discrepância pode indicar que o tratamento de imagem proposto pode estar degradando o *data set*, reduzindo o número de características presentes (como texturas das conchas e coloração natural dos animais). Dessa forma, a RNA teria maior dificuldade em segmentar os mariscos corretamente. Por outro lado, ambas redes treinadas apresentaram alta performance para a aplicação proposta que ainda pode ser otimizada através de *fine tuning* e abordagens centrada em dados.

4.3 Calibração e Inferência

A viabilidade da atividade de exploração do grupo de mariscos como *proxy* de qualidade da água foi avaliada através de análise estatística. A metodologia foi aplicada para os dois *data sets* (A e B) obtidos no Experimento 2 para verificação dos efeitos da temperatura no comportamento. Analogamente, o procedimento foi realizado para o *data set* obtido no Experimento

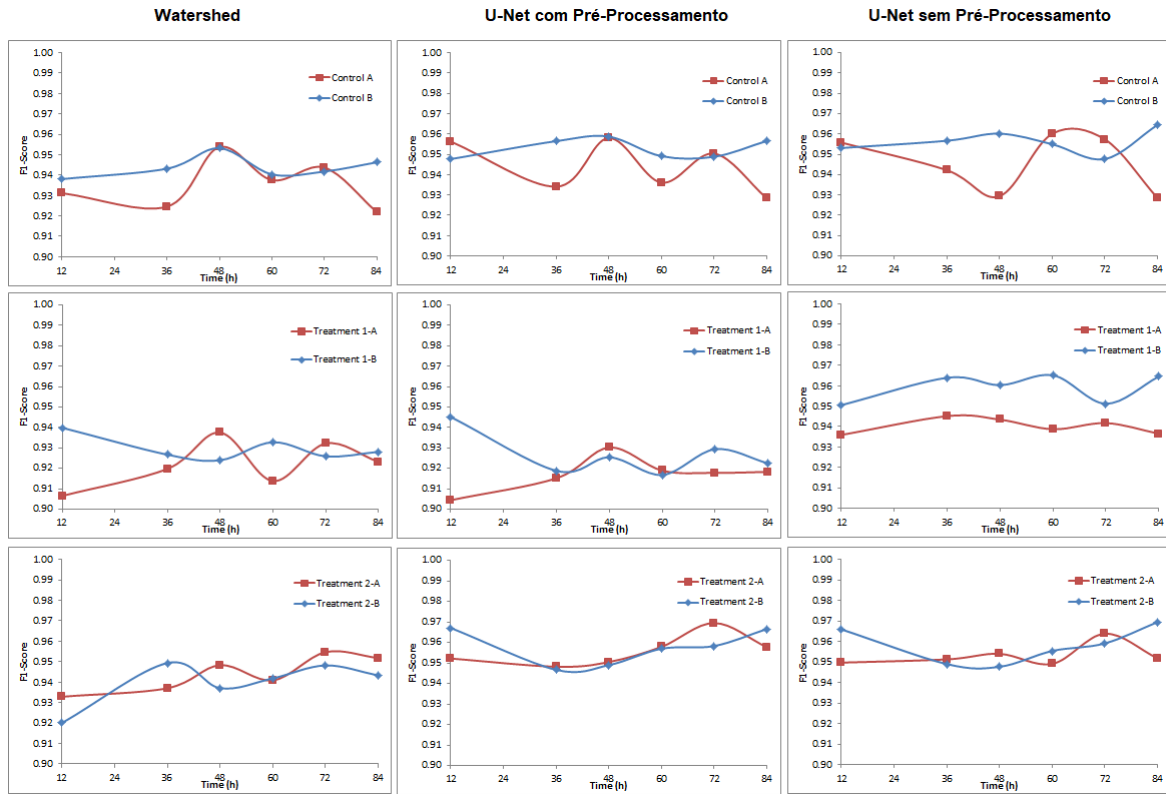


Figura 4.8: F_1 -score obtido por cada modelo de segmentação para cada intervalo de leitura. De forma geral, as performances entre modelos são próximas com exceção dos valores de F_1 -score médio para Tratamento 1, no qual a rede U-Net sem pré-processamento nos dados atinge resultados melhores em comparação com os demais modelos.

3 para análise comportamental frente a diferentes concentrações de contaminante.

Além disso, os parâmetros comportamentais propostos são empregados no desenvolvimento de classificadores inteligentes. Esses modelos, uma vez treinados e validados com os *data sets* apropriados, se tornam capazes de classificar novas observações de entrada. As próximas subseções apresentam resultados e discussões para essas abordagens.

4.3.1 Avaliação da Temperatura

As Figuras 4.9 e 4.10 apresentam os gráficos de área ocupada pelos animais ao longo das gravações durante o experimento 2 para os *data sets* A e B, respectivamente.

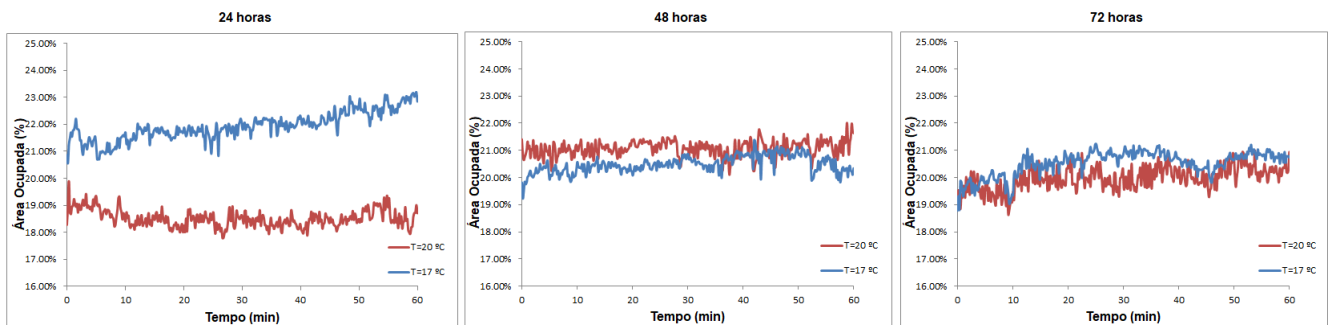


Figura 4.9: Área ocupada por mariscos durante o experimento 2-A

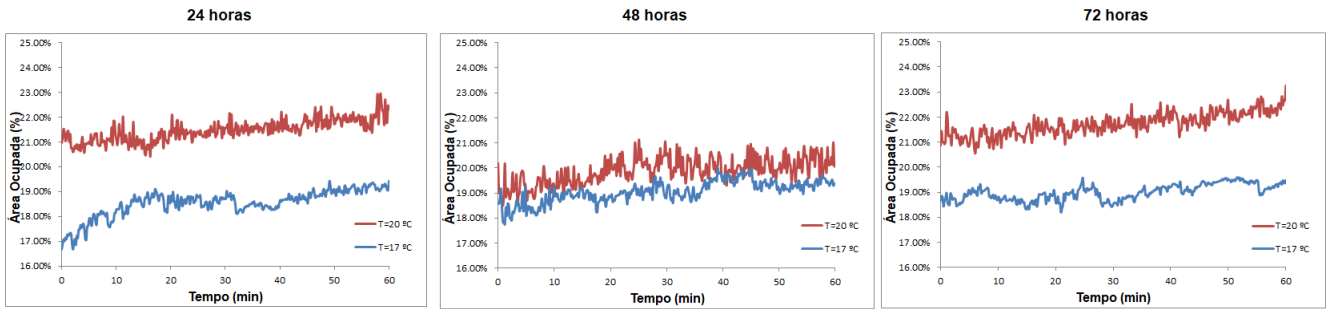


Figura 4.10: Área ocupada por mariscos durante o experimento 2-B

Nota-se que as temperaturas não provocam variações acentuadas na área total ocupada. Todas as curvas sugerem um leve aumento ou diminuição da área ocorrendo ao longo do período registrado. Pequenas flutuações ao longo de todas curvas também são evidentes. Essas alterações são provenientes de micro-movimentos do conjunto de animais. As atividades exploratórias foram calculadas com base na área ocupada ao longo do tempo para os *data sets* A (Figura 4.11) e B (Figura 4.12).

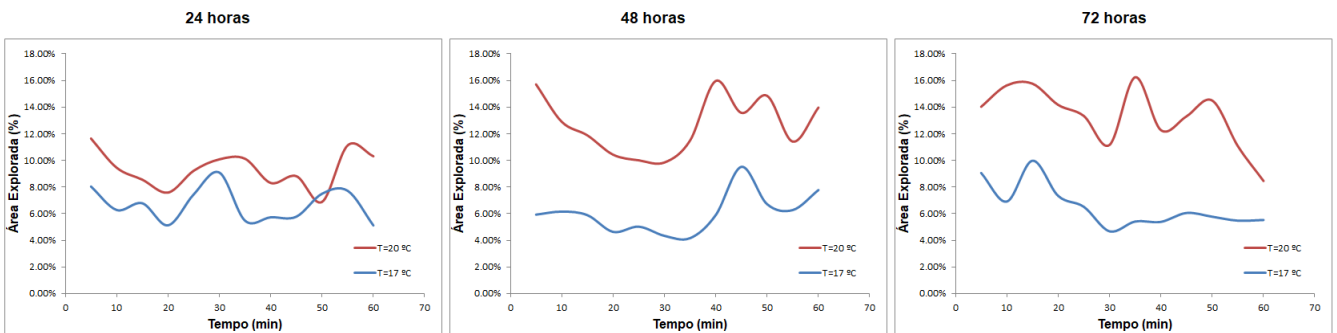


Figura 4.11: Área explorada por mariscos durante o experimento 2-A

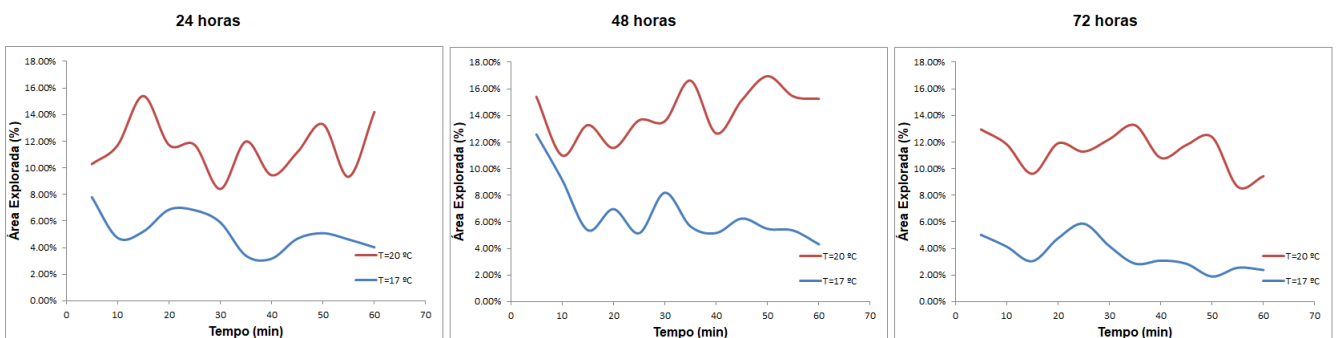


Figura 4.12: Área explorada por mariscos durante o experimento 2-B

Observou-se uma maior atividade de exploração nos aquários submetidos a 20 °C em comparação aos grupos de animais sujeitos a temperaturas mais amenas. Com exceção de um breve período aos 50 minutos de gravação em 24 horas do *data set* A, a área explorada no aquário menos frio foi consideravelmente maior. Essa diferença é evidenciada, também, pelo teste ANOVA (Tabela 3) que aponta diferenças significativas para todos os casos de teste.

Tabela 3: Teste ANOVA ($\alpha = 5\%$) para área explorada nos *data sets* de temperatura

Dataset	Período	df	F	P-valor	F crítico
A	24h	1	23,58	0,001	4,301
A	48h	1	76,41	0,001	4,301
A	72h	1	72,32	0,001	4,301
B	24h	1	78,32	0,001	4,301
B	48h	1	76,13	0,001	4,301
B	72h	1	203,56	0,001	4,301

Tabela 4: Acurácias atingidas para o *data set* de temperatura

Classificador	Acurácia Validação	Acurácia Teste
Fine Gaussian SVM	91,9%	93,2%
Medium Gaussian SVM	90,1%	90,9%
Weighted KNN	90,1%	95,5%

Para a construção de classificadores inteligentes, considerou-se o grupo submetido a 20 °C como controle, ao qual foi atribuído a classe "0". Os 3 modelos que atingiram maiores acurácias durante o processo de validação foram *fine gaussian SVM*, *medium SVM* e *weighted KNN*. Os valores de acurácia mantiveram-se entre 90,1% e 91,9% na validação e entre 90,9% e 95,5% no teste, conforme apresentado na Tabela 4. Suas matrizes de confusão e curvas ROC de teste são disponibilizadas na Figura 4.13.

Apesar de empatado como segunda maior acurácia de validação, o classificador *weighted KNN* atinge a maior acurácia durante testes. Além disso, esse também apresenta a maior AUC-ROC entre os modelos selecionados, atingindo 0,93. O presente classificador indica, através dessas métricas, que é possível alcançar boa performance na inferência de condições anormais de temperatura a partir da área total ocupada pelos mariscos e sua atividade de exploração.

4.3.2 Avaliação de Contaminante

O *data set* obtido no experimento 3 foi empregado para cálculo da área ocupada (Figura 4.14) e atividade exploratória (Figura 4.15), conforme a metodologia proposta.

A área ocupada pelo grupo Tratamento 1 nas primeiras 24h apresenta tendência similar ao observado para temperatura. Nota-se, dessa forma, uma diminuição na área ocupada pelos animais ao longo do tempo. O grupo Controle também passa por reduções, porém com mais variações ao longo do período. O Tratamento 2, por outro lado, sofre uma perturbação aos 10 minutos e se mantém quase constante até o final do período investigado. No período de 48h, o grupo Tratamento 1 apresenta maior oscilação por todo o intervalo, o qual é explicitado pelo aumento de maiores frequências (Figura 4.16). A semelhança entre as áreas de exploração por cada grupo toxicológico é analisada por meio do teste ANOVA (Tabela 5) e post-hoc (Tabela 6).

Através do teste estatístico, observou-se uma atividade de exploração similar entre os grupos

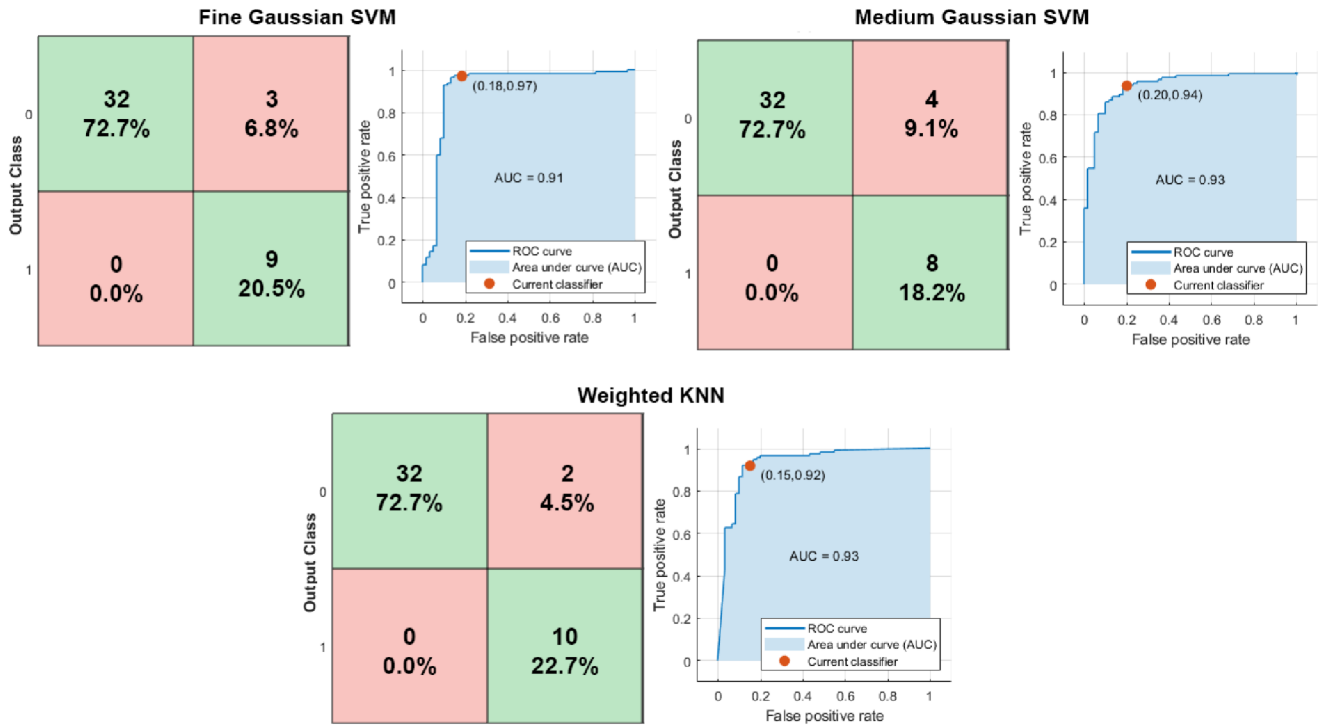


Figura 4.13: Matriz de confusão e curvas ROC para o *data set* de temperatura. Os valores de AUC-ROC obtidos variam entre 0,91 e 0,93 indicando ótima performance dos três classificadores em inferir o grupo temperatura em novas observações.

Tabela 5: Testes ANOVA ($\alpha = 5\%$) para *data set* DCOIT

Dataset	Período	df	F	P-valor	F crítico
DCOIT	24h	2	4,72	0,016	3,28
DCOIT	48h	2	110,10	0,001	3,28

de Controle e sujeitos à menor concentração de DCOIT nas primeiras 24 horas. Essa relação muda no período seguinte, no qual o Tratamento 1 passa a apresentar diferenças significativas com o Controle e semelhanças em relação ao grupo de animais expostos à maior concentração de DCOIT. Essa constatação indica que o contaminante promove estresse toxicológico ao animal possivelmente relacionada com o tempo de exposição.

Ao treinar classificadores, os modelos *fine gaussian SVM*, *fine KNN* e *medium KNN* atingiram as maiores acurácias na validação (Tabela 7)

O classificador *medium KNN*, apesar de ter atingido a terceira maior acurácia na validação, alcança o maior valor no teste (89.7%) e maior AUC-ROC (0.98). Observou-se ainda que as acurácias de teste obtidas foram abaixo das acurácias de validação, além de serem inferiores às atingidas pelo *data set* de temperatura.

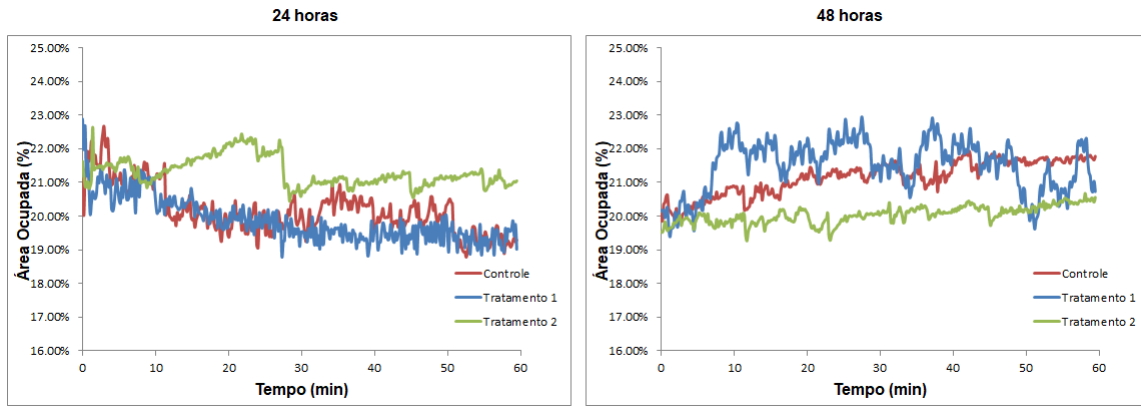


Figura 4.14: Área ocupada por mariscos durante o experimento 3 frente a diferentes concentrações de DCOIT

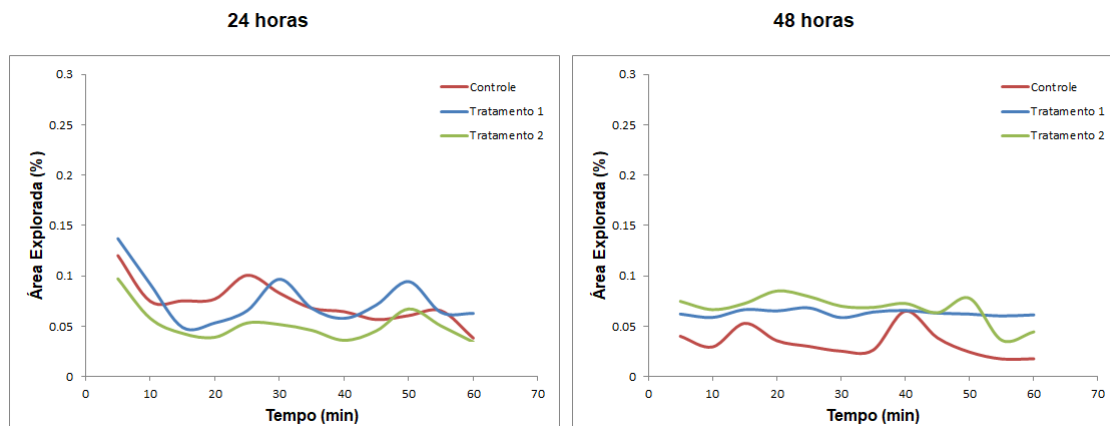


Figura 4.15: Área explorada por mariscos durante o experimento 3 frente a diferentes concentrações de DCOIT

4.4 Discussões

O uso de diferentes câmeras durante o Experimento 1 levou a importantes observações. Pelo uso simultâneo de modelos distintos, foi necessário a calibração da etapa de pré-processamento de forma a adequar todo conjunto de dados a atingir boa performance quando segmentados pelo algoritmo *watershed*, porém a rotina pode ter sido prejudicial para os modelos baseados em aprendizagem de máquina. Por esses modelos se basearem na identificação e reconhecimento de características da imagem (MO et al., 2022), as operações de pré-processamento realizadas podem prejudicar suas performances por eliminar parte da riqueza de detalhes contida nos dados. Esse fenômeno é observado nos dados de Tratamento 1, no qual a rede U-Net atuando com os dados tratados atinge F_1 -score menor que a mesma rede treinada com os dados originais. Dessa forma, nota-se que diferentes câmeras podem ser empregadas para aquisição de imagens de mariscos, entretanto é aconselhável não utilizar diferentes câmeras simultaneamente. Nos experimentos posteriores, o uso de câmeras GoPro eliminou possíveis interferências dessa origem nas leituras observadas.

A avaliação de qualidade de água foi realizada através de métodos estatísticos entre gru-

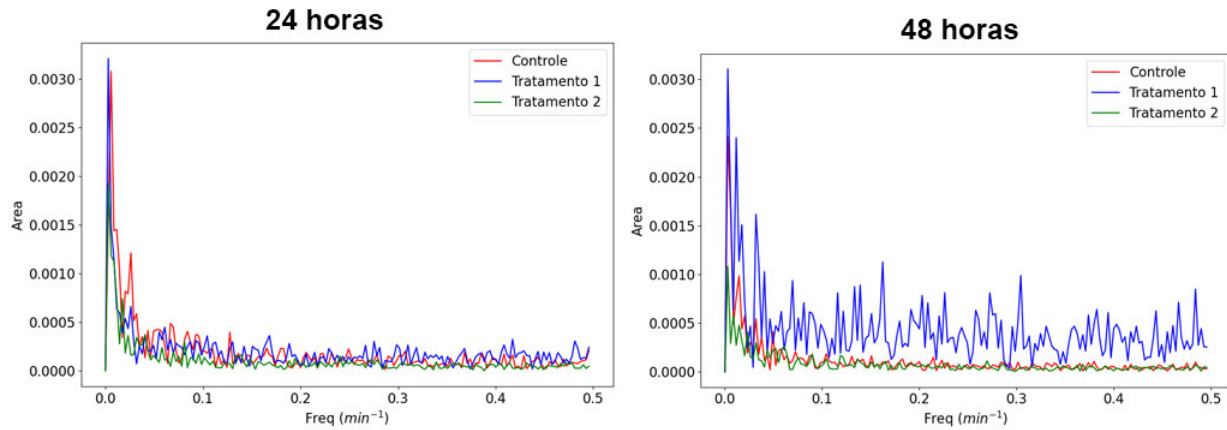


Figura 4.16: Séries de contaminante no domínio da frequência nos períodos 24 e 48 horas, respectivamente

Tabela 6: Testes post-hoc para *data set* DCOIT

Período	Grupo DCOIT	Média Abs	Q crítico	Diferença Significativa
24h	Controle x 0,01	0,0022	0,0184	Não há evidência
24h	0,01 x 0,13	0,0241	0,0184	Existe
24h	Controle x 0,13	0,0219	0,0184	Existe
48h	Controle x 0,01	0,0294	0,0101	Existe
48h	0,01 x 0,13	0,0446	0,0101	Existe
48h	Controle x 0,13	0,0340	0,0101	Existe

pos experimentais a cada período e do uso de classificadores, nos quais os *data sets* foram utilizados para treinamento e inferência de novos dados de entrada. A qualidade da água foi alterada de duas formas em ocasiões diferentes: Em um experimento, foram utilizados aquários em diferentes temperaturas enquanto em outro experimento foi inserido DCOIT no ambiente em diferentes concentrações mantendo os parâmetros físico-químicos do reservatório constantes. Os parâmetros comportamentais empregados na análise do efeito dessas variações no comportamento dos mariscos foram a área total ocupada e a atividade de exploração.

A análise da temperatura indica poucas variações na área total ocupada entre os grupos 20 °C e 17 °C, as quais apresentam diferenças sutis no percentual de área ocupada ao longo do período. Já a atividade exploratória se mostra menor para os grupos na temperatura mais baixa. Esse perfil de comportamento pode ser esperado quando se observa a natureza comportamental dos mariscos, que são menos ativos nos períodos mais frios do ano (KRAEUTER; CASTAGNA, 2001). A mudança comportamental foi evidenciada através da metodologia empregada indicando a capacidade do marisco de se adaptar em algum nível a diferentes temperaturas. Percebeu-se ainda evidências de que a alteração nos padrões comportamentais de mariscos pode ser *proxy* para identificação de mudanças de temperatura potencialmente causada por poluição térmica, por exemplo.

A mudança comportamental causada pela temperatura é identificada através do uso de

Tabela 7: Acurácias atingidas para o *data set* de contaminante

Classificador	Acurácia Validação	Acurácia Teste
Fine Gaussian SVM	94,8%	86,2%
Fine KNN	94,8%	82,8%
Medium KNN	91,3%	89,7%

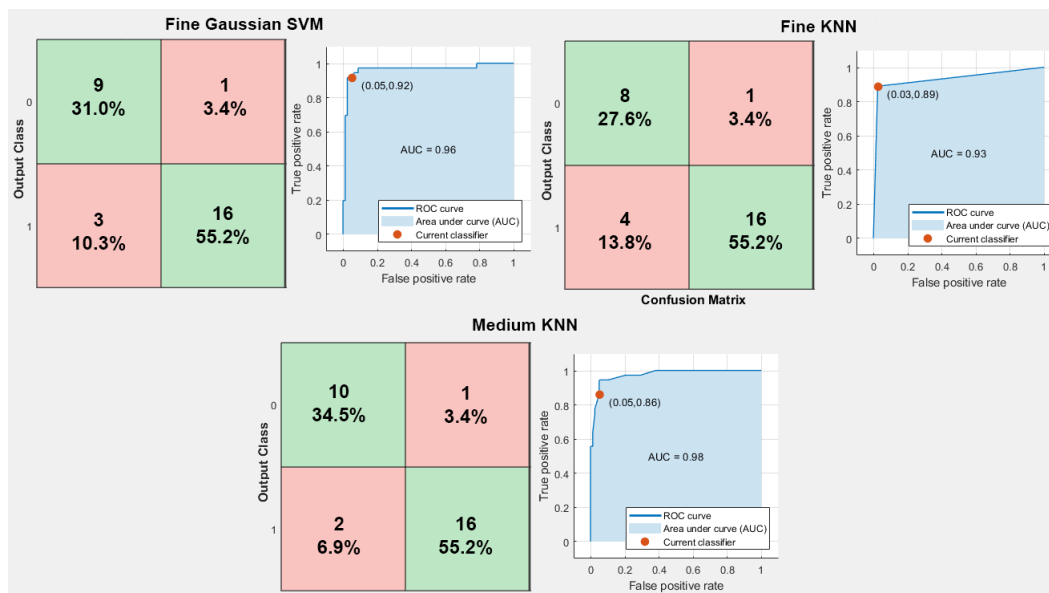


Figura 4.17: Matriz de confusão e curvas ROC para o *data set* de contaminante. Os três classificadores propostos indicam performance adequada para inferir novas observações, apesar do maior número de classificações erradas em comparação com os resultados obtidos para temperatura. A causa dessa maior taxa de erros pode estar relacionada principalmente com o comportamento similar dos animais nos grupos Controle e Tratamento 1 durante as primeiras 24h.

classificadores. Modelos treinados com os *data sets* atingiram boa performance durante teste com novas observações, com acurácias entre 90,9% e 95,5%. Esses resultados apontam predições satisfatórias ao empregar classificadores para inferir diferentes condições ambientais promovidos por temperatura (GUTERRES, 2021).

Para os experimentos com contaminante, no período 24 horas observa-se redução da quantidade de micro-movimentos no Tratamento 2 em relação aos demais grupos, o que pode indicar comportamento de evitação dos mariscos causado pelo contato com DCOIT. Por outro lado, o teste ANOVA para os grupos não apresenta evidências de diferenças entre Controle e Tratamento 1. Já no período 48 horas, a metodologia aponta diferenças significativas entre esses grupos, o que pode indicar efeito do tempo de exposição ao contaminante para início das reações comportamentais dos mariscos ((BALBI et al., 2021)) visto a baixa concentração de contaminante no ambiente (JORGE et al., 2016). A observação dos dados no domínio da frequência evidencia maiores oscilações do grupo Tratamento 1 no período de 48 horas potencialmente causadas por reação dos mariscos ao contaminante. No experimento realizado, essa mudança comportamental levou pelo menos 24 horas para ser percebida na concentração

0,13 mg DCOIT/L e pelo menos 48 horas no grupo 0,01 mg DCOIT/L. O tempo de reação ao contaminante pode ser mapeado com maior precisão em experimentos com concentrações intermediárias e maior frequência de aquisição, permitindo a calibração do sistema ao DCOIT.

O treinamento de classificadores para inferir a presença de contaminantes no ambiente obteve bons resultados, com acurácia entre 91,3% e 94,8%. Durante teste, sua performance foi inferior em relação aos modelos treinados para temperatura, com acurácias entre 82,8% e 89,7%. Apesar disso, os classificadores para contaminantes ainda apresentam alto AUC-ROC, que variaram entre 0,93 e 0,98. A diferença de performance entre esses classificadores e os modelos treinados para temperatura foi potencialmente provocada pelo tempo de reação dos mariscos no grupo Tratamento 1. Conforme identificado pela análise estatística nesse intervalo, o grupo de menor concentração não apresentava evidências de diferenças significativas com o grupo controle. Portanto, é esperável maior taxa de falsos negativos na saída do classificador.

5 CONCLUSÃO

A presente dissertação de mestrado teve como objetivo averiguar o uso de biossensores de qualidade da água baseados no grau de exposição de mariscos brancos (*Amarilladesma mac-troides*) a partir do desenvolvimento e utilização de modelos computacionais e metodologias não-invasivas. O processo baseou-se na observação comportamental de pés e sifões de mariscos. Esse o primeiro trabalho que se tem conhecimento a analisar o grau de exposição desse conjunto de tecidos dos bivalves de areia através de visão computacional. Trata-se, portanto, de um parâmetro comportamental inédito com potencial empregabilidade na construção de biossensores para monitoramento da qualidade da água. O modelo de três etapas proposto abrangeu desde a preparação de uma configuração de coleta de dados até sistemas inteligentes, os quais tem suas saídas como *proxys* para monitoramento aquático.

A estrutura de coleta de imagens apresentada foi realizada em experimentos laboratoriais. Ela se apresentou adequada para filmagem dos mariscos brancos. Os vídeos obtidos foram ajustados por uma rotina de pré-processamento e submetidos à etapa de segmentação. Nesta etapa, informações de área ocupada pelos animais foram extraídas ao longo do tempo. A validação dos modelos de segmentação *watershed* e aprendizado de máquina utilizou dados obtidos em um estudo de caso, a qual apresentou performance satisfatória entre 0,936 e 0,952 F_1 -score.

A análise comportamental dos mariscos foi realizada através da área total ocupada e da atividade de exploração dos mariscos, parâmetros até então não explorados pela literatura. Utilizando-se análise estatística e sistemas de classificadores inteligentes, foi possível relacionar mudanças comportamentais a diferentes condições aquáticas. Observaram-se variações significativas em função da temperatura, que podem ser utilizadas como *proxies* para detecção de poluição térmica, por exemplo. A reação dos mariscos à presença de DCOIT no ambiente também foi notável, embora esse fator esteja atrelado a uma relação entre tempo de exposição e concentração de contaminante. Sistemas de classificadores foram capazes de inferir as diferentes condições quando submetidos à classificação de novas observações, atingindo acurácias de até 95,5% para temperatura e até 89,7% para DCOIT.

Trabalhos futuros podem realizar estudos mais detalhados da diversidade de comportamento dos indivíduos, bem como a melhor calibração o biossensor através de experimentações com concentrações intermediárias de DCOIT e com menores intervalos de tempo entre cada filmagem. Além disso, estudos podem aprimorar o modelo proposto ao avaliar diferentes densidades de animais por aquário e realizar o monitoramento da ocupação espacial individualmente para cada marisco. Dessa forma, poderá ser identificada a relação entre concentração do contaminante e o tempo de resposta dos mariscos à presença do composto químico, além de permitir a análise comportamental de cada animal dentro dos seus grupos experimentais. A viabilidade do uso dos parâmetros área total ocupada e atividade de exploração pode ser explorada para outros parâmetros físico-químicos importantes no monitoramento ambiental, como pH, salinidade e

oxigênio dissolvido bem como servir de base para o uso de novos parâmetros comportamentais.

Portanto, a presente dissertação de mestrado contribuiu no desenvolvimento de biossensores baseados em mariscos verificando a viabilidade da análise comportamental desse indivíduos como *proxies* de qualidade da água. Através da análise inédita de padrões comportamentais baseados na exposição dos tecidos desses animais, foi possível calibrar modelos com potencial eficiência em aplicações de monitoramento aquático.

6 REFERÊNCIAS

- ANDRADE, H. et al. High frequency non-invasive (hfni) bio-sensors as a potential tool for marine monitoring and assessments. *Frontiers in Marine Science*, Frontiers, v. 3, p. 187, 2016.
- AQUADECT. *MOSSELMONITOR*. 2020. Disponível em: <http://www.mosselmonitor.nl/>.
- BALBI, T. et al. Immunological responses of marine bivalves to contaminant exposure: Contribution of the -omics approach. *Frontiers in Immunology*, PubMed Central, v. 12, 2021.
- BAMBER, S. D.; WESTERLUND, S. Behavioral responses of *Arctica islandica* (bivalvia: Arc-ticidae) to simulated leakages of carbon dioxide from sub-sea geological storage. *Aquatic Toxi-cology*, Elsevier, v. 180, 2016.
- BEUCHER, S.; MEYER, F. The morphological approach to segmentation: The watershed transformation. In: *Mathematical Morphology in Image Processing*. [S.l.]: Taylor Francis Group, 1993. p. 433–481.
- BHATI, B. S.; RAI, C. S. Analysis of support vector machine-based intrusion detection tech-niques. *Arabic Journal for Science and Engineering*, Springer, 2019.
- BIALONCZYK, A. D.; JERMACZ, L.; ZIELSKA, J. What scares a mussel? changes in valve movement pattern as an immediate response of a byssate bivalve to biotic factors. *Hydrobiologia*, Springer, v. 841, 2019.
- BRADSKI, G.; KAEHLER, A. *Learning OpenCV*. [S.l.]: O'Reilly, 2008. ISBN 9780596516130.
- BRÉHÉRET, A. *Pixel Annotation Tool*. 2017. <https://github.com/abreheret/PixelAnnotationTool>.
- BROWNLEE, J. *How Do Convolutional Layers Work in Deep Learning Neural Networks*. 2019. <https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/>.
- BURGER, W. Zhang's camera calibration algorithm: In-depth tutorial and implementation. 2016.
- BYRNE, R. A. et al. Behavioral and metabolic responses to emersion and subsequent reim-ersion in the freshwater bivalve, corbicula fluminea. *Biological Bulletin*, Marine Biological Laboratory, v. 178, 1990.
- CALDERÓN, J. F. E.; SIKORSKA, J. C.; DOUDA, K. High-frequency video analysis extends beyond the capabilities of valvometry in acute behavioral disturbance detection in bivalves. *Ecological Indicators*, Elsevier, v. 136, 2022.
- CAMPAS, M.; PRIETO-SIMON, B.; MARTY, J.-L. Biosensors to detect marine toxins: As-sessing seafood safety. *Talanta*, Elsevier, v. 72, 2007.
- CARVALHO, Y. B. M. *Patologias que afetam o marisco branco Mesodesma mactroides (Bival-via: Mesodesmatidae)*. Dissertação (Mestrado) — FURG, Brasil, 2015.
- CHAKI, J.; DEY, N. *A Beginner's Guide to Image Preprocessing Techniques*. [S.l.]: CRC Press, 2019.

- CORTES, C.; MOHRI, M. Auc optimization vs. error rate minimization. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2004. p. 313–320.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, Kluwer, v. 20, 1995.
- COSCARÓN, S. *La "almeja amarilla" (Mesodesma (T.) mactroides Deshayes) de la costa de la provincia de Buenos Aires*. [S.l.]: Agro, 1959.
- DONG, J. et al. A survey of smart water quality monitoring system. *Environmental Science and Pollution Research*, Springer, v. 22, n. 7, p. 4893–4906, 2015.
- ELGENDY, M. *Deep Learning for Vision Systems*. [S.l.]: Manning Publications, 2019.
- FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, Elsevier, v. 27, 2005.
- GEETHA, S.; GOUTHAMI, S. Internet of things enabled real time water quality monitoring system. *Smart Water*, v. 2, 2017.
- GIANUCA, N. M. Sobre un record de tamaño para *Mesodesma mactroides* deshayes, 1854 (mollusca: Pelecypoda). *Comunicaciones de la Sociedad Malacologica del Uruguay*, v. 4, p. 55–58, 1975.
- GOSLING, E. *Bivalve Molluscs - Biology, Ecology and Culture*. [S.l.]: Fishing News Books, 2003. ISBN 0852382340.
- GRAF, D. L.; CUMMINGS, K. S. Review of the systematics and global diversity of freshwater mussel species (bivalvia: Unionoida). *Journal of Molluscan Studies*, v. 73, p. 291–314, 2007.
- GUDIMOV, A. V.; KOMAROVA, E. P. Environmental control: from inertia of the standard biomonitoring to speed of the online one. *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, v. 625, 2021.
- GUO, G. et al. Knn model-based approach in classification. *On the Move to Meaningful Internet Systems*, 2003.
- GUTERRES, B. d. V. *Moluscos Bivalves como Biossensores de Poluição Aquática*. Brasil: [s.n.], 2021.
- HARTMANN, J. T. et al. Establishing mussel behavior as a biomarker in ecotoxicology. *Aquatic Toxicology*, Elsevier, v. 170, 2015.
- HASAN, K. et al. Diabetes prediction using ensembling of different machine learning classifiers. *IEEE Access*, IEEE, v. 8, 2020.
- HAYKIN, S. *Redes Neurais: Princípios e Prática*. Artmed, 2007. ISBN 9788577800865. Disponível em: <<https://books.google.com.br/books?id=bhMwDwAAQBAJ>>.
- HEIJDEN, F. v. d. *Image Based Measurement Systems: Object Recognition and Parameter Estimation*. [S.l.]: Wiley, 1994.
- HERRMANN, M. Population dynamics of the argentinean surf clams *Donax hanleyanus* and *Mesodesma mactroides* from open-atlantic beaches off argentina. 2008.

- HIGGINS, P. J. Effects of food availability on the valve movements and feeding behavior of juvenile *crassostrea virginica* (gmelin). i. valve movements and periodic activity. *Journal of Experimental Marine Biology and Ecology*, Elsevier, v. 45, n. 2, p. 229–244, 1980.
- JADON, S. A survey of loss functions for semantic segmentation. 2020.
- JANSSON, A. et al. Singing voice separation with deep u-net convolutional networks. *18th International Society for Music Information Retrieval Conference*, 2019.
- JENKINS, F. A.; WHITE, H. E. *Fundamentals of Optics*. [S.l.]: McGraw-Hill, 1976.
- JORGE, M. B. et al. Impaired regulation of divalent cations with acute copper exposure in the marine clam *Mesodesma mactroides*. *Comparative Biochemistry and Physiology, Part C*, Elsevier, v. 179, 2016.
- JOU, L.-J. et al. Synthesis and measurement of valve activities by an improved online clam-based behavioral monitoring system. *Computers and electronics in agriculture*, Elsevier, v. 90, p. 106–118, 2013.
- KRAEUTER, J. N.; CASTAGNA, M. *Biology of the Hard Clam*. [S.l.]: Elsevier, 2001.
- KRAMER, K. J. M.; JENNER, H. A.; ZWART, D. d. The valve movement response of mussels: A tool in biological monitoring. *Hydrobiologia*, Springer, v. 189, 1989.
- KUNCHEVA, L. *Combining Pattern Classifiers: Methods and Algorithms*. [S.l.]: Wiley, 2014.
- LIAO, C.-M. et al. Valve movement response of the freshwater clam *Corbicula fluminea* following exposure to waterborne arsenic. *Ecotoxicology*, Springer, v. 18, 2009.
- LIU, Z. et al. Computer vision-based concrete crack detection using u-net fully convolutional networks. *Automation in Construction*, v. 104, 2019.
- LYUBIMTSEV, V.; KHOLODKEVICH, S.; DRUZHININ, I. Measuring systems designed for working with living organisms as biosensors. features of their metrological maintenance. In: IOP PUBLISHING. *Journal of Physics: Conference Series*. [S.l.], 2019. v. 1379, n. 1, p. 012077.
- MARTIN. *Amarilladesma mactroides*. 2019. <<https://www.inaturalist.org/observations/25390600>>. Acessado em: 29-01-2021.
- MI, C. et al. Vision-based measurement: Actualities and developing trends in automated container terminals. *IEEE Instrumentation and Measurement Magazine*, IEEE, v. 24, 2021.
- MO, Y. et al. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, Elsevier, v. 493, p. 626–646, 2022.
- MORGAN, E. L. et al. Using bivalve mollusks as sensors in an early warning, automated biosensing system for water resource protection. *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen*, Taylor Francis Group, v. 29, p. 246–249, 2005.
- NAGAI, K. et al. Detecting the shellfish killer heterocapsa circularisquama (dinophyceae) by measuring bivalve valve activity with a hall element sensor. *Aquaculture*, Elsevier, v. 255, n. 1-4, p. 395–401, 2006.
- NAJAH, A. et al. An application of different artificial intelligences techniques for water quality prediction. *International Journal of the Physical Sciences*, Academic Journals, v. 6, 2011.

- NAYAR, R. Assessment of water quality index and monitoring of pollutants by physico-chemical analysis in water bodies: a review. *International Journal of Engineering Research and Technology*, v. 9, n. 01, 2020.
- NEWTON, T. J.; COPE, W. G. Biomarker responses of unionid mussels to environmental contaminants. p. 257–284, 2006.
- NUÑEZ, J. D. et al. Burrowing behavior of an infaunal clam species after siphon nipping. *Journal of Experimental Marine Biology and Ecology*, v. 459, 2014.
- POWERS, D. M. W. Evaluation: From precision, recall and f-factor to roc, informedness, markedness correlation. *Human Communication Science SummerFest*, 2007.
- PREIM, B.; BOTHA, C. Image analysis for medical visualization. In: *Visual Computing for Medicine*. [S.l.]: Morgan Kaufmann, 2014. p. 111–175.
- RAKHLIN, A.; DAVYDOW, A.; NIKOLENKO, S. Land cover classification from satellite imagery with u-net and lovász-softmax loss. In: *IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS*. [S.l.], 2018. p. 257–261.
- RANA, M. et al. Investigating data-driven approaches to understand the interaction between water quality and physiological response of sentinel oysters in natural environment. *Computers and Electronics in Agriculture*, Elsevier, v. 175, 2020.
- RIISGARD, H. Comment:physiological regulation versus autonomous filtration in filter-feeding bivalves: Starting points for progress. *Ophelia*, v. 54, 2001.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015.
- RUDER, S. *An overview of gradient descent optimization algorithms*. 2017. <https://arxiv.org/pdf/1609.04747.pdf>.
- SAHA, S. *A Comprehensive Guide to Convolutional Neural Networks - the ELI5 way*. 2018. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>.
- SANTAQUITERIA, J. R. et al. Semantic versus instance segmentation in microscopic algae detection. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 87, 2020.
- SANTOS, J. J. S. et al. Immunological profile of the yellow clam *Mesodesma mactroides* (mesodesmatidae) from the southern coast of rio grande do sul, brazil. *Journal of Aquatic Animal Health*, American Fisheries Society, v. 28, 2016.
- SCHÄFER, S. et al. Bioaccumulation in aquatic systems: methodological approaches, monitoring and assessment. *Environmental Sciences Europe*, v. 27, 2015.
- SEO, H. et al. Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications. *Medical Physics*, American Association of Physicists in Medicine, v. 47, 2020.
- SHIRMOHAMMADI, S.; FERRERO, A. Camera as the instrument: The rising trend of vision based measurement. *Instrumentation and Measurement Magazine*, IEEE, v. 17, p. 41–47, 2014.

- SONKA, M.; HLAVAC, V.; BOYLE, R. *Image Processing, Analysis, and Machine Vision*. [S.l.]: Cengage Learning, 2013. ISBN 9781133593607.
- SZELISKI, R. *Computer Vision: Algorithms and Applications*. [S.l.]: Springer, 2010.
- URBANOWICZ, R. J.; BROWNE, W. N. *Introduction to Learning Classifier Systems*. [S.l.]: Springer, 2017.
- VALLERO, D. A. *Waste*. [S.l.]: Academic Press, 2019.
- VENKATESAN, R.; LI, B. *Convolutional Neural Networks in Visual Computing: A Concise Guide*. [S.l.]: CRC Press, 2018.
- ZHANG, X. et al. A marker-based watershed method for x-ray image segmentation. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 113, p. 894–903, 2014.
- ZHANG, Z. Introduction to machine learning: k-nearest neighbors. *Annals of Translational Medicine*, v. 4, 2016.